FEDERAL HOUSING FINANCE AGENCY
Office of the Director

October 30, 2009

Honorable Christopher Dodd
Chairman
Committee on Banking, Housing,
and Urban Affairs
United States Senate
Washington, DC 20510

Honorable Barney Frank
Chairman
Committee on Financial Services
United States House of Representatives
Washington, DC 20515

Honorable Richard C. Shelby
Ranking Minority Member
Committee on Banking, Housing,
and Urban Affairs
United States Senate
Washington, DC 20510

Honorable Spencer Bachus
Ranking Minority Member
Committee on Financial Services
United States House of Representatives
Washington, DC 20515

Dear Chairmen and Ranking Minority Members:

Pursuant to section 1602 of the Housing and Economic Recovery Act of 2008, I am pleased to submit the enclosed report titled "Default Risk Evaluation in the Single-Family Mortgage Market." The report provides a detailed summary of recent research on the underwriting of single-family mortgages, particularly problems that arose during the height of the mortgage lending boom in the middle years of this decade.

The Federal Housing Finance Agency (FHFA) believes that sound underwriting is critical to the current recovery of mortgage markets and must be an essential foundation of the future prosperity of the industry. We believe the enclosed report provides useful insights for industry stakeholders and policymakers aiming to restore safe and sound underwriting practices in housing finance.

Sincerely,

Edward J. DeMarco

Edward J. DeMarco
Acting Director

Attachment

1700 G Street, N.W., Washington, D.C. 20552-0003 • 202-414-3800 • 202-414-3823 (fax)

DEFAULT RISK EVALUATION
IN THE SINGLE-FAMILY MORTGAGE MARKET

October 2009

Contents

[F] Adelino, Manuel, Kristopher Gerardi, and Paul S. Willen, "Why Don't Lenders Renegotiate More Home Mortgages? Redefaults, Self-Cures, and Securitization," *NBER Working Paper 15159, July 2009*.

[G] Harding, John P., Eric Rosenblatt, and Vincent W. Yao, "The Contagion Effect of Foreclosed Properties." July 13, 2009.

PREFACE

This Federal Housing Finance Agency (FHFA) report fulfills the requirement of Section 1602 of the Housing and Economic Recovery Act of 2008 that FHFA conduct a study of ways to improve the overall default risk evaluation used with respect to residential mortgage loans and report to Congress on the results of that study. To aid in the preparation of the report, FHFA and the Federal Deposit Insurance Corporation's (FDIC's) Center for Financial Research jointly selected seven papers for a public symposium held on September 16th, 2009. This report summarizes and evaluates those papers in the context of previous research and the comments provided by discussants at that symposium. The appendices to the report provide the papers themselves.

FHFA is grateful to the FDIC's Center for Financial Research and its director, Paul Kupiec, for co-sponsoring the September 2009 symposium. FHFA also gratefully acknowledges the contribution of Professor John Quigley of the University of California, Berkeley to the preparation of the report.

Edward J. DeMarco
Acting Director

October 2009

EXECUTIVE SUMMARY

This report reviews recent research on the methods of evaluating default risk in the market for single-family mortgages. The report is occasioned by the unusually rapid pace of technical change in the single-family mortgage industry in recent decades, the rapid expansion of that market during the early years of this century, and its subsequent collapse. Many related factors contributed to that collapse, including inadequacies in the institutional structure of the market, inaccuracies in the valuation of real property, flawed assessments of the creditworthiness of borrowers, and poor evaluations of linkages between general economic conditions and mortgage termination. Beyond those factors, widely-used benchmark scenarios for comparing mortgage performance with assumptions about worst-case economic conditions proved wildly optimistic.

Against that background, Section 1602 of the Housing and Economic Recovery Act of 2008 required the Federal Housing Finance Agency (FHFA) to conduct a study of "ways to improve the overall default risk evaluation used with respect to residential mortgage loans" and to report to Congress on the results of that study. This report fulfills that requirement. To aid in the preparation of the report, FHFA and the Federal Deposit Insurance Corporation's Center for Financial Research jointly selected seven papers for a public symposium held in September 2009. The body of this report summarizes those papers in the context of previous research and of the comments provided by discussants at that symposium. The appendices to the report provide the papers themselves.

The papers presented at the symposium were technical in nature and dealt with quite specific aspects of the evaluation and management of risk, especially default risk, in the single-family mortgage market. The papers were distinctive in that they presented and evaluated new research about the operation of the mortgage markets rather than merely providing historical narratives of the history of the past half-decade. They were also distinctive in that the technical analyses have direct policy application.

For example, for the analysis of the valuation of housing collateral, Austin Kelly's paper (see Appendix A) evaluates the potential benefits associated with access to the results of statistical or "automated" estimates of collateral value, as well as professional appraisals, in the origination process for single-family mortgages. In an experiment to determine the effectiveness of that supplemental information, statistical estimates of collateral value at origination were appended to mortgage origination data. It was clearly established that the additional information was useful in predicting subsequent mortgage termination, especially costly default by home purchasers. It would appear that more attention could be paid to the incorporation of such statistical information into the mortgage approval process. That additional information might be especially useful in the evaluation of applications for the refinance of mortgages on existing properties, especially when refinance includes a cash-out payment to property owners.

In the evaluation of the repayment capacity of borrowers in the single-family mortgage market, two of the papers selected for the symposium compared the experience of lenders who offered loans without the customary full documentation of the creditworthiness of borrowers. The papers analyzed loans issued to borrowers on the basis of their stated incomes and assets, without documentation, or else on the basis of no stated income at all. Historically, mortgages underwritten to those underwriting standards were available to those whose incomes varied considerably over time (e.g., small-time proprietors and entrepreneurs) and who consequently obtained loans with very low loan-to-value (LTV) ratios.

The analyses, by LaCour-Little and Yang (see Appendix B) and by Jiang, Nelson, and Vytlacil (see Appendix C), clearly establish that loans for which the stated (or inferred) income of the borrower was high relative to local average income also had much higher default risks. The latter paper addressed the channels of origination as well as the documentation standard. Jiang and her colleagues reported large increases in the delinquency probabilities for low-documentation loans initiated by brokers when compared to loans initiated in-house by bankers. Jiang, *et al* attributed that systematic variation to differences in mortgage originator preference (banker vs. broker) between sophisticated and unsophisticated borrowers. But differences in borrower clientele do not completely explain the higher delinquency rate for broker-originated, low-documentation loans. Thus, it appears that agency problems—misalignment of incentives for brokers and bankers—played a role in the higher delinquency rates observed.

Another research paper presented at the symposium also related to the problems of agency and the alignment of incentives noted above. Elul (Appendix D) examined a large loan-level data set, analyzing the riskiness of particular mortgage products: prime fixed-rate, prime adjustable-rate, and subprime loans. For prime mortgages, the results suggested that broker-originated and low-documentation loans were riskier, as were securitized loans.

The general finding that securitization increases the likelihood that mortgages will be available to borrowers who are worse credit risks but have the same easily observable characteristics highlights the distinction between “hard” and “soft” information and the role of that distinction in assessing default risk. With a routinized channel to sell mortgages based on “hard” data that is easily verifiable, such as FICO scores, lenders have weaker incentives to invest resources to uncover “soft” information about credit-worthiness (for example, employment prospects). That means that “soft” information will be under-supplied, benefiting some borrowers and mortgage holders, but harming others. Over time, if the extent of securitization based only on “hard” information increases, more of those borrowers whose “soft” information makes them less creditworthy will nevertheless obtain mortgages.

A paper by Ghent and Kudlyak (Appendix E) provides new findings about the role of delinquency judgments in affecting behavior in the mortgage market. The authors present a stylized economic model of the effect upon borrower and lender behavior in the mortgage market of a lender’s option to seek deficiency judgments in the event of a default. The model is highly stylized and specific. Yet the analysis convincingly demonstrates that the importance of recourse is not really reflected in the incidence of deficiency judgments, but rather in the effect of the threat of recourse on borrower behavior and lender response. Even if lenders seldom (or never) pursue deficiency judgments in court, losses are lower when the threat of recourse can be exercised credibly.

The authors also investigate empirically the importance of the availability to lenders of recourse on the default behavior of borrowers. The statistical results indicate that the effect of the value of the default option on the probability of exercise is significantly different in recourse and in non-recourse states. Recourse clearly decreases the effect of negative equity on the probability of exercising the default option, i.e. recourse dissuades some borrowers with negative equity from defaulting. However, the extent to which recourse offsets the effect of negative equity decreases as the value of the option increases.

The analysis establishes the importance of the threat of enforcement of deficiency judgments upon outcomes in the mortgage market—the extent of default, the type of mortgage termination, and the portfolios in which default is more likely. Those effects are large, and they persist even if the incidence of litigation to enforce deficiency judgments is neither large nor widespread.

Many have observed that the extent of renegotiation of existing mortgages in response to large reductions in asset values has been limited. When Congress established the Hope for Homeowners Program in the Housing and Economic Recovery Act of 2008, the Program elicited only a few hundred modification applications by the end of the year. The current environment is one in which we would expect the incentives for voluntary loan modification to be stronger, however. It is widely asserted that the principal barrier to mutually beneficial renegotiation between borrowers and lenders is securitization. When mortgages are securitized through the issuance of private-label mortgage-backed securities, the interests of owners of the various tranches of those securities are not perfectly aligned, and actions taken by the mortgage servicer need not make all investors better off without harming others.

The paper by Adelino, Gerardi and Willen (Appendix F) takes issue with that analysis of the nexus between securitization and loan modification. Using information on mortgage outcomes through the first quarter of 2009 (perhaps reflective of a different market environment than exists today), Adelino, *et al* argue that mortgage loan renegotiation is seldom in the economic interests of lenders. Even in the face of huge reductions in asset values after foreclosure, Adelino, *et al* claim that it is still typically in the interest of lenders to foreclose rather than to modify or write down a seriously delinquent mortgage. As reported by Adelino, *et al*, the cure rate of loans that are sixty days delinquent is thirty percent. Thus, if lenders modified all sixty-day delinquent loans, thirty percent of the lenders' resources would be "wasted" because those loans would have cured even without the resources spent on intervention. Similarly, the authors report that about thirty to forty-five percent of borrowers whose loans have been modified become seriously delinquent again within six months. Thus, the resources lenders spend on modifications of those loans would have been "wasted" as well. In summary, it is entirely possible that a general policy of foreclosure will allow lenders to recover more than a policy of modification, even if a foreclosure does cause the value of the dwelling to decline precipitously.

The logic and the empirical analysis reported by Adelino *et al* is suggestive. As discussed below, however, their results are inconsistent with those reported in a contemporaneous research paper based upon the same body of data. Also, notably, it is likely that the average cure rate estimated by Adelino and his associates is optimistic, given the general decline in the macro economy and further weakness in housing markets over the past year. The mortgage modification calculus changes with those lower cure rates and the likelihood that redefault rates may be much lower as a result of more significant reductions in borrower payments.

The final paper selected for the symposium, by Harding, Rosenblatt and Yao (Appendix G), provides an estimate of the size of the external effects of foreclosures on nearby property values. (Those externalities are not taken into account in the default calculus described by Adelino, *et al*.) That estimate is based upon a repeat sales index of house price trends, modified to identify nearby foreclosed properties. The paper provides evidence that foreclosures do have an independent causal effect upon local house values. The results are much more credible than those based upon hedonic house price models in which inferences may depend upon correlation of defaults with the unmeasured or unobservable characteristics of nearby properties and neighborhoods subject to common shocks. But the results of the careful analysis by Harding, *et al*, also provide compelling evidence that the effects of foreclosures are much smaller than those previously described. The authors conclude that "a million additional foreclosures would significantly affect three to five million homes not the forty million that has been estimated using earlier estimates of contagion effects."

The seven papers summarized in this report are based on extensive analyses of micro data. The reversals in the U.S. housing and single-family mortgage markets during the past three years have been manifest in millions of detailed transactions records at the micro level: mortgage loan origination and subsequent payment patterns, purchases of mortgages and their performance in portfolios and securities, and conditions in local housing markets and neighborhoods. That mass of data on individual transactions and outcomes is a fertile ground for analysis of behavior and decision-making by the actors in the housing and mortgage markets: mortgage borrowers and mortgage brokers, lenders, and investors.

That large body of evidence can be used not only to test microeconomic models of behavior, but also to help design programs and policies to make a widespread failure in U.S. housing and mortgage markets less likely in the future. Thus, the papers summarized in this report, by taking a first step toward understanding microeconomic behavior and its consequences during the crisis of the past several years, can inform discussion of options for improving U.S. mortgage markets going forward.

INTRODUCTION

The pace of technical change in the process of originating, distributing, bundling, financing, and servicing single-family mortgages has been phenomenal.[1] In the late 1970s, according to John Weicher (1994), more than half of the mortgages in the United States were originated under the "James Stewart" model. Local banks and thrift institutions mobilized the savings of local households and originated mortgages for those customers. After origination, those same institutions retained the debt and serviced it, collecting payments and guarding against delinquencies.

The rapid pace of technical change since the late 1970s has improved productivity in mortgage origination and distribution—by facilitating the division of labor in a massive way and the pooling of investment risks geographically and demographically. Systematic advances in underwriting and appraisal (e.g., automated valuation models), credit assessment (e.g., credit scoring), aggregation, and distribution (e.g., securitization) have encouraged specialization by function and have facilitated investments in mortgages tailored to the risk preferences of individual investors and financial institutions.

In some large part, statistical and behavioral modeling made the process of specialization more complete. Appraisal by an individual trained in local markets could be augmented, if not completely replaced, by a modern version of "computerized mass appraisal," relying upon simple statistical measures (e.g., hedonic models of house values) and a large sample of house sales, and firms specializing in that phase of the process grew and thrived during the past decade. Time-consuming inquiry into the creditworthiness of borrowers could be augmented, and for the most part replaced, by credit scores based on large samples of individuals, their incomes, their purchasing records and credit card transactions, and their debt repayment. And a number of firms prospered by developing proprietary methods to assess the creditworthiness of mortgage applicants using publicly available records and historical information on the purchasing and payment behavior of those applicants.

Finally, the pricing of pools of mortgages and mortgage-backed securities (MBS) could be informed by sophisticated behavioral models relating household and housing conditions to the likelihood of mortgage termination by prepayment or default. Investment banks employed well-trained analysts to develop models of the behavior of households in exercising options to terminate mortgages, and institutions used that information to design MBS, to price those contracts, and to trade them in world markets.

Advances in technology and technique led to an explosion in mortgage securitization, which reached an aggregate of about $2 trillion during 2006, and ultimately to a large-scale disturbance in the housing market itself. Depending on the precise measure used, U. S. house prices have fallen by somewhere between roughly 11

[1] Mortgages are loans or obligations collateralized by real estate. This report focuses on residential mortgages collateralized by properties with one to four units ("single-family mortgages" or simply "mortgages").

to 30 percent since their peak.[2] Meanwhile, the rate of serious delinquency for subprime mortgages is over 25 percent and the market for private-label MBS has been seriously disrupted.

The rapid pace of technical change and the unchecked increase in the size of the mortgage market may have contributed to that instability. Other factors that may have contributed to the decline include: (1) inadequacies in the institutional structure of the market;[3] (2) inaccuracies in the valuation of real property; (3) flawed assessments of the capacity for loan repayment; (4) poor evaluation of the linkages between economic conditions and mortgage termination; and (5) excessively optimistic "worst-case" scenarios adopted in mortgage pricing.

The division of labor facilitated by technical innovation has surely contributed to the distress in the mortgage market. The replacement of institutions rich in local knowledge—of property, of credit history—by specialists armed with general knowledge opens the way for severe problems of agency. One does not need to invoke fraud and criminality to conclude that the technical advances and the division of labor reduced the incentives for diligence in all aspects of the mortgage process.

We are now beginning to establish what went wrong and what can be done to prevent disaster in the future. To think intelligently about reform requires a clear understanding of the empirical evidence about many aspects of the mortgage process during the past decade. Researchers only now are beginning to develop that body of evidence.

The termination of mortgages after origination is closely related to the values of the options to prepay and to default that are imbedded in mortgage contracts, at least in the U.S. Those values, in turn, depend upon the course of housing prices, interest rates, and borrowers' capacities to repay. Economic and statistical models of those relationships have been widespread for a decade or more. (See Kau and Keenan, 1995, for an early synthesis of the economic model and Deng, *et al* 2000, for some empirics.) But none of that early literature addresses systematic mismeasurement of the key variables affecting termination. In statistical models of termination—probit models (Schwartz and Torous, 1989), duration models (Deng, 1997), and competing risks models (Pennington-Cross and Ho, 2006)—it is simply assumed that the economically important variables are measured without error.

---

[2] The Federal Housing Finance Agency's (FHFA's) "Purchase-Only" House Price Index, which is constructed using sales price data from home mortgages purchased or securitized by Fannie Mae and Freddie Mac, shows a roughly 11 percent decline from its peak in the second quarter of 2007 through the second quarter of 2009. Whereas the FHFA index generally shows the least significant decline among house price indexes, the S&P/Case-Shiller national home price index has generally shown the largest declines, with a measured decline of about 30 percent from its peak in the second quarter of 2006 through the second quarter of 2009.

[3] See, for example, Ashcraft and Schuermann (2007). Though the analysis in that paper nominally pertains to subprime mortgages and their securitization, the basic informational "frictions" discussed frequently have broader scope.

If the key exogenous variables in those statistical models are measured with error, that leads to biased estimates of *all* the parameters in the models. In that case, a theory of the systematic factors leading to measurement error is required to interpret the parameters. But there are few examples anywhere in applied economics where much attention is paid to measurement error in the independent variables. For the most part, recognition of measurement error is confined to the dependent variables, where such errors are less disruptive.[4] But the decline of the mortgage market, and ultimately the housing market, in 2007-2008, brings new attention to the models and measures that were used to predict borrower termination behavior—and, especially, borrower default—and hence to price pools of mortgages and MBS.

Reflecting heightened interest in the models and measures used to evaluate mortgage default risk, Section 1602 of the Housing and Economic Recovery Act of 2008 required the Federal Housing Finance Agency (FHFA) to "conduct a study of ways to improve the overall default risk evaluation used with respect to residential mortgage loans" and to report to Congress on the results of that study. The provision also directed FHFA, in conducting that study, to pay "[p]articular attention … to the development and utilization of processes and technologies that provide a means to standardize the measurement of risk." This report fulfills those requirements. In preparing the report, FHFA has presumed the phrase "processes and technologies" in Section 1602 to refer to the use of data and statistical techniques to measure mortgage default risk, rather than the business processes and information technologies used in mortgage lending.

The body of this report reviews important recent literature on the effects of mismeasurement of default risk on mortgage outcomes and pays some attention as well to research on the externalities that result from foreclosures on delinquent mortgage borrowers. To aid in the preparation of the report, FHFA and the Federal Deposit Insurance Corporation's Center for Financial Research jointly selected seven papers for a public symposium held in September 2009. The report summarizes those papers in the context of previous research and of the comments provided by discussants at that symposium. The appendices to the report provide the papers themselves. The remaining chapters of the report address the following topics:

- **The Value of Collatera**l. The report examines the technologies employed, the incentives of the various actors, and the empirical evidence showing the effects of mismeasurement of the value of homes that collateralize single-family mortgages.

- **Borrower Capacity for Repaymen**t. The report discusses the fundamental variables used in models of borrower payment capacity and behavior (e.g., credit score, income ratios) and the basic flaws of the models used during the recent boom period. Key research papers discussing the genesis and performance of reduced-documentation

[4] Indeed, in linear models at least, simple errors of measurement affect only the efficiency with which parameters are estimated.

loans are evaluated in depth. The widespread acceptance of low and no-documentation loans was one of the hallmarks of the boom period. The papers reviewed in the report provide clear evidence that the true risks associated with "liar loans" tended to be inadequately measured.

- **Securitization and Loan Credit Quality.** Extensive recent discussions have addressed the growth of and problems (theoretically) caused by mortgage securitization during the mortgage credit boom, particularly for subprime mortgages. The report addresses the evolving incentives of mortgage market actors, particularly those in different origination channels and summarizes recent (potentially telling) empirical evidence quantifying differences in loan performance across mortgage types and origination channels.

- **The Effects of Lender Options.** In evaluating the overall risk associated with mortgage lending, it is important for lenders to understand how institutional factors, including lender recourse options, affect the risk of borrower default. In addition, policy-makers can benefit (particularly in the current environment) from understanding the extent to which loan securitization may be hindering foreclosure prevention efforts. The report summarizes important recent literature addressing those topics.

- **Externalities of Mortgage Foreclosures.** The report considers research that seeks to measure the effects on property values of mortgages that terminate in default and foreclosure. The spillover externalities associated with home foreclosures—lower prices for adjacent homes—is an important issue in discussions of policies related to foreclosure prevention. Because foreclosure-caused price declines can themselves lead to additional home foreclosures, measurement of the spillover effect is also important to precise measurement of mortgage default risk for individual properties. For both of those reasons, the report summarizes the findings of a recent, very careful attempt to measure precisely the scope and magnitude of the externalities of foreclosures.

Inasmuch as the proper evaluation of mortgage risk is difficult even in the most stable housing markets, the symposium sponsored by FHFA and the FDIC, at which the papers summarized in this report were presented, elicited lively debate on virtually all topics, including the matters specifically addressed in the presented research. This report summarizes the comments and criticisms provided by paper discussants at the symposium in order to provide a more complete review of the current state of thinking on the relevant research topics.

THE VALUE OF COLLATERAL

Analyses of the relationship between "the value" of a house and an estimate of its value have a long history. Because houses are traded so infrequently, methods of imputation are necessary for many practical purposes (e.g., property tax assessment) and academic exercises (e.g., wealth estimates for the household sector). Early studies compared the properties of owners' estimates of value with appraisals (Kain and Quigley, 1972), largely because the Bureau of the Census relies upon owners' estimates for the valuation of single-family housing (Bucks and Pence, 2006, provide recent evidence).

The 1980s saw the first systematic attention to regression-based property valuation as a vehicle for imputing values to unsold properties. Publications such as the *Property Tax Journal* and the *Assessment Journal* publicized the application of statistical models to property valuation through the 1990s.

For the most part, those statistical models were assessed by their ability to forecast out-of-sample selling prices of houses. A theoretical paper by Shiller and Weiss (1999) illustrated the application of statistical or "Automated Valuation" models (AVMs) to mortgage underwriting and to the limitation of mortgage portfolio losses due to default. The key insight of the Shiller-Weiss analysis was the explicit linkage between lender profitability and the loan-to-value (LTV) ratios of loans, permitting the probability distribution for an appraisal to be summarized by the profitability of the underlying loan. Those ideas were quickly incorporated into the practical world of mortgage underwriting (see, for example, Elizabeth Mays' *Handbook of Credit Security*, 2001), and a variety of competing proprietary real estate valuation systems flourished.

Prior to Austin Kelly's paper[5] (see Appendix A), the only publicly available empirical application of the Shiller-Weiss theory was in work reported by LaCour-Little and Malpezzi (2003) using data supplied by a savings and loan institution in Alaska. LaCour-Little and Malpezzi computed the default hazards from micro data on a few hundred mortgages originated by that institution, relating default probabilities and losses given default to the extent to which the appraised value at origination deviated from a regression-based estimate. The results clearly indicated that there were costs to the lender—increased mortgage losses—as a result of lower-quality appraisals of collateral. Decreased appraisal quality (that is, over-appraisal of a property relative to a hedonic-based estimate) was significantly related to default, while under-appraisal had little effect. The authors also found that errors resulting from estimating collateral values by interpolation using a price index for the state (the Office of Federal Housing Enterprise Oversight [OFHEO] index[6] for Alaska) were quite large. The authors provided some sensible prescriptions for lenders, suggesting that the latter may want to compare appraisal information with statistically-based measures in a routine way.

[5] Kelly, Austin, "Appraisals, Automated Valuation Models, and Mortgage Default".

[6] Until July 2008, OFHEO, a predecessor agency to FHFA, published the house price index now released by FHFA and described in footnote two on page nine above.

The essential tradeoff between AVM estimates and traditional appraisals is well known. It is hard to pervert AVM methods merely to "confirm" the value estimates of market participants, but it is also hard to value aesthetic attributes of properties using AVM models. The comparison of methods is an empirical issue for which there is little prior evidence. Thus, the Kelly paper substantially increases our understanding of the link between appraisal systems and mortgage outcomes. Kelly analyzes a national sample of transactions about twenty times as large as that available to LaCour-Little and Malpezzi, as well as a sample for one city, Atlanta, that is about four times the size of the Alaska sample. The research design is based upon a sample of mortgages insured by the Federal Housing Administration (FHA) and originated between October 1999 and September 2002 whose performance was observed through June 2005. That sample was merged with data from the Department of Housing and Urban Development's (HUD's) Single-Family Data Warehouse indicating the disposition of each loans—the date of prepayment, delinquency or termination, or its survival until the end of the period. Most of the contracts had LTV ratios close to one hundred percent, and about 92 percent were fixed-rate, fixed-term loans. Four out of five loans had terminated by June 2005. (Recall the dip in mortgage interest rates in 2003, which made prepayment especially attractive.) Additional time-varying information—the course of interest rates, state unemployment rates, and state house prices—was appended.

Importantly, Kelly appended a variable representing a model-based appraisal for the date of origination where possible.[7] That match was possible for about eighty percent of the sample of 5,000 records. Beyond an estimate of market value, the AVM produced an estimate of the probability that the reported appraisal was within ten percent of the "true value" of the property. That is a measure of the "confidence" in the point estimate of valuation produced by the AVM. Altogether, the dataset described the initial conditions of each mortgage loan, including an appraised value, an AVM estimate of value and its confidence, time-varying factors in the economy, and the conditions of loan termination.

Kelly's paper reports a detailed analysis of the hazard of mortgage delinquency using those data. The analysis is based upon two basic specifications of the hazard relationship. One relies upon a well-known 2001 study by the Government Accountability Office (GAO) to aggregate the independent variables in the hazard model.[8] The second specification does not include the GAO-defined transformation of variables, but includes the LTV ratio directly, as well as measures of state house price appreciation. In both specifications, careful attention is paid to estimating baseline hazards and to accounting for the unobserved heterogeneity of borrowers.

The key variable in these duration-to-delinquency models is a measure of the deviation of the AVM estimate of value from the observed transaction price of the

---

[7] The AVM employed was actually a "cascade" of AVMs supplied by First American Loan Performance. The AVM system used public records data and a series of valuation models (hedonic, hybrid, and repeat sales) to estimate home values.

[8] The model uses the coefficients from the competing risks model estimated in 2001 to aggregate a large number of measures into a single "GAO risk" of default measure (GAO 2001).

property. For the national sample, the results suggest that that measure of imprecision and error in the initial valuation of properties at the time of mortgage origination is large and highly significant. Where the AVM estimate of home value is low relative to the transaction value (sales price), the duration model suggests a higher risk of subsequent delinquency. Similarly, the measure of "confidence" in the precision of the AVM estimate is inversely related to the risk of delinquency.

Evidently, properties that are more difficult to value are also those with higher delinquencies. That may reflect fewer transactions (and fewer "comparables") in certain neighborhoods or areas, or else the more idiosyncratic characteristics of some specific properties. Given the "black-box" nature of the AVM system, it is not possible to say much more.

The results for the single city are not as clear-cut, perhaps because the sample sizes are smaller by 75 percent. But the pattern of signs and the magnitudes of the coefficients for duration-to-delinquency models from the Atlanta sample are consistent with the results from the national sample of properties. Holding constant the other important determinants of mortgage delinquency—credit score, LTV ratio, etc.—erroneous and less-precise measures of the initial values of properties are associated with higher delinquencies and higher FHA claims. The quality of the initial estimate of collateral value is related systematically to the probability of subsequent delinquency and to the probability of an FHA claim. In fact, both measures of the initial quality of collateral, the professional appraisal and the AVM-estimate of value, provide complementary information about default propensities, as demonstrated for different samples of transactions histories.

Of course, the ratio of appraisal price to the transactions price in the FHA data analyzed by Kelly is truncated. An AVM estimate could not be obtained for more than a fifth of the sales, presumably those hardest to value using only regression methods. More importantly, transactions on properties with appraised values much below the value estimates by market participants are seldom recorded at all.[9]

As pointed out in the discussion by Michael LaCour-Little, the use of FHA mortgage origination data in the Kelly study makes the test of the importance of those appraisals on outcomes quite conservative. With more variation in LTV ratios at origination, there would have been more opportunity for variations in appraisals to affect outcomes. But more importantly, the application of Kelly's methodology to the refinance market could indicate a much greater value for AVM methods. It would be important to know whether the increases in cash-out refinancing and the subsequent default surge after 2006 are systematically related to lower appraisal quality and inaccurate assessments of borrower equity at the time the new mortgages were originated. The methodology employed by Kelly could be adapted to those circumstances.

---

[9] An important study of loan applications found that a low appraisal is one of the most important factors causing a loan application to be denied (see Green and LaCour-Little, 1998).

THE BORROWER'S CAPACITY FOR REPAYMENT

Historically, the capacity of mortgage borrowers to repay debt was established by proof of income (tax returns, W-2 forms, verification of employment, and paychecks) and evidence of assets.[10] The diffusion of credit scoring and the advantages of price discrimination (to borrowers as well as lenders) led to relaxation of those stringent standards, presumably in return for variations in the prices of mortgages. Work by Fortowsky and LaCour-Little (2002) suggested how and why variations in the documentation of borrower repayment capacity affected mortgage pricing. More recently, Chomsisengphet and Pennington-Cross (2006) reported a variety of factors that were used to define up to six "loan grades" by a particular mortgage lender. The combination of the LTV ratio of a loan and the credit score of the borrower determined the basic lending rate. In addition, applicants in lower loan grades paid an additional premium in interest rates. Presumably those characteristics were related to the default and loss information maintained by lenders.

Empirical work by Pennington-Cross and Ho (2006) provided some information on the implications of relaxed underwriting standards on loan losses. Pennington-Cross and Ho analyzed securitized loans from the Loan Performance Asset-Backed Securities database for 1998-2005. About 19 percent of a sample of hybrid loans and 38 percent of a sample of fixed-rate loans provided "limited or no documentation or down-payment sources." Their empirical findings documented the large effects of "limited documentation" on mortgage default. Their results suggested that the probability of default for limited documentation loans was higher by 47 percent for fixed-rate mortgages and by 43 percent for hybrid loans. Surprisingly, Pennington-Cross and Ho did not discuss (or even note) that finding.

The paper by LaCour-Little and Yang[11] (see Appendix B) provides more information on the behavior of borrowers with "no documentation" loans. Further, the paper provides almost the only evidence available to date on the importance of the "exaggeration" of income on mortgage outcomes. Finally, the authors provide some concrete suggestions about how lenders could reduce default risk in stated-income mortgages. LaCour-Little and Yang analyze a sample of some 218,000 mortgages originated and securitized by Bear Stearns and their affiliates during 2000-2007. For that sample, the authors have complete information about the payment types and other characteristics of the mortgages. They also analyze a subsample of the data—about 134,000 loans—for which it is possible to infer the income of the borrower.

LaCour-Little and Yang concentrate on comparisons of the behavior of loans made with full documentation of the income and assets of the borrowers ("full doc") with

[10] Andreas Lehnert, Federal Reserve Board economist and discussant at the September symposium, noted that documentation of income and assets is hardly a guarantee of the future repayment capacity of a borrower. Prior research has shown that even sophisticated econometric models of life-cycle income using detailed data are characterized by large unexplained residuals in medium-term forecasts of income (e.g., Feigenbaum and Li, 2008).

[11] LaCour-Little, Michael, and Jing Yang, "Taking the Lie out of Liar Loans".

loans made to borrowers who stated their incomes and assets, where those statements were not verified or documented ("stated doc"). In the Bear Stearns sample, "stated doc" loans were a negligible fraction of the sample in 2000, but by 2007 were about 46 percent. In further contrast, "stated doc" loans had substantially lower LTV ratios in 2000 (56.5 percent compared to 76.7 percent for "full doc" loans). By 2007, the average LTV ratio for "stated doc" loans was 73.1 percent, whereas the average LTV ratio for "full doc" loans was 75.3 percent.

The authors provide a rich description of the two categories of loans—by LTV ratio, credit score, and loan balance—over time. Of particular interest is an "inferred income ratio." The authors were able to calculate the borrower's income for about 134,000 mortgages in the Bear Stearns sample. The "inferred income ratio" is merely the ratio of that calculated income to the metropolitan average. When defaults are related to the inferred income ratio, the results are startling. For roughly three quarters of the "stated doc" loans, the income reported by the borrower was less than 150 percent of the metropolitan mean income. The default rate for that group of borrowers was 18.1 percent. For the other 25 percent of the "stated doc" loans, the default rate was about 31.4 percent (computed from Table 6). That presentation alone suggests that the euphemism of "liars' loans" for "stated doc loans" is apt.

The authors estimate multivariate models predicting the characteristics of loans that are more likely to be "full doc" or "stated doc." Loans are far more likely to be based upon "stated doc" when they were written as adjustable-rate contracts or when they included other alternative mortgage characteristics.

The results reported by LaCour-Little and Yang are consistent with those reported by Gerardi, *et al* (2009) using data from the First American CoreLogic database. Gerardi and his associates found that the typical subprime mortgage written in 2005 with low documentation was almost fifty percent more likely to default in the first year than were fully documented loans.

The estimation of logit models by LaCour-Little and Yang predicting the probability of default confirms the riskiness of "stated doc" loans reported by others. In four specifications of the default model (including LTV ratio, credit score, type of loan, loan balance, etc.), "full doc" loans are found to have default propensities much lower than for "no doc" mortgages, while "stated doc" loans have default propensities substantially larger than for "no doc" loans. The estimated differences by type of documentation are statistically significant and are material in magnitude.

Beyond that, the authors estimate a series of logit models of default including their constructed ratio—the ratio of borrower income inferred from the loan application to the metropolitan average income. The models are estimated in a number of variants to explore different sets of controls and to test for robustness. In all of those regressions, the ratio variable is highly significant, but the magnitude of the coefficient is roughly half-again as large for "stated doc" loans.

It is hard to escape the conclusion that stated-income loans facilitate the exaggeration of income and the subsequent increase in default propensities. At the point of means, the average loan would have a 51 percent higher default rate (24 percent instead of 16 percent) if it were presented as "stated doc" rather than "full doc."

The authors conduct a series of simulations in which the ratio of stated income to metropolitan average income is limited, thereby reducing the default propensities of "stated doc" loans. On that basis, they suggest that the ratio be limited and conclude, "Given these results, we think stated income lending can be a viable product."

Despite the careful and interesting work in this analysis, the conclusion appears dubious. If it were known that stated incomes above some limit were disqualifying in some markets, wouldn't clever "liars" simply state their incomes to be just below that limit?[12] Of course, a straightforward alternative to the institution of "stated doc" loans would be loan underwriting on the basis of full documentation (using two years or some other conventional period) of income, with any exceptions to debt-to-income guidelines made explicitly and documented as a matter of judgment by the underwriter.

It appears that it might be possible to use the LaCour-Little and Yang data to calculate the risk premiums charged by the lender for "stated doc" loans of various kinds (from the interest rate on the initial mortgage, not used in the analysis presented in the paper). If so, it would facilitate a more complete analysis of the costs and returns to the institution of mortgage lending on the basis of stated incomes and stated assets.

The paper by Jiang, *et al*[13] (see Appendix C) extends the analysis of low-documentation loans ("stated doc" and the like) to address the channels of origination as well as the credit quality of the loans. The authors analyze almost three quarters of a million mortgages written by a single bank, both "full doc" and "low doc" loans, that were originated by the bank itself or by an outside broker that sold the loans to the bank. The analysis highlights the differences between "full doc" and "low doc" loans, confirming and extending the analysis of LaCour-Little and Yang. But Jiang, *et al*, also analyze the alignment of incentives in the origination of mortgages between brokers and lenders and the potential for conflicting incentives between the originating bank and the financial institutions that purchase and securitize the loans.

The paper provides a comprehensive analysis of mortgage delinquency based upon a very large sample of loans, including information on contract terms, property characteristics, and the characteristics of borrowers, including extensive demographic information. The records on the loans begin in January 2004, and the performance of the loans is recorded up to early 2009. From the addresses of the properties, it was also possible to match local and neighborhood information to the loan records.

---

[12] As suggested in the paper by Keys, *et al* (2008) described below, there are many ways borrowers or originators could "work around" arbitrary limits, especially if they were "working together."

[13] Jiang, Wei, Ashlyn Nelson, and Edward Vytlacil, "Effects of Origination Channel and Information Falsification on Mortgage Delinquency."

As with the Bear Stearns data analyzed by LaCour-Little and Yang, the data analyzed by Jiang, *et al*, show a rapid expansion in low-documentation loans in the middle years of this decade. The Jiang, *et al*, analysis also indicates that a great deal of that expansion in lending was originated by brokers. By 2006, loans originated by brokers accounted for 94 percent of raw production, and “low doc” broker-originated loans accounted for 75 percent of all new production for this institution.

For loans classified by origination channel (bank, broker) and documentation type (“full doc”, “low doc”) the authors investigate mortgage termination using probit and hazard models. The sample sizes for those statistical analyses range between 31,000 and 425,000 observations. The differences between the implications of the models for those initiated by brokers and those initiated by the bank are striking, to say the least. The difference between the estimated delinquency probabilities from the probit models is more than ten percentage points; loans originated by brokers are far more risky. From the hazard specifications, it appears that, for “full doc” loans, the median predicted time to failure (delinquency) is roughly three times greater for bank-originated mortgages than for broker-initiated loans. For “low doc” loans, the median estimated duration of bank-initiated loans is about double that of broker-initiated loans.

In addition, the explanatory power of the statistical models for the “full doc” mortgages is substantially greater than those estimated for “low doc” loans, suggesting that there is “measurement error” in the variables used to predict termination of the latter. That measurement error could arise from simple falsification of the self-reported information. For the “full doc” loans, a higher income is associated with a lower propensity for delinquency and a longer time-to-failure. For “low doc” loans, the effect of income is either insignificant, or simply perverse.

A variable measuring the ratio of borrower income to that in the neighborhood (the postal code) is a more precise measure of the same concept employed by LaCour-Little and Yang at the metropolitan level. The comparisons and their implications are similar. In the “full doc” samples, borrower incomes are 3.6 and 3.3 times the local benchmark. In the “low doc” samples they average 4.3 and 3.8 times the neighborhood income. That suggests that variables are erroneously reported or falsified in the “low doc” samples of mortgages.

That is consistent with the findings when the models are estimated separately for samples of correspondent brokers and non-correspondents. The statistical results for the correspondent brokers are consistently closer to those of bank-originated loans than are those for the non-correspondent brokers. That similarity in the patterns of estimated parameters suggests a closer alignment of the incentives for “accuracy” for institutions that have ongoing relationships.

Jiang, *et al,* also analyze the choices of loan origination channel and documentation level made by borrowers. As demonstrated by LaCour-Little and Yang using Bear Stearns data, the choice of “low doc” loans is not associated with inexperience or irregular borrower conditions, but is rather associated with applications that "look good on paper" (p. 14) in terms of credit score, LTV ratio, etc.

But Jiang, *et al*, are also able to analyze the choice between a broker and a banker as the mortgage originator. Here, it seems quite clear that inexperienced or unsophisticated borrowers are far more likely to obtain mortgages through a broker. Borrowers with high debt, low income, or low credit score or who are young, female, or belong to a minority group are all systematically more likely to employ a broker rather than to apply and obtain their mortgages directly from the bank. Importantly, the channel of mortgage origination varies systematically over the typology of urban neighborhoods. The comparisons indicate quite clearly that broker originations are more likely in neighborhoods (census tracts) that have a smaller fraction of black and Hispanic households and younger populations. Thus, it appears from the analysis that minority households in non-minority census tracts are those most likely to employ mortgage brokers.

The authors attempt to sort out the effects on delinquency of the observable information about loans, borrowers, and properties, by channel of origination and by documentation requirements. Specifically, the authors employ the standard Oaxaca decomposition of effects to investigate the extent to which the higher delinquency rates of broker-initiated (and "low doc") loans are predictable from observables. The authors partition the differences in outcomes into that part attributable to differences in the distribution of the measured covariates in the samples and that part attributable to differences in the estimated coefficients, assuming the distribution of measured covariates is the same in the samples.

The comparison between delinquencies conditional upon origination channel suggests that most of the five-to-eight percentage point difference in delinquencies arises from differences in unobservable (or more precisely, unmeasured) characteristics. In sharp contrast, the comparison between broker and bank loans, conditional upon documentation, suggests that if brokers and banks had lent to borrowers with identical measured characteristics, a bit more than half of the difference in delinquency rates would have disappeared. The conclusion: "low doc" mortgages have a higher delinquency rate because borrowers are less carefully screened along hard-to-quantify characteristics ("unobservables"), while the broker origination channel has a higher delinquency rate because the characteristics of the borrowers and their properties ("observables") are also less creditworthy. Broker loans were made to those with higher credit risks based on measured and observable characteristics.

In the discussion of the two papers on "liars' loans" at the September symposium, discussants Andreas Lehnert and Robert Van Order both stressed regulatory reforms already enacted that became fully effective after September 30, 2009. The Housing and Economic Recovery Act of 2008 (P.L. 110-289) included several changes in the Truth in Lending Act (15 U.S.C. 1601 *et seq*) that will affect the position of "no doc" and "low doc" mortgages in the future. As of October 2009, if the interest rate on any particular loan exceeds the average prime offer rate for mortgages by 1.5 percent, then lenders *must* consider the borrower's ability to pay from sources other than the collateral itself. Specifically, for those high-interest-rate loans, lenders are *required* to verify the income and assets that they have relied upon to determine the borrower's capacity to repay.

SECURITIZATION AND LOAN CREDIT QUALITY

Notwithstanding the results about securitization, based on its timing relative to origination, reported by Jiang, *et al*, there is widespread concern that the boom in credit availability during the past decade and the increased securitization of mortgages led to declines in the standards for lending and in the screening of credit risks. For example, Dell'Ariccia, Igan, and Laeven (2008) analyzed micro data collected pursuant to the Home Mortgage Disclosure Act of 1975 (HMDA) for 2000 through 2006, establishing the extent to which increased mortgage delinquencies could be explained by the application of lending standards rather than changes in economic fundamentals. The authors analyzed variations in application denial rates and loan-to-income ratios for HMDA applicants aggregated to the metropolitan level, finding evidence of declines in lending standards at the times and in the places where the largest increases in loan applications occurred. They also found evidence that declines in standards were larger with increases in the number of lenders. Importantly, as it became easier to securitize mortgages, the behavior of lenders was affected, as lending standards declined more in areas with higher mortgage securitization rates. Among other changes, higher securitization rates were associated with higher LTV ratios; higher LTV ratios have been clearly associated with higher delinquencies and foreclosures (see especially Demyanyk and Van Hemert, 2009).

Evidence of a more direct link between increased securitization and decreased standards for screening loans is provided by Keys, Mukherjee, Seru and Vig (2010). The authors develop an identification strategy to focus on individual loans in the First American Loan Performance database. The authors take advantage of a rule of thumb, codified in a variety of guidelines issued by Fannie Mae and Freddie Mac, which implied that the securitization of a loan is much more likely if a threshold credit score provided by Fair, Isaac, and Co. (FICO) has been awarded. Because those scores cannot easily be manipulated by potential borrowers (in the short run at least), the analysis is concentrated on individual ("low doc") loans in the neighborhood of the threshold, a credit score of 620. Assuming that creditworthiness is constant in the neighborhood of the threshold, that other characteristics are not systematically related to the cutoff, and that additional screening is costly, the authors analyze the natural experiment in which the identical loan is scored at $[620-\varepsilon]$ or $[620+\varepsilon]$, where ε is very small.

Keys, *et al*, analyze the composition and performance of lenders' portfolios in the neighborhood of the threshold, finding that a portfolio that is more likely to be securitized has a default rate that is between about 10 percent and 25 percent higher than an otherwise identical unsecuritized portfolio. Those estimated differences are robust to alternative specifications and tests for selectivity. The authors also test the response to state laws that made securitization more costly and, thus, would have made the FICO score threshold irrelevant. When those laws were in effect, the default rates on the [+ε] side of the threshold were actually significantly lower than those on the [-ε] side. The implication of that careful work is that, when a loan can be sold easily to a securitizer, lenders will economize on screening costs, and screening becomes more lax.

However, a major problem in interpreting those results arises from considering the source of the data, the Loan Performance database. The data come from a population that consists entirely of securitized mortgages. Thus, the finding that [620-ε] loans outperform [620+ε] loans is also consistent with extra care exercised by the securitizers to insure that loans that could be questioned are indeed worthy enough to be securitized.

The paper by Elul[14] (see Appendix D) provides a more direct test of that general proposition by relying upon a loan-level dataset from LPS Analytics that covers almost 75 percent of the mortgage market between 2003 and 2007. The LPS dataset includes both loans sold to investors and securitizers and loans kept in the portfolio of the lender. Thus, the dataset includes micro data on individual loans, as does Jiang, *et al*, but it includes loans from a large number of lenders (without, however, being able to identify lenders individually).

Elul concentrates on four products (thirty-year fixed-rate loans, and three types of adjustable-rate mortgages), and he considers prime and sub-prime loans separately. He estimates hazard models of default for the large samples of LPS mortgages, sometimes restricting his sample to mortgages in the 25 largest metropolitan areas. The statistical models are based on millions of observations on payments.

For *prime fixed-rate* mortgages, the results indicate that broker-originated and "low doc" loans are riskier, holding credit score and LTV ratio constant. The results also show (confirming the results of Dell'Ariccia, *et al*) that loans originated in later years are riskier. The pattern of the dummy variables measuring securitization by year is important—there is a monotonic increase in the magnitude of the coefficients. *Ceteris paribus*, securitized mortgages went from being less risky than unsecuritized mortgages in 2003 to much riskier four years later.

For *prime adjustable-rate* mortgages, the results also suggest that securitized loans are riskier. That is true for all products, even those (three- and five-year hybrid ARMs) that lenders were much more likely to hold in portfolio.

For *sub-prime* loans, there seems to be no systematic relationship between securitization and increased default risk. That is true for both fixed-rate mortgages and for the adjustable-rate products. That result is quite inconsistent with those reported by Keys, *et al* (2008).

---

[14] Elul, Ronel, "Securitization and Mortgage Default: Reputation vs. Adverse Selection," Federal Reserve Bank of Philadelphia Working Paper 09-21.

As pointed out by Elul and discussant Anthony Sanders, the finding that subprime in-portfolio loans did not perform better than securitized subprime loans suggests two factors at work. First, subprime lenders' reliance on securitization made policies of "cherry picking" more risky for them. Second, investors were more careful in purchasing securities backed by loans that were defined as "subprime." Cherry picking was simply more difficult in the subprime market because it was difficult to distinguish households defined as subprime as a result of random events (e.g., medical expenditures), as an inevitable consequence of their professions (e.g., small-time proprietors), or as a result of their negligence or lack of financial discipline. But that difficulty also plagued investors investing in securitized subprime loans. The original lender would have had at least as good information as a potential investor about the reasons underlying any borrower's classification as a subprime credit risk. And the attractive yield spread on subprime securitization may have simply discouraged detailed investigation of loans in large pools.

The general finding—that securitization increases the likelihood that mortgages will be available to borrowers who are worse credit risks but have the same easily observable characteristics—highlights the distinction between "hard" and "soft" information and the role of that information in assessing credit risk. With a routinized channel to sell mortgages based upon "hard" data that are easily verifiable (such as FICO scores), lenders have weaker incentives to invest resources to uncover "soft" information about credit-worthiness (for example, employment prospects). That means that "soft" information will be under-supplied, benefiting some borrowers and mortgage holders, but harming others. Over time, if the extent of securitization based only on "hard" information increases, more of those whose "soft" information makes them less creditworthy will nevertheless obtain mortgages.

A provocative paper by Rajan, Serat, and Vig (2009) formalizes that logic and traces through its implications for statistical models of default. From that perspective, statistical models of default estimated in the period before the widespread diffusion of securitization will under-predict defaults during subsequent periods of high securitization, even if the models are based upon a consistent set of well-measured information. That is because the models will under-predict defaults for those borrowers for whom "soft" information is more valuable. The authors suggest that "the assessment of risk must be extra conservative" during periods in which market practices have changed sharply (i.e., during "regime changes").

THE EFFECTS OF LENDER OPTIONS

Recourse to Defaulted Borrowers

The incentives for default by borrowers depend not only on their equity positions in their properties and their repayment capacities, but also upon the regulations that govern the disposition of assets in the event of termination. The equity position of borrowers defines the instantaneous value of the default option. A negative equity position is a precondition for rational default, but it is not sufficient. That is so because by exercising the default option today, the homeowner foregoes the option to prepay or to default on the contract in the future. Thus, if the borrower has the capacity for mortgage repayment, she may rationally choose not to default in the face of negative equity. Indeed, some theories of mortgage default hold, and some empirical models find, that default typically occurs only when the borrower has negative equity and also suffers some "adverse life event" (see Vandell, 1995). Recent survey evidence from U.S. households, gathered in March 2009, suggests further that moral and social constraints play an important role in inhibiting "strategic default" by borrowers with negative equity (see Guiso, Sapienza, and Zingales, 2009).

The regulatory environment may also greatly affect the incentives for default. Depending upon national laws governing bankruptcy and state laws governing the recourse available to lenders, the borrower may suffer more than the loss of reputation in the event of default, and that may affect the calculus of default by borrowers. In the extreme, many European countries—e.g. the Netherlands—consider mortgage debts as personal obligations with full recourse. In other countries—e.g. Finland—those personal obligations may not even be discharged in bankruptcy. In the U.S. and Canada, those customs vary by state, and it has been observed that default rates vary systematically by state (Deng, *et al* 2000); that default rates are higher in Canadian provinces that do not permit deficiency judgments (Jones, 1993); and that mortgage loan amounts are smaller in states that permit deficiency judgments (Pence 2006).

The paper by Ghent and Kudlyak[15] (see Appendix E) provides a systematic analysis of the link between the availability of recourse by lenders across states and the default experience of borrowers in those states. An important contribution of the paper is the assembly of systematic information on the laws and practices of the states: the incidence of judicial foreclosures, the availability of lender recourse, and the elapsed time for an uncontested foreclosure.

The paper makes two analytical contributions. First, it presents a simple bargaining model of lender and borrower behavior to understand the role of recourse in affecting the behavior of homeowners with negative equity. Second, it relies upon the payment histories of a large sample of mortgages to estimate the quantitative importance of deficiency judgments in affecting the default behavior of borrowers.

The bargaining model is a straightforward description of the economic incentives that may cause borrowers and lenders to agree to a short sale, to agree to a deed in lieu of foreclosure, or that may cause a foreclosure (with a deficiency judgment against the borrower in states where this is permitted). In general, a borrower gains if she contests a foreclosure, slows the process, and, thus, lives longer in the house without paying rent. A borrower incurs lower costs if she agrees to a short sale or a deed in lieu of foreclosure than if her property is foreclosed. At foreclosure, the lender recovers less than the market

[15] Ghent, Andra C., and Marianna Kudlyak, "Recourse and Residential Mortgage Default: Theory and Evidence from US States," Federal Reserve Bank of Richmond Working Paper 09-10, available online at http://www.richmondfed.org/publications/research/working_papers/2009/pdf/wp09-10.pdf.

value of the property, recovering a larger fraction if he agrees to a deed in lieu of foreclosure than he would recover at a foreclosure sale. If the lender seeks a deficiency judgment after foreclosure, the borrower receives credit for the "fair market value" of the home.

The authors formalize those payoffs to borrowers and lenders under recourse and non-recourse. A comparison of the payoffs establishes that borrowers are more willing to agree to a short sale or a deed in lieu of foreclosure in those states that permit recourse. Lenders will also be less willing to accept either alternative to a foreclosure in states that permit recourse. The lender always prefers either alternative to a foreclosure in non-recourse states. In recourse states, the lender will foreclose if the deficiency recovery is large enough.

The possibility of a deficiency judgment will certainly make the borrower more likely to agree to a short sale or to a deed in lieu of foreclosure. The lender would prefer either of those outcomes to a foreclosure if the expected proceeds from that outcome exceeded the expected net proceeds from foreclosure. The expected net receipts from a foreclosure would include any expected recovery from a deficiency judgment, where that amount would equal the difference between the outstanding loan balance and the fair market value of the property.

The outcome of that bargaining scenario depends on the magnitudes involved, and the authors simulate payoffs and the implied strategies under a number of plausible conditions—variations in the time to be default, the fraction recovered by the lender, the initial LTV ratio—under recourse and non-recourse conditions. The simulations indicate that the presence of recourse will deter relatively well-off borrowers from strategic default. Recourse will also cause borrowers to default in ways that lower costs for lenders.

The importance of recourse and deficiency judgments is not really reflected in the *incidence* of deficiency judgments, but rather in the effect of the threat of recourse on borrower behavior and lender response. Even if lenders seldom (or never) pursue deficiency judgments in court, losses are lower when the threat of recourse can be exercised credibly.

The bargaining model of Ghent and Kudlyak can be criticized on a number of grounds. As noted by discussant Brent Ambrose, the bargaining game adopted by the authors makes no allowance for a borrower to "cure" a default (i.e., to stop the foreclosure process by paying all arrears). In an option-theoretic model, Ambrose and Buttimer (2000) have previously shown that option to be quantitatively important. More generally, the option-theoretic perspective recognizes that prepayment and default are interrelated. In a more general model, it can be demonstrated that the prepayment probability increases with the probability that a deficiency judgment will be enforced.

A more powerful criticism of the Ghent-Kudlyak model as presented is that all conclusions are drawn from a base case based on particular (and perhaps arbitrary) parameters. Beyond that special case, no information is presented about the existence of an equilibrium or the pure strategies employed by the players. For example, as noted by Ambrose, it would appear to make sense for a lender to pursue deficiency judgments only on some subset of borrowers.

Ghent and Kudlyak estimate the importance of the availability of recourse by lenders on the default behavior of borrowers using a large sample of data on the monthly payment decisions of almost three million borrowers. The authors estimate a default model that relates an estimate of the value of the default option, a variety of mortgage characteristics, and local (state) variables to the profitability of default, using a probit specification that also employs the age of the mortgage. The model also employs the spread between the current mortgage rate and the contract rate as a proxy for the value of the prepayment option.

The empirical analysis confirms that interest-only and adjustable-rate mortgages are riskier, as are second liens, high-LTV-ratio loans, and unseasoned mortgages. The statistical results indicate quite clearly that the effect of the value of the default option on the profitability of exercise is significantly different in recourse and non-recourse states. Recourse clearly decreases the impact of negative equity on the probability of exercising the default option. That effect declines as the value of the option increases.

The evidence indicates that the definition of recourse employed does not merely measure the many subtle differences between recourse states (like Texas) and non-recourse states. The analysis compares the results of the default model when state-specific variables are included to account for idiosyncratic differences among states and their populations. When state-specific variables are included in the model, the measure of recourse (which varies only among states) is still highly significant and economically important.

The specification of the model is such that the increased default in non-recourse states is highly non-linear; it depends upon the value of the default option. But on average it is a big number—about 21 percent at the mean. It also varies substantially by the value of the collateral at origination. For home values between $300,000 and $500,000, non-recourse states have default rates about 60 percent higher than for recourse states. By contrast, for homes valued between $500,000 and $750,000, non-recourse states' default rates are almost double those of recourse states.

Further analysis establishes that recourse has no effect upon the default probabilities for mortgages financed with securities guaranteed by the Government National Mortgage Association (Ginnie Mae) or loans acquired by Fannie Mae or Freddie Mac (which do not pursue deficiencies at a matter of course), but does affect defaults for privately-securitized and portfolio mortgages whose owners can and do (or at least credibly threaten to) sue for recourse. Recourse reduces mortgage terminations by default.

The analysis establishes the importance of the threat of enforcement of deficiency judgments upon outcomes in the mortgage market—the extent of default, the type of mortgage termination, and the portfolios in which default is more likely. Those effects are large, and they persist even if the incidence of litigation to enforce deficiency judgments is neither large nor widespread.

The theoretical and the empirical analysis by Ghent and Kudlyak is static. It analyzes the one-shot game between borrowers and lenders. It assumes that neither party forecasts changes in economic conditions, house values, or repayment capacity, in a systematic way. But the analysis also assumes that borrowers and lenders cannot agree to modify the terms of the mortgage in response to information revealed after the mortgage contract is signed. However, with reasonable transactions costs, the parties could agree to a loan modification that would make both better off. But they do not in the models presented by Ghent and Kudlyak, and that is why the threat of deficiency judgments matters.

Mortgage Modifications

It has been widely observed that the extent of renegotiation of existing mortgages in response to large reductions in the asset values of houses has been small. That fact was apparent when the Hope for Homeowners Program established in the Housing and Economic Recovery Act of 2008 produced only a few hundred applications for modifications by the end of that year.

It is well-established that the asset values of houses decline substantially after the termination of payments by delinquent borrowers and before properties are sold. Presumably, those losses arise because occupants have little incentive to spend for maintenance and for the prevention of deterioration after they stop making mortgage payments. Moreover, after foreclosure, vacant properties are more expensive to maintain. Beyond that, vacancies attract vandalism and theft. Estimates of loss given default vary widely, but it appears that about thirty percent of the value of housing assets “disappears” when a default occurs (see Qi and Yang, 2009, for a review of such estimates and their determinants).

Thus, there is ample “room for a deal” between mortgage borrowers and lenders to avert large social losses. Borrowers are clearly better off if terms can be renegotiated in response to sharp declines in asset values or reversals in the labor market. But given the large loss in asset values avoided by the prevention of a foreclosure, it would seem that renegotiation could frequently improve the economic circumstances of lenders as well. Of course, renegotiation is subject to a moral hazard. However, in circumstances where there are widespread and large declines in asset values, and where the value of contracted payments exceeds the value of the house, the moral hazard from re-contracting is probably not large.

Thus, the current environment is one in which we would expect the incentives for voluntary loan modification to be the strongest. It is widely asserted that the principal barrier to the trade that would make borrowers and lenders better off is the institution of securitization. When mortgages are securitized, the interests of owners of the various tranches are not perfectly aligned, and actions taken by the mortgage servicer need not make all investors better off. The avoidance of litigation may make servicers less willing to modify mortgage terms, even if they think the modifications are "in the best interest of the investor." Beyond this, Pooling and Servicing Agreements typically restrict the freedom of individual servicers to make loan modifications and may especially restrict the number of modifications made to an individual loan. For all those reasons, it has been generally presumed that securitization is the major obstacle to loan modification, and current federal government programs to support the write-down of mortgages have included substantial subsidies to servicers and lenders to induce participation.

The paper by Adelino, *et al*[16] (see Appendix F) takes issue with that analysis of the nexus between securitization and loan modification. Adelino, *et al*, argue that even without the potential contractual roadblocks to negotiation, and without substantial transactions costs from securitization, mortgage loan renegotiations are frequently not in the economic interests of lenders. The authors' empirical analysis and conclusions rely on data reflecting mortgage modifications made through late 2008 and mortgage performance information through early 2009 and, thus, do not necessarily reflect more recent and different market conditions. The paper's discussion of the basic calculus of modification decisions, however, is appropriate to all market regimes.

The story is simple and relies only on considering the options embedded in the mortgage contract. If a mortgage is modified today, the borrower still has the option to default tomorrow; if the mortgage is not modified today, the borrower has the option to continue paying tomorrow.

In assessing the benefits of modification, the authors use parameters estimated from loan-level data from the very large LPS dataset. Taking into account mortgage performance through the early part of 2009, the authors find that the cure rate of loans that are sixty days delinquent was thirty percent. From that finding, the authors inferred that thirty percent of the lender's resources spent on modifications would have been "wasted" because borrowers would have "cured" even without modification. Similarly, the authors' empirical evaluation suggests that, during the applicable evaluation period, between 30 and 45 percent of borrowers whose loans were modified become seriously delinquent again within six months. Thus, the lender's resources spent on those modifications would have been "wasted" as well. (Actually, the situation is worse. By postponing default in a declining housing market, the lender would lose more. And, by discouraging maintenance of the housing unit, the modification would contribute to its deterioration.) On balance, then, considering the cure and redefault rates that were evident in their dataset, the authors find that a policy of foreclosure would allow lenders

[16] Adelino, Manuel, Kristopher Gerardi, and Paul S. Willen, "Why Don't Lenders Renegotiate more Home Mortgages? Redefaults, Self-Cures, and Securitization," Public Policy Discussion Paper 09-4, Federal Reserve Bank of Boston available at http://www.bos.frb.org/economic/ppdp/2009/ppdp0904.pdf.

to recover more than a policy of modification, even if a foreclosure does cause the value of the dwelling to decline precipitously.

The authors' conclusion is, of course, intricately related to the "redefault" rates (i.e., the fraction of loans that become delinquent or are in foreclosure six months after modification) and cure rates estimated from the LPS data. That data source reports performance data for mortgage loans backing private-label securities and in investors' portfolios. The data source under-samples subprime mortgages that have been securitized. The LPS data also do not record loan modifications directly, but do provide mortgage terms monthly for all loans. Thus, in principle, the data are structured to allow changes in terms (i.e., loan modifications) to be identified. The authors test their algorithm for identifying loan modifications on an external dataset that does identify loan modification, and they claim it "works" well. (It appears, though, that the algorithm does not identify forbearance by lenders.)

When applied to the LPS dataset, the algorithm suggests that modifications have increased substantially in the recent past (i.e., between 2008:Q1 and 2008:Q4 the number of loans whose interest rate was reduced quadrupled and the number of loans whose terms were extended increased 250 percent). The data show few principal reductions but a large number of principal balance increases. The authors attribute the former to the fact that the underlying data source under-reports subprime loans; they attribute the latter to the addition of arrears to the mortgage balances of delinquent borrowers. But as a fraction of the sixty-day delinquent loans in the sample, less than three and a half percent received *any* concessionary loan modification within the year after the first delinquency. Only about 8.5 percent of those loans received any type of modification at all. Voluntary loan modification, at least among prime mortgages, was quite rare during the time period studied.

Moreover, the differences in the modification rates are quite small between securitized and portfolio mortgages. Delinquent mortgages in portfolio had a rate of concessionary modification of 3.2 percent; in MBS the rate was 2.6 percent. For all modifications, the rate was 8.7 percent in portfolio and 8.4 percent in MBS. In multivariate models relating the probability of concessionary modification of a delinquent loan to a set of control variables (LTV ratio, FICO score, etc.), mortgages backing private-label securities are 0.3 percent less likely (with a t-ratio of 1.7) to be modified. The difference is barely significant, but it is also tiny.

When the analysis is reproduced for various sub-samples—subprime, low FICO score, "full doc," etc.—the results are similar. There is no action at all. When the analysis is reproduced using thirty-day delinquencies, the results are essentially unchanged.

The authors then investigate the subsequent behavior of loans that have been modified. They find that the redefault rate is high; among loans that have been modified to produce lower monthly payments, between twenty and forty percent redefault. They also find that there are no significant differences between the "redefault rates" for mortgages held in portfolio and loans backing private-label securities.

When the modified loans are stratified—by subprime, low FICO score, "full doc," etc.—there is no statistical difference in the redefault rates. That suggests that there are no important differences in the renegotiation behavior of servicers and portfolio managers. Neither type of debt manager is obviously more careful or more successful in negotiating modifications of mortgages to benefit themselves or borrowers.

Finally, the authors compare the cure rates for mortgages backing private-label securities and in portfolio. The cure rates are large—more than three-and-a-half times the rate of loan modification—and there are differences between the cure rates for the two groups of borrowers. When stratified by group (subprime, fully-documented, etc.), the results generally indicate that loans backing private-label securities are more likely to cure. Presumably, other unobserved or unmeasured differences between the mortgages in portfolios and backing private-label MBS are important in affecting those differential outcomes.

The logic and the empirical analysis reported by Adelino, *et al*, is suggestive, and their conclusions have important policy implications, ranging from the design of the current administration's Making Home Affordable program to structural reform of the secondary mortgage market. Those results, however, are inconsistent with those reported in a contemporaneous paper by Piskorski, Seru and Vig (2009). Piskorski, *et al*, conduct an extensive analysis of modifications of distressed loans, suggesting that relative to the servicers of securitized loans, servicers of portfolio loans undertook actions that reduced foreclosure rates. In contrast to the results reported by Adelino, *et al*, Piskorski, *et al* conclude (with some caveats) that "it is natural to interpret our results as suggesting that securitization has imposed renegotiation friction…."

That discrepancy in findings is even more disturbing since both research papers are based upon the same underlying body of data, described as the "McDash Analytics" database in one paper and as the "Lender Processing Services" database in the other. Resolving those discrepancies should be a high-priority task.

Beyond the issue of fact, there is also an important issue of interpretation in assessing the implications of the work of Adelino, *et al*. It is certainly true that loan modifications are quite rare. However, as stressed by discussant Richard Brown, in the past two-and-a-half years, mortgage modifications have increased 2.5 times—as a percent of seriously delinquent loans—to roughly eight percent of net loans becoming sixty days past due. Brown also stressed that the cure rate on delinquent mortgages depends upon the general state of the economy. Thus, a thirty-percent average cure rate based upon recent history may overstate the likely cure rate for delinquencies at the end of 2009.[17]

[17] By way of comparison, the Federal Deposit Insurance Corporation (FDIC) assumed a fifteen percent cure rate in loans modified under its IndyMac program.

There is no disagreement between the authors of the paper and the discussant about the estimated redefault and cure rates presented in the paper. Nor is there disagreement about the appropriateness of the present value models used to value loan modifications.[18] But there remain reservations about the rhetoric employed by the authors in characterizing their results. Lender choices to make few modifications and very few that substantially lowered borrower payments may have seemed rational based on prior cure and redefault rates. But in the current very weak housing markets and weak broader economy, cure rates have fallen sharply and redefault rates are likely to be much lower with modifications that significantly lower borrower payments to a level affordable at current income.

EXTERNALITIES OF MORTGAGE FORECLOSURES

The calculus of the costs and benefits of loan modification to borrowers and mortgage lenders by Adelino, *et al*, neglects one crucial element for the analysis, namely the external costs of mortgage default. Even if the lender's calculation that he is better off by foreclosing on a loan rather than modifying its terms is correct, it may still be true that the uncompensated social costs of the default are large enough to make loan modification the socially-preferred choice. That depends crucially on the external costs imposed by the foreclosure of a property, a topic on which there is a great deal of opinion but little hard evidence.

It appears to be generally true that property values decline in the geographical area surrounding a foreclosed property. But that need not imply that there is any spillover from the value of a foreclosed property to that of its neighbors. Indeed, it is not at all obvious whether the existence of a foreclosed property causes a decline in the value of nearby homes or whether, instead, the homes in a neighborhood experienced a decline in values as a result of common external forces. Given the historical development of urban neighborhoods and their similarities in housing stock and in the demographic composition of the households who live there, it is not easy to distinguish between the effects of a common shock to households or to property values and the causal effect of a foreclosure. Does an observed decline in property values in the neighborhood of a foreclosure arise from a common decline in demand for these similar properties? Or from a decline in common labor-market conditions that would lead one homeowner to economize on home maintenance but would push another homeowner into default? Or does a foreclosure *cause* nearby property values to decline?

There is little credible evidence on that topic and little that would lead to confidence in conclusions. A widely-known paper by Immergluck and Smith (2006) illustrates the problem. Immergluck and Smith estimate hedonic models relating house prices to housing and location characteristics. The externality is measured by including the number and location of distressed properties as regressors. In that research design, it

[18] But there is, apparently, some disagreement about whether the analytical and empirical results reported by Adelino, *et al*, are novel, at least when compared to the internal documents of the FDIC.

is likely that the number and character of nearby foreclosures is correlated with a host of important, but unmeasured, characteristics of the property whose price is to be explained. That suggests that hedonic techniques to establish the importance of foreclosures on property values will systematically over-estimate their importance. Beyond that, if there is a time trend in the number of foreclosures and in the trend in house prices, it may be logically impossible to distinguish between them.

The paper by Harding, *et al*[19] (see Appendix G) sorts out the influence of correlations and omitted variables using repeat sales price measurements applied to a large sample of transactions in neighborhoods in seven cities. Their results suggest that nearby foreclosures do exert a *causal* influence on property values. But those results also suggest that the effects are much smaller than those presented by others, for example, the work of Immergluck and Smith. Moreover, the effects of foreclosures are greatly attenuated over space; in most cases, at least three quarters of the small effects are dissipated for homes more than five hundred feet from the foreclosed property.

The derivation and use of the weighted repeat sales (WRS) estimator of housing price indexes is well known.[20] The advantage of the WRS technique in this context is its treatment of the unobserved attributes of dwellings or locations that do not vary over time. Because the WRS estimator uses only information on individual changes in value over time, all of those unobserved or unmeasured attributes simply difference out.

The Harding, *et al*, paper relies upon the WRS specification to estimate the effect of foreclosures on property values by relying upon a large database of house sales and another database of mortgage transactions and outcomes during the period 1990 through 2007. The authors identify a sample of about three hundred postal codes in which the transactions database includes more than eighty percent of the foreclosures in the mortgage database. For those postal codes, the authors also have access to the repeat sales used to generate the FHFA metropolitan house price indices. For 140 of the postal codes, there were reasonably large samples of paired sales. Geocoding of the repeat sales and the foreclosures permitted the proximity of each sales pair to each foreclosure to be computed. The timing of each foreclosure relative to the timing of each of the repeat sales could be recorded as well.

That merged and matched body of data formed the basis for analysis. The Harding, *et al*, paper reports the results for postal codes in seven metropolitan areas, which include approximately 405,000 repeat sales. The authors partition the distance measures into four categories, and they define a series of 13 time intervals relative to the foreclosure and resale of a nearby property. A standard WRS model is estimated for each metropolitan area, augmented by variables measuring the number of foreclosures recorded at the first and the second sale. Externalities, “contagion effects,” are estimated for four distances, “rings,” separately for different measures of the timing of the

[19] Harding, John P., Eric Rosenblatt, and Vincent W. Yao, “The Contagion Effect of Foreclosed Properties”.

[20] See Case and Shiller (1989) and Bailey, Muth, and Nourse (1963).

foreclosure relative to the timing of the sale of the nearby house. The estimation results are complex, but they can be summarized by distance and timing.

For foreclosures within three hundred feet of a sale (that is, within a couple of parcels), the value of the sold property declines by about one percent from its value a year before the foreclosure to its value at the event itself. After the foreclosure, the value remains more or less the same. One year after the disposition of the foreclosed property, the sale price of the nearby property is still about one percent lower than it was a year before the default.

For properties located between three hundred and five hundred feet of a foreclosed property, the price effects are even smaller, and they are insignificant in some of the metropolitan areas. On average, there is little or no externality measured until about the time of the sale of the foreclosed property, at which the external effect may be a half percent or less of the house value. For foreclosures more than five hundred feet removed from a sale, there is virtually no discernible or significant effect.

In another series of specifications, the authors analyze the external effects of multiple nearby foreclosures, in less detail to be sure. Those results do suggest that the effects of multiple foreclosures are more or less linear, e.g., for a given distance from a subject property, three foreclosures would have roughly three times the impact of one foreclosure. But here again, the effects are greatly attenuated with distance.

Those general results survive a series of robustness checks. The most important of those is the use of instrumental estimates of the number of foreclosures in a two-stage least-squares estimation.

That analysis establishes quite clearly that foreclosures do have an independent causal effect upon local house values. The results are much more credible than those based upon hedonic models, in which inferences may depend upon correlations of defaults with the unmeasured or unobservable characteristics of nearby properties and neighborhoods subject to common shocks.

As indicated by discussant Paul Willen, within the research design adopted by the authors it is difficult to distinguish among three distinct causes of property decline arising from foreclosures: an increased supply of houses on the market; artificial or transient price declines because foreclosures are traded on the markets (“fire sales”); and the external effects of vacant and deteriorating properties. More general and powerful statistical models are also available to analyze the spatial and temporal effects of foreclosures.[21]

But the results of the careful analysis by Harding, *et al* provide compelling evidence that the effects of foreclosures are much smaller than those previously described. The authors conclude that “a million additional foreclosures would

[21] See, for example, Hwang and Quigley (forthcoming).

significantly affect three to five million homes not the forty million that has been estimated using earlier estimates of contagion effects." But even that seems to be an overstatement of the externalities associated with foreclosures, especially if the externalities are to be measured by realized capital losses rather than unrealized but transient reductions in estimates of asset values.

CONCLUSION

The reversals in the U.S. housing and single-family mortgage markets during the past three years have been manifest in millions of detailed transactions records at the micro level: mortgage loan origination and subsequent payment patterns, purchases of mortgages and their performance in portfolios and securities, and conditions in local housing markets and neighborhoods. That mass of data on individual transactions and outcomes is a fertile ground for analysis of behavior and decision-making by the actors in the housing and mortgage markets: mortgage borrowers and mortgage brokers, lenders, and investors.

That large body of evidence can be used not only to test microeconomic models of behavior, but also to help design programs and policies to make a widespread disruption in U.S. housing and mortgage markets less likely in the future. Thus, the seven papers summarized in this report, by taking a first step toward understanding microeconomic behavior and its consequences during the crisis of the past several years, can inform discussion of options for improving U.S. mortgage markets going forward.

References

Adelino, Manuel, Kristopher Gerardi, and Paul S. Willen, "Why Don't Lenders Renegotiate more Home Mortgages? Redefaults, Self-Cures, and Securitization," *NBER Working Paper 15159, July 2009*. [Appendix F]

Ambrose, Brent W., and Richard J. Buttimer, Jr. "Embedded Options in the Mortgage Contract," *The Journal of Real Estate Finance and Economics*, 21(2), 2000: 95-111.

Ashcraft, Adam B., and Til Schuermann. "Understanding the Securitization of Subprime Mortgage Credit," *Federal Reserve Bank of New York Staff Report*, 318, March 2008.

Bailey, Martin J., Richard F. Muth, and Hugh O. Nourse, "A Regression Method for Real Estate Price Index Construction," *Journal of the American Statistical Association*, 58, 1963: 933-942.

Bucks, Brian, and Karen Pence, "Do Homeowners Know Their House Values and Mortgage Terms?" FEDS Working Paper No. 2006-03, March 2006.

Case, Karl E., and Robert J. Shiller, "Prices of Single Family Homes Since 1970: New Indexes for Four Cities." *National Bureau of Economics Research Working Paper Series*, 2393, 1987: 45-56.

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," CEPR Discussion Paper No. 6683, February, 2008.

Demyanyk, Yuliya, and Otto Van Hemert, "Understanding the Subprime Mortgage Crisis," Forthcoming in *Review of Financial Studies.* Draft available online at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1020396.

Deng, Yongheng, "Mortgage Termination: an Empirical Hazard Model with Stochastic Term Structure," *The Journal of Real Estate Finance and Economics,* 14(3), 1997: 309-331.

Deng, Yongheng, John M. Quigley, and Robert Van Order, "Mortgage Terminations, Heterogeneity and the Exercise of Mortgage Options," *Econometrica*, 68(2), 2000: 275-308.

Elul, Ronel, "Securitization and Mortgage Default: Reputation vs. Adverse Selection," *Federal Reserve Bank of Philadelphia Working Paper 09-21*, September 22, 2009. [Appendix D]

Feigenbaum, James, and Geng Li, "A Nonparametric Characterization of Income Uncertainty over the Lifecycle," unpublished paper, July 14, 2008.

Foote, Christopher L., Kristopher Gerardi, and Paul S. Willen, "Negative Equity and Foreclosure: Theory and Evidence," *Journal of Urban Economics*, 64(2), 2008: 234-245.

Foote, Christopher, Kristopher Gerardi, Lorenz Goette, and Paul Willen, "Reducing Foreclosures: No Easy Answers," *National Bureau of Economic Research, Working Paper 15063*, June 2009. Available online at http://www.nber.org/papers/w15063.

Fortowsky, Elaine B., and Michael LaCour-Little, "Are Subprime Mortgages Fairly Priced? An Analytical Approach," Wells Fargo Home Mortgage Corporation, 2002.

Gerardi, Kristopher S., Andreas Lehnert, Shane M. Sherlund, and Paul S. Willen, "Making Sense of the Subprime Crisis," *Federal Reserve Bank of Atlanta, Working Paper 2009-2*, February 2009. Available online at http://www.frbatlanta.org/filelegacydocs/wp0902.pdf.

Ghent, Andra C. and Marianna Kudlyak, "Recourse and Residential Mortgage Default: Theory and Evidence from US States," *Federal Reserve Bank of Richmond Working Paper 09-10*, July 7, 2009. [Appendix E]

Government Accountability Office, *Mortgage Financing: FHA's Fund Has Grown, but Options for Drawing on the Fund Have Uncertain Outcomes*, GAO-01-460 (Washington, D. C. Feb. 28, 2001).

Green, Richard K., and Michael LaCour-Little, "Are Minorities or Minority Neighborhoods More Likely to Get Low Appraisals?" *Journal of Real Estate Finance and Economics*, 16(3), 1998: 301-315.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales, "Moral and Social Constraints to Strategic Default on Mortgages," *NBER Working Paper 15145*, July 2009.

Harding, John P., Eric Rosenblatt, and Vincent W. Yao, "The Contagion Effect of Foreclosed Properties." July 13, 2009. [Appendix G]

Hwang, Min, and John M. Quigley, "Housing Price Dynamics in Time and Space: Predictability, Liquidity, and Investor Returns," *Journal of Real Estate Finance and Economics*, forthcoming.

Immergluck, D., and G. Smith, "The External Costs of Foreclosure: The Impact of Single-Family Foreclosures on Property Values," *Housing Policy Debate*, 17(1), 2006: 57-79.

Jiang, Wei, Ashlyn Nelson, and Edward Vytlacil, "Liar's Loan? Effects of Origination Channel and Information Falsification on Mortgage Delinquency," September 2009. [Appendix C]

Jones, Lawrence D., "Deficiency Judgments and the Exercise of the Default Option in Home Mortgage Loans," *Journal of Law and Economics*, 36(1), 1993: 115-138.

Kain, John F., and John M. Quigley, "Note on Owner's Estimate of Housing Value," *Journal of the American Statistical Association*, 67(340), 1972: 803-806.

Kau, J. B., and D. C. Keenan, "An Overview of the Option-Theoretic Pricing of Mortgages," *Journal of Housing Research*, 6(2), 1995: 217-244.

Kelly, Austin, "Appraisals, Automated Valuation Models, and Mortgage Default," 2009. [Appendix A].

Keys, Benjamin J., Tanmoy Mukherjee, Amit Seru, and Vikrant Vig, "Did Securitization Lead to Lax Screening? Evidence from Subprime Loans," *Quarterly Journal of Economics*, 125 Forthcoming 2010.

LaCour-Little, Michael, and Stephen Malpezzi, "Appraisal Quality and Residential Mortgage Default: Evidence from Alaska," *Journal of Real Estate Finance and Economics*, 27(2), 2003: 211-233.

LaCour-Little, Michael, and Jing Yang, "Taking the Lie out of Liar Loans," August 30, 2009. [Appendix B]

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

Mays, Elizabeth, editor, *Handbook of Credit Scoring*, London: Global Professional Publishing, 2001.

Mian, Atif, and Amir Sufi, "The Consequences of Mortgage Credit Expansion: Evidence from the U. S. Mortgage Default Crisis," December 2008. Available online at http://ssrn.com/abstract=1072304.

Pence, Karen M., "Foreclosing on Opportunity: State Laws and Mortgage Credit," *The Review of Economics and Statistics*, 88(1), 2006: 177-182.

Pennington-Cross, Anthony and Giang Ho, “The Termination of Subprime Hybrid and Fixed Rate Mortgages,” Federal Reserve Bank of St. Louis, Working Paper 2006-042A, July 2006. Available online at http://research.stlouisfed.org/wp/2006/2006-042.pdf.

Piskorski, Tomasz, Amit Seru, and Vikrant Vig, “Securitization and Distressed Loan Renegotiation: Evidence from the Subprime Mortgage Crisis,” December 2008. Chicago Booth School of Business Research Paper No. 09-02. Available online at http://ssrn.com/abstract=1321646.

Qi, Min, and Xiaolong Yang, “Loss Given Default of High Loan-to-Value Residential Mortgages,” *Journal of Banking & Finance*, 33(5), 2009: 788-799.

Rajan, Uday, Amit Seru, and Vikrant Vig, “The Failure of Models That Predict Failure: Distance, Incentives and Defaults,” Chicago Booth School of Business Research Paper No. 08-19. Available online at http://ssrn.com/abstract=1296982.

Schwartz, E. S., and W. N. Torous, “Prepayment and the Valuation of Mortgage-Backed Securities,” *The Journal of Finance*, 44(5), 1989: 375-392.

Shiller, Robert J., and Allan N. Weiss, “Evaluating Real Estate Valuation Systems,” *Journal of Real Estate Finance and Economics*, 18:2, 1999: 147-161.

Vandell, K., “How Ruthless is Mortgage Default? A Review and Synthesis of the Evidence,” *Journal of Housing Research*, 6(2), 1995: 245-264.

Weicher, John C., “The New Structure of the Housing Finance System,” Federal Reserve Bank of St. Louis *Review*, 76(4), 1994: 47-65.

Appendices

Appendix A

Appraisals, Automated Valuation Models, and Mortgage Default

Austin Kelly
Associate Director, Division of Enterprise Regulation
Austin.kelly@fhfa.gov

Preliminary Draft. This paper represents the opinions of the author only, and cannot be construed as an expression of the opinions of the FHFA or the US government. The author would like to thank Michael LaCour-Little for helpful comments, Tom Boehm for sharing his Windows implementation of CTM software, and Jake Garcia of First American Real Estate Solutions, for providing and discussing the data used in this paper. Responsibility for remaining errors, if any, rests squarely on the author.

Appraisals, Automated Valuation Models, and Mortgage Default

AUSTIN KELLY
Associate Director, Division of Enterprise Regulation, Federal Housing Finance Agency. E-mail: Austin.Kelly@fhfa.gov

Abstract

Previous research has suggested the possibility that professional appraisals or econometric estimates of collateral value may be indicative of credit risk. This paper examines the issue by estimating the probability of a mortgage default (defined both as 90 day delinquency and as a claim on mortgage insurance) as a function of the difference between sales price of a home and the estimated value of the home at the time of the purchase, produced by both an appraisal and by an Automated Valuation Model (AVM). Logistic regression is used to estimate the quarterly hazard of a serious delinquency, or claim, as a function of a host of standard control variables, and the percent difference between the sales price and the appraisal and/or AVM estimate. The data consist of a nationally representative random sample of about 5,000 FHA insured single family mortgages endorsed in Fiscal Years 2000, 2001, and 2002, observed through January 31, 2006, and a sample of about 1,000 FHA loans from the Atlanta MSA in the same time period. The records are augmented with the results from an AVM. The difference between the sale price and the appraisal or AVM estimate is found to significantly increase the probability of delinquency, and increase the probability of foreclosure, significantly so in the national sample. Also, transactions that are valued with higher precision have lower default propensities. Additionally, the differences are found to increase loss given default in the small subset of loans that had completed the property disposition process.

1. Introduction and Literature Review

The idea that equity plays an important role in the homeowner's decision to default is longstanding in the academic literature[1]. Empirical estimates of the relationship between equity and default go at least as far back as Herzog and Earley (1970), and a firm theoretical underpinning for the decision to default was provided by Kau and Kim (1994). Equity can come in two flavors – initial equity in the form of the down payment when a home is purchased, and contemporaneous equity, which adds in price appreciation (or depreciation) post purchase, amortization, and sometimes changes in the market value of the mortgage balance. Research finds that contemporaneous equity has a strong influence on credit risk, and some papers, such as Harrison, Noordewier, and Yavas (2004) find that initial equity has a modest additional impact, over and above it's effect on contemporaneous equity, perhaps because it reflects the household's ability to save, or because it is more precisely measured than accumulated equity, which is usually measured as a state or MSA wide average. An exception is the modeling of FHA mortgages, where initial LTV is sometimes found to have little additional impact,

[1]For a recent review of the literature on mortgage credit risk, see US GAO (2005).

possibly because the vast majority of FHA loans are at or near the maximum LTV allowed by FHA. For examples, see Technical Analysis Center (2005) or US GAO (2001)[2].

Appraisers provide the estimate of value used in determining initial equity. A handful of papers have examined the role of appraisers in the underwriting process. Horne and Rosenblatt (1996) examine the distribution of the differences between appraised values and purchase prices. They find that differences between appraised values and sale prices are almost always less than one percent, and appraisals for less than the purchase price are extremely rare. LaCour-Little and Malpezzi (2001) estimates a model similar to the one in this paper. Using a stratified sample of 224 loans originated at a credit union in Alaska, they calculate the differences between appraisal and a hedonic estimate of property value, and estimate the effect of this difference on delinquency probabilities, controlling for some standard underwriting variables. This paper extends the work of LaCour-Little and Malpezzi in several dimensions. First, it examines a much larger and geographically representative group of loans. Second, it uses much more recent loans (2000 to 2002, versus 1980's originations for LC-L and M) and uses a set of control variables that reflect current underwriting practice, such as FICO scores. Third, it examines the probability that a loan will actually produce a loss, in addition to examining the probability of a serious delinquency. Although delinquency is a valuable early indicator of credit risk, papers such as Ambrose and Capone (2000) and Danis and Pennington-Cross (2005) find that delinquencies often cure, rather than lead to losses for lenders or insurers. Finally, it incorporates more information into the credit risk model, examining the quality of the hedonic estimate as measured by its standard error, in addition to considering the level of the difference.

Shiller and Weiss (1999) lay out a framework for evaluating the profitability of AVM deployment. This paper takes a step towards filling the data requirements of their framework, estimating the correlation between appraisal/selling price and AVM valuation, and demonstrating the effectiveness of AVM systems in predicting default, foreclosure, and loss severity.

The rest of the paper is laid out as follows. Section 2 describes the model in general terms. Section 3 describes the data. Section 4 discusses the estimation strategy and the CTM software used to estimate the model. Section 5 provides the results for 90 day delinquency, claims, prepayments, and loss given default. Section 6 offers concluding remarks and some observations concerning the relative predictive power of appraisals and AVMs.

[2]Both GAO and HUD have modeled FHA conditional termination rates as functions of LTV dummy variables, and find that default propensity rises sharply as LTV increases over ranges up to about 96% LTV. Beyond this point the relationship flattens out, perhaps because the bulk of FHA mortgages are at the maximum LTV allowed by program rules. Program rules allow slight variations in maximum LTV based on loan size and location in high or low closing cost states, and have varied over time. The fact that few loans are written for less than the maximum LTV means that, in high LTV ranges, the LTV variable is picking up location and vintage features along with “true” LTV effects.

2. Model

The focus of this paper is the effect of appraisal and AVM quality on the credit risk in mortgages. An appraisal is a measure of the "market value" of a property. In a highly liquid market with large numbers of identical commodities traded, this is a simple concept. In the housing market, with infrequently traded heterogeneous properties, market value is a more tenuous concept. To some extent, the fact that a buyer is willing to pay $X for a house sets $X as the market value, rendering an appraisal somewhat superfluous. From the perspective of the entity holding the credit risk on the mortgage (lender or, in this case, insurer), the most relevant concept might be the value that the second highest bidder is willing to spend on the property, a notion that mixes the concepts of "market value" and "liquidity." This is because the holder of the credit risk cares about the price at which the buyer could later sell the property, which determines the buyer's choice of prepayment or default in the face of trigger events, and determines the amount of recovery in case of default. This may be expressed as

$$\textit{Market Value} = \textit{Transaction Price} + \textit{Idiosyncrasies} \qquad (1)$$

where Market Value refers to the expected selling price if a property were immediately resold. Transaction Price is the price agreed upon by the buyer and seller. Idiosyncrasies represent any unique characteristics attached to the transaction, such as a buyer uniquely attracted to a particular property characteristic, or a seller motivated to sell exceptionally quickly, or, for that matter, fraud.

The appraisal process can provide an estimate of property value independent of the idiosyncratic circumstances that might cause a buyer to be the highest bidder. Single family appraisals are generally based upon the sale prices of comparable properties, with adjustments made for differences in characteristics between the property in question and the comparables, and with adjustments made for area-wide trends in price. An appraisal constitutes an estimate of the market value. Such an estimate may be biased or unbiased. Sources of bias to the high side are pressure from buyers, sellers, brokers, etc. who need an appraisal for at least the agreed upon price so that the transaction can take place. The holder of the credit risk on the transaction, for example, the insurer, would presumably wish to pressure appraisers for an accurate estimation, but in many cases the appraiser is hired by the lender, although the risk is borne primarily by the insurer.[3]

$$\textit{Appraisal Value} = \textit{Market Value} + \textit{Bias1} + \varepsilon_1 \qquad (2)$$

where Appraisal Value is the value assigned by an appraiser, Bias1 represents any possible tendency to assign a value other than the expectation of Market Value, and ε_1 is the inherent noise in any estimation process.

[3]FHA does maintain a list of approved appraisers, and can remove an appraiser from the list for fraud or unethical behavior, but it is not clear how effective this might be in the case of modest upward bias of the sort considered here. See US GAO (2004) for a discussion of FHA's role in monitoring appraisers.

An AVM produces a second estimate of the market value of the property. An AVM estimate may be less subject to bias, as AVM services are sold to a wide variety of parties, such as lenders, insurers, GSEs, or MBS investors, with no clear incentive to produce "high" or "low" estimates. On the other hand, AVMs constitute a mass-appraisal approach, rely upon generally available characteristics, and do not involve visits to properties to ascertain condition or incorporate local knowledge (the announcement of a factory closing or plans for a new transit stop), so that their variances may be much higher than the variances of appraisals.

$$\textit{AVM Value} = \textit{Market Value} + \textit{Bias2} + \varepsilon \qquad (3)$$

where AVM value is the value assigned by an AVM, Bias2 is the tendency (if any) for an AVM to produce a value other than the expectation of market value, and ε is the inherent noise in the AVM estimation process.

The relevant questions for a holder of mortgage credit risk are, 1) "does an appraisal contain any information helpful to the assessment of default propensities, and 2) "does an AVM estimate contain any information beyond that contained in an appraisal?" The latter will be the case if the mean square error of the AVM is not too large, relative to the mean square error of the appraisal, and if the correlation between the two errors is not too high. One way to test this proposition is to estimate equations such as

$$\textit{Prob(Default)} = \textit{fn(Appraisal, AVM Estimate, other risk variables)} \qquad (4)$$

$$\textit{Loss Given Default} = \textit{fn(Appraisal, AVM Estimate, other risk variables)} \qquad (5)$$

and test the coefficients on the Appraisal and AVM Estimate values.

Underwriters, and FHA guidelines in particular, generally take the minimum of the sale price or the appraised value as the denominator when calculating the loan-to-value ratio, used as a key indicator of default probability. Thus, the extent to which an appraisal exceeds the transaction price has no effect on the underwriting decision, or perceived degree of risk attached to the loan by the underwriter. An appraisal less than the transaction price has serious consequences, however, generally requiring an increase in the cash that the buyer has to bring to the table, or a decrease in the price received by the seller, or the failure of the transaction to go through. Thus appraisals may produce benefits in ways not captured by transaction data, either by preventing transactions on overpriced properties, or by triggering renegotiated prices.

AVMs are generally not used in FHA underwriting[4]. However, AVM estimates may provide an additional source of information on the value of the collateral; therefore on the level of credit risk for a given mortgage. The extra predictive power could be useful for risk monitoring on the part of FHA or other insurers, risk accounting, and for investor evaluations of portfolios of mortgages.

[4]Although FHA does sometimes use AVMs in post endorsement reviews.

3. Data

3.1. Loan Data

The data for this paper consist of a nationally representative sample of just over 5,000 FHA single family purchase money loans, endorsed in fiscal years 2000, 2001, and 2002, that is, from October 1999 to September 2002. Their performance is observed through June 2005. These loans were drawn by Concentrance Corp, a HUD contractor, for a HUD-sponsored study of down payment assistance.[5] This file is one of only two large random samples of seasoned FHA loans with FICO scores[6], as HUD only began the routine collection of FICO scores as part of their Single Family Data Warehouse (SFDW) in 2004. In addition to FICO scores, the file contained many fields from the SFDW, such as the initial LTV ratio, mortgage payment, borrower income, type of mortgage, term, interest rate, and street address of the borrower. This file was merged with a July 2005 extract of the SFDW containing dates for prepayment of the loans that paid off early, date of first 90-day delinquency reported by the lender, and date of claim for loans that terminated with a loss to FHA, and the loss (or, for 12 foreclosures, profit) for loans that had completed the property disposition process.

In addition to the national file, Concentrance drew 1,000 loan samples from each of three MSAs, Atlanta, Indianapolis, and Salt Lake City, over the same time period. For reasons discussed later, this paper focuses on the national sample, and reports analysis for the Atlanta sample.

The Concentrance samples were limited to loans with LTV ratios greater than 95%, as defined in HUD's SFDW. Since HUD's definition of LTV excludes the upfront mortgage insurance premium, which is generally rolled into the mortgage, in effect almost all of these loans had LTV ratios, as conventionally defined, greater than 96.5%, as HUD's upfront premium was 1.5% for most of the sample period. Loans with LTVs greater than 96.5% constitute almost 90% of FHA's purchase money loans, and constitute over 90% of FHA's claims. Because FHA allows some closing costs to be financed, and allows the financing of the upfront premium, FHA loans can, in some circumstances, slightly exceed 100% LTVs. In this sample almost 85% of the records had LTVs in the narrow range of 98% to 100%, and about 99% were between 95% and 101%, as conventionally defined.

The median price in the national sample was $110,000. About 99% of the loans were for a term of 30 years, with the remainder generally for 15. About 6% of the loans were for condominiums, and about 8% of the loans were 1 year ARMs, with the balance being fixed rate mortgages (FHA did not offer hybrid ARMs at that time). Just over 80% of the loans were to first time home buyers, and about 40% were in underserved area census tracts. See Table I for sample summary statistics.

---

[5]See Concentrance Consulting Group (2004).

[6]The other file was collected by HUD as part of their development of FHA's automated underwriting algorithm. The loan years covered precede the widespread deployment of AVM systems. See Cotterman (2004).

Table I.1

Summary Statistics

Variables	National Sample Mean	National Sample Sigma	Disclose State Sample Mean	Disclose State Sample Sigma	Atlanta Sample Mean	Atlanta Sample Sigma
Dependent						
Cumulative delinquent rate	12.47%		12.34%		16.40%	
Cumulative claim rate	4.90%		4.56%		9.64%	
Cumulative prepay rate	75.97%		77.95%		67.86%	
Loss severity rate	34%	21.30%	34%	22.70%	27%	12.10%
Time Invariant Independent						
frontend ratio	0.26	0.08	0.27	0.08	0.27	0.07
LTV ratio	0.99	0.01	0.99	0.01	0.99	0.01
FICO (/100)	6.55	0.61	6.55	0.61	6.43	0.58
NoFICO	8.2%		8.5%		7.3%	
reserves < 2 months	28.5%		28.3%		23.6%	
underserved area	40.5%		42.5%		40.3%	
condominium	6.1%		6.6%		4.1%	
first time buyer	81.8%		82.3%		82.0%	
ARM	7.6%		8.1%		10.1%	
Appraiseratio	0.02	0.06	0.02	0.07	0.02	0.03
Median	0.00	0.05	0.00	0.05	0.01	0.04
Avmratio	0.09	0.23	0.08	0.23	0.05	0.17
Median	0.04	0.41	0.03	0.56	0.02	0.29
Avmconfidence (/100)	0.75	0.15	0.75	0.14	0.80	0.00
Time Varying Independent						
GAOrisk	6.92	2.07	6.65	2.23	6.78	1.91
Growth	1.10	0.14	1.05	0.05	1.07	0.07
Number of Observations	3985		3403		1116	

Over 12% of the loans experienced at least one episode of serious delinquency by June 30, 2005[7]. About 5% of the loans resulted in a claim on FHA, generally through foreclosure, by January 31, 2006. For the few loans with a claim that had completed the property disposition process, the average loss was 34% of the original mortgage balance. Over 80% of the loans in the sample had terminated by the end of the observation window, either through prepayment or claim termination. Interest rates reached a local

[7]HUD provided an update file for claim and non-claim terminations through January 2006, but this file did not include delinquencies. Hence, claim regressions and statistics are though January 2006, while delinquency regressions are through June 2005. Additionally, HUD imposed a foreclosure moratorium for counties and parishes affected by hurricanes Katrina and Rita – loans in these areas still active in September 2005 are censored as of the end of September. This affected less than 1percent of the national sample.

minimum in 2003, and prepayment rates were fairly high for these cohorts.

3.2. External Data

These files were merged with several external sources to incorporate time-varying covariates for the hazard analysis. State level unemployment rates were obtained from BLS, the state level constant quality house price index was obtained from OFHEO, 30 year fixed-rate mortgage rates were taken from Freddie Mac's Primary Mortgage Market Survey, and one-year Treasury rates were taken from the Fed.

Finally, the addresses in the file were submitted to First American Real Estate Solutions in July 2005, for the purpose of appending the results of their AVM models to each transaction. First American is a large vendor of real estate data, and maintains several AVMs that are in wide commercial use among mortgage lenders. First American was given the address, an "as-of" date, defined as 30 days prior to the settlement date of the mortgage, and indicators for whether the property was 1-unit or 2- to 4-unit, and whether the property was a condominium. First American used a "cascading AVM" approach. A loan was first submitted to their PASS model, a hedonic model[8]. If there was insufficient data to produce a high confidence estimate, it was then submitted to HPA, which First American describes as a primarily hedonic model, with some repeat sale index hybrid elements. About 95% of the successfully valued loans had results from either PASS or HPA. In about 5% of the cases, if these failed to produce high confidence estimates, the loan was submitted to PB6 or VP4, which First American describes as neural net models.

For each successfully valued loan, First American returned the predicted mean value from the model, as-of the month prior to settlement, a high value, defined as the 90th percentile estimate from the AVM, and a Confidence Score, defined as the probability that the true value is within 10% of the AVM produced value. Confidence Scores ranged from 40 to 98, with a median of 78. Records with Confidence Scores below 40 were deemed failures by First American and values were not returned.

First American returned values for 3,985 loans in the national sample of 5,101, a success rate of 78%. A record might not be valued for several reasons. There could be a problem merging the address fields in the HUD database to the First American database. There could be too few transactions of a property type in an area to allow for a high confidence estimate. Additionally, 8 states are "non-disclosure" states[9] – that is, local officials do not release property transaction data to firms such as First American. While First American gets some data from lenders on transactions in these states, they have less than complete coverage, leading to low hit rates in most of these states. Unfortunately, both Utah and Indiana are non-disclosure states, leading to questionable model fits for two of the three MSAs that were also sampled. Georgia is a disclosure state, and the

[8]First American regards the details of their AVM algorithms as proprietary. For a general discussion of hedonic, repeat sale index, and hybrid models, see Case (1991). For a discussion of neural net models, see Kershaw and Rossinni (1999).

[9]They are Alaska, Indiana, Kansas, Mississippi, Missouri, New Mexico, Texas, and Utah. Texas is a middle case – although First American does not get data from local officials, they have arrangements with the MLS in some Texas MSAs and get nearly complete data for those parts of the state.

Atlanta sample had a hit rate of 95%, with a median Confidence Score of 85. Hence, analysis is done for the Atlanta MSA, but not for Indianapolis or Salt Lake City, where hit rates were under 70%. For the national sample, results are presented for both the "full" national sample – that is, the full sample of loans for which First American produced an AVM value, and for a subset (3,403 loans) that includes only loans made in disclosure states. For this subset of the national sample, hit rates are about 85%, and the median Confidence Score is 80.

The mean of the ratio of the AVM value to the sale price is 1.09, substantially exceeding 1 (Table I). However, as both sale price and AVM value are random variables, being noisy signals of a "true value," this number represents the mean of a ratio of two random variables. Assuming the variables are both Normal, the result has a Cauchy distribution, with an undefined expectation. In general, the mean of a ratio is not the ratio of the means. For that reason, the median AVM ratio may be preferred as an indicator for the ability of the AVM values to match the transaction prices. The median AVM ratio is 1.04 for the national sample, 1.03 for the disclosure states, and 1.02 for the Atlanta sample.

As the off-the-shelf commercial AVM models used here are proprietary, it is impossible to ascertain why the AVM values tend to exceed the transaction prices, but there are at least two (non-exclusive) possibilities. The years 2000 to 2002 were noteworthy for double digit annual appreciation in housing prices. If the timing of the valuation process is off by even a month the result could be an upward bias of a percentage point. Another possibility is a form of selection bias. As these are all FHA mortgages, and the FHA program has a maximum loan amount, there will be some tendency for cheaper properties to have higher probabilities of being financed with an FHA mortgage, conditional on the observed characteristics of the property. The fact that the disclosure states have a smaller bias than the non-disclosure states lends some credence to this explanation, as the greater the ability to model the price with covariates, the smaller is the problem from unobservables correlated with FHA status. The fact that the Atlanta sample has the smallest bias is also consistent with these stories, as Atlanta had lower than average appreciation, and has a lower than average median house price, meaning that the FHA ceiling is a binding constraint in fewer cases. So long as the unobservables are uncorrelated with default, the selection problem would not produce an upward bias in the significance of the AVM ratio. To the extent that such unobservables may be positively correlated with default the problem would lead to an underestimate of the impact of AVM value. For example, properties adjacent to a nuisance, not accounted for in the AVM model, may be assigned an AVM value above the FHA ceiling but may sell for less than the FHA ceiling. If such properties had higher default propensities, then the high AVM ratio observations would be associated with default in a sample of only FHA properties, weakening the ability to measure the true effect of high AVM ratios in lessening the probability of default.

4. Estimation Strategy

The loan records from the Concentrance study were merged with recent data on claims and delinquencies. The merged loan records were then used to produce loan quarter records, which were merged with time-varying data, such as house price appreciation, unemployment, and interest rates. To measure the influence of appraisals on default propensity, a variable was created equal to the percent difference between the appraised value and the purchase price of the house. To measure the influence of AVM value on default propensity, a variable was created equal to the percent difference between the AVM estimate and the purchase price of the house. Additionally, the confidence score attached to the AVM estimate was entered as a regressor in some specifications.

$$AppraiseRatio = (appraisal\text{-}price)/price \tag{6}$$

$$AVMRatio = (AVM\text{-}price)/price \tag{7}$$

Two strategies were employed in choosing other covariates for the logistic regression. In one, time-invariant variables of the type used in FHA's TOTAL scorecard automated underwriting system were chosen[10]. These are FICO score[11], LTV at origination, an indicator for whether the borrower will have at least 2 months of reserves after closing, and the Front End ratio. These variables were augmented by other loan, borrower, and property variables that might influence credit risk, such as indicators for first-time home buyers and properties in underserved areas. A time-varying covariate is also included to measure post-origination price appreciation. This is defined as the state level percentage change in the OFHEO price index, measured quarterly. For the first two quarters of the loan's life, this value is set to 1; starting with the third quarter, the value is calculated as the ratio of the price index 2 quarters prior to the current quarter and the price index at origination (the claim process is fairly lengthy for FHA loans). Additionally, 3 time splines constitute the baseline hazard; the nodes for the splines are at 2 years, and 3.5 years.

The second strategy was designed to control for more covariates, despite the relatively small sample size (less than 4,000 in the national sample and 1,116 in the Atlanta sample[12], after dropping loans that could not be valued by the AVM). In 2001, GAO estimated competing risk hazard models using millions of FHA loans originated between 1975 and 1999[13]. Explanatory variables for credit risk included LTV at origination, an estimate of contemporaneous LTV, geographic controls for Census division and judicial foreclosure states, contemporaneous unemployment rates, and, for ARM loans, changes in payments over time. A similar spline structure was used for the baseline hazard. Separate models were run for 30-year fixed, investor, 15-year fixed, and ARM loans. The coefficients from this prior study were combined with the Concentrance data to form a mortgage score, and this score (GAOrisk), was used as an independent variable along

[10]FHA releases general information about TOTAL, such as the characteristics that influence the score, but regards precise details, such as the definition of the reserves or FICO score variables, or functional form, as proprietary.
[11]About 8% of the borrowers did not have a FICO score. For these cases, the median FICO score for the sample was inserted, and a dummy variable (NOFICO) was set to 1. The results, therefore, show the extent to which borrowers without a FICO score are riskier than borrowers with a median score.
[12]The Atlanta sample consists of just over 1,000 records from the MSA sample and about 150 records from the National sample that represented loans in the Atlanta MSA.
[13]The model is documented in US GAO (2001).

with important variables not in the GAO model, such as FICO score and reserves.

In order to estimate the effect of the source of the down payment on claim and delinquency propensities, the instantaneous conditional claim (or delinquency) rate was modeled using James Heckman's CTM program (Yi, Walker and Honoré, 1985). Prepaid loans were treated as censored on the date of prepayment. The hazard rate framework was chosen to allow for the inclusion of time varying covariates, such as post origination price appreciation.

CTM (Continuous Time Models) is a Fortran based package with a long history in labor econometrics. It estimates competing risk termination models with a flexible (Box-Cox) parametric baseline hazard, and allows for the choice of any of several parametric forms of unobserved heterogeneity, or Heckman-Singer non-parametric heterogeneity (Yi, Walker, and Honoré 1985). Unobserved heterogeneity is usually referred to in mortgage modeling as “burnout” - the tendency for some loans to terminate faster than observationally similar loans, so that conditional termination rates fall over time, despite unchanging conditions. Essentially, borrowers who are “slow terminators” for some reason not observed by the econometrician remain in the pool after all the “fast terminators” have left.

CTM was first applied to mortgage analysis in GAO's third report on the actuarial soundness of the FHA single family program (GAO 1996), and has also been used to model FHA multifamily mortgage terminations (Ondrich and Huang 2002). Regressions incorporating unobserved heterogeneity have also been estimated with other routines. For example, Stanton (1995) estimates a single termination risk model of prepayment with a gamma heterogeneity distribution, and Deng, Quigley, and VanOrder (2000) estimate a competing risk model with Heckman-Singer non-parametric heterogeneity using McCall's software program[14].

CTM estimates an equation of the form

$$h_{ij}(t_{ij}\ \{x(u)\}_0^{\infty}, \theta) = \exp\{\gamma_{ij0} + \Sigma(t_{ij} + \tau_{ijk})\beta_{ijk} + \tau_{ij}(t^{\tau} - 1)/\tau \quad + c_{ij}\theta$$

where i indexes the origination state (active loan), j indexes the destination state, default or prepayment, t is time (measured in days divided by 100), tau and beta are independent variables and their coefficients, lambda is the Box-Cox parameter for the baseline hazard, and the c's and thetas are the points of support for the non-parametric heterogeneity distribution and their coefficients (factor loadings).

The final regressions were of the form

$$L(Default_{\,t,\,t+1}/Survivor_t) = Risk\ Covariates_t,\ AVM\ Ratio,\ AVM\ Confidence,\ Appraise\ Ratio,\ Time,\ Heterogeneity) \qquad (8)$$

[14] A non parametric baseline with competing risks and unobserved heterogeneity, as in McCall’s program, has to be estimated with some care, as unreliable results may be obtained from singularities. See Ridder and Woutersen (2003).

5. Estimation Results

Tables II.1, II.2, II.3, and II.4 present logistic regression results for the national sample, with 90 day delinquency[15] as the dependent variable. The next set of tables, III.1, III.2, III.3, and III.4 show results with claim as the dependent variable. For each set of tables, there are two specifications, one using the set of regressors used by FHA in its TOTAL scorecard, and one using the GAOrisk variable, both augmented with extra variables. Each specification is repeated twice, once for all observations with an AVM result, and once for observations in disclosure states only. The first columns of results in each table show regressions with just the Appraisal ratio, the second with just the AVM ratio, while the third column shows results when both variables are entered, and the last shows the results when both variables are entered, along with the confidence level associated with the AVM.

At the bottom of each table is the distribution of the heterogeneity parameters. At least two points of support are estimated. The location of the first point of support is fixed at zero, and that of the second is fixed at one. Any additional points of support are constrained to lie in this interval. Along with the location of (additional) points of support, the cumulative probability to that point is estimated. There is no formal way to determine the number of heterogeneity parameters to include in an estimation; usual practice is to start with two points of support, and continue adding until the convergence routine fails. With the national termination samples, 3 points of support were the most that could be estimated.[16] When the number of observations were limited, as in the Atlanta sample, it was sometimes the case that only two points of support could be estimated. While CTM jointly estimates the default parameters, the prepayment parameters, and the heterogeneity distribution, in the interest of space the prepayment parameters are presented only for the national sample, and later in the paper.

[15]The dependent variable indicates 90-day delinquency, or other "bad outcomes" such as the initiation of foreclosure proceedings or a loss mitigation foreclosure alternative. Although lenders are required to report delinquencies to FHA after 90 days, sometimes a delinquency is never reported but the loan appears as a claim or claim alternative. In over 90% of the "delinquencies" in this file, the event is 90-day delinquency.
[16]The same number of support points are in the GAO 1996 model and Denq, Quigley, VanOrder (2000).

Table II.1 **Delinquency**
National All State Sample / GAORISK Specification

Parameter	Estimate	Std. Error	T	Estimate	Std. Error	T	Estimate	Std. Error	T
intercept	6.122	0.610	10.039	6.665	0.663	10.059	6.663	0.667	9.989
gamma	0.166	0.102	1.633	0.153	0.096	1.593	0.153	0.096	1.593
lambda	-1.418	0.440	3.221	-1.463	0.448	3.265	-1.462	0.448	3.264
GAOrisk	-0.022	0.029	0.758	-0.018	0.030	0.595	-0.018	0.030	0.598
FICO	-1.242	0.081	15.371	-1.249	0.081	15.484	-1.249	0.081	15.456
frontend	1.763	0.607	2.905	1.803	0.607	2.969	1.805	0.610	2.958
noFICO	0.617	0.140	4.394	0.625	0.141	4.439	0.625	0.141	4.437
reserve	-0.014	0.097	0.148	-0.004	0.097	0.041	-0.004	0.097	0.040
underserved	0.057	0.089	0.644	0.049	0.089	0.550	0.048	0.089	0.546
firsttime	0.029	0.133	0.218	0.036	0.133	0.268	0.036	0.134	0.268
condo	-0.130	0.204	0.637	-0.098	0.204	0.482	-0.098	0.204	0.481
factor_loading	-0.162	0.356	0.456	-0.235	0.360	0.652	-0.234	0.361	0.650
Appraise_ratio	-0.562	0.833	0.674	0.000	0.000	0.000	0.041	0.884	0.046
AVM_ratio	0.000	0.000	0.000	-0.487	0.209	2.334	-0.491	0.217	2.266
AVMconfidence	0.000	0.000	0.000	-0.655	0.283	2.314	-0.654	0.284	2.302
Heterogeneity Distribution									
	Cum_Prob	**Location**		**Cum_Prob**	**Location**		**Cum_Prob**	**Location**	
	0.15	0		0.14	0		0.14	0	
	0.48	0.62		0.46	0.61		0.46	0.6	
	1	1		1	1		1	1	

Table II.2 **Delinquency**
National All State Sample / TOTAL Scorecard Specification

Parameter	Estimate	Std. Error	T	Estimate	Std. Error	T	Estimate	Std. Error	T
intercept	10.765	4.145	2.597	10.490	4.146	2.530	10.506	4.210	2.496
gamma	0.878	0.216	4.064	0.839	0.210	4.001	0.839	0.210	3.999
lambda	-0.357	0.227	1.570	-0.387	0.230	1.683	-0.387	0.230	1.680
FICO	-1.252	0.081	15.468	-1.254	0.081	15.513	-1.254	0.081	15.505
frontend	2.347	0.618	3.794	2.332	0.616	3.783	2.334	0.619	3.772
noFICO	0.598	0.141	4.251	0.603	0.141	4.287	0.603	0.141	4.284
reserve	-0.003	0.098	0.028	0.004	0.098	0.041	0.004	0.098	0.043
underserved	0.089	0.089	1.002	0.079	0.089	0.890	0.079	0.089	0.885
ARM	-0.353	0.190	1.857	-0.362	0.191	1.897	-0.362	0.191	1.895
firsttime	0.046	0.133	0.344	0.052	0.133	0.395	0.052	0.133	0.394
condo	-0.020	0.219	0.093	0.010	0.220	0.045	0.010	0.220	0.044
growth	-4.424	0.732	6.046	-4.403	0.733	6.005	-4.403	0.734	5.997
LTV	0.142	4.069	0.035	0.878	4.068	0.216	0.859	4.141	0.207
factor_loading	0.034	0.384	0.089	-0.093	0.385	0.243	-0.091	0.385	0.237
Appraise_ratio	-0.649	0.850	0.764	0.000	0.000	0.000	0.050	0.896	0.056
AVM_ratio	0.000	0.000	0.000	-0.542	0.213	2.543	-0.547	0.220	2.490
AVMconfidence	0.000	0.000	0.000	-0.497	0.280	1.775	-0.496	0.281	1.764
Heterogeneity Distribution									
	Cum_Prob	**Location**		**Cum_Prob**	**Location**		**Cum_Prob**	**Location**	
	0.15	0		0.15	0		0.14	0	
	0.44	0.61		0.42	0.6		0.42	0.6	
	1	1		1	1		1	1	

Table II.3 **Delinquency**
National Disclosure State Sample / GAORISK Specification

Parameter	Estimate	Std. Error	T	Estimate	Std. Error	T	Estimate	Std. Error	T
intercept	6.180	0.682	9.067	6.774	0.751	9.017	6.752	0.756	8.933
gamma	0.142	0.100	1.429	0.127	0.092	1.375	0.131	0.094	1.395
lambda	-1.514	0.486	3.115	-1.581	0.499	3.169	-1.564	0.494	3.170
GAOrisk	-0.034	0.037	0.925	-0.033	0.038	0.885	-0.035	0.038	0.928
FICO	-1.262	0.089	14.109	-1.267	0.089	14.194	-1.266	0.089	14.176
frontend	1.993	0.664	3.000	2.021	0.666	3.035	2.046	0.670	3.054
noFICO	0.615	0.152	4.042	0.621	0.153	4.070	0.617	0.153	4.041
reserve	-0.016	0.107	0.148	-0.004	0.107	0.042	-0.002	0.107	0.021
underserved	0.056	0.096	0.581	0.061	0.096	0.637	0.059	0.096	0.614
firsttime	0.061	0.149	0.411	0.066	0.149	0.445	0.066	0.150	0.440
condo	-0.221	0.225	0.983	-0.188	0.226	0.832	-0.186	0.226	0.822
factor_loading	-0.082	0.370	0.222	-0.200	0.371	0.538	-0.181	0.373	0.485
Appraise_ratio	-0.271	0.838	0.323	0.000	0.000	0.000	0.460	0.911	0.506
AVM_ratio	0.000	0.000	0.000	-0.501	0.237	2.113	-0.554	0.248	2.229
AVMconfidence	0.000	0.000	0.000	-0.674	0.314	2.147	-0.663	0.315	2.106
Heterogeneity Distribution									
	Cum_Prob	**Location**		**Cum_Prob**	**Location**		**Cum_Prob**	**Location**	
	0.14	0		0.13	0		0.13	0	
	0.48	0.62		0.45	0.61		0.45	0.61	
	1	1		1	1		1	1	

Table II.4 **Delinquency**
National Disclosure State Sample / TOTAL Scorecard Specification

Parameter	**Estimate**	**Std. Error**	**T**	**Estimate**	**Std. Error**	**T**	**Estimate**	**Std. Error**	**T**
intercept	8.092	4.855	1.667	7.695	4.863	1.583	7.798	4.926	1.583
gamma	0.921	0.240	3.832	0.846	0.231	3.661	0.847	0.231	3.663
lambda	-0.325	0.239	1.36	-0.380	0.248	1.53	-0.380	0.248	1.53
FICO	-1.277	0.089	14.29	-1.274	0.089	14.31	-1.273	0.089	14.3
frontend	2.642	0.674	3.919	2.600	0.673	3.863	2.612	0.676	3.862
noFICO	0.590	0.153	3.867	0.591	0.152	3.879	0.588	0.153	3.856
reserve	0.002	0.108	0.015	0.010	0.108	0.090	0.011	0.108	0.103
underserved	0.064	0.096	0.670	0.068	0.096	0.704	0.066	0.096	0.689
ARM	-0.397	0.200	1.99	-0.394	0.201	1.97	-0.393	0.201	1.96
firsttime	0.079	0.151	0.526	0.086	0.150	0.574	0.086	0.150	0.571
condo	-0.063	0.244	0.26	-0.030	0.246	0.12	-0.031	0.246	0.13
growth	-4.703	0.796	5.91	-4.567	0.795	5.74	-4.561	0.796	5.73
LTV	3.259	4.767	0.684	3.906	4.761	0.821	3.778	4.839	0.781
factor_loading	0.041	0.412	0.100	-0.111	0.410	0.27	-0.103	0.410	0.25
Appraise_ratio	-0.476	0.849	0.56	0.000	0.000	0.000	0.253	0.915	0.276
AVM_ratio	0.000	0.000	0.000	-0.498	0.242	2.06	-0.529	0.253	2.1
AVMconfidence	0.000	0.000	0.000	-0.438	0.307	1.43	-0.433	0.308	1.41
Heterogeneity Distribution									
	Cum_Prob	**Location**		**Cum_Prob**	**Location**		**Cum_Prob**	**Location**	
	0.14	0		0.13	0		0.13	0	
	0.43	0.62		0.41	0.60		0.41	0.6	
	1	1		1	1		1	1	

Table II.5 **Delinquency**
Atlanta MSA / GAORISK Specification

Parameter	Estimate	Std. Error	T	Estimate	Std. Error	T	Estimate	Std. Error	T
intercept	5.453	1.096	4.975	5.467	1.166	4.690	5.462	1.175	4.649
gamma	0.127	0.143	0.884	0.135	0.152	0.893	0.130	0.147	0.881
lambda	-1.709	0.802	2.130	-1.669	0.797	2.093	-1.696	0.804	2.110
GAOrisk	-0.055	0.052	1.065	-0.057	0.052	1.109	-0.057	0.052	1.102
FICO	-1.210	0.151	8.008	-1.213	0.149	8.144	-1.210	0.152	7.981
frontend	3.825	1.133	3.375	3.772	1.133	3.328	3.787	1.134	3.340
noFICO	0.712	0.229	3.106	0.715	0.230	3.104	0.713	0.231	3.084
reserve	-0.218	0.182	1.197	-0.217	0.182	1.196	-0.216	0.182	1.190
underserved	0.200	0.146	1.367	0.191	0.148	1.292	0.192	0.148	1.299
firsttime	-0.061	0.213	0.287	-0.064	0.213	0.302	-0.063	0.213	0.295
condo	0.548	0.348	1.576	0.537	0.346	1.552	0.544	0.347	1.569
factor_loading	0.488	0.446	1.095	0.517	0.450	1.151	0.493	0.453	1.089
Appraise_ratio	-0.944	2.265	0.417	0.000	0.000	0.000	-0.658	2.305	0.286
AVM_ratio	0.000	0.000	0.000	-0.344	0.554	0.621	-0.317	0.560	0.566
AVMconfidence	0.000	0.000	0.000	0.043	0.567	0.076	0.038	0.573	0.067
Heterogeneity Distribution									
	Cum_Prob	**Location**		**Cum_Prob**	**Location**		**Cum_Prob**	**Location**	
	0.37	0		0.37	0		0.37	0	
	1	1		1	1		1	1	

Table III.1 Claim Rate
National All State Sample / GAORISK Specification

Parameter	Estimate	Std.Error	T	Estimate	Std. Error	T	Estimate	Std. Error	T
intercept	1.657	1.096	1.512	3.115	1.184	2.630	3.170	1.198	2.646
gamma	0.376	0.273	1.378	0.367	0.267	1.374	0.393	0.275	1.430
lambda	-2.167	0.808	2.681	-2.176	0.816	2.668	-2.112	0.791	2.669
GAOrisk	0.208	0.043	4.790	0.216	0.047	4.638	0.220	0.045	4.848
FICO	-0.904	0.141	6.417	-0.947	0.142	6.685	-0.949	0.141	6.713
frontend	2.128	0.984	2.164	2.579	0.983	2.624	2.423	0.997	2.429
noFICO	0.929	0.210	4.423	0.918	0.215	4.271	0.943	0.214	4.397
reserve	-0.153	0.168	0.914	-0.113	0.169	0.668	-0.127	0.170	0.746
underserved	-0.124	0.148	0.837	-0.152	0.150	1.018	-0.144	0.149	0.967
firsttime	-0.162	0.207	0.784	-0.159	0.209	0.761	-0.153	0.207	0.739
condo	-0.246	0.393	0.626	-0.091	0.399	0.228	-0.115	0.397	0.290
factor_loading	-1.067	0.860	1.241	-1.193	0.883	1.351	-1.017	0.852	1.194
Appraise_ratio	-4.767	2.721	1.752	0.000	0.000	0.000	-4.460	2.743	1.626
AVM_ratio	0.000	0.000	0.000	-1.087	0.422	2.574	-0.797	0.430	1.852
AVMconfidence	0.000	0.000	0.000	-1.873	0.479	3.912	-1.947	0.482	4.037

Heterogeneity Distribution

	Cum_Prob	Location		Cum_Prob	Location		Cum_Prob	Location	
	0.17	0		0.19	0		0.19	0	
	0.47	0.58		0.48	0.59		0.49	0.59	
	1	1		1	1		1	1	

Table III.2 **Claim Rate**
National All State Sample / TOTAL Scorecard Specification

Parameter	Estimate	Std. Error	T	Estimate	Std. Error	T	Estimate	Std. Error	T
intercept	11.410	8.840	1.291	14.125	8.982	1.572	11.374	8.783	1.295
gamma	1.623	0.373	4.349	1.558	0.361	4.310	1.589	0.368	4.319
lambda	-0.853	0.344	2.481	-0.881	0.351	2.510	-0.864	0.350	2.472
FICO	-0.942	0.137	6.886	-0.966	0.137	7.060	-0.965	0.137	7.071
frontend	3.187	1.027	3.104	3.375	1.014	3.320	3.346	1.021	3.277
noFICO	0.888	0.207	4.290	0.875	0.211	4.140	0.885	0.211	4.198
reserve	-0.107	0.168	0.638	-0.081	0.168	0.480	-0.088	0.169	0.520
underserved	-0.038	0.149	0.252	-0.059	0.149	0.400	-0.051	0.149	0.343
ARM	-0.928	0.397	2.339	-0.919	0.390	2.350	-0.933	0.395	2.360
firsttime	-0.156	0.210	0.744	-0.138	0.209	0.660	-0.137	0.208	0.657
condo	-0.201	0.430	0.467	-0.116	0.435	0.266	-0.084	0.436	0.192
growth	-4.603	0.866	5.314	-4.513	0.874	5.160	-4.497	0.872	5.154
LTV	-2.940	8.814	0.334	-4.525	8.935	0.506	-1.758	8.764	0.201
factor_loading	-0.310	0.821	0.377	-0.530	0.831	0.637	-0.385	0.817	0.472
Appraise_ratio	-4.340	2.858	1.518	0.000	0.000	0.000	-3.867	2.895	1.336
AVM_ratio	0.000	0.000	0.000	-0.978	0.433	2.260	-0.777	0.438	1.773
AVMconfidence	0.000	0.000	0.000	-1.479	0.470	3.150	-1.511	0.473	3.195
Heterogeneity Distribution									
	Cum_Prob	**Location**		**Cum_Prob**	**Location**		**Cum_Prob**	**Location**	
	0.15	0		0.15	0		0.15	0	
	0.43	0.59		0.42	0.58		0.42	0.58	
	1	1		1	1		1	1	

Table III.3 **Claim Rate**
National Disclosure State Sample / GAORISK Specification

Parameter	Estimate	Std. Error	T	Estimate	Std. Error	T	Estimate	Std. Error	T
intercept	2.054	1.296	1.585	4.054	1.460	2.778	4.155	1.510	2.751
gamma	0.103	0.207	0.496	0.101	0.203	0.498	0.101	0.204	0.495
lambda	-4.128	2.548	1.620	-4.143	2.567	1.614	-4.135	2.585	1.599
GAOrisk	0.223	0.058	3.831	0.204	0.065	3.150	0.221	0.065	3.391
FICO	-1.028	0.165	6.217	-1.063	0.167	6.349	-1.077	0.170	6.346
frontend	2.694	1.095	2.461	3.305	1.103	2.997	3.102	1.126	2.756
noFICO	1.085	0.233	4.667	1.053	0.236	4.467	1.094	0.238	4.603
reserve	-0.059	0.185	0.319	0.009	0.189	0.047	-0.010	0.189	0.052
underserved	-0.134	0.165	0.811	-0.122	0.168	0.726	-0.126	0.167	0.754
firsttime	-0.105	0.237	0.441	-0.130	0.240	0.539	-0.104	0.238	0.437
condo	-0.459	0.472	0.972	-0.266	0.478	0.556	-0.279	0.476	0.585
factor_loading	-1.116	0.875	1.275	-1.327	0.921	1.440	-1.272	0.914	1.392
Appraise_ratio	-5.467	3.048	1.794	0.000	0.000	0.000	-5.139	3.102	1.657
AVM_ratio	0.000	0.000	0.000	-1.287	0.481	2.675	-0.906	0.490	1.848
AVMconfidence	0.000	0.000	0.000	-2.404	0.544	4.423	-2.526	0.560	4.512
Heterogeneity Distribution									
	Cum_Prob	**Location**		**Cum_Prob**	**Location**		**Cum_Prob**	**Location**	
	0.17	0		0.16	0		0.17	0	
	0.46	0.57		0.46	0.57		0.46	0.57	
	1	1		1	1		1	1	

Table III.4 **Claim Rate**
National Disclosure State Sample / TOTAL Scorecard Specifications

Parameter	Estimate	Std. Error	T	Estimate	Std. Error	T	Estimate	Std. Error	T
intercept	9.481	10.399	0.912	14.725	10.104	1.457	10.257	10.215	1.004
gamma	1.367	0.437	3.130	0.979	0.390	2.513	1.208	0.414	2.919
lambda	-1.284	0.554	2.319	-1.662	0.680	2.444	-1.421	0.600	2.366
FICO	-1.045	0.160	6.548	-1.081	0.168	6.434	-1.068	0.160	6.680
frontend	3.917	1.159	3.380	4.040	1.149	3.516	4.121	1.147	3.592
noFICO	1.051	0.229	4.583	1.059	0.243	4.362	1.041	0.232	4.485
reserve	0.014	0.188	0.076	0.055	0.192	0.286	0.037	0.190	0.194
underserved	-0.057	0.166	0.344	-0.028	0.169	0.165	-0.039	0.166	0.238
ARM	-0.896	0.428	2.095	-0.810	0.424	1.912	-0.802	0.430	1.865
firsttime	-0.132	0.248	0.531	-0.094	0.246	0.385	-0.101	0.241	0.420
condo	-0.403	0.512	0.786	-0.313	0.522	0.600	-0.242	0.517	0.468
growth	-4.712	0.919	5.127	-4.048	0.907	4.463	-4.374	0.925	4.728
LTV	-0.389	10.348	0.038	-4.389	9.961	0.441	0.191	10.187	0.019
factor_loading	-0.416	0.890	0.467	-1.412	1.058	1.330	-0.688	0.893	0.770
Appraise_ratio	-4.722	3.250	1.453	0.000	0.000	0.000	-4.401	3.282	1.341
AVM_ratio	0.000	0.000	0.000	-0.995	0.481	2.068	-0.709	0.492	1.441
AVMconfidence	0.000	0.000	0.000	-2.636	0.544	4.849	-2.164	0.536	4.042
Heterogeneity Distribution									
	Cum_Prob	**Location**		**Cum_Prob**	**Location**		**Cum_Prob**	**Location**	
	0.15	0		0.17	0		0.14	0	
	0.42	0.58		0.46	0.56		0.41	0.57	
	1	1		1	1		1	1	

Table III.5 **Claim Rate**
Atlanta MSA / GAORISK Specification

Parameter	Estimate	Std. Error	T	Estimate	Std. Error	T	Estimate	Std. Error	T
intercept	-4.485	2.439	1.839	-1.787	1.871	0.955	-1.571	1.827	0.860
gamma	0.965	0.418	2.309	0.995	0.648	1.536	1.049	0.661	1.587
lambda	-1.412	0.818	1.725	-1.481	1.029	1.439	-1.435	0.995	1.442
GAOrisk	0.447	0.094	4.772	0.340	0.074	4.603	0.337	0.073	4.608
FICO	-0.564	0.212	2.659	-0.581	0.212	2.742	-0.594	0.214	2.781
frontend	1.839	1.575	1.168	1.509	1.524	0.990	1.414	1.531	0.924
noFICO	0.707	0.312	2.266	0.560	0.312	1.798	0.569	0.314	1.808
reserve	-0.169	0.252	0.672	-0.143	0.246	0.581	-0.135	0.248	0.542
underserved	0.062	0.212	0.294	0.050	0.211	0.238	0.033	0.213	0.153
firsttime	0.106	0.300	0.351	0.079	0.300	0.265	0.075	0.302	0.248
condo	0.166	0.533	0.311	0.270	0.490	0.552	0.235	0.506	0.464
factor_loading	2.398	1.522	1.575	1.106	0.768	1.440	1.183	0.772	1.533
Appraise_ratio	0.737	3.271	0.225	0.000	0.000	0.000	2.292	2.734	0.839
AVM_ratio	0.000	0.000	0.000	-0.850	0.849	1.001	-1.000	0.872	1.147
AVMconfidence	0.000	0.000	0.000	-0.149	0.837	0.183	-0.146	0.857	0.171
Heterogeneity Distribution									
	Cum_Prob	**Location**		**Cum_Prob**	**Location**		**Cum_Prob**	**Location**	
	0.14	0		0.37	0		0.38	0	
	1	1		1	1		1	1	

Signs were as expected for both the appraisal and AVM variables. When the dependent variable was 90-day delinquency, the appraisal ratio had the expected sign but a fairly low significance level. The AVM ratio had the correct sign, and was always significant at 5 percent in one-tailed tests in the national samples. When the dependent variable was the claim hazard, AVM, considered separately, was always significant in one-tailed tests at 5 percent, and Appraisal was often significant. When entered together, the AVM variable was predictive of claim rates at one-tailed significance levels of 0.02 to 0.08, depending on the specification and sample. The Appraisal variables are sometimes significant. There is a modest degree of multicollinearity between the appraisal and AVM ratios; the r-squared between them is about 0.4. The AVM ratio had higher significance, but a slightly smaller standardized Beta. The coefficients on the Appraisal ratio were about 6 times the magnitude of the AVM ratio coefficients, but the AVM ratio's standard deviation was 4 times that of the Appraisal ratio, 6 percentage points for Appraisal vs. 23 percentage points for the AVM ratio[17].

The confidence score associated with the AVM was also highly significant, indicating that harder to value properties have higher claim rates. This result may be consistent with the work of Rachlis (1992) who hypothesized that neighborhoods with few transactions would result in appraisal problems that could lead to higher risk and less lending activity. The result is not consistent with Rachlis' hypothesis if lenders rationed credit in such neighborhoods on unobservables, but is consistent with lenders rationing on observables that are in the claim regression equation, and may also be consistent with lenders rationing in such neighborhoods with conventional loans, but not with FHA loans where the government, not the lender (or PMI), bears the bulk of the credit risk.

Results were mostly favorable for other covariates. The FICO score has a very strong effect with the expected sign, as does the Front end ratio, and the measure of post-origination price appreciation. The GAOrisk variable is also positive and highly significant for claim, but not delinquency, prediction. Significance levels and goodness of fit statistics are generally better for the specification using the GAOrisk variable, indicating the usefulness of capturing risk characteristics using a mortgage score in small samples where including a large number of covariates might not be feasible. LTV is not significant, presumably because there is so little variation in LTV in this sample of very high LTV loans. Reserves are also not significant; again few FHA borrowers have significant reserves after closing. The indicators for first-time homebuyers, condominium loans, and loans in underserved areas were not significant, but there was no theoretical expectation for a particular sign for these variables.

The heterogeneity results are similar to those found in GAO (1996) or Deng, Quigley, and VanOrder (2000). For the national sample, the model estimates that there are three categories of borrowers, with about 10% to 20% in the very slow prepayment category, about 30% to 40% in the medium speed prepayment category, and the remainder in the rapid prepayment category. Because the factor loadings are opposite in sign for the claim

[17]The standard deviation may overstate the degree to which appraisals differ from price, as it is driven largely by a few positive outliers. About half of appraisals are exactly the sale price, and the interquartile range is 1.9%. For AVMs, the interquartile range for the difference between AVM and transaction price is 17.5%, or 9 times as large. The AVM and Appraisal ratios were truncated at +/- 50% of price.

and prepayment regressions, borrowers who are fast prepayers are predicted to be slow claim terminators, a result consistent with adverse selection at time of prepayment. The Box-Cox baseline hazard parameter, lambda, is negative and generally about -1, implying that a baseline of the form 1/time gives the best fit to the data, a remarkably sensible form for the baseline, as it allows a rapidly rising hazard in the early part of a loan's life followed by an essentially flat hazard. Except for GAO (1996) which finds a similar form, to the best of my knowledge no one has used such an inverse transform for a baseline mortgage termination hazard.

One potential disadvantage to working with conditional hazard rates is the potential for the competing risk of prepayment to influence the default regression results. It would be possible, for example, for appraisal or AVM estimates to have an impact on conditional claim rates, but not on unconditional claim rates, if low appraisals or AVM estimates resulted in higher prepayment rates. The conditional claim rates would be high, not because claims were high, but because survival was low. To test for this possibility, the conditional prepayment rate was also modeled as a function of standard prepayment variables, such as the ratio of book-to-market value of the mortgage (splined at 1), standard underwriting variables, and appraisal and AVM estimate variables. The logistic regression results for prepayment are in Table IV.1. The AVM ratio has a statistically significant impact on prepayment rates, but the coefficients on both AVM and Appraisal ratios are small in magnitude, indicating that the conditional claim rate results reflect higher claims, and not merely lower survival. To establish this, a simulation was run in which the average conditional claim rates and conditional prepayment rates for the full national sample were integrated over 5 years to produce a 5-year claim rate. The coefficients from the claim regressions were used to adjust the conditional claim rates upward, assuming that the AVM estimate was 10% below the sale price. For example, using the full sample GAOrisk coefficient of 0.7, the 5-year claim rate increased by 6.5%. When the AVM Ratio coefficient from the prepayment regression was also used to adjust the conditional prepayment rate upward, the 5-year claim rate rose by 4.5%. Therefore, the increase in prepayment rate acted as a partial offset to the elevated conditional claim rates.

Table IV.1 All	**Prepay**								
Parameter	**Estimate**	**Std. Error**	**T**	**Estimate**	**Std. Error**	**T**	**Estimate**	**Std. Error**	**T**
intercept	-17.05	1.12	15.22	-17.49	1.11	15.75	-17.45	1.11	15.79
gamma	0.36	0.08	4.52	0.37	0.08	4.48	0.38	0.08	4.47
lambda	-0.88	0.14	6.22	-0.88	0.14	6.20	-0.87	0.14	6.09
GAOrisk	-0.05	0.02	2.08	-0.05	0.02	2.13	-0.05	0.02	2.08
FICO	0.39	0.05	7.81	0.41	0.05	8.21	0.41	0.05	8.24
frontend	1.45	0.40	3.61	1.30	0.40	3.23	1.30	0.40	3.23
noFICO	-0.31	0.10	3.05	-0.34	0.10	3.34	-0.34	0.10	3.31
reserve	0.09	0.06	1.38	0.08	0.06	1.21	0.08	0.06	1.22
underserved	-0.23	0.06	3.76	-0.21	0.06	3.51	-0.21	0.06	3.50
ARM	0.95	0.14	6.57	0.89	0.14	6.24	0.89	0.14	6.22
firsttime	-0.30	0.08	3.88	-0.29	0.08	3.42	-0.29	0.08	3.81
Condo	0.01	0.12	0.11	0.08	0.12	0.63	0.08	0.12	0.63
releqphi	4.32	0.29	14.91	4.24	0.29	14.61	4.26	0.29	14.62
releqplo	7.31	0.83	8.78	7.17	0.83	8.59	7.14	0.84	8.54
Factor_loading	3.78	0.48	7.81	3.60	0.43	8.44	3.56	0.39	9.19
Appraise_ratio	0.14	0.59	0.22	0.00	0.00	0.00	0.50	0.63	0.79
AVM_ratio	0.00	0.00	0.00	-0.25	0.17	1.45	-0.29	0.18	1.60
AVMconfidence	0.00	0.00	0.00	1.07	0.20	5.42	1.06	0.20	5.37

Table V.1	All		Disclose		Atlanta	
	rsq=.185 n=163		Rsq=.215 N=129		Rsq=.166 N=83	
Variable	Estimate	T	Estimate	T	Estimate	T
intercept	1.474	0.640	1.930	0.700	1.380	0.570
LTV	-0.656	0.270	-1.206	0.410	-1.011	0.400
FICO	-0.023	0.730	-0.021	530	-0.025	0.900
noFICO	0.094	1.920	0.100	1.710	-0.016	0.370
appreciation	-0.324	1.540	-0.331	1.380	0.152	0.560
noterate	0.032	1.460	0.043	1.680	-0.008	0.470
Log origination$	-0.002	4.990	-0.003	4.910	-0.001	2.170
Appraise_ratio	-0.183	0.300	-0.063	0.080	0.539	0.800
AVM_ratio	-0.193	2.610	-0.228	2.490	-0.346	2.920
AVMconfidence	0.300	0.030	0.079	0.620	0.138	1.190

Turning to loss given default, OLS regressions indicate that loss rates, defined as the dollars lost on a defaulted loan divided by the original mortgage balance, are a function of the AVM estimate of the property value (Table V.1). In the national samples, the higher is the AVM estimate, the lower is FHA's percentage lost. Original mortgage amount and post-origination price appreciation are also significant determinants of losses, with smaller losses in faster appreciating states, and smaller (percent) losses on larger loans, consistent with a substantial fixed cost component of total losses (foreclosure costs, for example). Appraisal differences have the right sign, but are small in magnitude and never close to significance.

In the Atlanta sample, neither the AVM nor the Appraisal ratios are a significant predictor of 90-day delinquency, although the AVM ratio always gets the right sign (see Table II.5).[18] Neither ratio is a significant predictor of claim rates, although the AVM ratio gets the right sign with an asymptotic T statistic near 1 (Table III.5). In Atlanta, the AVM ratio is also a significant predictor of loss rates, while the appraisal difference is not (Table V.1). Presumably the small number of claims in the Atlanta sample (107 out of a sample of 1116) limits the ability of any model to predict claim rates or losses effectively.

6. Conclusions

AVM estimates are predictive of both claim and delinquency propensities. Appraisal ratios also have predictive power for claims. Examined separately, each is useful as a predictor of the claim propensity of a mortgage. Entered together, the correlation between the two estimates is weak enough that each serves as a useful indicator of credit

[18] In the interest of space, only the full model for the GAOrisk specification is included for the Atlanta results. Other results were similar, with AVM values predictive of risk and appraisal ratios insignificant, and with the GAOrisk specification slightly outperforming the TOTAL scorecard specification.

risk, although the significance levels are higher on the AVM estimate when both are in the regression.

The confidence measure attached to the AVM estimate also serves as a predictor of credit risk. Properties that are easier to value have lower credit risk, even after conditioning on a host of standard underwriting variables. Additionally, AVM estimates are a significant predictor of loss given default, an important but often ignored dimension of credit risk.

Much of the value of an appraisal presumably comes prior to origination, in preventing transactions at prices far above market value, or contributing to the renegotiation of price prior to closing. The results here should not be taken to imply that appraisals have less value than AVMs, only that appraisal values have less post-origination predictive power than do AVMs.

These results confirm the value of econometric estimates of property value first found by LaCour-Little and Malpezzi, using a more recent, larger, and nationally representative sample, and focusing on claims and losses, not just delinquency. Additionally, this work demonstrates the utility of commercial, off-the-shelf, AVM estimates for predicting credit risk. Finally, the results demonstrate the usefulness of appraisal estimates in the prediction of claim propensities, over and above the information contained in AVMs.

References

Ambrose, Brent, and Capone, C. "The Hazard Rates of First and Second Defaults." *Journal of Real Estate Finance and Economics* 20, 2000 275-293

Case, Bradford, Pollakowski's, H., and Wachter, S. "On Crossing Among House Price Index Methodologies" *AREUEA Journal*, v. 19 # 1991 pp. 286-307

Horne, David K. and Rosenblatt, E. "Property Appraisals and Moral Hazard" working paper June 1996

Concentrance Consulting Group "Audit of Loans with Down Payment Assistance" prepared for the U.S. Department of Housing and Urban Development, Washington, D.C. Feb. 6, 2004

Cotterman, Robert "Analysis of FHA Single Family Default and Loss Rates" a report for the Office of Policy Development and Research, US Department of Housing and Urban Development, Washington, D.C. March 5, 2004

Danis, Michelle A. and Anthony Pennington-Cross. "A Dynamic Look at Subprime Loan Performance." Federal Reserve Bank of St. Louis. Working Paper 2005-029A. May 2005

Deng, Y.J, Quigley, M., and VanOrder, R. "Mortgage Terminations, Heterogeneity, and the Exercise of Mortgage Options." *Econometrica* 2000, pp. 275-307

Harrison, David, Noordewier, T. and Yavas A. "Do Riskier Borrowers Borrow More?" *Real Estate Economics* V. 32 #3, Fall 2004 pp. 385-411

Herzog, John, and Earley, J. "Home Mortgage Delinquency and Foreclosure" New York, NBER, 1970

Huang, Wenyi and Ondrich, J. "Pay, Stay, or Walk Away: A Hazard Rate Analysis of FHA Insured Multifamily Mortgages." *Journal of Housing Research* 13 (1) 2002 pp. 85-117

Kau, James, and Kim, T. “Waiting to Default: The Value of Delay” *AREUEA Journal*, V. 22 #3, 1994, pp.539-551
Kershaw, Paul, and Rossini, P. “Using Neural Networks to Estimate Constant Quality Home Price Indices” Fifth Annual Pacific Rim Real Estate Society Conference, Kuala Lumpur, Malaysia, January 1999
LaCour-Little, Micheal, and Malpezzi, S. “Appraisal Quality and Residential Mortgage Default: Evidence from Alaska” *Journal of Real Estate Finance and Economics*, v. 27 #2, 2003
Rachlis, Mitch “Impact of Location on Risk in Urban Markets” George Washington University dissertation, Washington, DC, 1992
Ridder, Gert and Woutersen, T. M. “The Singularity of the Information Matrix of the Mixed Proportional Hazard Model” *Econometrica* v. 71 #5, 2003, pp. 579-1590
Shiller, Robert and Weiss, A. “Evaluating Real Estate Valuation Systems” *Journal of Real Estate Finance and Economics*, V. 18 #2, 1999, pp. 147-161
Stanton, R. “Ruthless Prepayment and the Valuation of Mortgage Backed Securities.” *Review of Financial Studies* 8 (3) 1995 pp. 677-708
Technical Analysis Center, Inc. with Integrated Financial Engineering, Inc. "An Actuarial Review of the Federal Housing Administration Mutual Mortgage Insurance Fund for Fiscal Year 2005," prepared for the U.S. Department of Housing and Urban Development, Fairfax, VA: Oct. 14, 2005
US GAO, Mortgage Financing: “FHA Has Achieved Its Home Mortgage Capital Reserve Target”, GAO/RCED-96-50 (Washington, D.C. Apr. 12, 1996)
US GAO, Mortgage Financing: “FHA's Fund Has Grown, but Options for Drawing on the Fund Have Uncertain Outcomes”, GAO-01-460 (Washington, D.C. Feb. 28, 2001)
US GAO, Single Family Housing: “HUD's Risk Based Oversight of Appraisers Could be Enhanced”, GAO-05-14 (Washington, D.C. Nov. 5, 2004)
US GAO, Mortgage Financing: “Actions Needed to Help FHA Manage Risks from New Mortgage Products”, GAO-05-194 (Washington, D.C. Feb. 11, 2005)
Yi, Kei-Mu, Walker, L., and Honoré, B. “CTM: A User’s Guide” University of Chicago, 1985.

Appendix B

Taking the Lie Out of Liar Loans

Michael LaCour-Little
Finance Department
5133 Mihaylo Hall
California State University
Fullerton, CA 92834-6848
Tel 657-278-4014
Fax 657-278-2161
Email: mlacour-little@fullerton.edu

and

Jing Yang
Finance Department
5149 Mihaylo Hall
California State University
Fullerton, CA 92834-6848
Tel 657-278-2924
Fax 657-278-2161
Email: jyang@fullerton.edu

August 30, 2009

Taking the Lie Out of Liar Loans

Abstract

We examine stated income loans originated by Bear Stearns affiliates during the recent housing market run-up and market collapse. After showing the extent to which these loans have higher default rates than do fully documented loans after controlling for other risk factors, we develop a measure for the extent of likely income over-statement. We then simulate a loan origination process that rejects stated income loan applications with high degrees of likely over-statement and calculate the reduction in default rates that might have been achieved had such an algorithm been in place.

1 Introduction

The recent surge in default and foreclosure rates in the recent U.S. residential mortgage has prompted considerable research with declining housing prices and negative equity areas of particular focus[1]. In addition, policymakers and academic researchers have focused on the proliferation of risky lending contracts including adjustable-rate instruments (ARMs), subprime and Alt-A quality credit (see GAO, 2006, Campbell and Cocco, 2003, Pavlov and Wachter, 2006, and LaCour-Little and Yang, 2009). Such loans may increase default risk by inducing greater borrowing when the principal repayment obligations can be deferred (e.g. interest-only or pay-option ARMs) or when future interest rate risk is involved (e.g. ARMs in general). While lax underwriting and misaligned incentives in the mortgage securitization process have been broadly blamed for the market meltdown, the explicit effect of reduced loan documentation has not been widely studied. In contrast, a recent web-search of the term "liar loan" produces 59,000 entries, most of which appear tied to popular press and blogger accounts of the mortgage crisis.

In this paper we examine 218,000 home purchase loans originated during 2000 - 2007 and securitized by Bear Stearns and affiliates, with loan information updated by May 2009.[2] We find that percentage of loans with full documentation dropped significantly from 42.4% to 21.6% over this time period, while those with reduced documentation (including stated-income, stated-assets, no-income, no-asset, or no- ratio, more on these categories later) increased dramatically from 11.5% to 69.3 %. For loans with self-reported information on key variables (such as income or assets), which we call "stated-doc" loans, almost all are ARMs and 74% are classified as Alt-A. For loans omitting information on traditional underwriting variables, including income, assets, and front-end

[1] As of the first quarter of 2009, a record 9.1% of home mortgages were in either default or the foreclosure process (Inside Mortgage Finance, 2009).

[2] Bear Stearns collapsed and was sold to JP Morgan in March 2008.

and back-end ratios, which we will call "no-doc" loans, almost all are ARMs and 97% are Alt-A. Stated-doc loans were also an important component of the subprime loan segment. According to Inside Mortgage Finance MBS Database, "about 32 percent of subprime mortgages securitized in the first four months of 2007 were originated with stated income, no documentation or so called no ratio underwriting, in which borrower income is not considered."[3] We believe our study here is the first to explore in detail the effects of loan documentation on default risk and analyze the underlying mechanisms that drive these results.

Residential mortgage loan applicants typically encounter varied documentation requirements with subtle differences across lenders and loan programs[4]. The following are some major categories: (1) "full doc", the strictest type under which the borrower must provide proof of income and perhaps income tax returns, W-2's, paycheck with YTD earnings information, verification of employment, and evidence of assets; (2) "lite doc", a common requirement for subprime loans in which the borrower provides bank statements to document income in lieu of paycheck stubs, W-2's, and 1040's, as well as a proof for employment; (3) SIVA (stated income/verified assets) in which borrower income is stated on the loan application, employment is verified, and assets are verified; (4) SISA (stated income/stated assets), where both income and assets are stated but not verified while employment is verified; (5) NORA (no ratio) in which employment and assets are verified, income appears on the application but debt ratios are not calculated; (6) NINA (no income/no asset) in which neither income nor assets are listed on the application but employment is still verified; and (7) "no doc", for which assets, income, and employment are all omitted from the loan application.

Our focus here is the stated-income category (amusingly referred to as "self certified" in the U.K.).

[3]See article "Regulators Keep Pressing for Tougher Standards In Subprime Market; Stated Income Under Fire" Inside Mortgage Finance 24 (21), May 25, 2007.

[4]A "loan program" may reflect a particular investor's underwriting standards so that, for example, an application for a "low doc" loan slated for sale to Freddie or Fannie may be quite different from one slated for sale to a private-label conduit.

Traditionally, these loans were intended for self employed borrowers or those with seasonable income for which income is hard to document or verify. In the U.K. context, lenders reportedly sometimes also find it efficient to "fast track" borrowers with high credit scores and professional occupations with self certification (Cohen, 2009). In addition, lenders may have incentives to encourage brokers to solicit stated-income loans because such loans may produce "excessive rates and penalties" (Harney, 2009). The gist of this argument is that reduced documentation allowed unqualified borrowers to enter the market and/ or allowed those with marginal credit to qualify for unaffordable amounts, contributing to the elevated default rates we now observe.

The academic research on the topic of loan documentation has been limited. LaCour-Little (2007) confirms the traditional relationship posited between self-employment and use of reduced documentation loan programs using single-lender data from 2002; however, subsequent loan performance is not evaluated. Courchane (2007) uses a very large multi-lender dataset of 2004-2005 originations to estimate endogenous switching regressions to examine the effect of demographic and risk factors on loan pricing (as measured by the annual percentage rate or APR). She reports a 16 basis point premium for loans without full documentation in the subprime market segment but a 7 basis point price reduction in the prime loan category. Again, subsequent loan performance is not addressed. Pennington-Cross and Ho (2010, forthcoming) examines performance of both fixed and adjustable rate subprime mortgages using multi-lender data on securitized loans and reports that reduced documentation level is associated with both greater default and greater prepayment risk. The magnitude of the low doc effect is roughly a 40% increase in the marginal probability of early mortgage termination, whether by default or prepayment.

One study that explicitly addresses the issue is Rajan, Seru and Vig (2008), who focus on what they term "hard versus soft information" in the context of asset securitization. If a piece of information can be readily documented or easily verified, it is hard information. Credit score is a good example (an objective measure that may be obtained at low cost in a matter of seconds). In contrast, a piece of information that is hard to document or verify is soft information, e.g. the risk of future borrower job loss. The authors argue that securitization makes it more difficult for the lenders to collect soft information due to their greater distance from the loan origination process. As a result, their increased reliance on hard information will produce moral hazard in differentiating the qualities of borrower who have the same hard information data but heterogeneous soft information, increasing default risk. Our study is consistent with their finding, showing that under mortgage securitization, lenders may have a tendency to reduce their reliance on even hard information, further weakening screening efficiency, aggravating the default risk.

In this study, we use recently updated (by May, 2009) data on residential mortgage loans originated during 2000-2007 and securitized by Bear Stearns and affiliates[5] to investigate the impact of loan documentation type on default risk, focusing on the stated income category. The data includes over 218,000 observations for which there is complete information on major loan characteristics such as loan balance, monthly payment, LTV, and credit score, which we call the "full sample", and a subsample with over 134,000 loans that have sufficient information from which to infer borrower income, which we call the "restricted sample". With the full sample, we find that stated-income is more widely used with alternative mortgage products and ARMs than other documentation types. Without full-documentation, especially when stated-income or stated-assets programs are used, we find the default rate is significantly higher, even after controlling for other risk factors such as credit score, local housing market affordability, housing cost growth rate, and capital market conditions. Using the restricted subsample we find that, as compared to full-doc loans, stated-doc loans are associated with a significantly higher ratio between inferred borrower income and local

[5] More information on the data is available at www.emcmortgage.com

average income measured at the MSA level. We find the higher the ratio of borrower income to the local average income, the higher the default rate. This pattern is even more pronounced in the stated-doc categories compared to full-doc loans. Results suggest that lenders may be able to reduce default risk by setting limits on the ratio between (verified or stated) borrower incomes and local average income levels. In other words, too large a discrepancy between borrower income and local average income suggests a suspicious transaction in which the borrower (and/or broker) may have exaggerated income to qualify for a loan that is greater than they can really afford.

In summary, our study here contributes to the broader literature on residential mortgage loan performance by providing the first comprehensive study on the effects of loan documentation levels. Moreover, given our analysis we can simulate alternative loan origination policies to quantity the potential extent to which current defaults might have been avoided.

The structure of the paper is as follows. The next section explains the data and the major empirical methodologies we employ. Section 3 presents results of the empirical analysis. Section 4 presents conclusions and planned extensions.

2 Data and Methodology

2.1 Data

Our study relies on data from three sources: (1) loan-level data to identify factors such as loan type, documentation type, borrower credit score, and LTV; (2) MSA-level, for data on local housing market conditions such as housing price levels and the MSA median household income; (3) national-level, for data describing capital market conditions, such as the yield curve.

The loan-level data consists of loans securitized by Bear Stearns and its affiliates during 2000-2007, restricting the sample to home purchase purpose, single-family dwelling units, with loan terms of thirty years. After deleting observations with missing data, we create a dataset of 218,589 loans, henceforth the "full sample". As income per se is not an available data element, we infer it from loan payment (PITI) and front-end ratio. We hypothesize that likely "income exaggeration" can be measured by the ratio of the inferred borrower income to MSA median household income. Of course, for the "no ratio" and "no doc" categories there is no front-end ratio from which to infer borrower income, so we discard these observations. There are also a small number of loans for which the ratio of borrower to MSA median household income is extremely high and view these as outliers that may be suspect. We initially delete those with a ratio greater than 10, but will come back to these later in the paper. The final subsample consists of 134,174 loans or about 61% of the full sample. We call this the "restricted sample". We calculate loan age (in months) as of April 2008 for non-defaulted loans and based on the date the loan was referred to a foreclosure attorney for defaulting loans. Thus we observe loan age at default or the point of data censoring.

Documentation level takes various forms in our data. We begin by classifying loans into three broad categories: (1) "full-doc", for loans with "full" marks in documentation type descriptions; (2) "stated-doc", for loans with "stated" marks in documentation type descriptions, including loans with stated-income, stated-income/stated-assets, and stated-income but verified assets; and (3) "no-doc", for loans with "no" marks in documentation type description, including loans with no-income, no-assets, no-ratio, and no-documentation at all. In the full sample, these three categories comprise 30%, 44% and 18% of the loans, respectively. In the restricted sample, the full-doc and stated-doc comprise 37.5% and 54.2% of the loans, respectively, and given the limited count of no-doc loans, our major comparison within the restricted sample data is between the stated-doc loans and the full-doc loans. Documentation type was missing in 8% of the loans in the full sample, and about 8% of the loans in the restricted sample, so we discard these observations.

We also include MSA-level variables to control for market-specific factors. We include local housing price levels, which we measure with the publicly available OFHEO HPI; the 5-year average annual growth rate in MSA HPI; the MSA-median household income; local housing affordability, measured by the ratio between the MSA-median household income and the concurrent MSA HPI; and so forth. Finally, we include several capital market condition indicators as additional control variables, including the slope of the yield curve, the return on equity markets, and the level of mortgage rates, all measured as of the date of loan origination. The slope of the yield curve is calculated as the ratio of the 10-year Treasury bond rate and the 2-year Treasury note rate. The return on equity markets is measured by the 1-year return of S&P500 index. The level of mortgage rate is measured by the contract rate on 30-year, fixed-rate conventional home mortgage commitments, based on the Freddie Mac Primary Mortgage Market Survey data.

2.2 Methodology

The major questions we address are the following: (1) Does the lack of full documentation create additional default risk? (2) Is there evidence of "income exaggeration" among stated-documentation borrowers? (3) Does "income exaggeration" among stated-doc loan borrowers further increase default risk?

2.2.1 Issue one: does the lack of full documentation create additional default risk?

We develop the following three specifications of logit regressions of loan default[6], to explore this issue using the full sample data:

$$D_{default} = \alpha_f + \beta_f \; D_{fulldoc} + \sum_{j=1}^{\upsilon} \gamma_{fj} V_j + \vartheta_f, \tag{1}$$

$$D_{default} = \alpha_s + \beta_s \; D_{stateddoc} + \sum_{j=1}^{\upsilon} \gamma_{sj} V_j + \vartheta_s, \tag{2}$$

$$D_{default} = \alpha_n + \beta_n \; D_{nodoc} + \sum_{j=1}^{\upsilon} \gamma_{nj} V_j + \vartheta_n. \tag{3}$$

$D_{default}$ is the default dummy, which takes on the value of 1 if the data shows that the loan has defaulted; $D_{fulldoc}$, $D_{stateddoc}$ and D_{nodoc} are the dummies for full-doc, stated-doc and no-doc types, respectively; set V contains control variables that are expected to also affect default probability, including loan-level factors such as credit score, MSA-level factors such as local housing market affordability, and capital market factors such as 1-year return in S&P 500 index; α_f, α_s and α_n are the intercepts; β_f, β_s and β_n are the coefficients for the documentation type dummies; γ_{fj}, γ_{sj} and γ_{nj} ($j = 1, ..., \upsilon$) are coefficients for control variables; and finally, ϑ_f, ϑ_s and ϑ_n are the error terms. We will test the following hypothesis:

- [Hypothesis 1] Stated-doc and no-doc loans are more likely to default than full-doc loans. In regressions 1, 2 and 3, this means that the stated-doc dummy and the no-doc dummy positively affect the default dummy, while the full-doc dummy negatively affects the default dummy, that is, $\beta_f < 0$, $\beta_s > 0$ and $\beta_n > 0$.

[6] We measure default by the indicator variable "Referred to foreclosure attorney" contained in the data. There is also a field indicating the date the loan was referred to the foreclosure attorney, so we can determine loan age at time of default. Other authors have used the first instance of a 90-day delinquency, the occurrence of the filing of a notice of default, or similar measures intended to capture serious loan delinquency and pending foreclosure. None of these definitions implies that the loan actually proceeds to a foreclosure sale, of course, as the borrower may always reinstate the loan, pay off the loan, and/or sell the property prior to auction date. Capturing those outcomes in detail is important for measurement of loss severity, as opposed to default rates, which is our focus here.

2.2.2 Issue two: is there evidence of "income exaggeration" by stated-documentation borrowers?

We develop the following simple regressions, to explore this issue using the restricted sample data:

$$Incratio = \kappa_f + \rho_f \, D_{fulldoc} + \varepsilon_f, \tag{4}$$

$$Incratio = \kappa_s + \rho_s \, D_{stateddoc} + \varepsilon_s, \tag{5}$$

where $Incratio$ is the inferred-MSA median income ratio, that is, the ratio between inferred borrower income and the MSA median household income; $D_{fulldoc}$ and $D_{stateddoc}$ are the dummies for full-doc and stated-doc, respectively; κ_f and κ_s are the intercepts; ρ_f and ρ_s are the coefficients for the documentation type dummies; and finally, ε_f and ε_s are the error terms. We will test the following hypothesis:

- [Hypothesis 2] Stated-doc loans have higher inferred-MSA median income ratio than full-doc loans. In regressions 4 and 5, this means that the income ratio is positively affected by the stated-doc dummy while negatively affected by the full-doc dummy, that is, $\rho_f < 0$ and $\rho_s > 0$.

2.2.3 Issue three: does "income exaggeration" among stated-doc borrowers further increase default risk?

To explore this issue, we first develop the following regression for the full-doc subsample and the stated-doc subsample in the restricted sample:

$$D_{default} = \pi + \mu \, Incratio + \sum_{j=1}^{k} \varpi_j S_j + \xi, \tag{6}$$

where $D_{default}$ is the default dummy; $Incratio$ is the inferred-MSA median income ratio; π is the intercept; set S contains control variables that are expected to also affect default probability; μ is the coefficient for $Incratio$; ϖ_j $(j = 1, ..., v)$ are the coefficient for control variables; and finally, ξ is the error terms. We will test the following hypothesis:

- [Hypothesis 3] The default rate is increasing in the inferred-MSA median income ratio for the stated-doc subsample, with a stronger effect than that for the full-doc subsample. In regression 6, this means that with the stated-doc subsample data, $\mu > 0$, and in addition, μ is positive and larger in magnitude than when estimated using the full-doc subsample data.

We further explore the effect of income ratio on stated-doc loans' default risk by conducting a "income ratio cap" sensitivity analysis, using the restricted sample. We analyze the default rate of stated-doc loans versus that of full-doc loans, setting different upper boundaries on the income ratio. We will call these "income ratio caps". For a specific cap $\widehat{H_i}$ $(i = 1, ..., n)$, we extract a subsample with exclusively stated-doc loans, and a subsample with exclusively full-doc loans. With n different caps, we could get n cap-specific stated–doc-loan subsamples, and n cap-specific full–doc-loan subsamples. We calculate the cap-corresponding mean default rate for loans with each documentation type, and run the following regressions for stated-doc loans and also for full-doc loans:

$$\overline{D_{default}} = \phi + \theta \widehat{H} + \delta. \tag{7}$$

$\overline{D_{default}}$ is the cap-specific mean default rate; $\widehat{H}$ is the inferred-MSA median income ratio cap; ϕ is the intercept; θ is the coefficient; and δ is the error term. We will test the following hypothesis:

- [Hypothesis 4] A more restrictive policy on income exaggeration (a lower income ratio cap) will reduce stated-doc loan default risk, and the effect is stronger among stated-doc loans than for full-doc loans. In regression 7, this means that the average default rate will be positively affected by the income ratio cap for the stated-doc loans, and the effect is stronger than for full-doc loans, that is, $\theta > 0$ for the stated-doc subsample, and θ is more positive and larger in magnitude for the stated-doc subsample than for the full-doc subsample.

Testing these four hypotheses will be the main focus of our empirical analysis. Along the way, we explore related issues such as the relationship between doc type and other loan characteristics, interactions between product type and doc type, and variation with housing market and capital market conditions. The following section presents results.

3 Empirical Results

3.1 Descriptive Statistic Results

3.1.1 Time trend

We first examine time trends in the data for loans originated during 2000-2007. Figure 1 illustrates the time trend for the loans in the full sample. Full-doc type remained the dominant category with a market share (in terms of loan number) consistently above 40% (peaking with a 60.2% in 2004), until 2005 (roughly the peak of the housing bubble). After 2005, its market share declined sharply. By 2007, its market share was only 21.6%, lower than the shares of other two types (46.0% for stated-doc and 23.3% for no-doc). In contrast, stated-doc loans were very rare before 2002, but grew at an accelerating pace becoming the dominant category in 2005 with over a 40% market share. No-doc type also experienced fast growth since 2005 and was the second largest category by 2007. These results suggest a regime shift somewhere around the year of 2005, concurrent with the rise of alternative mortgage products (AMPS). The interaction between these two patterns will be addressed later.

Figure 1 illustrates that 2005 is also a turning point for loan origination where we observe that stated-doc loans had become the riskiest category with mean default rates exceeding those of the other two documentation types. Among loans originated in 2000, the full-doc, stated-doc and no-doc default rates (by May, 2009) are 37.3%, 0.0% and 14.7%, respectively, while among loans originated in 2007, these rates were 6.0%, 10.7% and 10.4%, respectively. This suggests that the rapid growth in the stated-doc type increased risk, consistent with Hypothesis 1.

The increase in the relative riskiness of stated-doc loans cannot be explained by credit score and original LTV alone, the two traditional risk factors associated with default. As illustrated in Figure 1, stated-doc loans originated in 2000 had mean FICO scores of 739, as compared to 593 for full-doc borrowers, and 631 for no-doc borrowers. But this pattern changed dramatically over time. For loans originated in 2003, mean FICO scores had converged to 638 versus 645 and 650. From 2003 to 2005, the credit scores of all-type loan borrowers increased consistent with lender reliance on hard information, as in Rajan, Seru and Vig (2008). Since then, the credit scores of stated-doc and no-doc borrowers have stayed rather stable while the scores of full-doc borrowers have declined slightly.

In terms of original LTV, the stated-doc loans used to be the "safest" loans, with mean original LTV of only 56.5%, as compared to 76.7% (full-doc) and 73.1% (no-doc), for loans originated in 2000. For loans originated in 2007, however, full-doc, stated-doc and no-doc had mean original LTVs of 75.3%, 73.1% and 65.2%, respectively, a substantial increase for the stated-doc loans.

Given these patterns, how can we explain the higher default rates of stated-doc loans? We address this question with the restricted sample. Figure 2 shows that on average, loans in the restricted sample are similar as those in the full sample: as compared to full-doc loans, stated-doc loans migrated from lower- to higher-default risk loans, which cannot be explained by credit scores and original LTV ratios. However, stated-doc loans do show higher ratios between borrower income and

MSA median household income than do full-doc loans since 2001, a pattern consistent with income exaggeration. Income exaggeration may shift lending to less qualified borrowers. This provides some support for the implication of Hypothesis 2 that stated-doc loans are associated with income exaggeration as well as Hypothesis 3 that income exaggeration may by the reason that stated-doc loans become the riskiest category over time.

3.1.2 Descriptive statistics

We next present descriptive statistics on loan characteristics by documentation level for loans originated during the entire origination period 2000-2007.

Table 1 shows results for the full sample. Stated doc is the most common category (44%), followed by full-doc (30%) and then no-doc (18%). The distribution of doc type is further illustrated in Figure 3. Most of the loans are ARMs, while FRMs comprise less than one percent. The majority (63%) are ALT-A, a category which overlaps with AMPS, which comprise 43% of the loan population. About one-third of all loans are subprime.

Cross-documentation type comparisons are highlighted in Table 2. In general, doc type seems to interact with LTV, credit score and loan type, affecting default rates. While no-doc loans have, on average, the highest credit scores, the lowest LTV, the lowest subprime percentage, the highest ALT-A and AMPS percentages, the most recent origination years, and relatively larger loan size. This pattern suggests higher underwriting standards for such loans, consistent with their lower mean default rates. Likewise, full-doc loans have the lowest borrower credit score, the highest LTV, the highest subprime percentage, the lowest ALT-A and AMPS percentages, tend to be older, and smaller in loan size. This suggests underwriting standards that require lower quality loans to provide more documentation. This category has the lowest mean default rate by May,

2009. Finally, stated-doc loans have features in between the other two categories, except that the loan size and the default rate are the largest across the three groups. This patterns suggests that lenders were able to successfully screen high and low-quality borrowers but may be relatively less effective in underwriting those in the middle-range, particularly since in that category borrowers could misrepresent their income.

Turning to the restricted sample, we report additional information including inferred-MSA median income ratio, local housing market conditions such as housing cost and affordability, as well as capital market conditions. Table 3 shows that, with the restricted sample as a whole, stated-doc is still a much more frequent category (54%) than full-doc (36%). In general, inferred income is a little lower than the local area median, so the median inferred-MSA median income ratio lower is 0.81, although the mean is a little bit higher at 1.10.

Table 4 compares stated-doc with full-doc loans. Across the two groups, stated-doc loans have much higher mean inferred-income to MSA median ratio than do full-doc loans, 1.20 versus 0.94, with the difference significant at 1% level. As was the case with the full sample, stated-doc loans also have a significantly higher default rate than do full-doc loans. The income ratio difference supports Hypothesis 2, and the coexistence of an income ratio difference and a default risk difference are consistent with Hypothesis 3. Furthermore, Table 4 also shows that stated-doc loans are more concentrated in areas with higher and more rapidly increasing housing costs and areas with lower affordability, and become more frequently used when stock market returns were higher or yield curve was flatter, that is, when the capital market is less constrained.

3.1.3 Statistics by income ratio range

Since findings thus far with respect to the inferred income-MSA median are interesting, we further explore loan characteristics and the doc type effects by examining loans in each range of this income ratio. To explore the issue more fully, we enlarge the restricted sample to include those loans with inferred-MSA median income ratio >10, so the sample size increases from 134,174 loans to 135,119 loans.

More than 95% of the loans have income ratios lower than 3.5, which we will then use as the first cutoff in income ratio range analysis. Table 5 and Figure 4 show the statistics for the two income ratio ranges: $[0, 3.5)$ and $[3.5, \infty)$. We focus on the three most important variables, default rate, original LTV, and credit score. With respect to default risk, from income ratio range $[0,3.5)$ to range $[3.5, \infty)$, full-doc loans' mean default rate has a modest increase (from 15.93% to 17.99%), while the stated-doc loans' mean default rate jumps more dramatically (from 20.89% to 29.95%). With respect to original LTV, the two groups' mean LTV ratios not only increase but do so roughly proportionately (62.18% to 72.43% for full-doc loans, and 61.31% to 73.16% for the stated-doc loans). This suggests that LTV may not be a reason for the cross-doc type variation in default rate. Finally, with respect to credit score, full-doc loans' average credit score has a noticeable increase from 659 to 688, while mean credit score for stated-doc loans' shows a smaller increase from 683 to 699. This suggests that the association between the stated-doc loan default risk and the inferred income-MSA median ratio may be due to income exaggeration by loan applicants.

We further investigate these patterns by dividing the sample into ten income ratio ranges. Results are shown in Table 6 and Figure 5. Visual inspection suggests that an income ratio 1.5 is an important inflection point. Above that point, default risk will increase dramatically especially for stated-doc loans; and in addition, original LTV also increases dramatically. In contrast, credit score is stable, until the income ratio exceeds 2.5. This divergence between again suggests that the relationship between stated-doc default risk and income ratio may be due to income exaggeration.

In summary, our analysis suggests that by setting maximum allowable income ratio ranges, lenders might have been able to reduce default risk among stated income borrowers. When a borrower declares usually high income (relative to the local average income level) they could be either rejected for the stated-doc loan or required to switch to a full-doc loan, potentially requiring a smaller loan amount. Anecdotally, some lenders employed analogous techniques by comparing stated income to average reported income within the borrower's occupation. An interesting extension would be to compare default rates across stated-income lenders where some employed such a screening mechanism and others did not. We do not have this degree of detail available with our current data set.

3.2 Regression results

3.2.1 Documentation type regressions

We begin by investigating which loan types are more likely to have certain documentation types. Using the full sample, we run a logit regression for each doc type dummy variable on major loan characteristics including original LTV, credit score, whether it is a FRM and whether it is an AMP. We employ two model specifications to control for multi-co linearity among explanatory variables. As shown in Table 7, a loan is more likely to be stated-doc when it is ARM and/or AMPS, as the stated-doc logit regression has a large and negative coefficient for the FRM dummy (-2.406), and the largest positive coefficient for the AMPS dummy (0.611, as versus -0.650 with full-doc, and 0.536 with no-doc). In addition, a relatively higher LTV combined with a relatively higher credit score is also related to use of stated-doc. These results suggest that stated-doc loans are used by borrowers with moderate-level LTVs and credit scores, but widely used in conjunction with aggressive lending instruments such as ARMs and AMPS.

3.2.2 Default regressions (for Hypothesis 1)

Using the full sample, we run a default probability regressions 1, 2 and 3 to test Hypothesis 1 that stated-doc and no-doc loans are more likely to default than full-doc loans. Comparing the regressions, we will examine whether the stated-doc dummy and the no-doc dummy positively affect the default dummy, while the full-doc dummy negatively affects the default dummy, that is, $\beta_f < 0$, $\beta_s > 0$ and $\beta_n > 0$. We develop four model specifications to explore the issue, with the results shown in Table 8. In Specification 1, after controlling multi-co linearity, we include original LTV, credit score, documentation type dummies and loan origination year dummies as explanatory variables. The results are in consistent with Hypothesis 1, with the full-doc dummy coefficient β_f negative (-0.527), and the stated-doc dummy coefficient β_s and the no-doc dummy coefficient both positive (0.352 and 0.139). Interestingly, the coefficient on stated-doc is larger than that on no-doc. The full-doc coefficient is negative, confirming that greater loan documentation reduces default risk. With respect to other factors, we find that as expected, default rates are increasing in original LTV and decreasing in credit score. In Specification 2, we add more risk factors such as loan balance, a dummy variable for FRM and AMP contract type. Previous results continue to hold, with β_f negative, while β_s and β_n both positive and β_s larger in magnitude than β_n. Collectively results strongly support Hypothesis 1: lack of full documentation boosts default risk. As a robustness test, we replace the original year dummies with a loan age dummy, forming Specifications 3 and 4, which generate essentially similar results as the previous two specifications.

Given these results, we can simulate default rates. For instance, at the sample mean of borrower characteristics (original LTV of 63%, credit score of 676, and loan age of 29 months and using

Specification 3 coefficients) we find that the average loan will have a 11.10% default rate if fully documented, while 18.35% if stated documented, or 17.83% if without documentation. This indicates a 65% increase in default rate by switching from full-doc to stated-doc, while a 61% increase in default rate by switching from full-doc to no-doc. A similar simulation using coefficients from Specification 4 indicates a 56% (full-to-stated) increase and a 46% (full-to-no) increase in default rate. Note these results are slightly higher than the estimates in Pennington-Cross and Ho (2010) discussed previously.

3.2.3 Inferred-MSA median income ratio regression (for Hypothesis 2)

We then use the restricted sample to run the inferred income-MSA median ratio regressions 4 and 5, to test Hypothesis 2 that stated-doc loans have higher ratio between borrower income and local average income compared to full-doc loans. Across the two regressions we will examine whether the stated-doc dummy positively affects the income ratio while the full-doc dummy negatively affects the income ratio, that is, $\rho_f < 0$ and $\rho_s > 0$. Table 9 reports results, with the dependent variable defined as the natural logarithm of inferred income-MSA median income ratio. It shows a significantly positive coefficient for the stated-doc dummy, ρ_s (0.203), while a significantly negative coefficient for the full-doc dummy, ρ_f (-0.242). Results confirm results in Table 4 that stated-doc loans are used by ostensibly higher income households (compared to full-doc loans); however, that higher income may be exaggerated.

3.2.4 Cross-subsample difference in default regression (for Hypothesis 3)

From the restricted sample, we can extract subsamples of full-doc and stated-doc loans. Using these subsamples we run regressions 6 to test Hypothesis 3 that default rate is increasing in the inferred-MSA median income ratio for the stated-doc subsample, but that any such effect is not as large in the full-doc subsample. Results appear in Table 10. Again we include a comprehensive set of risk factors including loan-level characteristics, MSA-level housing market conditions, and national-level capital market factors, and develop five model specifications which control for multico linearity differently. Across all specifications results support Hypothesis 3, with the coefficient of the natural logarithm of inferred-MSA median income ratio positive for both stated-doc loan and full-doc loan subsamples, but the magnitude of the coefficient is over 35% greater for stated-doc loans.

Examining other risk factors, compared to full-doc loans, stated-doc loan default rates are more sensitive to housing cost (+), housing price appreciation (+) and affordability (-), while less sensitive to credit score (-), suggesting that relaxing documentation constraints will increase risk during difficult housing market conditions.

Again calculating implied default rates at sample means using Specification 6 (the only one that includes loan age), we find that an average loan will have a 16% default rate if fully documented, while 24% if stated documented, indicating a 51% increase in default rate by switching from full-doc to stated-doc.

3.2.5 Effect of inferred-MSA median income ratio cap (for Hypothesis 4)

Finally and again using the two doc type subsamples, we run a regression 7 to test Hypothesis 4 that more restrictive policy on income exaggeration will reduce stated-doc default risk, and this effect is stronger for stated-doc than for the full-doc loans. In other words, we will examine whether average default rates are negatively affected by the income ratio cap (that is, the upper boundary of the allowed range of inferred income-MSA median ratio) for the stated-doc loans and whether any such effect is stronger than is the case for full-doc loans.

We first create a data set for this regression. For each possible income ratio cap we calculate the loan counts, mean default rate, and mean inferred income-MSA median ratio for the all-loan sample, the full-doc and the stated-doc subsamples. We then vary the cap level, to generate a series of loan counts, mean default rates, and mean income ratios for each sample, which will be used later for running the regression 7. The data thus created are summarized in Table 11. As the income ratio cap increases, all three variables (loan count, mean default rate, and mean income ratio) in general increase faster for the stated-doc subsample than for the full-doc subsample. These effects are further highlighted in Figure 6. Interestingly, when the cap is higher, the default rate difference between the two documentation types becomes more significant. For the stated-doc loans, an income ratio cap of 1.5 will reduce the average default rate from 21.37% to 18.09% (a 18% drop), a cap of 2 will reduce the default rate to 19.70% (a 8% drop), and a cap of 2.5 will reduce to the default rate to 20.32% (a 5% drop). While these absolute levels of default are still unacceptably high, results help illustrate the effect of controlling the reliance on stated income that is likely exaggerated.

We then use these data to run regression 7 for stated-doc and full-doc loans. As shown in Table 12-Specification 2, we confirm that the cross-documentation type difference in the effect of inferred-MSA median income ratio cap on default risk is statistically significant, with the cap coefficient as 0.005 for the stated-doc loans while only 0.001 for the full-doc loans, supporting Hypothesis 4 that applying a restrictive policy on likely income exaggeration will help reduce default risk particularly for stated-doc loans. Similar results are apparent from Specifications 1 and 3.

In summary, regression results provide strong support for our four major hypotheses, confirming the importance of loan documentation discipline and management to reduce default risk.

4 Conclusions

In this paper we have examined the effects of documentation type on default risk. Although loan documentation requirements have changed dramatically in recent years, their contribution to increasing rates of residential mortgage default has not been rigorously analyzed. We believe our study is the first to focus on this issue. We do so using a large database on home purchase loans securitized by Bear Stearns and affiliates over the recent period 2000-2007.

We find that reduced levels of documentation do increase the likelihood of default after controlling for other risk factors. The problem is particularly acute for stated-doc loans, which are offered to lower quality borrowers (as measured by credit score and LTV) compared to no-doc loans, though they are higher in quality than full-doc loans. Simulation based on our default regression models suggests an over 50% increase in default rate when a loan with average characteristics switches from full-doc to stated-doc, a result that is slightly higher than the 40% increase in default rate for subprime loans reported by Pennington-Cross and Ho (2010), though our methods are not exactly comparable. The reason that these mid-quality borrowers perform worse than objectively worse borrowers because lenders allowed them simply state, as opposed to verify, income or assets, while not allowing objectively lower quality borrowers to do so. We find evidence of income exaggeration in the stated doc category and show that the degree of likely exaggeration is related to default risk. We also show that limiting the ratio between stated income and local average household income may reduce default risk. Simulation suggests that when the ratio is limited to 2.5, stated-doc loans should be less risky than full-doc with objectively worse credit score and LTV characteristics. Given these results, we think stated income lending can be a viable mortgage product; however, careful risk management is essential to mitigate inherent risks.

Further research efforts on this topic may involve replicating the current analysis using data through at least year-end 2008, potentially measuring local area income at a finer level of geography and replicating the analysis using multi-lender data, where available.

References

Campbell, John Y. and Joao Cocco. 2003. Household Risk Management and Optimal Mortgage Choice . Quarterly Journal of Economics 118(4): 1449-1494.

Cohen, Norma. 2009. Banks Still Lending at Many Times Earnings. Financial Times, January 31, 2009.

Courchane, Marcia. 2007. The Pricing of Home Mortgage Loans to Minority Borrowers: How Much of the APR Differential Can We Explain? Journal of Real Estate Research 29(4): 399-440.

Inside Mortgage Finance Publications. 2007. Regulators Keep Pressing for Tougher Standards In Subprime Market; Stated Income Under Fire. Inside Mortgage Finance, May 25, 2007.

Inside Mortgage Finance Publications. 2009. Default Rates Still Rising. Inside Mortgage Finance, June 5, 2009.

General Accountability Office. 2005. Alternative Mortgage Products.

Harney, Kenneth. 2009. Congress Takes A Series Look At Reforming the Mortgage Market. Washington Post, April 4, 2009.

LaCour-Little, Michael. 2007. The Home Purchase Mortgage Preferences of Low-and-Moderate Income Households". Real Estate Economics 35(4):265-290.

LaCour-Little, M., and J. Yang. 2009. Pay Me Now or Pay Me Later: Alternative Mortgage Products and the Mortgage Crisis. California State University at Fullerton Working Paper.

Pavlov, A. and S. Wachter. 2006. Aggressive Lending and Real Estate Markets. University of Pennsylvania Working Paper.

Pennington-Cross, Anthony and Giang Ho. 2010. The Termination of Subprime Hybrid and Fixed Rate Mortgages. Real Estate Economics 38(1), forthcoming.

Rajan, Uday, Amit Seru and Vikrant Vig. 2008. The Failure of Models That Predict Failure: Distance, Incentives and Defaults. University of Chicago Working Paper.

Turnbull, Stuart M., Crouhy, Michel and Jarrow, Robert A.,The Subprime Credit Crisis of 07(July 9, 2008). Available at SSRN: http://ssrn.com/abstract=1112467

United States Government Accountability Office. 2006. Alternative Mortgage Products . Report to the Chairman, Subcommittee on Housing and Transportation, Committee on Banking, Housing, and Urban Affairs, U.S. Senate. GAO-06-1021.

Figure 1 Time Trend of Loan Characteristics (Full Sample)

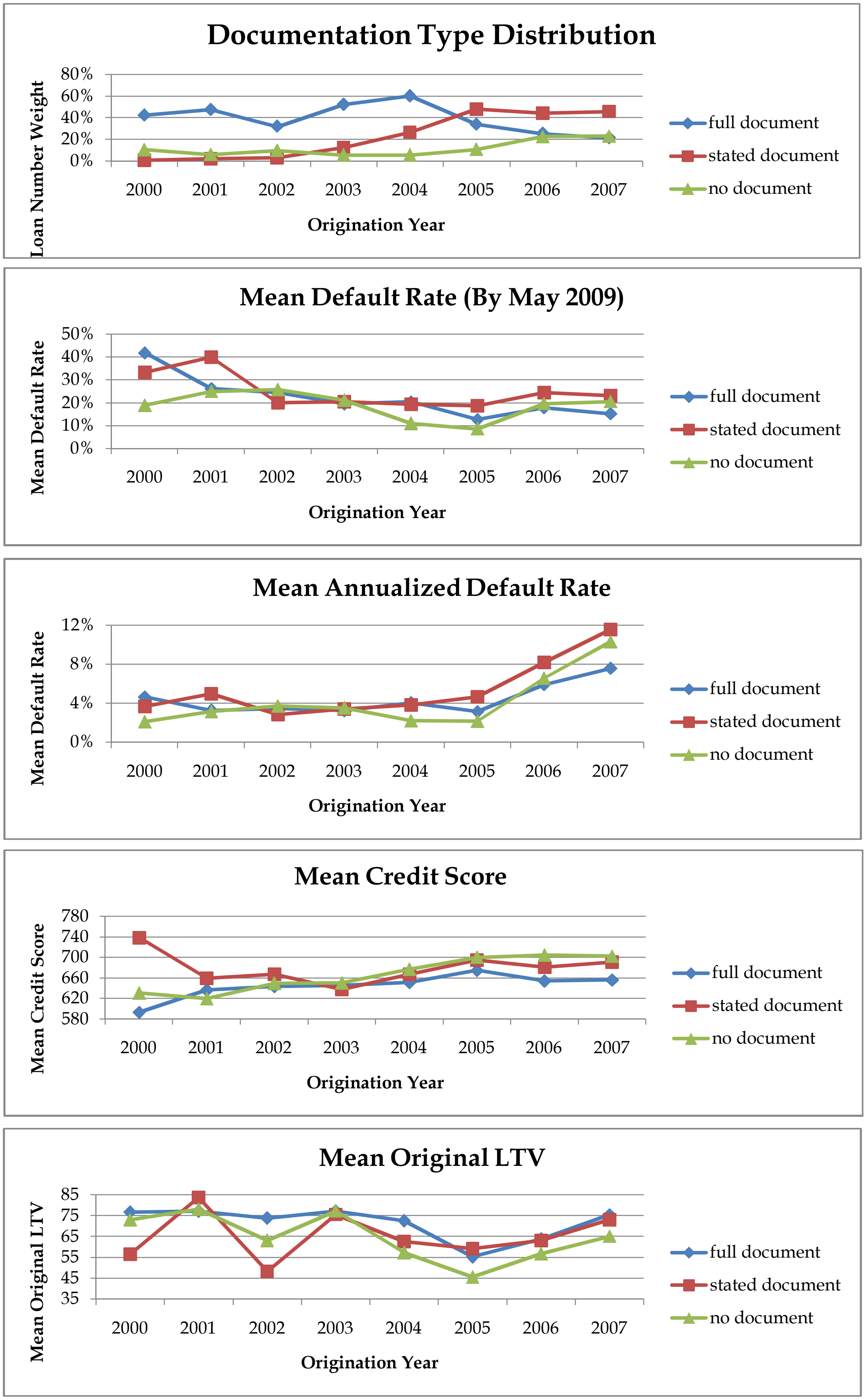


Figure 2 Time Trend of Loan Characteristics (Restricted Sample)

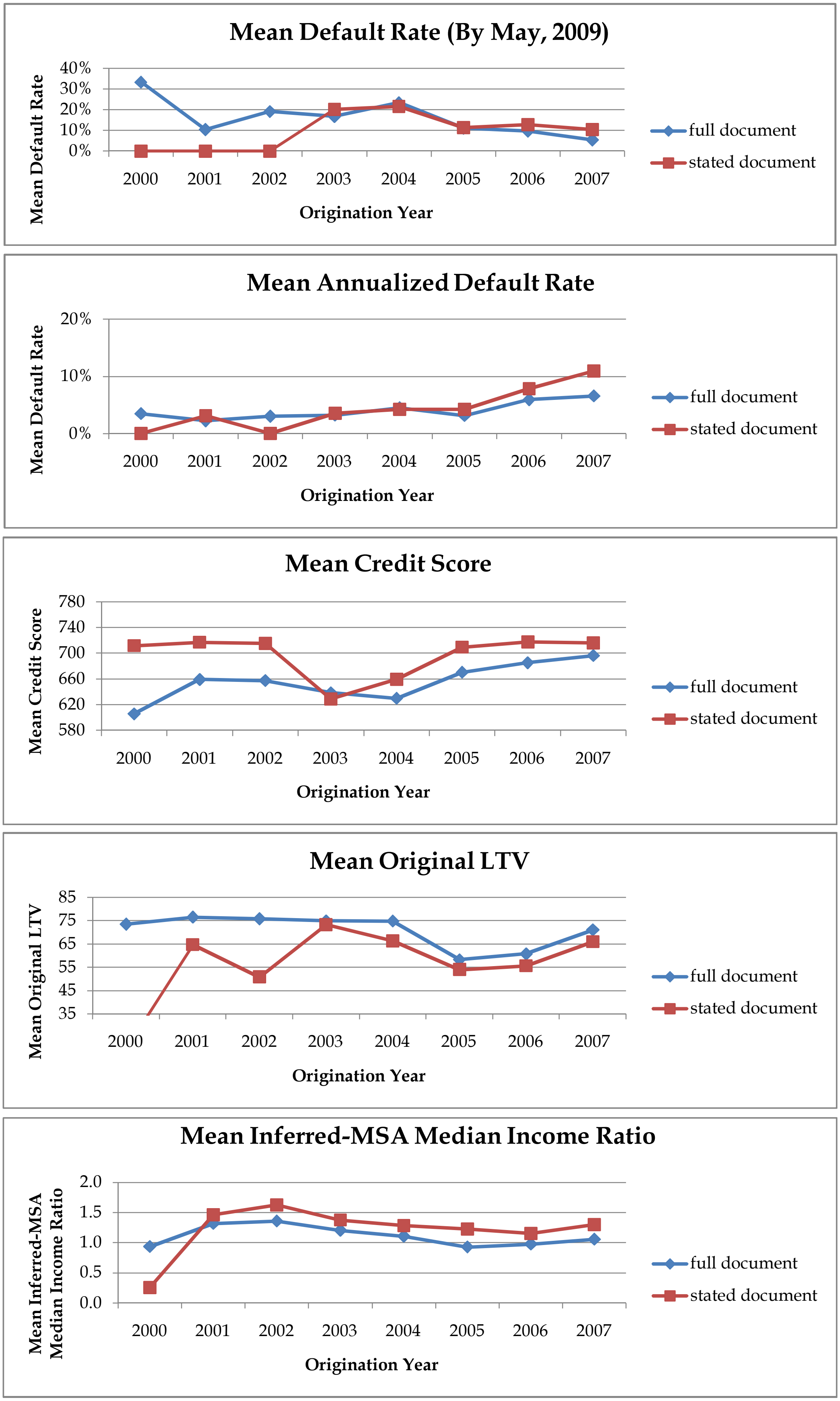


Table 1 Descriptive statistics for the full sample as a whole

Variable	N	Mean	Median	Minimum	Maximum	Std Dev
loan balance	218,589	209612	155072	4050	6375000	196972
original LTV	218,589	63.06	80	1	100	28.40
origination year	218,589	2005.6	2006	2000	2007	0.98
credit score	218,589	676.07	683	356	843	68.00
default	218,589	0.20	0	0	1	0.40
FRM	218,589	0.003	0	0	1	0.056
prime	218,589	0.04	0	0	1	0.20
subprime	218,589	0.31	0	0	1	0.46
ALT-A	218,589	0.63	1	0	1	0.48
AMPS	218,589	0.43	0	0	1	0.49
full-document	218,589	0.30	0	0	1	0.46
stated document	218,589	0.44	0	0	1	0.50
no document	218,589	0.18	0	0	1	0.38
loan age	218,589	28.61	25	4	1299	50.56

Figure 3 Documentation Type Distribution in the Full Sample

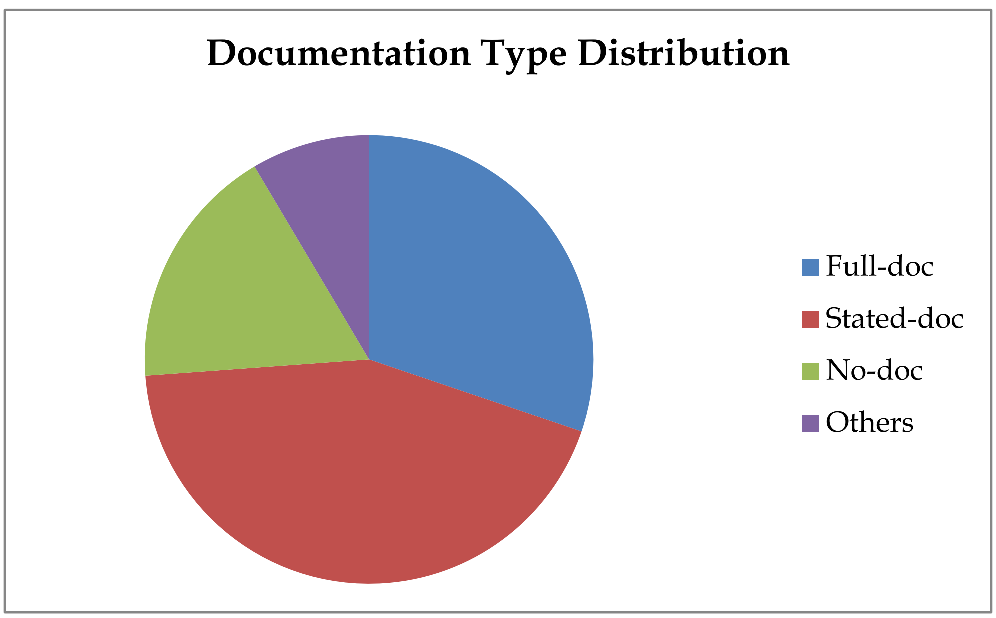


Table 2 Descriptive statistics for the subsamples of the full sample

Panel A

	Full-Document Subsample				Stated-Document Subsample						
Variable	N	Mean	Median	Std Dev	N	Mean	Median	Std Dev	Mean Diff.	t-stat	p-value
loan balance	66,032	171136	126810	161535	95377	234185	184800	214504	63049	67.30	<.0001
original LTV	66,032	63.67	80	28.93	95,377	63.47	80	27.75	-0.20	-1.40	0.1606
origination year	66,032	2005.4	2005	1.07	95,377	2005.7	2006	0.79	0.37	75.48	<.0001
credit score	66,032	660.70	663	72.39	95,377	686.63	693	62.84	25.93	74.62	<.0001
default	66,032	0.16	0	0.37	95,377	0.22	0	0.42	0.06	30.18	<.0001
FRM	66,032	0.0005	0	0.02	95,377	0.0001	0	0.01	-0.0004	-4.31	<.0001
prime	66,032	0.06	0	0.24	95,377	0.02	0	0.14	-0.04	-40.30	<.0001
subprime	66,032	0.47	0	0.50	95,377	0.23	0	0.42	-0.24	-100.72	<.0001
ALT-A	66,032	0.42	0	0.49	95,377	0.74	1	0.44	0.32	132.83	<.0001
AMPS	66,032	0.30	0	0.46	95,377	0.52	1	0.50	0.22	92.63	<.0001
loan age	66,032	32.23	28	60.24	95,377	27.26	25	48.58	-4.97	-17.59	<.0001

Panel B

	Stated-Document Subsample				No-Document Subsample						
Variable	N	Mean	Median	Std Dev	N	Mean	Median	Std Dev	Mean Diffe	t-stat	p-value
loan balance	95,377	234185	184800	214504	38559	227197	171200	211506	-6988	-5.45	<.0001
original LTV	95,377	63.47	80	27.75	38,559	56.50	75	29.34	-6.97	-39.99	<.0001
origination year	95,377	2005.7	2006	0.79	38,559	2005.9	2006	0.79	0.20	41.12	<.0001
credit score	95,377	686.63	693	62.84	38,559	702.54	701	48.23	15.91	49.88	<.0001
default	95,377	0.22	0	0.42	38,559	0.18	0	0.38	-0.04	-18.99	<.0001
FRM	95,377	0.0001	0	0.01	38,559	0.0002	0	0.01	0.0001	1.02	0.3062
prime	95,377	0.02	0	0.14	38,559	0.02	0	0.13	0.00	-1.92	0.0546
subprime	95,377	0.23	0	0.42	38,559	0.02	0	0.12	-0.22	-144.24	<.0001
ALT-A	95,377	0.74	1	0.44	38,559	0.97	1	0.18	0.23	135.56	<.0001
AMPS	95,377	0.52	1	0.50	38,559	0.56	1	0.50	0.03	11.60	<.0001
loan age	95,377	27.26	25	48.58	38,559	23.56	21	26.69	-3.69	-17.76	<.0001

Panel C

	Full-Document Subsample				No-Document Subsample						
Variable	N	Mean	Median	Std Dev	N	Mean	Median	Std Dev	Mean Diffe	t-stat	p-value
loan balance	66,032	171136	126810	161535	38559	227197	171200	211506	56061	44.95	<.0001
original LTV	66,032	63.67	80	28.93	38,559	56.50	75	29.34	-7.17	-38.35	<.0001
origination year	66,032	2005.4	2005	1.07	38,559	2005.9	2006	0.79	0.57	97.50	<.0001
credit score	66,032	660.70	663	72.39	38,559	702.54	701	48.23	41.84	111.95	<.0001
default	66,032	0.16	0	0.37	38,559	0.18	0	0.38	0.01	6.00	<.0001
FRM	66,032	0.0005	0	0.02	38,559	0.0002	0	0.01	-0.0003	-2.97	0.0029
prime	66,032	0.06	0	0.24	38,559	0.02	0	0.13	-0.04	-37.61	<.0001
subprime	66,032	0.47	0	0.50	38,559	0.02	0	0.12	-0.46	-223.73	<.0001
ALT-A	66,032	0.42	0	0.49	38,559	0.97	1	0.18	0.55	257.34	<.0001
AMPS	66,032	0.30	0	0.46	38,559	0.56	1	0.50	0.26	83.34	<.0001
loan age	66,032	32.23	28	60.24	38,559	23.56	21	26.69	-8.66	-31.96	<.0001

Table 3 Descriptive statistics for the restricted sample as a whole

Variable	N	Mean	Median	Minimum	Maximum	Std Dev
Ln(loan balance)	134,174	11.78	12	9	15	0.98
original LTV	134,174	63.09	80	2	100	28.83
origination year	134,174	2005.7	2006	2000	2007	0.83
credit score	134,174	669.18	675	411	843	68.67
default	134,174	0.20	0	0	1	0.40
inferred-MSA median-income ratio	134,174	1.10	0.81	0.00	9.99	1.09
MSA median household income	134,174	53817.89	53336	25701	84318	9184.99
MSA HPI	134,174	238.70	237	87	366	60.39
MSA ratio of household income to HPI	134,174	238.90	233	82	664	66.61
5-year historical growth in MSA median household income	134,174	0.14	0	0	0	0.08
5-year historical growth rate in MSA HPI	134,174	0.70	1	0	1	0.39
level of 30 Year FRM	134,174	6.19	6	0	9	0.36
yield curve slope	134,174	1.08	1	0	3	0.23
SP500 1-year return	134,174	0.09	0	0	0	0.05
FRM	134,174	0.00004	0	0	1	0.01
prime	134,174	0.03	0	0	1	0.16
subprime	134,174	0.42	0	0	1	0.49
ALT-A	134,174	0.55	1	0	1	0.50
AMPS	134,174	0.37	0	0	1	0.48
full-document	134,174	0.36	0	0	1	0.48
stated document	134,174	0.54	1	0	1	0.50
loan age	134,174	26.07	25	4	133	10.14

Table 4 Descriptive statistics for the subsamples of the restricted sample

Variable	Full-Document Subsample				Stated-Document Subsample				Comparisons		
	N	Mean	Median	Std Dev	N	Mean	Median	Std Dev	Mean Diff.	t-stat	p-value
Ln(loan balance)	48,819	11.62	12	0.94	72,505	11.87	12	1.01	0.25	43.5	<.0001
original LTV	48,819	62.42	80	29.62	72,505	61.86	80	28.68	-0.56	-3.26	0.0011
origination year	48,819	2005.5	2006	0.91	72,505	2005.8	2006	0.72	0.33	67.89	<.0001
credit score	48,819	659.93	662	70.84	72,505	683.33	689	63.04	23.41	58.96	<.0001
default	48,819	0.16	0	0.37	72,505	0.21	0	0.41	0.05	23.72	<.0001
inferred-MSA median-income ratio	48,819	0.94	1	0.93	72,505	1.20	1	1.16	0.26	43.57	<.0001
MSA median household income	48,819	52926.04	52713	8836.05	72,505	54714.79	53708	9465.45	1788.75	33.59	<.0001
MSA HPI	48,819	224.84	217	57.46	72,505	249.63	259	60.02	24.79	72.37	<.0001
MSA ratio of household income to HPI	48,819	248.82	248	66.73	72,505	231.73	220	65.45	-17.09	-44.08	<.0001
5-year historical growth in MSA median household inc	48,819	0.12	0	0.08	72,505	0.15	0	0.08	0.03	52.95	<.0001
5-year historical growth rate in MSA HPI	48,819	0.63	1	0.38	72,505	0.74	1	0.39	0.11	48.66	<.0001
level of 30 Year FRM	48,819	6.14	6	0.40	72,505	6.23	6	0.30	0.09	42.07	<.0001
yield curve slope	48,819	1.13	1	0.30	72,505	1.04	1	0.15	-0.09	-58.56	<.0001
SP500 1-year return	48,819	0.09	0	0.05	72,505	0.10	0	0.04	0.01	19.25	<.0001
FRM	48,819	0.00004	0	0.01	72,505	0.00000	0	0.00	-0.00004	-1.41	0.1573
prime	48,819	0.04	0	0.20	72,505	0.01	0	0.12	-0.03	-26.96	<.0001
subprime	48,819	0.53	1	0.50	72,505	0.26	0	0.44	-0.26	-93.97	<.0001
ALT-A	48,819	0.43	0	0.50	72,505	0.72	1	0.45	0.29	104.13	<.0001
AMPS	48,819	0.28	0	0.45	72,505	0.49	0	0.50	0.21	76.55	<.0001
loan age	48,819	28.14	27	11.22	72,505	24.61	24	8.65	-3.54	-58.81	<.0001

Table 5   Default analysis in two ranges of the inferred-MSA income ratio

Range Index	Range of Inferred-MSA Income Ratio	Loan Counts			Default Rate			Defaulted Loan Counts			Original LTV			Credit Score		
		All	Full-doc	Stated-doc	All	Full-doc	Stated-doc	All	Full-doc	Stated-doc	All	Full-doc	Stated-doc	All	Full-doc	Stated-doc
1	**[0, 3.5)**	129,275	47,716	69,211	19.55%	15.93%	20.89%	25,273	7,601	14,458	62.69	62.18	61.31	668	659	683
2	**[3.5, ∞)**	5,844	1,401	3,841	27.64%	17.99%	29.95%	1,615	252	1,150	73.21	72.43	73.16	690	688	699
	Total	**135,119**	**49,117**	**73,052**				**26,889**	**7,853**	**15,609**						

Figure 4   Statistics for Two Ranges of Inferred-MSA Median Income Ratio

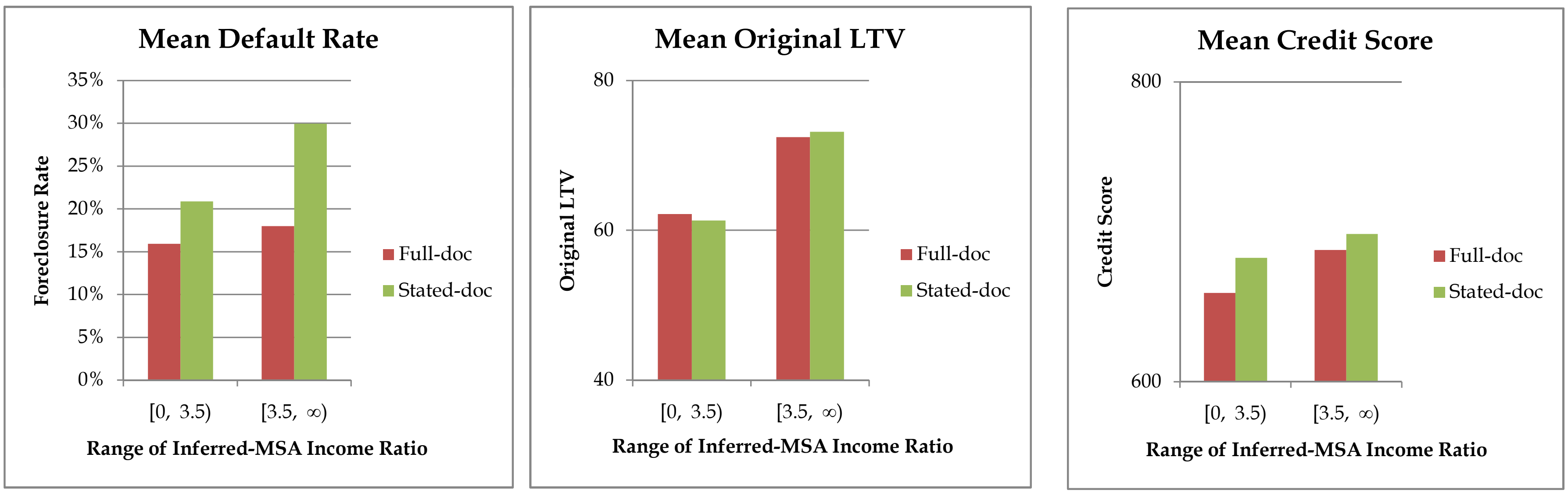


Table 6 Default analysis in ten ranges of the inferred-MSA income ratio

Range Index	Range of Inferred-MSA Income Ratio	Loan Counts			Default Rate			Defaulted Loan Number			Original LTV			Credit Score		
		All	Full-doc	Stated-doc	All	Full-doc	Stated-doc	All	Full-doc	Stated-doc	All	Full-doc	Stated-doc	All	Full-doc	Stated-doc
1	[0, 1.5)	106,444	41,247	55,109	17.59%	15.22%	18.09%	18,719	6,279	9,968	60.02	60.14	57.88	668	659	683
2	[1.5, 2.5)	17,590	5,137	10,679	28.75%	20.99%	31.85%	5,057	1,078	3,401	75.28	75.47	74.77	668	659	681
3	[2.5, 3.5)	5,241	1,332	3,423	28.56%	18.39%	31.81%	1,497	245	1,089	74.68	73.89	74.58	681	675	691
4	[3.5, 4.5)	2,131	486	1,447	27.69%	16.26%	30.27%	590	79	438	74.25	73.32	74.16	687	688	695
5	[4.5, 5.5)	1,086	230	744	28.55%	21.30%	31.72%	310	49	236	73.66	72.45	73.99	688	686	697
6	[5.5, 6.5)	661	166	439	27.08%	18.07%	29.61%	179	30	130	71.55	71.50	71.67	691	691	697
7	[6.5, 7.5)	439	97	288	28.47%	20.62%	27.78%	125	20	80	73.88	75.05	72.98	690	681	705
8	[7.5, 8.5)	305	72	189	27.21%	18.06%	28.57%	83	13	54	73.82	71.38	74.29	687	689	697
9	[8.5, 9.5)	195	34	132	31.28%	20.59%	33.33%	61	7	44	71.63	74.71	71.08	694	677	708
10	[9.5, ∞)	1,027	316	602	26.00%	17.09%	27.91%	267	54	168	71.49	70.72	71.01	695	690	704
	Total	**135,119**	**49,117**	**73,052**				**26,888**	**7,854**	**15,608**						

Figure 5 Statistics for Ten Ranges of Inferred-MSA Median Income Ratio

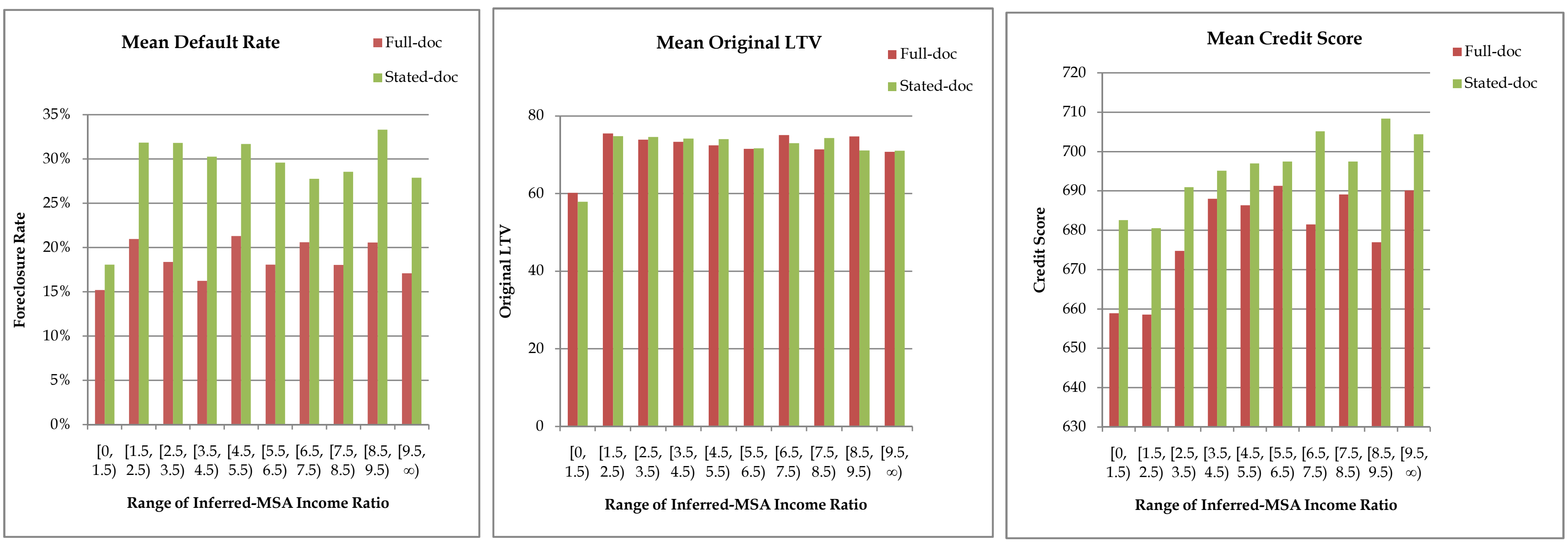


Table 7 Documentation type logit regression results: for the full sample

Variable	Full-Documentation coef.	p-value		Stated-Documentation coef.	p-value		No-Documentation coef.	p-value	
[Specification 1]									
intercept	-0.928	<.0001	***	-0.533	<.0001	***	-1.529	<.0001	***
FRM dummy	-3.165	<.0001	***	-2.406	<.0001	***	-2.193	<.0001	***
AMPS dummy	-0.650	<.0001	***	0.611	<.0001	***	0.536	<.0001	***
origination year dummies	Yes			Yes			Yes		
[Specification 2]									
intercept	1.814	<.0001	***	-3.062	<.0001	***	-5.838	<.0001	***
original LTV	-0.001	<.0001	***	0.004	<.0001	***	-0.009	<.0001	***
credit score	-0.005	<.0001	***	0.004	<.0001	***	0.008	<.0001	***
origination year dummies	Yes			Yes			Yes		

Table 8 Default logit regression results: for the full sample

Variable	coef.	p-value	coef.	p-value	coef.	p-value
[Specification 1]						
intercept	-1.565	<.0001 ***	-1.206	<.0001 ***	-1.605	<.0001 ***
original LTV	0.034	<.0001 ***	0.035	<.0001 ***	0.035	<.0001 ***
credit score	-0.004	<.0001 ***	-0.004	<.0001 ***	-0.004	<.0001 ***
full-document dummy			-0.527	<.0001 ***		
stated document dummy	0.352	<.0001 ***				
no document dummy					0.139	<.0001 ***
origination year dummies	Yes		Yes		Yes	
[Specification 2]						
intercept	-1.027	<.0001 ***	-0.744	<.0001 ***	-1.018	<.0001 ***
Loan balance	0.000	<.0001 ***	0.000	<.0001 ***	0.000	<.0001 ***
original LTV	0.032	<.0001 ***	0.033	<.0001 ***	0.032	<.0001 ***
credit score	-0.005	<.0001 ***	-0.005	<.0001 ***	-0.005	<.0001 ***
FRM dummy	0.007	0.945	-0.232	0.028 **	-0.013	0.903
AMPS dummy	0.180	<.0001 ***	0.160	<.0001 ***	0.206	<.0001 ***
full-document dummy			-0.472	<.0001 ***		
stated document dummy	0.306	<.0001 ***				
no document dummy					0.079	<.0001 ***
origination year dummies	Yes		Yes		Yes	
[Specification 3]						
intercept	-1.458	<.0001 ***	-1.068	<.0001 ***	-1.520	<.0001 ***
original LTV	0.034	<.0001 ***	0.035	<.0001 ***	0.035	<.0001 ***
credit score	-0.004	<.0001 ***	-0.004	<.0001 ***	-0.004	<.0001 ***
full-document dummy			-0.546	<.0001 ***		
stated document dummy	0.359	<.0001 ***				
no document dummy					0.184	<.0001 ***
loan age	0.001	<.0001 ***	0.001	<.0001 ***	0.001	<.0001 ***
[Specification 4]						
intercept	-0.829	<.0001 ***	-0.539	<.0001 ***	-0.821	<.0001 ***
Loan balance	0.000	<.0001 ***	0.000	<.0001 ***	0.000	<.0001 ***
original LTV	0.032	<.0001 ***	0.033	<.0001 ***	0.032	<.0001 ***
credit score	-0.005	<.0001 ***	-0.005	<.0001 ***	-0.005	<.0001 ***
FRM dummy	-0.004	0.964	-0.283	0.005 ***	-0.092	0.358
AMPS dummy	0.171	<.0001 ***	0.145	<.0001 ***	0.202	<.0001 ***
full-document dummy			-0.479	<.0001 ***		
stated document dummy	0.304	<.0001 ***				
no document dummy					0.109	<.0001 ***
loan age	0.001	<.0001 ***	0.001	<.0001 ***	0.001	<.0001 ***

Table 9 Inferred-MSA median-income ratio regression result

Dependent variable is the natural logarithm of inferred-MSA median income ratio

Variable	coef.	p-value	coef.	p-value
intercept	-0.166	<.0001	0.033	<.0001
full-document dummy			-0.242	<.0001
stated document dummy	0.203	<.0001		
origination year dummies	Yes		Yes	

Table 10 Default logit regression results: restricted sample

	Full-Doc		Stated-Doc			
Variable	Estimate	p-value	Estimate	p-value	Difference	p-value
[Specification 1]						
intercept	3.388	<.0001 ***	3.589	<.0001 ***	0.201	<.0001 ***
credit score	-0.008	<.0001 ***	-0.006	<.0001 ***	0.001	<.0001 ***
Ln (inferred-MSA median-income ratio)	0.558	<.0001 ***	0.764	<.0001 ***	0.206	<.0001 ***
MSA ratio of household income to HPI	-0.001	<.0001 ***	-0.003	<.0001 ***	-0.002	<.0001 ***
SP500 1-year return	0.674	0.0153 **	0.989	0.0003 ***	0.315	0.4194
origination year dummies		Yes		Yes		
[Specification 2]						
intercept	3.166	<.0001 ***	3.434	<.0001 ***	0.268	<.0001 ***
credit score	-0.008	<.0001 ***	-0.006	<.0001 ***	0.001	<.0001 ***
Ln (inferred-MSA median-income ratio)	0.574	<.0001 ***	0.806	<.0001 ***	0.232	<.0001 ***
yield curve slope	-0.073	0.3448	-0.630	<.0001 ***	-0.558	0.0002 ***
SP500 1-year return	0.764	0.0063 ***	1.227	<.0001 ***	0.463	0.2380
origination year dummies		Yes		Yes		
[Specification 3]						
intercept	2.921	<.0001 ***	2.187	<.0001 ***	-0.734	<.0001 ***
credit score	-0.008	<.0001 ***	-0.007	<.0001 ***	0.001	<.0001 ***
Ln (inferred-MSA median-income ratio)	0.565	<.0001 ***	0.775	<.0001 ***	0.211	<.0001 ***
MSA HPI	0.001	0.0004 ***	0.003	<.0001 ***	0.002	<.0001 ***
SP500 1-year return	0.704	0.0112 **	0.961	0.0004 ***	0.257	0.5099
origination year dummies		Yes		Yes		
[Specification 4]						
intercept	3.024	<.0001 ***	2.626	<.0001 ***	-0.397	<.0001 ***
credit score	-0.008	<.0001 ***	-0.007	<.0001 ***	0.001	<.0001 ***
Ln (inferred-MSA median-income ratio)	0.563	<.0001 ***	0.776	<.0001 ***	0.213	<.0001 ***
FRM dummy	-8.725	0.9209		***		
5-year historical growth rate in MSA HPI	0.160	<.0001 ***	0.512	<.0001 ***	0.353	<.0001 ***
SP500 1-year return	0.702	0.0114 **	1.015	0.0002 ***	0.314	0.4199
origination year dummies		Yes		Yes		
[Specification 5]						
intercept	2.783	<.0001 ***	1.645	<.0001 ***	-1.138	<.0001 ***
credit score	-0.008	<.0001 ***	-0.006	<.0001 ***	0.001	<.0001 ***
Ln (inferred-MSA median-income ratio)	0.574	<.0001 ***	0.804	<.0001 ***	0.230	<.0001 ***
level of 30 Year FRM	0.048	0.2217	0.177	0.0004 ***	0.129	0.0433 **
SP500 1-year return	0.772	0.0056 ***	1.619	<.0001 ***	0.847	0.0352 **
origination year dummies		Yes		Yes		
[Specification 6]						
intercept	3.256	<.0001 ***	2.437	<.0001 ***	-0.818	<.0001 ***
credit score	-0.008	<.0001 ***	-0.006	<.0001 ***	0.001	<.0001 ***
Ln (inferred-MSA median-income ratio)	0.580	<.0001 ***	0.797	<.0001 ***	0.217	<.0001 ***
loan age	0.013	<.0001 ***	0.025	<.0001 ***	0.013	<.0001 ***

Table 11 The effects of inferred-MSA median income ratio cap

Inferred-MSA Income Ratio Cap	Loan Numbers				Mean Default Rate						Mean Inferred-MSA Median-Income Ratio					
	All	Full-doc	Stated-doc	Stated-Full Difference	All	Full-doc	Stated-doc	Stated-Full Difference	Difference P-value		All	Full-doc	Stated-doc	Stated-Full Difference	Difference P-value	
1.5	106,444	41,247	55,109	13,862	17.59%	15.22%	18.09%	2.86%	<.0001	***	0.695	0.649	0.725	0.076	<.0001	***
2	118,314	44,883	62,182	17,299	18.72%	15.68%	19.70%	4.02%	<.0001	***	0.798	0.736	0.839	0.103	<.0001	***
2.5	124,034	46,384	65,788	19,404	19.17%	15.86%	20.32%	4.46%	<.0001	***	0.864	0.784	0.914	0.131	<.0001	***
3	127,321	47,240	67,914	20,674	19.42%	15.92%	20.70%	4.78%	<.0001	***	0.912	0.819	0.971	0.152	<.0001	***
3.5	129,275	47,716	69,211	21,495	19.55%	15.93%	20.89%	4.96%	<.0001	***	0.947	0.843	1.014	0.171	<.0001	***
4	130,501	47,995	70,045	22,050	19.62%	15.92%	21.01%	5.09%	<.0001	***	0.973	0.860	1.046	0.186	<.0001	***
4.5	131,406	48,202	70,658	22,456	19.68%	15.94%	21.08%	5.15%	<.0001	***	0.996	0.874	1.074	0.199	<.0001	***
5	132,047	48,333	71,101	22,768	19.72%	15.94%	21.15%	5.21%	<.0001	***	1.014	0.885	1.096	0.212	<.0001	***
5.5	132,492	48,432	71,402	22,970	19.75%	15.96%	21.19%	5.23%	<.0001	***	1.028	0.894	1.114	0.220	<.0001	***
6	132,872	48,517	71,662	23,145	19.78%	15.97%	21.22%	5.25%	<.0001	***	1.041	0.902	1.131	0.229	<.0001	***
…				…				…						…		
10	134,174	48,819	72,505	23,686	19.86%	15.98%	21.32%	5.34%	<.0001	***	1.104	0.942	1.205	0.263	<.0001	***
100	135,015	49,070	73,001	23,931	19.90%	16.00%	21.36%	5.36%	<.0001	***	1.228	1.059	1.330	0.271	<.0001	***
1000	135,079	49,101	73,030	23,929	19.90%	15.99%	21.36%	5.36%	<.0001	***	1.411	1.294	1.502	0.208	<.0001	***
∞	135,119	49,117	73,052	23,935	19.90%	15.99%	21.37%	5.38%	<.0001	***	2.266	2.270	2.394	0.124	<.0001	***

Figure 6 The Illustrations of Effects of Inferred-MSA Income Ratio Cap

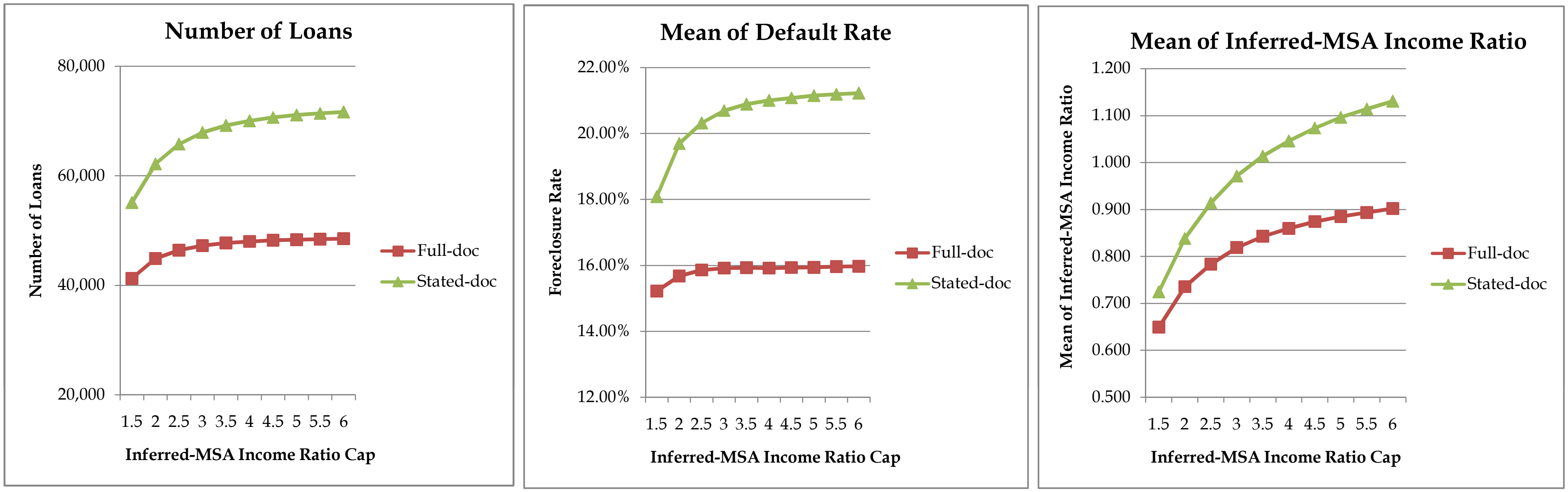


Table 12 Regressions results for the effects of inferred-MSA median income ratio cap

Variable	All Loans		Full-Doc		Stated-Doc		Stated-Full Difference	
	Estimate	p-value	Estimate	p-value	Estimate	p-value	Estimate	p-value
[Specification 1] Regression of loan number sample mean								
intercept	108714	<.0001 ***	42205	<.0001 ***	56182	<.0001 ***	13978	<.0001 ***
Cap	4735.079	0.002 ***	1250.703	0.003 ***	3019.976	0.001 ***	1769.273	0.001 ***
Adj. R-square		0.699		0.648		0.722		
[Specification 2] Regression of default rate sample mean								
intercept	0.179	<.0001 ***	0.154	<.0001 ***	0.185	<.0001 ***	0.031	<.0001 ***
Cap	0.004	0.004 ***	0.001	0.017 ***	0.005	0.004 ***	0.004	0.002 ***
Adj. R-square		0.629		0.468		0.636		
[Specification 3] Regression of inferred-MSA median income ratio sample mean								
intercept	0.665	<.0001 *	0.640	<.0001 *	0.682	<.0001 *	0.043	0.002 ***
Cap	0.070	<.0001 ***	0.049	0.000 ***	0.083	<.0001 ***	0.033	<.0001 ***
Adj. R-square		0.876		0.835		0.892		

Appendix C

Liar's Loan?

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

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

This Draft: September 2009

---

[1] The authors thank a major national mortgage bank for providing the data and assistance in data processing and the National Science Foundation (NSF Grant #SES-0851428) for financial support. Comments and suggestions from Vyacheslav Fos, Chris Mayer, Daniel Paravisini, Tomasz Piskorski, David Scharfstein, Amit Seru, Bob Van Order, and seminar/conference participants at Columbia, Georgia State, Kansas City Federal Reserve Bank, the NBER 2009 Summer Institute, and the FDIC 2009 Mortgage Symposium, have contributed to this draft. The authors also thank Erica Blom, Sunyoung Park, and Mike Tannenbaum for excellent research assistance.

[2] Corresponding author. Columbia Business School, Finance and Economics Division, Tel: 212 854 9002, Email: wj2006@columbia.edu.

[3] Indiana University, School of Public and Environmental Affairs, Tel: 812 855 5971, Email: ashlyn@indiana.edu.

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

Liar's Loan?

Effects of Origination Channel and Information Falsification on Mortgage Delinquency

ABSTRACT

This paper presents a comprehensive predictive model of mortgage delinquency using a unique dataset from a major national mortgage bank containing all of its loan origination information from 2004 to 2008. Our analysis highlights two major agency problems underlying the mortgage crisis: an agency problem between the bank and mortgage brokers that results in lower quality broker-originated loans, and an agency problem between banks and borrowers that results in information falsification by borrowers of low-documentation loans--known in the industry as "liars' loans"--especially when originated through a broker. We also document significant differences in loan performance by race/ethnicity that cannot be explained by observable risk factors or loan pricing.

The recent crisis in the housing and mortgage debt market has drawn considerable attention from regulators and market participants. A decade-long boom in the housing market and related financial sectors was followed in 2007 by a market bust with falling house prices and a rapid increase in mortgage defaults and foreclosures. The nationwide delinquency rate on subprime loans reached 39% by early 2009, more than seven times the level in 2005.[5] Those caught in the crisis included large financial institutions that experienced sharp expansion in, and profited from, their exposure to mortgage loans. The crisis that started from the mortgage market quickly spread to other financial markets and throughout the economy.

We use the experience of a major national mortgage bank to uncover the determinants of the mortgage crisis and the evolution of the crisis at a micro level. The particular bank provides an ideal context for the study by presenting a representative and yet amplified version of the boom-and-bust cycle experienced by the national mortgage sector in the last decade. First, the bank was among the nation's top ten mortgage banks in 2006 and was one of the fastest growing players in the mortgage market, specializing in low- and no-documentation loans (nicknamed "liars' loans," which constitute a large portion of the Alt-A loans) while also providing full-documentation loans (about 30% of their total loan originations). Second, the bank suffered one of the largest losses in the industry since the 2007 crisis.

[5] Source of information: LPS Applied Analytics website: http://www.lpsvcs.com/NewsRoom/IndustryData/Pages/default.aspx . Delinquency is commonly defined as payment delinquency of 60 days or more, including foreclosure.

Loans issued by the bank since the beginning of 2004 reached a cumulative delinquency rate of 28% by early 2009; approximately half of these delinquent loans were in the state of short sale or foreclosure. Finally, the borrowers and properties underlying the bank's loans during our sample period have fair representations in all 50 states. Therefore, lessons from this particular bank have general implications for the national mortgage market.

The proprietary data set represents the most detailed and disaggregated data sets so far in the mortgage loan literature. Our data set consists of all 721,767 loans that the bank originated between January 2004 and February 2008. We have all of the information that the bank collected at the time of loan origination, as well as monthly performance data for each loan through January 2009. Our data set includes not only information about the loan (pricing, loan product, and other contractual terms) and the property (address, appraisal value, owner occupancy status, etc.), but also about the borrowers demographic characteristics (race, age, gender, etc.) and economic conditions (income, cash reserves, employment status, etc.). Finally, we are able to use the property address information to match about three-quarters of the loans to community attributes such as demographics and business opportunities in narrow localities.

Our sample is divided into four distinct subsamples by a two-way sorting. The first sorting variable is the loan origination channel: whether a loan is originated directly by the bank or through a third party originator (such as a mortgage broker or a correspondent; henceforth, we simply call this category brokered loans). The second sorting variable is the loan documentation level: whether a loan is originated with full documentation of borrowers' economic conditions or with various reduced levels of documentation (including no documentation). Throughout the paper we refer to the four subsamples (with the initial letters capitalized) as: Bank/Full-Doc; Bank/Low-Doc; Broker/Full-Doc; Broker/Low-Doc. The Bank (Broker) subsamples include both Bank/Full-Doc and Bank/Low-Doc (Broker/Full-Doc and Broker/Low-Doc) subsamples, and the Full-Doc (Low-Doc) subsamples are defined analogously.

Our empirical analysis uncovers two types of agency problems in mortgage lending which constitute the fundamental causes of high loan delinquency rates, and by extension, the mortgage crisis. The first agency problem lies between the bank and its mortgage brokers. We find that loans in the Broker subsamples have delinquency probabilities that are 10-14 percentage points (or more than 50%) higher than the Bank subsamples, a manifestation of the misalignment of incentives for brokers who issue loans on the bank's behalf for commissions but do not bear the long-term consequences of low-quality loans. A binary decomposition attributes three-quarters of the Bank-Broker delinquency gap to differences in observable borrower characteristics, and the remaining quarter to differences due to unobserved heterogeneity. Hence, the higher delinquency rates among brokered loans are explained largely by broker penetration of borrower pools that were of observably worse quality (according to credit score, loan-to-value ratio, income, etc.) than the borrower pools penetrated by the bank.

Within each origination channel, the Low-Doc subsample exhibits worse performance than the Full-Doc subsample, and the difference in delinquency is 5-8 percentage points. The same decomposition method reveals that unobserved heterogeneity explains nearly 100% of this difference. In contrast to the Broker channel, the Low-Doc channel does not necessarily compromise lending standards along the observable metrics, but suffers from less careful verification 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). This relation highlights the second agency problem that lies between the lender and the borrower, where the latter could hide or even falsify unfavorable information, especially in the context of lax screening and verification procedures.

We provide evidence of borrower information falsification at both individual variable and aggregate levels. First, we find that both the in-sample goodness-of-fit and the out-of-sample predictive power of our delinquency prediction model are about 50% higher for the Full-Doc subsamples than for the Low-Doc subsamples. These differences suggest that borrower information collected for low-documentation loans is of lower quality, either in terms of inaccurately recorded data or intentionally falsified information, thereby compromising the ability of such information to predict delinquency. Second, certain variables--notably income--exhibit weak or even perverse relations to delinquency probabilities among low-documentation loans. These weak or perverse relations are especially evident in the Broker/Low-Doc subsample, where brokers both apply looser lending standards and are less diligent in verifying borrower information. The most plausible explanation for this observed pattern is information falsification. Through further analysis, we conservatively estimate that the median magnitude of income exaggeration is about 20% among low-documentation borrowers.

Finally, we document significantly higher delinquency rates among Hispanic and black borrowers. The differences in delinquency rates--4 to 11 percentage points higher for Hispanics and 3 to 4 percentage points higher for blacks, relative to white borrowers--are not explained by the full set of individual risk factors collected at loan origination, or by differences in loan pricing. Our analysis--which includes far more detailed data than that used in prior research on the relationship between race/ethnicity and credit--does not support a finding of discrimination, whereby minorities are subjected to higher lending standards or higher pricing for given financial products. Rather, the findings suggest that systematic differences between white and minority borrowers--such as information and experience disparities resulting from a lack of prior home buying experience or exposure to mainstream financial institutions--may explain these delinquency differences.

Our paper builds on a fast-growing literature on the mortgage crisis,[6] and most closely relates to a few recent empirical papers exploring the causes of the mortgage crisis using large sample micro-level archival data. Mian and Sufi (2008) identify the effects of the increase in the supply of mortgage credit on fueling the housing bubble between 2001 and 2005, and on the subsequent large increase in mortgage defaults. Demyanyk and Van Hemert (2008) and Keys, Mukherjee, Seru, and Vig (2008) both use data from LoanPerformance, a provider of performance data on securitized loans. 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. Deng, Quigley, and Van Order (2000) analyze mortgage termination risk using large sample of loans purchased by the Federal Home Loan Mortgage Corporation. Commercial or government agency loan databases mentioned above usually do not include borrower demographic characteristics, detailed loan contractual terms, or location (address) information, and usually only include securitized loans. Some earlier papers (e.g., Munnell, Tootell, Browne, and McEneaney (1996)) obtain demographic information from government data sources, such as those reported for compliance with the Home Mortgage Disclosure Act (HMDA). However, loan performance and detailed location information are absent from these data sources, as are certain central economic variables such as the borrowers' credit scores and the loan-to-value ratio.

The contribution of this paper can be summarized as follows. First, the unique dataset allows us to present the most comprehensive and updated predictive model of delinquency in the literature. The comprehensive list of predictors—including data on loan contract terms, property characteristics, and borrower attributes—not only afford us a better understanding of the determinants of loan delinquency, but also provide us an accurate calibration of the information possessed by the bank, thereby facilitating analyses of the moral hazard and adverse selection problems in the loan market. Moreover, with loan performance information updated to early 2009, we are able to capture the full effect of the crisis on the mortgage market. Second, we model the borrower choice of loan types and quantify the agency problems arising from the broker origination channel and from information falsification among low-documentation loans to the current mortgage crisis. Finally, we find evidence of a race/ethnicity effect in mortgage loan performance, underlining the need to examine mortgage lending practices--such as those that disadvantage less experienced borrowers--that may disparately impact minorities.

The rest of the paper is organized as follows. The next section provides a detailed data description. Section II contains a comprehensive analysis of predictive models of loan delinquency. Section III models borrowers' choices of loan origination channel and documentation level, and decomposes the cross-subsample differences in delinquency rates into two components: one reflecting

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

observable borrower characteristics or lending standards, and another reflecting unobservable borrower heterogeneity. Section IV documents and quantifies borrower information falsification among low-documentation loans. Section V discusses the relationship between race/ethnicity and loan performance. Finally, Section VI concludes.

I. Data and Sample Overview

A. Data Sources and Description

As described in the prior section, our proprietary data set contains 721,767 loans funded by the bank between January 2004 and February 2008. Our sample includes prime, Alt-A, and subprime mortgages.

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), loan purpose (such as home purchase or refinance), origination channel (broker versus bank-originated), and documentation requirements.

Property data used in our analysis includes the property address, whether the property will be owner-occupied and used as a primary residence or used as an investment property/second home, and home appraisal value. Borrower data includes protected class demographic variables collected under the Home Mortgage Disclosure Act (HMDA) such as race, ethnicity, gender, and age, as well as all financial and credit information collected at origination: income, cash reserves, expenditures, additional debts, bankruptcy and/or foreclosure status at loan origination, credit score,[7] employment status, employment tenure (months in current job), 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 delinquency status: whether the loan payments are current or delinquent, the number of days delinquent, and whether the property is in a state of short sale or foreclosure.

[7] Fair Isaacs Co. developed the first nationwide, general purpose credit scoring model and released the eponymous FICO score in 1989. Since then, each of the three major credit-reporting bureaus--Equifax, Experian, and TransUnion--have developed proprietary credit scoring models and jointly developed the VantageScore to compete with FICO. Most mortgage lenders use these scores as the primary measure of borrower credit risk. While there is some variation across the models used by the three credit bureaus--depending on the specific credit events reported to and/or collected by each bureau--the credit score used in this study is numerically comparable and analytically equivalent to the FICO score.

We are able to use the recorded property addresses to match approximately three-quarters of the loans to community attributes such as demographics and business opportunities in narrow localities. Using the ArcGIS geo-coding software and Decennial Census geographic boundary files, we match the property addresses to their census tract, zip code, metropolitan statistical area (MSA), and county. The geographic distribution (at the county level) of the properties in our sample is plotted in Figure 1; the sample properties have fair representations in all 50 states, and their distribution is roughly proportional to population density.

[Insert Figure 1 here]

We also obtain the following information at the census tract level from the Decennial Census and the Bureau of Labor Statistics: population count, median age of the residents, percent of residents who are black or Hispanic, and unemployment rate. In addition, we obtain zip-code level average household income information from the Internal Revenue Service's Individual Master File system.

B. Sample Overview

During the sample period, the bank experienced substantial changes in the composition of its loans and borrowers, as did the national mortgage market. Figure 2 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 precipitous decline starting in the second half of 2007.

[Insert Figure 2 here.]

Figure 2 also shows that the rapid expansion in loan production was driven almost exclusively by increased loan originations through the broker channel, and expansion of low-documentation loans through the broker channel in particular. Broker-originated loans represented 73% of all loan originations in the first half of 2004, increasing to 94% by the second half of 2006; while broker low-doc loans accounted for 39% of originations in early 2004, they comprised 75% of loan originations by late 2006.

Cumulative delinquency rates progressively and substantially increased over the time period in our sample; at 18 months after origination, only 6.7% of loans originated in the first half of 2004 were ever more than 60 days delinquent, as compared to 23.9% of loans originated in the second half of 2007. Demyanyk and Van Hemert (2008) document a similarly deteriorating trend for subprime loans from 2001-2006 using the LoanPerformance database.

We define all of the variables used in this paper in Table 1 Panel A, and we report their mean, median, and standard deviation values at a semi-annual frequency in Table 1 Panel B.

[Insert Table 1 here.]

The time trends in the key determinants of delinquency reflect changes in housing prices, the loosening of lending standards during the boom period (2005 - 2006), and the subsequent tightening of loan underwriting guidelines by the bank starting in 2007. Mean loan-to-value ratios (LTV, the ratio of loan amount to the property's appraised value) decreased from 69% in late 2004 to 65% in early 2007 before climbing to 77% in early 2008, mostly varying inversely with housing prices. Median borrower credit score was 707 in early 2004, ranged from 689-694 in 2006 through early 2007, and subsequently increased in 2007. Simultaneously, median reported income increased from $5,500 per month in early 2004 to $6,500 in late 2006, before trending downward. The growth in borrowers' incomes through the end of 2006 may result from the booming economy as well as from borrower income falsification on low-documentation loans. Statistics on borrower job tenure exhibit a U-shape: median job tenure (a proxy for job stability) decreased from 60 to 50 months at the peak of the boom, before bouncing back to 60 months at the end of the sample period.

The housing boom welcomed many first-time homebuyers to the mortgage market. In early 2004, only 7.6% of borrowers in the sample were first-time homebuyers, a figure that climbed to 18.1% by late 2006. As the housing market collapsed and lenders tightened standards, the percent of first-timers fell to 12.7% by the end of 2007. During the sample period, black and Hispanic borrowers gained a significantly higher share of new loan originations. In early 2004, they represented 4.5% and 7.5% of the borrower population, respectively; by early 2007, the percentages were 8.9% and 23.3%. More strikingly, the proportion of blacks and Hispanics who were first-time borrowers increased from 10.3% in early 2004 to more than 25% in late 2006. The national mortgage market experienced a similar increase during the same period in the percentage of first-time homebuyers and the expansion of credit to minority households, who were disproportionately first-time homebuyers. According to national HMDA data on home purchase loans,[8] 6.6% (10.8%) of borrowers were black (Hispanic) in 2004; the numbers increased to 8.7% (14.4%) in 2006.

C. Sample Representativeness

Given that our analyses build on information from one bank, it is natural to ask how representative this sample is and to what extent our results can be generalized. The large mortgage bank under analysis operated under an "outsource origination to distribution" business model, wherein nearly 90% of loans were broker-originated, and 72% of loans were originated by non-exclusive brokers. These figures are considerably higher than those for mortgage banks with more traditional models; for example, a Wall Street Journal article in 2007 estimates that brokers originate around 60% of all home loans.[9] In

[8] Source of information: http://www.ffiec.gov/hmdaadwebreport/NatAggWelcome.aspx.
[9] See "Mortgage Brokers: Friends or Foes?" by James Hagerty, *The Wall Street Journal*, May 30, 2007.

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 75-91% reported in "Inside Mortgage Finance" for subprime and Alt-A loans during the same period.[10]

We further compare our 2004-2008 sample average statistics (reported in Table 1 Panel B) to those covered by McDash Analytics, the most comprehensive commercial database on mortgage performance, to assess whether the loan and borrower profiles in our bank sample are representative of the general mortgage market. The comparison dataset is used in recent studies such as Pikorski, Seru, and Vig (2009).[11] Our sample exhibits a comparable LTV, higher loan amount (about 15% higher on average), and lower credit score (about 5-8 points lower).[12] Finally, low-documentation loans represent just 20% of loans in the McDash database, but represent 70% of our sample. The difference is due to the lender's specialization in low-documentation loans.

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

In summary, the bank in our analysis pursued an aggressive expansion strategy relying heavily on broker originations and low-documentation loans in particular. The strategy allowed the bank to grow at an annualized rate of over 50% from 2004 to 2006. Such a business model is typical among the major players that enjoyed the fastest growth during the housing market boom and incurred the heaviest losses during the downturn. By January 2009, the delinquency rate among the bank's outstanding loans approached 26%; while this figure is significantly higher than the industry average of 10.4%, the delinquency rate of subprime loans is comparable to the industry subprime average of 39%.[14]

This particular bank experienced a representative and yet amplified version of the boom-bust cycle experienced in the mortgage industry overall, thereby providing unique insights into the agency problems underlying the mortgage crisis. To avoid generalizing on empirical relations that emerge from

---

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

[11] We thank Amit Seru for providing the summary statistics for this dataset.

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

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

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

the bank's particular loan composition, we conduct our analyses on subsamples partitioned by loan type (origination channel and documentation level), rather than on the pooled sample.

II. Prediction of Loan Delinquency: Model Specification

A. General Framework

The most important question in the mortgage literature is how to predict delinquency. We estimate two predictive models of delinquency, where we maintain the standard definition of delinquency as the borrower being at least 60 days behind in payment, or being in a more serious condition of default (such as short sale or foreclosure). Our first model uses probit regressions to predict the occurrence of delinquency for individual loans at any point in time during the sample period; our second model uses duration analysis to predict the length of time between loan origination and the first occurrence of delinquency.

While our sample includes all loans issued by the bank from January 2004 to February 2008, our performance data is updated through January 2009. Figure 3 plots the cumulative delinquency rates (since origination) of loans by origination date, in half-year intervals. It shows that loans originated during 2006 (2004) have the highest (lowest) cumulative delinquency rates, and more recently originated loans have higher delinquency rates during the first year of their lives.

[Insert Figure 3 here.]

The covariates in our regression analysis include loan contract terms,[15] borrower financial conditions, and borrower demographics. We partition the sample into four subsamples through a two-by-two sorting as outlined in the previous section: Bank/Full-Doc, Bank/Low-Doc, Broker/Full-Doc, and Broker/Low-Doc. All analyses throughout the paper, unless otherwise stated, control for loan origination year fixed effects and report standard errors that are robust to heteroskedasticity and within-cluster correlation of observations at the MSA level[16] to account for common shocks to real estate markets in the same MSA. The effective number of observations for the purpose of computing standard errors of estimated parameters is on the order of the number of clusters, which is 983 in the full sample. Finally, we use the 5% level as the criterion for statistical significance.

We do not include interest rates as regressors in our delinquency analysis because of two major complications. First, interest rates are endogenous to delinquency propensity. Second, our current dataset includes only initial and current interest rates, which may not be informative of the long-term interest rate

[15] Loan maturity is not included in the list of regressors due to a lack of variation; 30-year loans comprise 93% of our sample (the majority of the remainder are 15-year and 40-year loans).
[16] For observations where an address cannot be matched to any MSA, we form the clusters at the state level.

for variable-rate loans originated in recent years. We leave the full analysis of loan pricing to a separate paper. However, in Section V we consider the effects of interest rate on the differential delinquency rates across demographic groups.

B. Probit Analysis

The probit regression specification is as follows:

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

In equation (1), $Delinquency_i^*$ is the underlying propensity of delinquency, and $Delinquency_i$ is an indicator variable for actual delinquency.

We conduct the analysis separately for each of the four subsamples, and report the results in Table 2. We report the estimated coefficients of the probit model (β) and standard errors robust to clustering at the MSA level. We also report estimates of the average partial effects (APE), where the APE is defined as:

$$APE = E\left(\partial \Pr(Delinquency_i = 1 \mid X_i) / \partial X_i\right). \quad (2)$$

Our estimates of the APE are the empirical analog to the expression above:

$$\widehat{APE} = \hat{\beta}\frac{1}{n}\sum_{i=1}^{n}\phi\left(X_i\hat{\beta}\right), \quad (3)$$

where $\phi(\bullet)$ is the standard normal probability density function. The APE associated with a covariate is determined by both the underlying sensitivity of delinquency propensity to this covariate (β) and the sample distribution of all covariates (the sample average of $\phi(X\beta)$).

[Insert Table 2 here.]

C. Duration Analysis

In our duration analysis, we define the start of a spell as when the loan is originated; the failure of the spell is when the loan first becomes delinquent, and the duration of the spell is the time from loan origination to the first incident of delinquency. The duration of the spell is right censored if the loan is in good standing at the end of our sample period (the end of January 2009). The duration time is parameterized as follows:

$$\ln(t_j) = X_j\beta + \varepsilon_j\,. \quad (4)$$

We adopt the log-logistic distribution (very close to the log-normal distribution) for the "accelerated time" $\tau_j = \exp(-X_j\beta)t_j$. Accordingly, (4) can be re-expressed as:

$$\ln(t_j) = X_j\beta + \ln\left(\tau_j\right). \tag{5}$$

Moreover, the survival function is:

$$S\left(t_j\right) = \left[1 + \left\{\exp(-X_j\beta)t_j\right\}^{1/\gamma}\right]^{-1}. \tag{6}$$

In this model, the coefficient β has a semi-elasticity interpretation; that is, $\beta = \partial[\ln(t)] / \partial X$. A positive coefficient means that a higher value of the covariate is associated with a longer time to delinquency or equivalently a *lower* propensity to default within any given time span.

It is worth noting that the parameter γ in the survival function (6) provides flexibility on the duration dependence of the model, which is an attractive feature of the log-logistic specification. If $\gamma \geq 1$, the hazard rate is monotonically decreasing. That is, the instantaneous propensity to delinquency (conditional on the loan being in good standing up to that time) decreases over time. If $\gamma < 1$, then the hazard increases and then decreases over time. Moreover, a lower γ value is associated with a later peak in the higher hazard rate and a higher overall hazard rate for any given value of *Xβ*.

We estimate separately for the four subsamples the duration model using the maximum likelihood method; the results are reported in Table 3. In addition to reporting the estimated β coefficients and their standard errors, we also report the marginal effect of a one-unit change in the covariate (from the mean values) on the expected median duration of the spell (according to the survival function given by (6)).

[Insert Table 3 here.]

Though the probit and duration analyses are closely related, they examine somewhat different aspects of the propensity to delinquency. In the probit analysis, all loans that are delinquent at any point in time during the sample period are treated the same. While the probabilistic results are intuitive, they do not capture the accuracy of duration, i.e., the time from origination to delinquency. On the other hand, a duration analysis does not distinguish a pool of loans with a low occurrence of quick delinquency from another pool of loans with higher delinquency rates but where delinquency tends to occur among more seasoned loans. For these reasons, the two sets of results complement one another. When they are mutually consistent, our discussion will focus on the probit results because they are easier to interpret. The following sections provide a detailed discussion of the results from both tables, along with additional analyses.

III. Loan Types and Attribution of Differences in Delinquency

This section discusses the differences in loan performance across loan type: origination channel (Bank vs. Broker) and documentation level (Full-Doc vs. Low-Doc). We further analyze two related issues: First, which covariates determine a borrower's choice of loan type? Second, how can we decompose the differential delinquency rates across loan types into differences due to observable characteristics versus unobserved heterogeneities?

A. Differences in Loan Performance by Loan Type

A prominent feature of our results is that broker-initiated loans exhibit much higher delinquency rates than bank-initiated loans, as evidenced by the subsample summary delinquency rates at the bottom of Table 2. The difference in the probabilities is greater than 10 percentage points, a difference that is statistically and economically highly significant, indicating serious conflicts of interest in the brokerage channel where the loan originators' incentive is to maximize fees and commissions without bearing the long-term consequences of low-quality loans. [17]

The contrasts among subsamples are even more striking in the duration model. The median duration times (in months) reported at the bottom of Table 3 reveal that a loan originated with full documentation by the bank has a median life of 25 years (300 months) before delinquency; the same median lifetime drops steeply to 8.4 years for Bank/Low-Doc loans, and to 7.9 years for Broker/Full-Doc loans. Finally, the median life is a mere 4.6 years for Broker/Low-Doc loans.

The comparison of the delinquency propensity between Bank/Low-Doc and Broker/Full-Doc loans is not straightforward. While the former have a considerably lower overall delinquency rate, their median time to delinquency is comparable to the latter (the difference is not statistically significant). Moreover, the γ estimate (reported at the bottom of Table 3) is in fact smaller for the Bank/Low-Doc subsample than for the Broker/Full-Doc subsample, indicating a higher hazard rate in the former, conditional on covariates. Such a combination implies that, conditional on delinquency, the borrowers from the Bank/Low-Doc channel go into delinquency more quickly. Plausibly, a borrower who will default quickly after loan origination should be easier to screen out than a borrower who defaults years into the life of the loan. Therefore, low documentation leads to financing some of the more "obvious" low-quality borrowers.

B. Choice of Loan Origination Channel and Documentation Level

[17] Using data on loans originated in Florida in 2002, LaCour-Little (2009) shows that brokered loans tend to have higher interest rates (about 20 basis points) than loans available directly from retail lenders. Alexander, Grimshaw, McQueen, and Slade (2002) document that brokered loans originated during 1996-1999 in a multi-lender sample were 15% more likely to be delinquent than loans in the same sample that were originated through the retail channel of the banks. The two studies do not contain the level of borrower detail in this study.

Differences in loan performance by loan type raise the question of how borrowers select into different types of loans. Theoretically, a borrower living in any location can apply for a loan directly from the bank. In regions where the bank does not have branch operations, the loan application can be completed via phone or internet. Therefore, obtaining a loan from a broker represents a choice made by the borrower, or a lack of knowledge about available alternatives. The same can be said for choosing a low documentation loan. Table 4 reports our model results in two panels. Panel A uses only loan and borrower characteristics as regressors, while Panel B adds neighborhood characteristics to the list of covariates. The sample size for Panel B is about 25% smaller due to the additional data requirement.

[Insert Table 4 here.]

Column 1 of Table 4 Panel A indicates that the following variables predict a higher likelihood that a borrower will obtain a loan from a broker rather than from the bank: high debt level, original purchase (as opposed to refinance), first lien, first-time owner, owner-occupied, low income, low credit score, female borrower, minority borrower, young borrower, short employment tenure, and self-employed. All non-white racial groups favor the Broker channel in comparison to whites. 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. This relation calls attention to the issue of irresponsible lending--lending without due regard to ability to pay, to poorly informed borrowers--as analyzed by Bond, Musto, and Yilmaz (2008) and Inderst (2006).

The variables that predict choosing a low-doc loan have the following contrasts with those that predict choosing a broker. First, borrowers with low loan-to-value (LTV) ratios but high loan size are more likely to choose low documentation. Second, first-time owners and those purchasing owner-occupied properties are less likely to choose low documentation. Third, borrowers with high income and credit scores tend to choose low documentation. Fourth, black borrowers do not appear disproportionately in low documentation loans, while Hispanic and Asian borrowers do. Finally, age is not correlated with documentation level. To summarize, low documentation loans do not necessarily attract less-experienced borrowers. The most prominent summarizing feature of these borrowers seems to be that they are "good on paper." That is, borrowers who have favorable "hard" information (i.e., information that is quantifiable and could potentially be verified, such as LTV, prior mortgage experience, high income, and high credit score) choose low documentation.

Prior research has shown that lending practices and borrower characteristics are correlated with neighborhood characteristics (e.g., Calem, Gillen, and Wachter (2004), Nelson (2009)). Table 4 Panel B reports the relation between neighborhood characteristics and the respective likelihoods that a borrower will select the broker channel or apply for a low-doc loan. The model's regressors include average per capita income (*Avgincome)* at the zip code level, and also include the following regressors at the census tract level: Log population size (*Population*)[18], percentage of residents who are black (*Pctblack*) and Hispanic (*Pcthisp*), median age (*Medage*), and unemployment rate (*Unemprate*). All regressors included in the model reported in Table 4 Panel A are also included in the model reported in Panel B, but are not tabulated for economy of space.

Brokers seem to predominate in neighborhoods with low minority representations and young residents. The combination of results from Panels A and B indicates that minority households in non-minority neighborhoods are the prime clients of mortgage brokers. Low documentation loans, on the other hand, are significantly more popular in minority neighborhoods and in booming neighborhoods (with low unemployment rates) with young populations.

C. Decomposition of Pairwise Subsample Differences in Delinquency

When researchers try to examine the effect of a variable, they often include the variable as a regressor and estimate its contribution in explaining the outcome. Following this logic, we could estimate a regression model that includes loan type as a regressor:

$$Delinquency_i^* = X_i\beta + \lambda LoanType_i + \varepsilon_i, \tag{7}$$

where *LoanType* indicates the origination channel or documentation status. We refrain from conducting such an analysis because a specification like (7) is meant to capture a "treatment effect," where the relevant question is: if two *ex ante* identical borrowers--along both observable and unobservable dimensions--were assigned to different loan types, how would their delinquency propensity differ *ex post*?

We argue that there is no conventional "treatment effect" of the loan types in our context because all loans are serviced by the bank, regardless of the origination channel and documentation level. As a result, any difference in the outcome that is correlated with loan type should be attributed solely to the "selection effect"; that is, borrowers of different observable and unobservable characteristics are attracted to different loan types, and such characteristics are correlated with delinquency propensities.

The dichotomy between observable qualities and unobserved heterogeneities has implications for understanding why delinquency rates vary across subsamples. For example, if the higher delinquency rates in the Broker subsamples are predictable from observed characteristics (such as LTV and credit score), we could conclude that the Broker channel serves an observably lower-quality clientele, or applies looser lending standards than the Bank channel. If unobserved heterogeneity is responsible for the difference, then we infer that the Broker channel is subject to more severe adverse selection among potential borrowers along unobserved or unquantifiable dimensions (such as income stability, or hidden expenditures), presumably because mortgage brokers are less diligent than bank employees in using

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

additional hard or soft information to screen borrowers. The same logic applies to the Full-Doc/Low-Doc comparison.

We apply a non-linear version of the Blinder-Oaxaca (1973) decomposition to the probit model to separate the effects of observable qualities from the effects of unobserved heterogeneities. Let $D = \{0, 1\}$ be the index for the two subsamples for comparison, and let Y be the indicator variable for loan delinquency. More specifically, we will compare loans from the Bank ($D = 0$) and Broker ($D = 1$) channels, controlling for the documentation level, and loans issued as Full-Doc ($D = 0$) and Low-Doc ($D = 1$), controlling for the origination channel. We obtain coefficient estimates for all subsamples from the probit model as specified in equation (1) and reported in Table 2.

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

$$\begin{aligned} &E(Y \mid D=1) - E(Y \mid D=0) \\ &= \left\{ E\left[\Phi(X\beta^0) \mid D=1\right] - E\left[\Phi(X\beta^0) \mid D=0\right] \right\} + \left\{ E\left[\Phi(X\beta^1) - \Phi(X\beta^0) \mid D=1\right] \right\}, \end{aligned} \tag{8}$$

or as:

$$\begin{aligned} &E(Y \mid D=1) - E(Y \mid D=0) \\ &= \left\{ E\left[\Phi(X\beta^1) \mid D=1\right] - E\left[\Phi(X\beta^1) \mid D=0\right] \right\} + \left\{ E\left[\Phi(X\beta^1) - \Phi(X\beta^0) \mid D=0\right] \right\}. \end{aligned} \tag{9}$$

Equations (8) and (9) are numerically different but employ the same logic. The left sides of the equations are the difference in the expected value of the outcome variable (delinquency) between the two subsamples. The right sides of the equations feature a sum of two terms. In labor economics, the first term is called the "endowment effect"; that is, the difference in the outcome due to different distributions of the covariates (the X variables) in the two subsamples. The difference due to the endowment is isolated by using the same set of coefficients for both subsamples. The second term is called the "coefficient effect" (in a production function, the coefficients are also referred to as "returns to factors") and estimates the hypothetical difference in delinquency if the two subsamples had identical covariate distributions but the coefficients remained different. The coefficient effect encompasses two possibilities: a differential sensitivity of the outcome to the covariates in the underlying model, or the effects of missing variables that spill over to the remaining covariates. Both possibilities reflect unobserved heterogeneity.

Equations (8) and (9) differ only because they use a different subsample as the "base" sample. There is no *a priori* argument to favor using one subsample versus the other as the base, so we report both sets of results in Table 5. Table 5 Panel A reports the comparison of Full-Doc (D = 0) versus Low-Doc (D = 1) loans separately for the Bank and Broker channels. The total difference (the left sides of the above equations) is reported in the bottom row, and is, by construction, 100% of the difference. The "Low-Doc sample as benchmark" comparison applies equation (8) and uses the D = 1 subsample as the base; the "Full-Doc sample as benchmark" comparison applies equation (9) and uses the D = 0 subsample as the base. The t-statistics are based on standard errors obtained through the block bootstrap clustered at the MSA level.[19]

[Insert Table 5 here.]

The two sets of results are qualitatively similar, so we focus on the first set of results (equation (8)) for discussion. Conditional on the Bank (Broker) channel, Low-Doc loans have, on average, a delinquency rate that is 4.8 (8.0) percentage points higher than for Full-Doc loans. Almost 100% of this difference should be attributed to the "coefficient effect". The estimated "endowment effect" is small and is not statistically significant; if anything, the "endowment effect" indicates that Low-Doc loans are of slightly better observed quality. We conclude that Low-Doc loans are just as "good on paper" as Full-Doc loans, but encompass more adverse selection along unobserved dimensions.

The comparison between Bank and Broker loans conditional on documentation level (reported in Table 5 Panel B) offers a different picture. Here, the endowment effect accounts for three-quarters (over half) of the total difference in delinquency rates between Bank and Broker loans using the Broker (Bank) subsample as the base sample. Put differently, if the bank and its brokers had loaned to borrowers of the same *observable* quality, more than half of the difference in the incremental delinquency rate between the Broker and Bank subsamples (10.4 percentage points for Full-Doc, and 13.6 percentage points for Low-Doc) would have disappeared.

The implications stemming from the higher delinquency rates among Broker and Low-Doc loans are markedly different. 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 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). On the other hand, the Broker channel--while also lacking incentives for careful screening--penetrated a borrower pool that was of significantly worse quality, even by observable, quantifiable, and potentially verifiable standards.

The following hypothetical example illustrates the differences in borrower profiles across loan type. Suppose Borrower A has a high credit score and high income but has major withholding from his income (such as alimony); Borrower B has high income that is difficult to verify (because he is self-employed) or is unwilling to reveal his true income (because of tax reasons); and Borrower C has a low credit score and does not have a stable job or income. Our analysis predicts that borrowers A and B are more likely to choose low-doc loans, while Borrower C is more likely to approach (or be approached by) a mortgage broker.

[19] The conventional delta method for computing standard errors does not apply. The estimator is a function of the model coefficients that depends on the sample distribution of covariates, and thus is a stochastic function of the coefficients. In contrast, the delta method applies when the estimator is a nonstochastic function of the model coefficients.

Among all borrower characteristics, credit score has the highest predictive power for delinquency and is verified for full-documentation as well as for low- or no-documentation loans. Exploring the relationship between credit score and other covariates sheds additional light on the composition of borrowers in different subsamples. The results we report in Table 6 confirm our interpretation of results in Tables 4 and 5. We find that Low-Doc borrowers have, on average, higher credit scores than Full-Doc borrowers. Moreover, credit score and reported income and cash reserves are strongly related in the Full-Doc subsamples, but the relation is much weaker in the Low-Doc subsamples. The fact that reported income and cash reserves may not be certified in the Low-Doc subsample may explain their weakened relationship with credit score, an issue we discuss in more detail in Section IV.

[Insert Table 6 here.]

An examination of credit scores by race reveals that average credit scores are highest among Asian and white borrowers, and lowest among Hispanic and black borrowers. Hispanic borrowers who obtain loans directly from the bank have credit scores that are comparable to those of white borrowers, but those who obtain loans through a broker have credit scores that are on average 2-5 points lower. Black borrowers have average credit scores that are 14-27 points lower than white borrower credit scores, across all subsamples. Section V offers a more detailed analysis of these race/ethnicity effects.

Last, the time trend of credit scores, as shown by the year dummy variable coefficients, is informative; while Bank loans saw steady improvement in credit scores over time from 2004-2008, credit scores for Broker loans deteriorated from 2004-2007, and only recovered in 2008. The findings provide evidence that the bank pursued a growth strategy which relied on penetrating marginal borrowers through the broker channel.

D. Differences within the Broker Channel

We differentiate within the Broker channel between pure brokers and correspondents. Pure brokers act as matchmakers and submit loan applications to a variety of banks for competitive pricing. In contrast, the correspondents in our sample have long-term, established, and near-exclusive relationships with the bank for at least one product type (such as prime loans) and abide by the bank's particular underwriting guidelines in exchange for expedited loan processing. Correspondents in our sample close loans in their own name using a warehouse line of credit advanced by the bank, and then quickly re-sell the loans to the lending bank. Due to the longer and more exclusive relationships, the incentives of the correspondents are more aligned with that of the bank than pure brokers.

To examine the difference between the two groups of brokers, we estimate the probit model (equation (1)) for correspondents and non-correspondents separately, interacted with the Full-Doc/Low-Doc sorting. The double sorting produces four subsamples. We report the results in Table 7.

[Insert Table 7 here.]

A comparison of Table 7 to Table 2 confirms our conjecture. The patterns revealed in the Correspondent subsamples are always between those of the Bank subsamples and those of the Non-Correspondent subsamples, and tend to be closer to the former. For example, total delinquency rates for Correspondent loans are marginally higher than for Bank loans (5 percentage points higher for both Full-Doc and Low-Doc loans), but are much lower than for the Non-Correspondent subsamples (5.7 percentage points lower for Full-Doc, and 15.1 percentage points lower for Low-Doc). Also, there are more commonalities in the relations between loan performance and individual covariates among the Bank and Correspondent subsamples than among the Correspondent and Non-Correspondent subsamples.

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

The "liar's loan" problem includes various forms of borrower information falsification, possibly at the encouragement of brokers who have stronger incentives to close deals than to screen applicants. Such falsification appears primarily among low- or no-documentation loans, where much of the recorded information is self-reported without strict verification. Anecdotal evidence[20] suggests that the following falsifications are among the most common: exaggerating income or assets, hiding other major expenditures, and claiming that properties purchased for investment/speculation purposes will be owner-occupied as primary residences.

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

A. Model Predictive Power across Different Loan Types

Inaccurately recorded loan and borrower characteristics, whether due to unintentional mistakes or due to intentional falsification, will attenuate the empirical relationship between these variables and loan performance, thereby compromising the model's fit and predictive power. Because the bank services and

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

maintains records for all loans in our sample, there is no obvious reason to believe that incidences of random data recording error should vary systematically across the subsamples after loan origination. This leaves intentional falsification (including hiding) of information as the most plausible explanation for differences in model predictive power across loan type.

In Tables 2 and 3, we observe that the goodness-of-fit (i.e., the in-sample model predictive power) is indeed substantially different across the four subsamples. More specifically, the two Full-Doc subsamples have much higher pseudo R-squared statistics (22.1% and 18.2% for Bank and Broker subsamples using probit, or 17.4% and 16.2% using duration, respectively) than the two Low-Doc subsamples (13.6% and 14.6% using probit, or 14.1% and 14.5% using duration), indicating higher quality explanatory variables in the Full-Doc subsamples. Here the reported pseudo R-squared is $(1-\ln L/\ln L_0)$, where *lnL* is the maximized log likelihood value of the probit or duration model using all covariates, and lnL_0 is the maximized log likelihood value of the same model on the same sample, but with a constant as the sole regressor.

The pseudo R-squared discussed above is the most popular goodness-of-fit measure for non-linear models for which there are no obvious empirical analogs to the residuals. Nevertheless, it suffers from two major drawbacks. First, it does not have an interpretation as intuitive as the R-squared metric for linear models, which indicates the percent of variation explained. Second, the in-sample goodness-of-fit should not be equated with model predictive power. When economic agents (the bank or mortgage brokers) make decisions, their predictions are based on information revealed at the time, without knowledge of the full sample. Therefore, an out-of-sample prediction method is more appropriate for our research purposes, because it avoids the look-ahead bias. With these two issues in mind, we develop the following "excess percentage of correct predictions" measure to assess the predictive power of the probit model.

Let P_i denote the predicted probability of delinquency for the *i*-th observation, where the prediction is made out-of-sample (to be described in more detail later). Let Y_i denote an indicator variable for delinquency, and let $\bar{p}$ denote a cutoff value. Then the objective to maximize "correct predictions" can be expressed without loss of generality as:

$$S=\omega S_1+\left(1-\omega\right)S_2-\alpha=\omega\Pr(P_i\geq\bar{p}\mid Y_i=1)+\left(1-\omega\right)\Pr(P_i<\bar{p}\mid Y_i=0)-\alpha \tag{10}$$

for some $\omega\in(0,1)$, which reflects the relative importance of a type-I error (failure to predict a delinquent loan) and a type-II error (mistakenly predicting that a non-delinquent loan will be delinquent); α is a constant representing the maximum probability of obtaining a correct prediction with a random guess . The maximization of (10) has a unique solution of $\overline{p}$:[21]

$$\overline{p} = \frac{(1-\omega)E(Y)}{\omega\left[1-E(Y)\right]+(1-\omega)E(Y)} . \tag{11}$$

A natural choice of ω is 1/2, where the objective function weights the two types of prediction errors equally. Under such a criterion, equation (11) simplifies to $\overline{p} = E(Y)$, with the corresponding empirical analog being the sample frequency of delinquency revealed at the time of the evaluation.[22] According to this rule, we classify a loan as "predicted to be delinquent" if the out-of-sample predicted probability exceeds the time-adapted sample frequency of delinquency.

Such a classification method has the desirable feature of coinciding with the likelihood ratio rule if the probit model is correctly specified. Let f^D (f^{ND}) be the density functions of the predicted probability of delinquency for the subsample of loans that are *ex post* delinquent (non-delinquent). Then for any value v, $f^D(v) > f^{ND}(v)$ if and only if $v > E(Y)$, as long as the model is correctly specified, i.e., as long as equation (1) holds with the residual ε normally distributed. In other words, the two density functions $f^D(v)$ and $f^{ND}(v)$ have a single crossing at $v = E(Y)$. As a result, $P_i > E(Y)$ implies that the *i*-th observation is more likely to be drawn from the subsample of *ex post* delinquent loans than from that of the *ex post* non-delinquent loans, and therefore should be classified as "predicted to be delinquent" based on the relative likelihood. The opposite applies when $P_i < E(Y)$.[23]

Finally, the percentage of correct predictions should be judged against the benchmark of a non-informative model, which produces correct predictions half of the time in expectation when ω = 1/2. As a result, we set α = 1/2 in equation (10) to obtain the "excess percentage of correct predictions."[24]

We use the following empirical procedure to calculate the out-of-sample excess percentage of correct predictions. First, we divide each of the four subsamples into semi-year segments by the loan origination date, and pick one semi-year segment at a time to measure the accuracy of the model predictions. We call this the "test sample/period." Second, for each "test period," we use all information available up to just before the test period to estimate the model in equation (1) without the year dummy variables[25]; we call this the "estimation sample/period." It is important to emphasize that not only do the loans in the estimation sample have to be originated before the test period, but their delinquency status must also be assessed at the beginning of the test sample period. Third, we apply the predictive model

[21] The proof of equation (11) is in the appendix.
[22] Another natural choice of ω is *Pr[Y=1]=E(Y)*, which would lead to maximizing the un-weighted fraction of predictions correctly predicted. Under such a criterion, equation (11) simplifies to 1/2.
[23] The proof of this argument is in the appendix.
[24] For general values of ω, the corresponding α parameter is equal to *max(1- ω, ω)*.
[25] Time dummy variables should be omitted from any out-of-sample predictions because they are not applicable for future samples.

using the coefficients estimated from the estimation sample on the test sample to form the predicted probability of delinquency. Finally, equation (10) formulates the calculation of the final measures.

Table 8 reports the percentage of correct predictions by subsample for each semi-year, separately for S_1, S_2, and S as defined in equation (10). The test periods start from the first half of 2005 to allow for a prior estimation period.

[Insert Table 8 here.]

Two patterns evident in the table warrant further discussion. First, loan documentation type--not loan origination channel--is the key determinant of the model's predictive power. Figure 4 depicts model predictive power by plotting the time series of the excess percentage of correct predictions (S) by loan type. The model's predictive power in the Bank/Full-Doc and Broker/Full-Doc subsamples is indistinguishable in each semi-year; the same can be said about the model's predictive power in the Bank/Low-Doc and Broker/Low-Doc subsamples. More importantly, the model's predictive power in the Full-Doc subsamples is substantially higher than for the Low-Doc subsamples. The across-time averages are as follows: Bank/Full-Doc (17.2%), Bank/Low-Doc (11.5%), Broker/Full-Doc (18.1%), and Broker/Low-Doc (11.1%). Such a contrast suggests that low documentation loans may allow some borrowers to falsify information in order to qualify for loans or obtain more favorable loan terms. As a result, some of the variables in the regressions could contain measurement errors, compromising their predictive power.

[Insert Figure 4 here.]

Second, the predictive power of the model--especially for the Full-Doc subsamples—declined from 2005 to 2006, before rebounding slightly in 2007. This trend suggests that loans originated during the boom period experienced positive shocks in delinquency that could not be predicted by their characteristics based on information available at the time of loan origination. Rajan, Seru, and Vig (2009) also find that the predictive power of credit score and LTV deteriorated during the high securitization period.[26] The difficulty in predicting loan performance based on observed characteristics for loans originated in 2006 indicates the bank may not have been aware it was originating low-quality loans during that time period; this explains why the bank did not tighten its lending standards until 2007, when it began to incur losses from loans originated during the boom.[27]

B. Evidence of Borrower Information Falsification from Individual Variables

[26] We find that the deterioration in model predictive power is more prominent among full-documentation loans, while Rajan, Seru, and Vig (2009) found it to be stronger among low documentation loans. The difference could be due to our use of a larger set of covariates in the prediction and a different metric of model predictive power, and our use of a more recent sample which begins and ends later than theirs.

[27] Please also see Figure 3.

B1. Overview

The model's lower predictive power for Low-Doc subsamples relative to Full-Doc subsamples provides strong evidence that the information recorded for low-documentation loans is of lower quality. The lower predictive power is an aggregate measure of the quality of the recorded information, but it does not reveal which particular variables are mis-measured. We now present evidence that borrowers of low-documentation loans tended to falsify particular variables, especially income. We find that such falsification is especially prominent among Broker/Low-Doc loans.

Due to both incentives and the reporting system, falsification is most likely to occur in the following variables. First, borrowers purchasing a second home or investor 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), lenders are unable to verify occupancy status for home purchase loans at origination. Occupancy fraud is often cited as a major contributor to the surge in delinquencies, as borrowers became over-leveraged from holding multiple mortgages.

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

B2. Income Falsification

The coefficients on *Income* in Tables 2 and 3 support the hypothesis that reported income was often falsified by borrowers of Low-Doc loans.[28] In the Full-Doc subsamples, higher income is significantly and negatively associated with delinquency, as measured by both lower probability of delinquency and longer duration to delinquency conditional on all other attributes. However, the sign on the *Income* coefficient switches in the Low-Doc sub-samples. Moreover, the coefficients are particularly strong in the Broker/Low-Doc subsample where higher income is associated with significantly higher propensity for delinquency. The most plausible explanation for this contrast is that, when income is not verified, higher income (conditional on all other attributes) may more often be the result of exaggeration

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

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.

The dummy variable for missing income information, *IncomeMiss*, offers corroborative evidence. In the Bank/Full-Doc and Broker/Full-Doc subsamples, only 0.6% and 0.9% of the observations have missing income information, and in these subsamples missing income information does not predict loan performance. Thus, in the Full-Doc subsamples, the sporadic cases of missing income information most likely result from data recording error and not from falsification. In contrast, income is missing for 10.3% and 9.2% of the observations in the Bank/Low-Doc and Broker/Low-Doc subsamples, respectively. Missing income information significantly predicts higher delinquency propensity in the Broker/Low-Doc subsample, where missing income information is associated with a 4.7 percentage point increase in the probability of delinquency, or an 8 month reduction in the time from loan origination to delinquency. The same effect is present but not significant in the Bank/Low-Doc subsample. Thus, purposefully not reporting income information is a low-documentation-only phenomenon. Presumably, borrowers with low or irregular incomes in the Low-Doc subsamples are more likely than comparable High-Doc borrowers to exaggerate or omit their incomes on the loan application.[29]

In comparing Table 2 and Table 7, it is worth noting that the various perverse relations discussed above for broker-originated loans are mostly driven by non-correspondent brokers. This evidence suggests that correspondents are far less likely to encourage or accommodate borrower information falsification than non-correspondents because the former have stronger reputation concerns due to their exclusive or long-term relationships with the bank.

What is the magnitude of income falsification by borrowers when income is self-reported? While we are not able to pin down the exact number for any individual, it is possible to form some conservative estimates for the average extent of income falsification based on the following identifying assumption:

$$E(Income^{*} \mid X = x, \text{Low-Doc}) \leq E(Income^{*} \mid X = x, \text{Full-Doc})\,; \tag{12}$$

where *Income** denotes the borrower's true income, and *X* denotes a vector of borrower characteristics. Formally, equation (12) is implied by the condition that *Pr(Full-Doc|X, Income*)* is non-decreasing in *Income**.

All that is required for equation (12) to hold is a relative preference ordering: if Borrower A's true income is more favorable than Borrower B with similar characteristics, then on average A should not have a stronger preference for low-documentation loans than B. In general, such an assumption is

[29] 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.

plausible because a high certified income is more likely to result in lower interest rates or more favorable loan terms on full-documentation loans, while some of these benefits are forfeited in low-documentation loans because the sensitivity of loan pricing to uncertified income is lower. Self-reported income could still materially affect the qualification of the loan application, providing an incentive for falsification.

The only group for whom equation (12) may plausibly not hold is the self-employed. Self-employed borrowers disproportionately choose low-documentation loans (see the more detailed analysis in Section III and the results in Table 4), not necessarily because they want to exaggerate their income but because their income is often difficult to certify (e.g., they do not have W-2 forms) or they do not wish to reveal their true cash flows for tax reasons. We therefore exclude the self-employed from our estimation of the extent of income exaggeration among borrowers of low-documentation loans.

Our first estimate of the extent of income exaggeration comes from simply comparing borrower income (at the household level) to the average income of the neighborhood where the property is located. We obtain the average per capita adjusted gross income information at the zip code level from the Internal Revenue Service's Individual Master File (IMF) system for the years 2004, 2005, and 2006. A zip code area has, on average, 2,326 households, 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 the neighborhood average income per capita 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 (12) implies that the average degree to which low-documentation borrowers exaggerate their income is at least 16%-19%, if their true income stands at a ratio to their neighborhood average that is no higher than their full-documentation counterparts.

A more refined estimate incorporates borrower demographics in addition to neighborhood attributes to proxy the 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, with the error term mean independent of covariates conditional on documentation status. Then such a function could be estimated reliably using the sample of full-documentation loans because there should be no systematic bias in the recorded income given that it is certified; hence, $Income \approx Income^*$. Below is the regression output from full-documentation loans, where the dependent variable is the reported (and certified) household monthly income in $1,000 units and the t-statistics are reported below the coefficients.

$$
\begin{aligned}
Income = {} & 0.014 * CreditScore - 0.846 * Female + 0.651 * \ln(Age) - 0.416 * Hispanic \\
& [18.01] \qquad [-16.49] \qquad [13.31] \qquad [-1.92] \\
& - 0.430 * Black + 0.575 * Asian + 0.051 * AvgIncome - 0.030 * Unemprate \\
& [-4.31] \qquad [5.04] \qquad [4.40] \qquad [-2.15] \\
& + 0.131 * Y2005 + 0.373 * Y2006 + 0.299 * Y2007 + 0.010 * Y2008 \\
& [2.58] \qquad [5.40] \qquad [4.76] \qquad [0.096]
\end{aligned}
\tag{13}
$$

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

All coefficients in the above regression are intuitive. Older borrowers and borrowers with higher credit scores tend to have higher income. Female borrowers have lower income on average.[30] Black and Hispanic borrowers have lower income on average than white borrowers, and Asian borrowers as a group have the highest income. In addition, borrower income is significantly and positively correlated with the zip-code area average income (*AvgIncome*) and negatively correlated with the census tract unemployment rate (*Unemprate*). Finally, overall borrower income grew from 2004 (the omitted year in the regression) to 2006, and then decreased afterwards.

Resorting to the identifying assumption of (12)--which presumes that the error term from regression (13) is not positively correlated with Low-Doc status--we can estimate the upper bound for the expected true income of low-documentation borrowers by applying the estimated coefficients from (13) to the covariates of these borrowers. We generate an "income exaggeration" variable to capture the difference between the reported *Income* and the estimated *Income**. We find that in dollar terms the 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. Given that these estimates err on the conservative side, the data suggest serious income falsification among low-documentation borrowers from the benchmark of full-documentation borrowers.

The correlations between estimated true income, estimated income exaggeration and loan performance are all highly statistically significant, and reveal more about the incentives for and consequences of income falsification. First, the correlation between the estimated true income and estimated income exaggeration in percentage terms is -7.9%, indicating a stronger incentive to inflate income when the true income is lower. Second, the correlation between the estimated true income and *ex post* delinquency is -23.5%, recovering the normal inverse relationship between income and delinquency in the Low-Doc subsample that was perverted using reported income (see Tables 2 and 3). Finally, as expected, the correlation between estimated income exaggeration and *ex post* delinquency is positive at

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

8.2%. In other words, delinquency risk increases when borrowers inflate income to obtain a loan beyond their true means.

B3. Evidence of Other Information Falsification

Additional important variables for delinquency prediction, which potentially can be falsified in the absence of certification, are *OwnerOccupied* (a dummy variable for whether the property is owner-occupied as a primary residence) and *CashResv* (the borrower's cash reserves in multiples of the monthly mortgage payment, in logs). Mortgages on owner-occupied properties are usually considered to be safer than properties purchased as investments or second homes; the latter are often purchased by borrowers who have higher combined leverage and who have a lower cost of "walking away" from a mortgage that has negative equity value. Cash reserves help pull households through temporary negative income shocks without disrupting mortgage payments.

The coefficients on both variables in Tables 2 and 3 reveal patterns that are generally intuitive: owner occupancy and high levels of cash reserves are associated with significantly lower delinquency propensity. However, the coefficients that represent sensitivity of delinquency propensity to both variables are stronger in the Full-Doc subsamples than in the Low-Doc subsamples,[31] and the difference is more evident using the duration method than the probit model. Hence, there is some evidence that Low-Doc borrowers may not always truthfully report owner occupancy status and cash reserves, thereby lowering the explanatory power of these variables. Yet the evidence is weaker and less conclusive than that regarding income falsification.

Our conversations with bank officials yield two explanations for the higher quality of cash reserve information relative to income information in explaining delinquency. First, borrowers and brokers have better information about how income affects loan qualification and pricing, so they have a stronger incentive to falsify income. Second, verification of assets is often better than that of income because asset statements are more available than proof of income for a large group of borrowers, especially those who are self-employed or cash compensated.

V. "The Color of Credit:" Race/Ethnicity and Loan Performance

There is a large body of research dedicated to exploring disparate impact on minorities in credit markets and in the mortgage market in particular. A common challenge in this line of research is

[31] For this context, we resort to the comparison of the β coefficients in equations (1) and (2), rather than the partial effects. This is because the partial effects are a function of both the sensitivity of the outcome to the regressor (the coefficients) and the subsample average outcomes (delinquency rates). See equation (3).

distinguishing between the effects of disparate impact and discrimination, because most researchers pursuing this question do not have access to the full set of variables to predict loan pricing and performance (see Ross and Yinger (2002) for a full analysis of challenges in identifying racial discrimination in the mortgage market).

As an example, the landmark Boston Fed Study (Munnell, Tootell, Browne, and McEneaney (1996)) found that race strongly predicted loan approval among applicants even after controlling for a long list of personal characteristics and individual risk factors, though their estimated race effects were smaller than those found in earlier studies employing a smaller set of control variables. Yet their study did not include other important covariates--such as credit score--which strongly predict loan performance, and did not have information on *ex post* loan performance. Thus, the study was unable to conclude whether the disparate loan approval rates across race resulted from legitimate economic considerations or from discrimination. Our findings complement this line of prior research by including additional covariates and by relating loan performance to race/ethnicity.

In the full sample, the ranking of delinquency rates by race/ethnicity is as follows: white (24.7%), Asian (27.1%), black (37.4%), and Hispanic (40.2%). Controlling for observable characteristics, the black-white (2.8 to 5.2 percentage points) and Hispanic-white (5.9 to 8.3 percentage points) differences are statistically significant at the 1% level in all four subsamples, while the Asian-White differences (-1.1 to 1.1 percentage points) are not significant even at the 10% level. Notably, the difference in the delinquency rates between white and black/Hispanic borrowers is more than 50% higher in the Broker subsamples than in the Bank subsamples.

We must also control for loan pricing in order to attribute these delinquency differences to race/ethnicity. If certain racial/ethnic groups pay higher interest rates conditional on other characteristics, then the heavier payment burden could cause higher delinquency. Such a concern is warranted by prior research on consumer financing. Charles, Hurst, and Stephens (2008) show that blacks pay significantly higher rates when financing a new car, in large part because blacks are more likely to use more expensive financing companies. Similarly, Ravina (2008) finds that black borrowers in an online lending market pay rates that are over 100 basis points higher than comparably risky white borrowers. Much of the difference can be attributed to favorable interest rates obtained in same-race lender-borrower pairings and the underrepresentation of black lenders. In the context of mortgage lending, price differences could occur by pricing a given product differently for borrowers from different demographic groups, but more likely occurs through steering uninformed borrowers into more costly products, such as subprime loans, when more attractive products are available; or through aggressive negotiation strategies used by brokers to enhance their fees and commissions (known in the industry as yield spread premiums).

We next examine the determinants of interest rates to assess the importance of the pricing effect. While we focus on the race/ethnicity variables, we also include all other variables that appear in the delinquency analysis (reported in Tables 2 and 3). Our sample includes both fixed- and adjustable-rate loans, and we have information on the initial and current (updated to 2008) rates. To ensure the rates are comparable across observations, we analyze the following two dependent variables on select samples: the current interest rate on the full sample, and the initial interest rate for loans originated in 2004 and 2005 that have not incurred a rate change up to 2008. The second sample is meant to approximate a sample of fixed-rate loans. We conduct the analysis separately for different loan types, and report the results in Table 9.

[Insert Table 9 here.]

We find no evidence that black or Hispanic borrowers pay higher initial or current interest rates on bank-originated loans, conditional on observable individual risk factors. However, among broker-originated loans, black borrowers appear to pay higher rates, on the order of 10-16 basis points, while there is no clear evidence that Hispanic borrowers are subject to higher loan pricing. While the coefficients on *Black* are significant and positive in the Broker subsamples, the magnitudes are much lower than those documented in other credit markets (e.g., Charles, Hurst, and Stephens (2008) and Ravina (2009)). The estimated gender effect is insignificant throughout, both in terms of loan pricing and loan performance. Our results are closer to findings in Courchane (2007) and Haughwout, Mayer, and Tracy (2009) that there is no significant adverse pricing by race, ethnicity, or gender in the pricing of mortgage credit after controlling for other observable differences.

Our data suggest that loan pricing is an unlikely explanation for the higher delinquency rates observed among black and Hispanic borrowers. Black borrowers exhibit higher delinquency rates relative to white borrowers, even for bank-originated loans for which we find no evidence of unfavorable pricing. The average (median) unpaid balance on loans among black borrowers is $185,000 ($150,000). Thus, the estimated black-white difference in interest rate among broker-originated loans--10-16 basis points--amounts to an additional monthly payment of $15-$25 (or $13-$20) using the mean (or median) balance. It is unlikely that such a difference could be pivotal in loan delinquency. Moreover, Hispanic borrowers exhibit the highest delinquency rates in our sample among all demographic groups, although there is no evidence that they face unfavorable interest rates in comparison to other groups.

Previous work sheds light on the unobserved risk factors that are correlated with race/ethnicity variables. First, blacks and Hispanics have lower savings rates on average than whites of similar age, education and income (Blau and Graham (1990), Charles, Hurst, and Roussanov (2007)). As a result, they accumulate less wealth (often difficult to measure), making them more vulnerable to adverse economic shocks. Second, minorities are less likely to have family or relatives who can help when they have trouble meeting their mortgage payments (Yinger, 1995). Third, Guiso, Sapienza, and Zingales (2009) offer an interesting explanation for the highest delinquency rates observed among Hispanic borrowers. Based on survey data, the authors find that Hispanics are much less likely (between 18 and 27 percentage points) than blacks or whites to feel morally or socially obligated to continue paying their mortgages when the equity value is significantly below zero.

Historically, policymakers and researchers concerned with mortgage lending discrimination have focused on two key issues: unequal access to credit (i.e., disparities in loan approvals and denials) and pricing disparities. While we do not examine differences in mortgage approvals by race, our analysis suggests that the housing boom fueled a rapid expansion of credit among Hispanic and black borrowers.[32] Moreover, the share of first-time borrowers among black and Hispanic households grew from 10% in early 2004 to 25% in late 2006. In addition, we find little evidence of pricing discrimination as a cause for loan delinquency. Taken together, the findings suggest that market dynamics and credit expansionary practices during the sample period may have alleviated some of the inequalities in credit access and pricing. Yet the *ex post* loan performance data suggests that such credit expansion was achieved largely through lowered lending standards, particularly among brokers originating low-documentation loans. The persistence of Hispanic and black race effects in the delinquency models raises further questions, including whether such borrowers were well-informed about the mortgage process and possessed the requisite experience and knowledge to continue making their mortgage payments in full and on time.

VI. Conclusion

This paper uses a unique, proprietary data set from a major national mortgage bank to examine how mortgage loan performance relates to loan origination channel, documentation level, and borrower demographics. Our research aims to identify and quantify the micro-level fundamental causes of the mortgage crisis, and highlights two agency problems. The first agency problem arises between the bank and its mortgage brokers, who originate observably lower quality loans. We find that brokered loans are more than 50% more likely to be delinquent than bank-originated loans, and that approximately three-quarters of this difference can be attributed to lower borrower/loan quality based on observable risk factors. The second agency problem arises between lenders and borrowers, and results in borrower information falsification among low-documentation loans, especially when issued through a broker. We

[32] Recall from Table 1 that Hispanic borrowers experienced the fastest growth in newly originated loans during our sample period, followed by black borrowers.

find poor model predictive power and strong evidence of information falsification among low-documentation loans.

Our analysis raises the question of why this major mortgage bank—as well as other market players—allowed such deterioration in borrower and loan quality to persist before tightening its lending standards. A plausible explanation is that the expansion of the secondary mortgage market and the ease of loan securitization weakened the bank's incentive to screen borrowers by allowing the bank to offload risk. We refer the readers to Keys, Mukherjee, Seru and Vig (2008) for an analysis on the relation between loan performance and the *ex ante* probability of loan securitization, and to Jiang, Nelson, and Vytlacil (2009) for a contrast between the *ex ante* and *ex post* relation of the two.

Appendix:

1. Proof of equation (11):

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

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

$$f^D(v) = \frac{vf(v)}{E(Y)};\ f^{ND}(v) = \frac{(1-v)f(v)}{1-E(Y)}, \tag{14}$$

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

Equation (14) implies:

$$\Pr(P \geq \bar{p} \mid Y=1) = \int_{\bar{p}}^{1} f^D(v)dv = \int_{\bar{p}}^{1} \frac{vf(v)}{E(Y)}dv = \frac{[1-F(\bar{p})]}{E(Y)}E(P \mid P \geq \bar{p}). \tag{15}$$

Similarly,

$$\Pr(P < \bar{p} \mid Y=0) = \int_{0}^{\bar{p}} f^{ND}(v)dv = \int_{0}^{\bar{p}} \frac{(1-v)f(v)}{1-E(Y)}dv = \frac{F(\bar{p})}{1-E(Y)}E(P \mid P < \bar{p}). \tag{16}$$

We obtain (11) by substituting (15) and (16) into (10).

2. Proof that equation (11) satisfies the likelihood ratio property:

Using equation (14) and the fact *Var(Y) = E(Y)[1-E(Y)]*, we have:

$$f^D(v) - f^{ND}(v) = f(v)Var(Y)[v - E(y)]. \tag{17}$$

Thus,

$$\begin{aligned} f^D(v) - f^{ND}(v) &> 0 \text{ if } v > E(y), \\ &= 0 \text{ if } v = E(y), \\ &< 0 \text{ if } v < E(y). \end{aligned} \tag{18}$$

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

References

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

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

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

Blau, Francine and John Graham, 1990, Black-White Differences in Wealth and Asset Composition, *Quarterly Journal of Economics*, 105, 321-339.

Blinder, Alan, 1973. Wage Discrimination: Reduced Form and Structural Variables, *Journal of Human Resources,* 8, 436-455.

Bond, Philip, David Musto, and Bilge Yilmaz, 2008, Predatory Mortgage Lending, *Journal of Financial Economics,* forthcoming.

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

Charles, Kerwin Kofi, Erik Hurst, and Melvin Stephens Jr. 2008, Rates for Vehicle Loans: Race and Loan Source, *American Economic Review*, 98:2, 315-320.

Charles, Kerwin Kofi, Erik Hurst, and Nikolai Roussanov, 2007, Conspicuous Consumption and Race, NBER Working Paper, No 13392.

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

Courchane, Marsha. 2007. The Pricing of Home Mortgage Loans to Minority Borrowers: How Much of the APR Differential Can We Explain? *Journal of Real Estate Research* 29(4): 399-440.

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

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

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

Guiso, Luigi, Paola Sapienza, and Luigi Zingales, 2009, Moral and Social Constraints to Strategic Default on Mortgage, working paper, University of Chicago.

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

Inderst, Roman, 2006, "Irresponsible Lending" with a Better Informed Lender, *Economic Journal*, forthcoming.

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

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

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

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

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

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

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

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

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

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

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

Ravina, Enrichetta, 2009, Love and Loans: The Effect of Beauty and Personal Characteristics in Credit Markets, Working Paper, Columbia University.

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

Ross Stephen and John Yinger, 2002, *The Color of Credit: Mortgage Discrimination, Research Methodology, and Fair-Lending Enforcement*, the MIT Press: Boston, U.S.

Yinger, John, 1995, *Closed Doors, Opportunities Lost: The Continuing Costs of Housing Discrimination*, The Russell Sage Foundation.

Figure 1. Geographic Distribution of Properties in the Sample

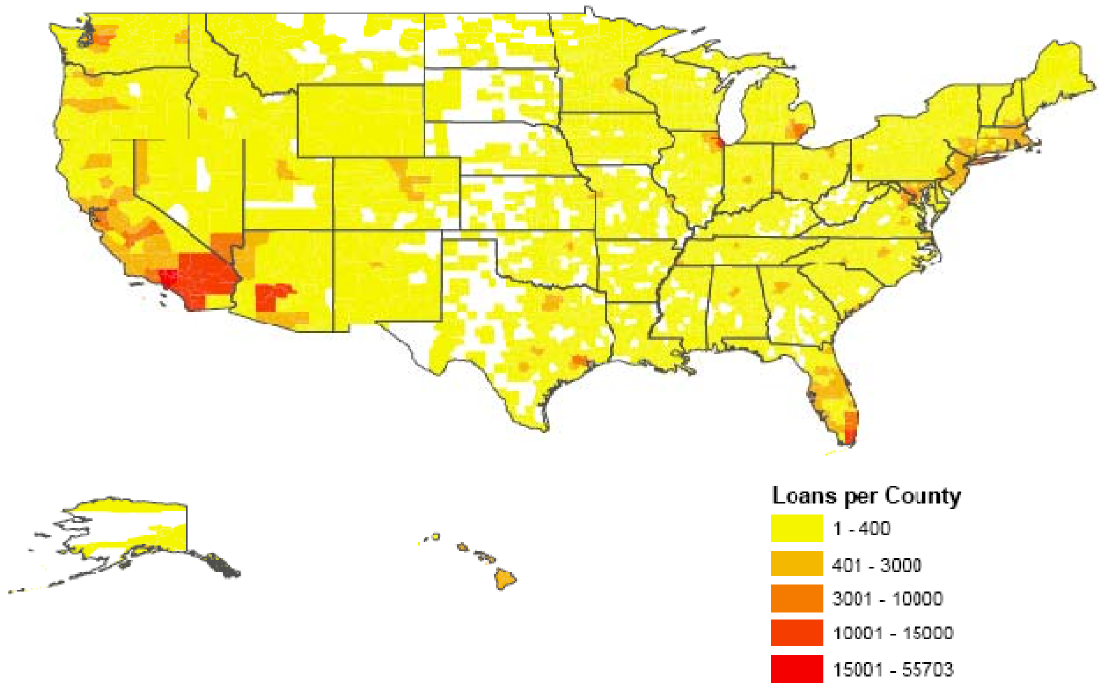


Figure 2. Number of Loans and Composition by Semi-Year: 2004-2008

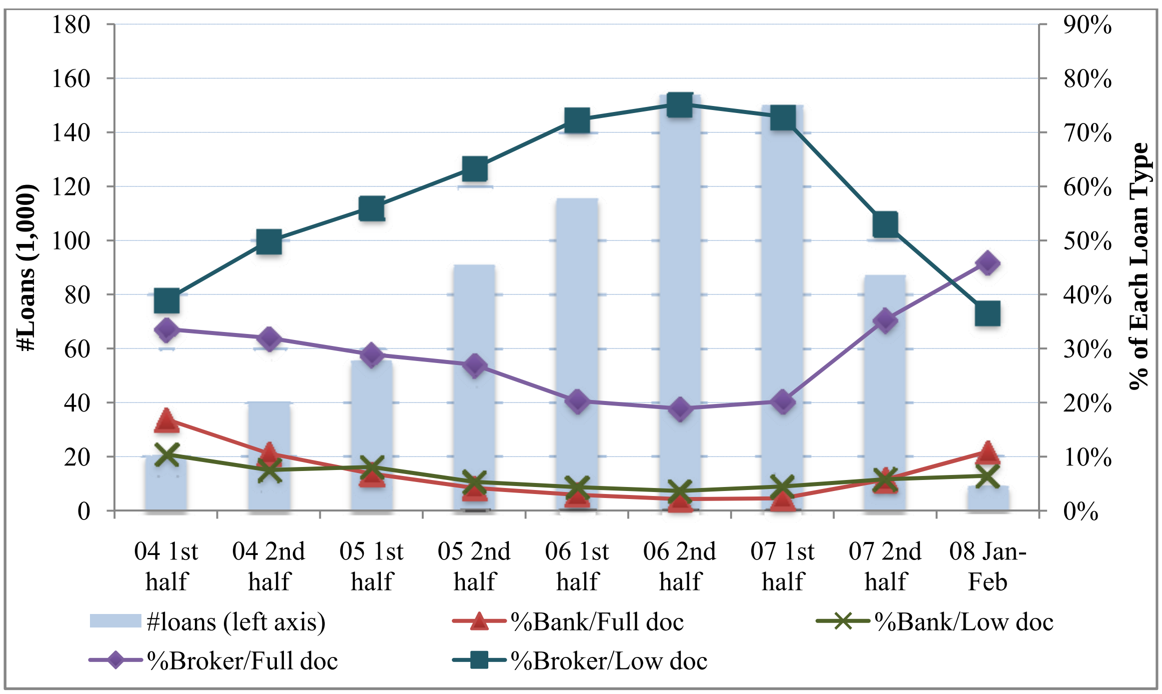


Figure 3. Delinquency Rates since Loan Origination by Semi-Year: Updated to January 2009

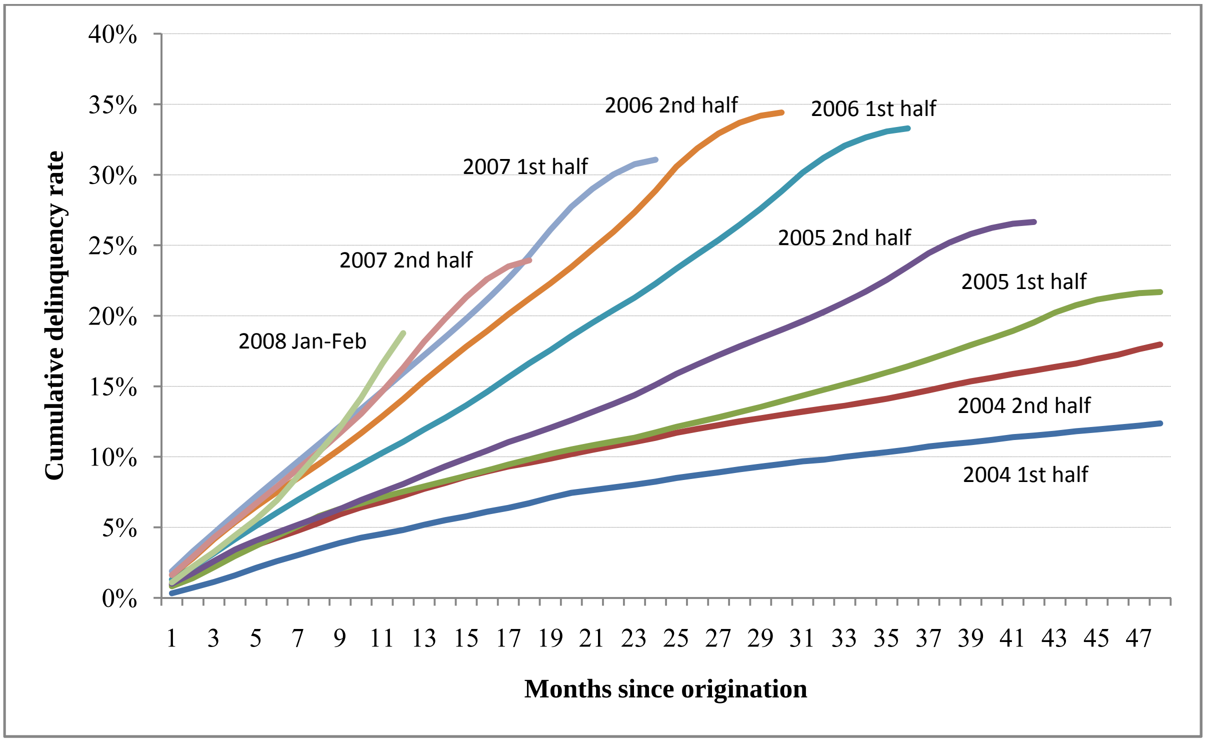


Figure 4. Time Series of Out-of-Sample Model Predictive Power by Loan Type

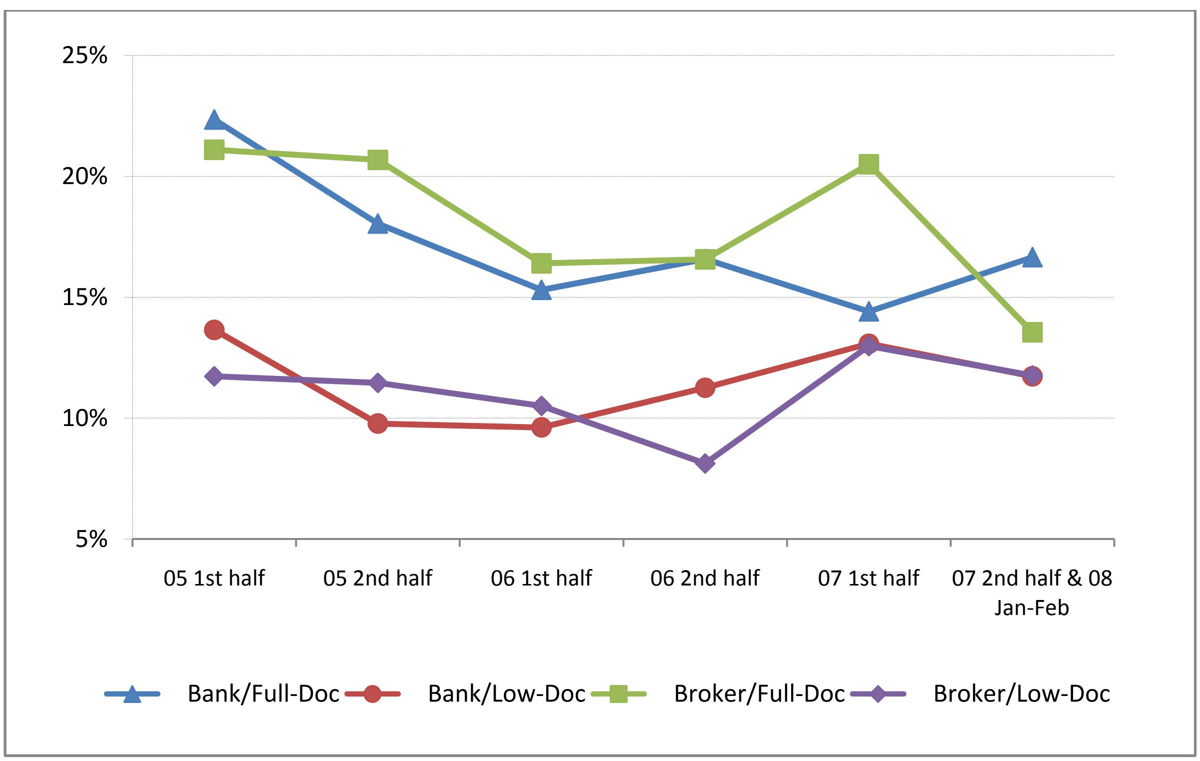


Table 1. Variable Definitions and Summary Statistics

Panel A: Definitions of main variables

	Definition
AddLTV	The ratio of additional loans (including from other banks) secured to the property to the property value
Age	Age of the borrower
Asian	Dummy variable = 1 if the borrower is Asian
Avgincome	Average income per capita of the census tract where the property is located
Black	Dummy variable = 1 if the borrower is black
Cashresv	Cash reserves, in multiples of monthly mortgage payments
Delinquency	Dummy variable for delinquency, defined as being at least 60 days behind in payment
Female	Dummy variable = 1 if the borrower is female
CreditScore	Borrower's credit score
CurrRate	The current interest rate (updated in February 2008) on the loan
FirstTimeOwner	Dummy variable = 1 if the borrower is a first-time mortgage borrower
Hispanic	Dummy variable = 1 if the borrower is Hispanic
Income	Monthly income of the borrower in $1,000
IncomeMiss	Dummy variable = 1 if the income information is missing
InitialRate	Initial interest rate on the mortgage
Loan	Total loan amount
LTI	Loan-to-income ratio, the percentage of monthly gross income that is used to pay for the mortgage
LTV	Loan-to-value ratio
Medage	Median age of residents in the census tract where the property is located
OneBorrower	Dummy variable = 1 if there is only one borrower on the mortgage
OwnerOccupied	Dummy variable = 1 if the property is the owner's primary residence
Pctblack/Pcthisp	Proportion of black/Hispanic households in the census tract where the property is located
Population	Population size of the census tract where the property is located
PrepayPenalty	Dummy variable = 1 if there is hard prepayment penalty in the loan contract
Refinance	Dummy variable = 1 if the mortgage is for refinancing
Secondlien	Dummy variable = 1 if the mortgage is a second-lien
SelfEmploy	Dummy variable = 1 if the borrower is self-employed
Subsample1	Bank/Full-Doc subsample
Subsample2	Bank/Low-Doc subsample
Subsample3	Broker/Full-Doc subsample
Subsample4	Broker/Low-Doc subsample
Tenure	Number of months that the borrower has been employed in the current job
TenureMiss	Dummy variable = 1 if the tenure information is missing
Unemprate	Unemployment rate in the census tract where the property is located

Panel B: Summary statistics

This table reports the mean, median and standard deviation (the first, second, and third line of each variable) values of the major variables by semi-year from 2004 to early 2008. Their definitions are in Panel A. "1st" and "2nd" indicate the first and second half of each year.

	04 1st	04 2nd	05 1st	05 2nd	06 1st	06 2nd	07 1st	07 2nd	08 Jan-Feb
Age (years) (average)	44.9	44.2	44	43.3	42.9	42.9	43.8	45.6	45.42
(median)	44	43	43	42	42	42	43	45	45
(std. dev.)	11.6	12.5	12.4	12.3	12.6	12.7	12.7	12.6	12.6
Credit score	703	697.5	698.8	696.9	692.2	695.1	695.5	697.5	699.5
	707	700	699	695	689	692	694	698	701
	60	61.3	58.3	56.2	53	53.6	55.9	59.3	62.2
Income ($1,000, monthly)	7.1	6.9	6.9	7.3	7.7	8	8.7	8	7.3
	5.5	5.5	5.5	5.8	6.3	6.5	6.3	5.7	5.6
	8.6	11.7	8.6	22.1	9.4	10.5	187.1	21.6	10.1
Initial rate	5.3%	5.5%	4.9%	5.4%	6.1%	6.7%	7.2%	7.2%	6.8%
	5.6%	6.0%	5.8%	6.0%	6.6%	6.9%	6.9%	7.0%	6.8%
	1.7%	2.3%	2.5%	2.7%	3.0%	2.8%	2.2%	1.2%	0.8%
Loan size in $1,000	235.7	230.9	253.7	263.9	260.2	271.3	275.5	296	282.1
	192	192	217	225	223.2	231.1	232	251.8	256.5
	165.1	162.9	166.1	187.1	189.5	202.8	214	223.1	177.2
Loan-to-income	25.5%	24.4%	25.3%	25.0%	25.3%	26.8%	31.2%	32.4%	34.7%
	23.1%	22.2%	22.9%	23.0%	23.0%	24.9%	30.7%	32.0%	33.7%
	13.8%	13.3%	13.3%	13.4%	14.0%	14.6%	14.5%	15.3%	15.0%
Loan-to-value	69.3%	70.6%	72.6%	69.5%	67.2%	66.9%	65.3%	74.6%	77.1%
	75.0%	78.1%	79.4%	79.4%	79.2%	79.8%	78.4%	80.0%	80.0%
	16.9%	18.3%	14.7%	19.3%	21.8%	22.6%	24.1%	19.0%	17.3%
Tenure (months)	100.2	94	90.7	85.2	81.6	80.8	83.2	91	93.8
	60	60	60	57	50	50	51	60	60
	98.3	95.1	92.4	90	86.7	86.3	89.4	95.2	94.6

	04 1st	04 2nd	05 1st	05 2nd	06 1st	06 2nd	07 1st	07 2nd	08 Jan-Feb
%Asian	5.1%	5.7%	5.6%	5.8%	5.5%	4.8%	5.0%	5.2%	4.2%
%Black	4.5%	6.3%	6.7%	7.2%	8.1%	8.4%	8.9%	9.9%	10.3%
%Black & Hispanic that are first-time owners	10.3%	13.5%	14.1%	19.7%	23.9%	25.2%	24.5%	17.6%	20.9%
%Female	28.8%	32.6%	30.8%	32.5%	32.8%	34.2%	34.7%	35.4%	36.0%
%First-time owner	7.6%	10.8%	11.3%	14.7%	16.8%	18.1%	17.5%	12.7%	15.5%
%Hispanic	7.5%	10.6%	13.1%	15.6%	20.0%	19.2%	23.3%	21.8%	23.5%
%Owner occupied	86.1%	84.2%	84.8%	84.4%	84.6%	86.7%	85.4%	81.8%	88.3%
%Refinance	71.2%	56.0%	59.3%	54.7%	55.4%	54.7%	58.7%	66.9%	65.3%
%Self-employed	18.2%	18.5%	18.2%	18.0%	20.2%	19.7%	21.4%	22.6%	20.5%

Table 2. Delinquency Prediction: Probit Analysis

The dependent variable is loan delinquency, and the estimation method is probit. The definitions of all variables are given in Table 1 Panel A. Reported are the coefficients (Coef), t-statistics (t-stat) that adjust for clustering at the MSA level, and the average partial effects (APE). At the bottom of the table, we report the sample frequency of delinquency, the pseudo R-squared, the number of observations and the number of clusters (at the MSA level).

	1			2			3			4		
	Bank/Full-Doc			Bank/Low-Doc			Broker/Full-Doc			Broker/Low-Doc		
	Coef	t-stat	APE	Coef	t-stat	APE	Ceof	t-stat	APE	Coef	t-stat	APE
LTV	1.693	14.61	36.15%	2.480	19.41	56.35%	2.028	17.81	50.99%	3.021	19.15	91.45%
AddLTV	1.467	7.24	31.32%	1.566	7.44	35.57%	1.665	15.98	41.84%	2.975	24.28	90.05%
Loan (log)	0.113	4.08	2.42%	0.178	7.23	4.04%	0.214	8.83	5.38%	0.252	8.78	7.64%
SecondLien	0.245	1.78	5.22%	0.729	6.23	16.56%	0.498	8.07	12.52%	0.297	3.79	9.00%
Refinance	-0.046	-1.08	-0.97%	-0.038	-1.32	-0.86%	-0.050	-2.15	-1.25%	0.097	5.49	2.94%
PrepayPenalty	0.111	2.1	2.37%	0.028	0.7	0.63%	0.005	0.26	0.12%	0.082	6.38	2.49%
FirstTimeOwner	-0.186	-4.2	-3.97%	-0.072	-1.17	-1.63%	-0.010	-0.61	-0.24%	-0.054	-3.81	-1.62%
OwnerOccupied	-0.259	-5.31	-5.53%	-0.275	-8.18	-6.24%	-0.350	-13.75	-8.79%	-0.281	-10.31	-8.51%
OneBorrower	0.267	12.81	5.70%	0.346	15.34	7.87%	0.292	19.32	7.34%	0.298	17.07	9.03%
Income (log)	-0.108	-6.91	-2.30%	0.023	1.32	0.53%	-0.064	-4.33	-1.61%	0.041	4.75	1.26%
IncomeMiss	-0.033	-0.28	-0.71%	-0.006	-0.13	-0.14%	-0.160	-2.97	-4.02%	0.155	6.98	4.71%
CashResv	-0.047	-5.61	-1.01%	-0.027	-3.61	-0.60%	-0.090	-17.94	-2.27%	-0.069	-16.12	-2.10%
CreditScore	-0.009	-53.89	-0.18%	-0.008	-31.84	-0.17%	-0.008	-49.91	-0.21%	-0.007	-71.41	-0.21%
Female	-0.043	-1.71	-0.93%	-0.014	-0.75	-0.32%	-0.003	-0.2	-0.07%	0.003	0.34	0.08%
Hispanic	0.276	5.5	5.89%	0.219	3.78	4.98%	0.391	7.75	9.83%	0.275	10.55	8.33%
Black	0.129	2.74	2.76%	0.156	2.75	3.55%	0.167	5.16	4.21%	0.120	4.53	3.64%
Asian	-0.053	-0.52	-1.13%	-0.052	-1.05	-1.18%	0.022	0.69	0.55%	0.037	1.25	1.12%
Age (log year)	-0.089	-3.65	-1.90%	0.020	1.04	0.45%	-0.020	-1.64	-0.50%	0.005	0.57	0.16%
Tenure(log month)	-0.018	-2.01	-0.38%	-0.045	-5.25	-1.02%	-0.012	-1.87	-0.30%	-0.035	-6.95	-1.06%
TenureMiss	-0.072	-1.16	-1.54%	-0.174	-4.01	-3.95%	-0.251	-7.56	-6.32%	-0.266	-11.52	-8.07%

	1			2			3			4		
	Bank/Full-Doc			Bank/Low-Doc			Broker/Full-Doc			Broker/Low-Doc		
	Coef	t-stat	APE	Coef	t-stat	APE	Ceof	t-stat	APE	Coef	t-stat	APE
SelfEmploy	-0.001	-0.03	-0.03%	0.053	2.82	1.20%	0.051	2.44	1.29%	0.000	-0.01	0.00%
2005	0.007	0.2	0.15%	0.155	3.99	3.52%	0.026	0.9	0.65%	0.138	5.28	4.18%
2006	0.018	0.49	0.39%	0.170	4.21	3.86%	0.053	1.13	1.34%	0.265	6.3	8.03%
2007	-0.188	-3.88	-4.02%	0.108	2.14	2.46%	-0.078	-1.48	-1.95%	0.167	3.85	5.05%
2008	-0.263	-4.07	-5.61%	-0.039	-0.49	-0.90%	-0.167	-2.8	-4.20%	0.048	0.88	1.46%
Constant	2.7901	7.82		0.231	0.75		1.00	2.53		-1.51	-4.17	
%Delinquency and (Pseudo) R-suqared		0.132	0.221		0.180	0.136		0.236	0.182		0.316	0.146
# obs and # clusters		31,408	807		35,553	778		166,402	963		425,181	949

Table 3. Delinquency Prediction: Duration Analysis

The dependent variable is duration between loan origination and delinquency (or censored at the end of the sample) in months, and the estimation method is a duration model with a log-logistic distribution for the accelerated time. The definitions of all variables are given in Table 1 Panel A. Reported are the coefficients (Coef), t-statistics (t-stat) that adjust for clustering at the MSA level, and changes in the median survival time for a one unit change in a covariate while holding all other covariates at their sample mean levels (∂t/∂x). At the bottom of the table, we report the *gamma* coefficient and its standard error, the sample mean of median survival time and its standard deviation, the number of observations, and the pseudo R-squared. The number of clusters (at the MSA level) is the same as in Panel A.

	1			2			3			4		
	Bank/Full-Doc			Bank/Low-Doc			Broker/Full-Doc			Broker/Low-Doc		
	Coef	t-stat	∂t/∂x	Coef	t-stat	∂t/∂x	Ceof	t-stat	∂t/∂x	Coef	t-stat	∂t/∂x
LTV	-2.730	-15.02	-804.84	-2.838	-23.39	-284.86	-2.641	-19.51	-249.19	-3.267	-34.08	-178.40
AddLTV	-1.773	-5.49	-522.68	-1.516	-9.27	-152.13	-1.952	-15.67	-184.20	-2.774	-35.27	-151.45
Loan (log)	-0.135	-3.31	-39.84	-0.193	-8.00	-19.33	-0.263	-9.4	-24.77	-0.237	-9.04	-12.95
SecondLien	-0.683	-3.07	-149.30	-0.953	-8.02	-63.67	-0.726	-8.94	-51.86	-0.553	-8.52	-24.73
Refinance	0.092	1.41	26.53	0.032	1.04	3.16	0.098	3.49	9.23	-0.087	-5.99	-4.79
PrepayPenalty	-0.149	-1.79	-41.13	0.014	0.32	1.45	0.064	2.77	6.16	-0.022	-1.98	-1.19
FirstTimeOwner	0.299	4.05	101.03	0.035	0.55	3.58	0.001	0.03	0.06	0.052	4.88	2.91
OwnerOccupied	0.338	4.83	89.73	0.262	8.32	24.87	0.430	10.51	34.62	0.302	9.58	14.87
OneBorrower	-0.417	-12.2	-121.26	-0.378	-14.23	-40.34	-0.370	-24.23	-35.91	-0.299	-13.89	-17.67
Income (log)	0.165	7.21	48.78	-0.009	-0.46	-0.93	0.079	3.86	7.43	-0.040	-4.89	-2.21
IncomeMiss	0.070	0.37	21.37	0.059	1.08	6.01	0.241	3.16	25.57	-0.156	-7.37	-8.26
CashResv	0.076	5.67	22.42	0.026	3.09	2.58	0.129	16.18	12.21	0.075	13.07	4.12
CreditScore	0.015	33.81	4.35	0.009	17.82	0.89	0.012	25.22	1.14	0.007	23.77	0.39
Female	0.072	1.77	21.46	0.012	0.56	1.18	0.005	0.32	0.50	-0.001	-0.15	-0.06
Hispanic	-0.393	-4.91	-98.47	-0.203	-3.65	-18.85	-0.385	-8.47	-31.83	-0.181	-10.12	-9.45
Black	-0.157	-2.08	-43.22	-0.166	-3.19	-15.44	-0.229	-5.38	-19.86	-0.149	-5.42	-7.65
Asian	0.063	0.39	19.15	0.077	1.48	8.06	0.010	0.26	0.98	0.000	0.00	-0.01
Age (log year)	0.187	4.71	55.06	-0.014	-0.71	-1.43	0.039	2.53	3.66	-0.014	-1.79	-0.77

	1			2			3			4		
	Bank/Full-Doc			Bank/Low-Doc			Broker/Full-Doc			Broker/Low-Doc		
	Coef	t-stat	∂t/∂x	Coef	t-stat	∂t/∂x	Ceof	t-stat	∂t/∂x	Coef	t-stat	∂t/∂x
Tenure(log month)	0.025	1.61	7.22	0.045	5.00	4.56	0.009	1.26	0.86	0.028	6.36	1.55
TenureMiss	0.116	1.17	35.85	0.163	3.41	17.38	0.324	9.29	33.03	0.232	9.83	13.42
SelfEmploy	0.038	0.45	11.42	-0.047	-2.24	-4.70	-0.066	-2.27	-6.05	-0.008	-0.86	-0.45
2005	-0.208	-3.45	-58.24	-0.383	-9.17	-34.90	-0.190	-5.07	-17.06	-0.335	-15.5	-16.59
2006	-0.495	-8.25	-127.96	-0.694	-17.66	-61.72	-0.486	-8.4	-41.77	-0.750	-22.47	-38.71
2007	-0.660	-8.28	-170.11	-1.058	-20.27	-91.32	-0.743	-9.98	-64.47	-1.031	-27.05	-50.05
2008	-0.850	-8.76	-173.24	-1.350	-14.5	-75.94	-0.943	-11.44	-58.94	-1.400	-25.28	-41.57
Constant	-1.905	-3.02		3.284	8.35		1.430	2.51		5.192	15.61	
Gamma and std. err		0.866	0.021		0.614	0.021		0.747	0.018		0.577	0.021
Median time (months)		300.1			100.4			94.4			54.6	
#obs and (Pseudo) R-squared		31,400	0.174		35,550	0.141		166,399	0.162		425,180	0.145

Table 4. Choice of Loan Origination Channel and Documentation Level

Panel A: Without neighborhood information

The dependent variable is the choice of broker channel, that of low documentation, and that of the combination of two. The estimation method is probit. The definitions of all variables are given in Table 1 Panel A. Reported are the coefficients (Coef), t-statistics (t-stat) that adjust for clustering at the MSA level, and the average partial effects. At the bottom of the table, we report the average of the dependent variables (the sample frequency of the choices), and the pseudo R-squared.

	1			2			3		
Dep. Var.	Broker			Low-Doc			Broker Issue/Low Doc		
	Coef	t-stat	APE	Coef	t-stat	APE	Coef	t-stat	APE
LTV	0.372	5.14	5.67%	-0.773	-6.82	-19.68%	-0.506	-5.18	-14.91%
AddLTV	3.730	14.92	56.90%	0.411	3.31	10.48%	1.088	8.17	32.09%
Loan (log)	0.088	3.02	1.34%	0.220	11.46	5.61%	0.171	7.69	5.04%
SecondLien	-1.896	-10.84	-28.92%	-0.160	-2.17	-4.06%	-0.497	-5.92	-14.67%
Refinance	-0.146	-5.26	-2.23%	-0.053	-2.2	-1.35%	-0.092	-5.26	-2.72%
FirstTimeOwner	0.331	16.21	5.05%	-0.047	-2.71	-1.19%	-0.004	-0.26	-0.12%
OwnerOccupied	0.126	3.25	1.92%	-0.047	-3.07	-1.20%	0.082	3.1	2.42%
OneBorrower	0.217	18.38	3.31%	0.507	37.65	12.92%	0.449	39.55	13.24%
Income (log)	-0.038	-3.49	-0.57%	0.241	15.01	6.14%	0.218	12.55	6.43%
IncomeMiss	0.128	3.28	1.95%	2.270	56.13	57.80%	1.606	34.22	47.36%
CashResv	-0.015	-1.84	-0.23%	0.003	0.88	0.07%	-0.003	-0.85	-0.09%
CreditScore	-0.001	-14.22	-0.02%	0.002	13.93	0.05%	0.001	9.10	0.03%
Female	0.027	3.67	0.41%	0.150	11.1	3.82%	0.125	10.61	3.70%
Hispanic	0.448	13.53	6.84%	0.432	6.67	11.00%	0.475	8.66	14.02%
Black	0.439	15.57	6.70%	-0.030	-1.15	-0.77%	0.059	2.14	1.75%
Asian	0.485	18.38	7.41%	0.368	18.51	9.37%	0.443	25.73	13.06%
Age (log year)	-0.039	-3.72	-0.59%	0.001	0.06	0.01%	-0.013	-1.87	-0.39%
Tenure(log month)	-0.017	-4.55	-0.26%	-0.055	-9.6	-1.40%	-0.055	-9.53	-1.62%
TenureMiss	0.540	13.61	8.24%	-0.350	-9.66	-8.92%	-0.176	-4.8	-5.18%
SelfEmploy	0.208	8.8	3.18%	1.036	48.57	26.38%	0.775	27.6	22.85%
2005	0.339	12.98	5.18%	0.261	14.42	6.65%	0.306	16.99	9.01%
2006	0.443	12.73	6.77%	0.542	30.15	13.81%	0.544	26.01	16.05%
2007	0.419	17.44	6.40%	0.263	16	6.70%	0.318	15.55	9.38%
2008	0.196	5.3	2.99%	-0.329	-12.57	-8.38%	-0.234	-7.63	-6.89%
Constant	0.188	0.57		-4.213	-18.4		-3.577	-15.06	
E(Dependent Variable)		0.898			0.700			0.646	
#obs and (Pseudo) R-squared		658,544	0.149		658,544	0.265		658,544	0.201

Panel B: With neighborhood information

This table is identical to that in Panel A, with the addition of neighborhood covariates at the census tract or zip-code level. The definitions of all variables are given in Table 1 Panel A. For the economy of space and to avoid repetition, only results regarding the neighborhood variables are reported.

	1			2			3		
Dep. Var.	Broker			Low-Doc			Broker Issue/Low Doc		
	Coef	t-stat	APE	Coef	t-stat	APE	Coef	t-stat	APE
Population(log)	-0.006	-0.98	-0.10%	0.005	0.91	0.12%	-0.001	-0.14	-0.02%
Pctblack	-0.079	-4.69	-1.23%	0.055	2.97	1.39%	0.027	1.38	0.80%
Pcthisp	-0.055	-1.65	-0.86%	0.175	8.88	4.46%	0.143	6.41	4.22%
Medage	-0.002	-3.18	-0.03%	-0.001	-2.67	-0.03%	-0.002	-3.19	-0.04%
Avgincome	0.000	-0.51	0.00%	0.000	0.46	0.00%	0.000	0.66	0.00%
Unemprate	-0.001	-0.29	-0.01%	-0.006	-2.19	-0.15%	-0.007	-2.44	-0.19%
Other controls included?	Y			Y			Y		
E(Dep Var)		89.9%			69.9%			64.6%	
#obs and (Pseudo) R-squared		491,816	0.116		491,816	0.268		491,816	0.200

Table 5. Non-Linear Blinder-Oaxaca Decomposition of Differences in Delinquency Rates

This table reports the non-linear Blinder-Oaxaca (1973) decomposition to the probit model. The total difference in delinquency rates between two subsamples is decomposed into an "endowment effect" and a "coefficient effect" using equations (8) (using the high average outcome subsample as the base) and (9) (using the low average outcome subsample as the base).

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

	Bank			Broker		
	Difference	t-stat	Percentage	Difference	t-stat	Percentage
Low-Doc sample as benchmark						
Endowment Effect	-0.06%	-0.10	-1.20%	-0.89%	-1.62	-11.10%
Coefficient Effect	4.87%	9.13	101.20%	8.91%	12.84	111.10%
Full-Doc sample as benchmark						
Endowment Effect	-2.10%	-2.37	-43.71%	-2.84%	-4.15	-35.47%
Coefficient Effect	6.91%	8.12	143.71%	10.86%	12.94	135.47%
Total Difference	4.81%	5.37	100%	8.02%	8.05	100%

Panel B: Comparison of Bank and Broker subsamples

	Full-Doc			Low-Doc		
	Difference	t-stat	Percentage	Difference	t-stat	Percentage
Broker sample as benchmark						
Endowment Effect	7.84%	8.09	75.69%	10.40%	12.16	76.67%
Coefficient Effect	2.52%	9.46	24.31%	3.16%	8.76	23.33%
Bank sample as benchmark						
Endowment Effect	6.12%	10.28	59.06%	6.93%	9.88	51.10%
Coefficient Effect	4.24%	5.74	40.94%	6.63%	9.45	48.90%
Total Difference	10.35%	10.51	100%	13.56%	13.99	100%

Table 6. Projections of Credit Score on Other Borrower Characteristics

The dependent variable is credit score, and the estimation method is OLS. The definitions of all variables are given in Table 1 Panel A. We report the coefficients (coef) and t-statistics (t-stat) that adjust for clustering at the MSA level.

	1		2		3		4	
	Bank/Full-Doc		Bank/Low-Doc		Broker/Full-Doc		Broker/Low-Doc	
	coef	t-stat	coef	t-stat	coef	t-stat	coef	t-stat
Income (log)	14.91	14.72	3.28	4.43	15.98	17.03	5.86	11.82
IncomeMiss	27.58	6.19	11.25	4.05	32.62	14.45	21.97	17.72
CashResv	13.84	14.45	8.43	14.84	16.77	45.48	8.16	30.13
Female	-11.13	-9.23	-5.16	-8.15	-5.61	-10.37	-3.12	-15.37
Hispanic	0.84	0.36	-2.12	-2.46	-5.39	-4.24	-2.30	-3.41
Black	-23.30	-13.79	-14.31	-7.76	-27.72	-20.47	-18.10	-17.48
Asian	14.03	3.46	9.19	6.02	14.10	11.49	8.35	10.23
Age (log year)	-0.44	-0.29	5.19	7.14	-4.05	-5.33	2.42	5.38
Tenure(log month)	0.90	2.49	1.33	4.58	-1.05	-3.49	0.27	1.67
TenureMiss	11.27	4.9	12.93	8.16	46.20	30.33	19.31	27.43
SelfEmploy	-1.53	-0.76	-3.52	-4.16	-1.78	-1.69	-6.87	-15.09
2005	6.37	4.69	1.45	1.24	-4.71	-3.89	-2.92	-5.57
2006	11.79	4.93	1.96	1.37	-9.58	-9.7	-7.39	-10
2007	18.68	8.43	7.71	5.94	-9.28	-7.07	-4.45	-4.72
2008	18.43	8.38	34.05	13.06	-14.04	-7.68	14.70	16.21
Constant	635.69	151.87	659.44	136.96	647.13	190.28	668.15	391.9
Average Credit Score	683.86		702.10		686.55		699.75	
#obs and R-squared	31,464	0.194	35,685	0.087	168,046	0.202	429,481	0.073

Table 7. Delinquency Analysis: Correspondent and Non-Correspondent Brokers

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

	1			2			3			4		
	Correspondent/Full Doc			Correspondent/Low Doc			Non-Correspondent/Full Doc			Non-Correspondent/Low Doc		
	Coef	t-stat	APE	Coef	t-stat	APE	Ceof	t-stat	APE	Coef	t-stat	APE
LTV	2.035	14.57	45.51%	3.154	16.05	90.78%	2.000	17.35	51.12%	2.973	20.65	91.02%
AddLTV	1.781	8.18	39.84%	3.260	18.82	93.82%	1.707	16.98	43.62%	2.929	25.83	89.66%
Loan (log)	0.140	3.3	3.13%	0.264	6.45	7.61%	0.232	9.22	5.94%	0.255	9.15	7.81%
SecondLien	0.278	2.05	6.22%	0.220	2.31	6.33%	0.490	7.65	12.53%	0.301	3.74	9.23%
Refinance	0.012	0.31	0.26%	0.128	3.72	3.69%	-0.062	-2.55	-1.57%	0.087	5.41	2.66%
PrepayPenalty	0.059	1.37	1.31%	0.068	2.99	1.95%	-0.015	-0.76	-0.38%	0.076	5.52	2.34%
FirstTimeOwner	-0.120	-3.86	-2.69%	-0.098	-4.12	-2.83%	0.003	0.18	0.07%	-0.048	-3.68	-1.47%
OwnerOccupied	-0.366	-8.23	-8.20%	-0.226	-5.64	-6.51%	-0.348	-13.17	-8.90%	-0.288	-10.89	-8.82%
OneBorrower	0.211	8.38	4.73%	0.281	14.01	8.08%	0.299	20.29	7.65%	0.300	17.5	9.18%
Income (log)	-0.059	-2.21	-1.32%	0.020	1.02	0.57%	-0.065	-4.04	-1.67%	0.043	5.24	1.32%
IncomeMiss	-0.042	-0.22	-0.94%	0.053	1.14	1.52%	-0.166	-2.91	-4.24%	0.174	7.8	5.32%
CashResv	-0.095	-8.92	-2.13%	-0.096	-14.94	-2.75%	-0.087	-17.22	-2.23%	-0.063	-13.39	-1.92%
CreditScore	-0.008	-31.1	-0.19%	-0.007	-50.26	-0.20%	-0.008	-45.74	-0.21%	-0.007	-66.9	-0.21%
Female	0.015	0.56	0.33%	0.014	1.15	0.41%	-0.007	-0.47	-0.18%	-0.001	-0.07	-0.02%
Hispanic	0.323	10.01	7.24%	0.360	10.93	10.36%	0.379	7.6	9.69%	0.254	10.11	7.76%
Black	0.152	3.99	3.40%	0.103	3.63	2.98%	0.167	5.19	4.26%	0.127	4.79	3.88%
Asian	0.075	1.49	1.67%	0.143	4.39	4.11%	0.010	0.29	0.25%	0.012	0.42	0.38%
Age (log year)	-0.021	-0.7	-0.47%	0.028	2.23	0.82%	-0.019	-1.58	-0.48%	0.004	0.42	0.12%
Tenure(log month)	0.007	0.74	0.15%	-0.015	-2.03	-0.44%	-0.018	-2.3	-0.45%	-0.041	-8.77	-1.24%
TenureMiss	-0.009	-0.15	-0.19%	-0.129	-4.13	-3.72%	-0.307	-7.87	-7.84%	-0.302	-11.95	-9.24%
SelfEmploy	0.033	0.62	0.74%	0.022	1.06	0.63%	0.057	2.51	1.45%	-0.001	-0.09	-0.03%
2005	0.049	0.97	1.11%	0.154	4.53	4.43%	0.028	0.92	0.71%	0.137	5.31	4.20%
2006	0.031	0.57	0.68%	0.247	5.7	7.10%	0.067	1.34	1.72%	0.271	6.29	8.30%
2007	-0.071	-1.17	-1.58%	0.161	3.28	4.63%	-0.082	-1.54	-2.10%	0.166	3.86	5.07%

	1			2			3			4		
	Correspondent/Full Doc			Correspondent/Low Doc			Non-Correspondent/Full Doc			Non-Correspondent/Low Doc		
	Coef	t-stat	APE	Coef	t-stat	APE	Ceof	t-stat	APE	Coef	t-stat	APE
2008	-0.153	-0.97	-3.42%	-0.094	-0.63	-2.70%	-0.189	-3.15	-4.83%	0.037	0.69	1.13%
Constant	1.706	3.02		-1.888	-4.44		0.83	1.99		-1.484	-3.93	
%Delinquency and (Pseudo) R-squared		0.189	0.171		0.180	0.161		0.246	0.184		0.331	0.143
# obs and # clusters		25,666	657		88,778	672		140,736	955		336,403	936

Table 8. Out-of-Sample Model Predictive Power: Across Loan Types and Over Time

This Table reports the "excess percentage of correct predictions" as defined in equation (10). We reporte percentage of delinquent loans correctly predicted (S_1), percentage of non-delinquent loans correctly predicted (S_2), and the total percentage of correct predictions in excess of 50% (S). Each measure is reported for the full sample, and separately for the four sub-samples, and is reported by semi-annual intervals, together with the all-sample average.

	05 1st half	05 2nd half	06 1st half	06 2nd half	07 1st half	07 2nd half & 08 Jan-Feb	All-Time Average
% Delinquency correctly predicted (S_1)	54.2%	54.4%	52.2%	42.5%	52.9%	46.1%	50.4%
Bank/Full-Doc	57.6%	46.6%	43.3%	46.7%	40.9%	50.1%	47.5%
Bank/Low-Doc	51.9%	43.3%	42.5%	48.3%	51.4%	44.0%	46.9%
Broker/Full-Doc	61.6%	67.3%	58.9%	56.6%	66.9%	54.2%	60.9%
Broker/Low-Doc	50.7%	51.0%	51.4%	39.7%	50.4%	41.5%	47.4%
% Non-Delinquency correctly predicted (S_2)	76.3%	73.9%	71.9%	76.9%	75.7%	78.9%	75.6%
Bank/Full-Doc	87.1%	89.5%	87.3%	86.5%	88.0%	83.2%	86.9%
Bank/Low-Doc	75.4%	76.2%	76.8%	74.3%	74.8%	79.5%	76.2%
Broker/Full-Doc	80.6%	74.1%	73.9%	76.5%	74.1%	72.9%	75.4%
Broker/Low-Doc	72.7%	71.9%	69.7%	76.6%	75.6%	82.0%	74.7%
Total Excess % of correct prediction (S)	15.3%	14.2%	12.1%	9.7%	14.3%	12.5%	13.0%
Bank/Full-Doc	22.4%	18.0%	15.3%	16.6%	14.4%	16.7%	17.2%
Bank/Low-Doc	13.7%	9.8%	9.6%	11.3%	13.1%	11.7%	11.5%
Broker/Full-Doc	21.1%	20.7%	16.4%	16.6%	20.5%	13.6%	18.1%
Broker/Low-Doc	11.7%	11.5%	10.5%	8.1%	13.0%	11.8%	11.1%

Table 9. Race/Ethnicity and Interest Rates

This table examines the determinants of interest rates, with borrower race/ethnicity as the main variable. In columns 1-4, the dependent variable is the initial interest rate, and the sample includes loans initiated in 2004 and 2005 that have not incurred an interest rate change by 2008 (a proxy for fixed-rate loans). In columns 5-8, the dependent variable is the current interest rate, and we use the full loan sample. Interest rates are expressed in percentage points. We report the coefficients, t-statistics (in brackets) that adjust for clustering at the MSA level.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep. Variable	Initial Rate for Approximately Fixed-Rate Loans				Current Rate			
Subsample	Bank/ Full-Doc	Bank/ Low-Doc	Broker/Full-Doc	Broker/Low-Doc	Bank/Full-Doc	Bank/Low-Doc	Broker/Full-Doc	Broker/Low-Doc
Hispanic	-0.018	0.049	0.001	-0.019	-0.009	0.008	0.082	-0.059
	[-0.72]	[1.02]	[0.03]	[-1.09]	[-0.24]	[0.21]	[2.29]	[-3.21]
Black	-0.103	0.016	0.130	0.162	-0.058	0.027	0.101	0.115
	[-2.16]	[0.37]	[4.81]	[6.31]	[-1.51]	[0.70]	[5.11]	[5.69]
Asian	0.070	0.093	-0.049	-0.074	0.027	0.039	0.027	0.008
	[1.31]	[2.34]	[-2.40]	[-4.61]	[0.68]	[1.10]	[1.55]	[0.51]
Female	0.051	0.060	0.000	0.018	0.009	0.007	0.006	-0.001
	[3.07]	[3.20]	[0.03]	[2.55]	[0.57]	[0.59]	[0.89]	[-0.35]
LTV	0.751	0.804	0.628	1.165	1.256	1.542	1.229	1.784
	[14.29]	[13.13]	[9.08]	[30.02]	[24.73]	[22.04]	[15.01]	[26.49]
AddLTV	0.378	0.068	-0.480	0.039	-0.230	-1.093	-1.274	-1.577
	[2.83]	[0.48]	[-4.64]	[0.38]	[-1.14]	[-4.61]	[-12.60]	[-17.06]
Loan (log)	-0.378	-0.234	-0.384	-0.235	-0.566	-0.366	-0.444	-0.211
	[-17.13]	[-8.09]	[-19.35]	[-11.44]	[-19.47]	[-19.44]	[-29.07]	[-15.58]
SecondLien	2.400	2.425	3.466	3.715	2.011	3.494	3.734	4.908
	[16.40]	[23.94]	[29.66]	[45.09]	[6.95]	[15.16]	[36.45]	[53.56]
Refinance	-0.109	-0.403	-0.130	-0.111	-0.597	-1.074	-0.103	-0.023
	[-4.75]	[-13.76]	[-5.60]	[-5.67]	[-14.44]	[-25.19]	[-6.10]	[-0.95]
PrepayPenalty	0.145	0.027	-0.143	-0.114	0.453	0.336	0.262	0.088
	[5.59]	[1.07]	[-7.46]	[-9.38]	[13.75]	[13.22]	[9.28]	[3.86]
FirstTimeOwner	-0.169	-0.288	0.040	0.045	-0.500	-0.627	0.047	-0.017
	[-3.43]	[-4.55]	[2.59]	[3.83]	[-11.89]	[-18.38]	[3.01]	[-1.74]
OwnerOccupied	-0.468	-0.368	-0.384	-0.363	-1.225	-0.931	-0.441	-0.227
	[-12.95]	[-10.79]	[-16.70]	[-17.52]	[-14.04]	[-12.68]	[-24.67]	[-8.62]
OneBorrower	0.030	0.066	-0.003	0.065	-0.003	0.021	-0.035	0.054
	[2.12]	[4.52]	[-0.36]	[5.67]	[-0.30]	[1.35]	[-3.22]	[6.19]

Dep. Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Initial Rate for Approximately Fixed-Rate Loans				Current Rate			
Subsample	Bank/ Full-Doc	Bank/ Low-Doc	Broker/Full-Doc	Broker/Low-Doc	Bank/Full-Doc	Bank/Low-Doc	Broker/Full-Doc	Broker/Low-Doc
Income (log)	0.043	0.063	0.021	0.009	0.104	0.176	0.118	0.033
	[3.74]	[3.21]	[2.43]	[1.55]	[6.27]	[10.78]	[10.72]	[5.01]
IncomeMiss	0.108	0.044	-0.090	0.125	0.241	0.068	-0.035	0.030
	[1.11]	[1.18]	[-1.22]	[5.81]	[3.19]	[1.56]	[-0.53]	[1.46]
CashResv	-0.008	-0.039	-0.052	-0.065	-0.031	-0.104	-0.050	-0.090
	[-1.75]	[-3.17]	[-4.34]	[-11.24]	[-5.69]	[-10.79]	[-7.46]	[-32.95]
CreditScore	-0.007	-0.004	-0.008	-0.005	-0.007	-0.005	-0.009	-0.006
	[-21.87]	[-13.89]	[-20.69]	[-19.86]	[-25.75]	[-18.54]	[-43.89]	[-26.60]
Age (log year)	0.071	0.020	0.117	0.068	0.016	0.027	0.100	0.100
	[3.51]	[1.12]	[12.49]	[8.75]	[1.00]	[1.89]	[17.54]	[21.43]
Tenure	-0.008	-0.016	0.013	0.004	-0.002	-0.018	0.006	0.008
(log month)	[-1.19]	[-2.57]	[1.98]	[1.38]	[-0.35]	[-3.26]	[1.44]	[2.47]
TenureMiss	0.030	-0.026	-0.284	-0.160	-0.039	-0.177	-0.462	-0.3847
	[0.58]	[-0.75]	[-7.40]	[-4.96]	[-1.05]	[-6.67]	[-16.08]	[-16.61]
SelfEmploy	0.175	0.083	0.045	0.001	0.173	0.043	0.006	-0.053
	[5.48]	[3.01]	[2.61]	[0.06]	[6.21]	[2.37]	[0.38]	[-6.63]
2005	0.206	0.144	0.300	0.164	0.284	0.384	0.226	0.231
	[13.83]	[6.38]	[9.52]	[7.89]	[11.58]	[12.61]	[8.07]	[11.01]
2006					0.946	0.983	0.704	0.686
					[36.70]	[28.83]	[31.88]	[30.25]
2007					0.640	0.673	0.453	0.495
					[16.25]	[16.36]	[18.79]	[17.96]
2008					0.329	0.328	0.084	0.202
					[5.46]	[5.32]	[1.94]	[7.20]
Constant	15.084	12.203	16.228	12.219	19.107	15.173	17.477	12.739
	[34.54]	[24.55]	[36.07]	[38.77]	[48.47]	[59.25]	[52.18]	[46.19]
Observations	12,774	10,169	41,655	72,047	31,408	35,553	166,402	425,181
R-squared	0.525	0.357	0.697	0.665	0.522	0.564	0.576	0.593

Appendix D

WORKING PAPER NO. 09-21

SECURITIZATION AND MORTGAGE DEFAULT: REPUTATION VS. ADVERSE SELECTION

Ronel Elul
Federal Reserve Bank of Philadelphia

First version: April 29, 2009
This version: September 22, 2009

Securitization and Mortgage Default: Reputation vs. Adverse Selection[1]

Ronel Elul[2]

Federal Reserve Bank of Philadelphia

First Version: April 29, 2009

This version: September 22, 2009

Preliminary

Abstract

The academic literature, the popular press, and policymakers have all debated the securitization's contribution to the poor performance of mortgages originated in the run-up to the current crisis. Theoretical arguments have been advanced on both sides, but the lack of suitable data has made it difficult to assess them empirically. We examine this issue by using a loan-level data set from LPS Analytics, covering approximately three-quarters of the mortgage market from 2003-2007 and including both securitized and non-securitized loans. We find evidence that privately securitized loans do indeed perform worse than similar, non-securitized loans. Moreover, this effect is concentrated in prime mortgage markets; for example, a typical prime ARM loan originated in 2006 becomes delinquent at a 20 percent higher rate if it is privately securitized, ceteris paribus. By contrast, subprime loan performance does not seem to be worse for most classes of securitized loans.

Introduction

The recent dramatic increase in mortgage default rates, particularly for subprime loans, has led many to blame securitization. Simply put, the argument is that since the majority of subprime loans were securitized, issuers had less incentive to screen those loans, and this

---

[1] The author thanks Mitchell Berlin, Philip Bond, Paul Calem, Larry Cordell, Will Goetzmann, David Musto, Leonard Nakamura, Amit Seru, Anthony Sanders, Nick Souleles, and Paul Willen, as well as participants at the Wharton Macro Finance Lunch, the FDIC Mortgage Default Symposium, and the Yale Financial Crisis Conference. I am particularly indebted to Ted Wiles for outstanding research support.

[2] Research Department, Federal Reserve Bank of Philadelphia, Ten Independence Mall, Philadelphia, PA 19106. Email: ronel.elul@phil.frb.org. Tel: (215) 574-3965. The views expressed in this paper are those of the author and do not necessarily represent policies or positions of the Federal Reserve Bank of Philadelphia nor the Federal Reserve System. This paper is available free of charge at: www.philadelphiafed.org/research-and-data/publications/working-papers/.

encouraged a decline in lending standards. This argument has featured prominently in the popular press and has also been echoed by policymakers.[3] For example, the recently released U.S Treasury report on regulatory reform notes that "[t]he lack of transparency and standards in markets for securitized loans helped to weaken underwriting standards." and the report goes on to propose that issuers be required to maintain a 5 percent stake in any securitization. The argument has also found support in recent academic work, for example, Dell'Ariccia, Igan, and Laeven (2008), Mian and Sufi (2009), and Keys, et al. (2009).[4]

On the other hand, others (most prominently, Gorton, 2008) have pointed out that issuers retained substantial exposure even after mortgages were securitized. Some of this was explicit, since issuers often continued to service mortgages they had sold, or they retained senior tranches of CDOs containing these mortgages. But it was also implicit; for example, Higgins and Mason (2004) document instances in which issuers of credit card ABS have taken back non-performing loans (Higgins and Mason, 2004). More generally, Gorton and Souleles (2007) show that prices paid by investors in credit card ABS take into account issuers' ability to bail out their ABS. Thus, issuers' incentives need not necessarily be misaligned with those of investors. This view is also supported by earlier work on the securitization of prime mortgages. (See Ambrose, et al. , 2005, who found that securitized loans tended to perform *better* than similar non-securitized loans.)

One difficulty with most of the recent academic work is that the data used do not allow researchers to determine whether individual loans are in fact securitized. Dell'Ariccia, Igan, Laeven (2008) and Mian and Sufi (2009) instead use local-level aggregate securitization rates, an approach that makes interpreting their results difficult, since it is difficult to distinguish the effect of securitization from that of other local conditions.

Keys, et al. (2009) use loan-level data, but only for securitized loans (from the Loan Performance ABS database). So they must use an instrumental variables approach to characterize loans that are "harder" to securitize (those with credit scores just below 620) and find that these loans are indeed less likely to default, ceteris paribus. While this is indeed an ingenious approach, several issues arise. First, this instrument is rather weak, since many subprime MBS

[3] http://www.financialstability.gov/docs/regs/FinalReport_web.pdf
[4] See also Nadauld and Sherlund (2009), who argue that house price appreciation facilitates securitization, and that underwriters tended to purchase lower credit-quality subprime mortgages in those ZIP codes that experienced the highest house price appreciation.

did indeed contain substantial numbers of loans below this cutoff. For example, in the New Century securitization studied by Ashcraft and Schuermann (2008), 57 percent of all loans have FICO scores below 620. Furthermore, work by Krainer and Laderman (2009), and others, suggests that this “620-discontinuity” appears to affect the performance of *non-securitized* loans in equal measure.[5] Relative to this paper, however, a key limitation of their approach is that they can only examine the effect of securitization for a very narrow subset of loans — those in the neighborhood of their cutoff. And, indeed, they find a significant effect only for a small subsample of loans — those with low or no documentation of income. By contrast, our approach allows us to examine a much broader segment of the mortgage market.[6]

In this paper we take a more direct approach, one that avoids many of these problems. We use the LPS data set, which includes both securitized loans and those held in portfolio by the original lender. We find evidence that for prime mortgages, private securitized loans indeed perform worse than portfolio loans; for instance, for loans originated in 2006, the two-year default rate is at least 15 percent higher, on average. Given the large number of prime loans that were originated over this period, this difference in default rates is economically significant. By contrast, securitized subprime loans do not appear to have defaulted at higher rates than similar non-securitized loans. As we discuss below, this relative difference in performance between prime and subprime loans may be driven by two factors. First, subprime loans are likelier to have been subject to greater scrutiny by investors, whereas prime loans would have been presumed to be of higher quality, thereby reducing the scope for adverse selection. In addition, as we discuss below, very few subprime loans were actually held in portfolio, further reducing the benefit to the lenders from cream-skimming and also increasing lenders’ risk from doing so.

Our analysis also breaks down the effect of securitization by origination year. We find some evidence that this effect grows over time. In particular, for the largest segment of the market, prime FRMs, securitized loans originated in earlier years perform no worse, and sometimes better, than similar non-securitized loans. However, this effect decreases over time, and beginning with the 2006 vintage, such loans actually become delinquent at higher rates than

[5] Some of these criticisms are addressed, at least in part, by additional analyses that they undertake in the paper. In particular, they also examining the introduction, and repeal, of anti-predatory lending laws in Georgia and New Jersey. The results of this latter analysis are consistent with those of their primary approach; during the period that these laws were in force, loans with credit scores slightly above 620 default a higher rates than those with scores slightly below.

[6] A final issue with their approach is that credit scores themselves are not completely exogenous and are subject to manipulation.

non-securitized loans. Our interpretation is that while in earlier years reputational effects were sufficient to sustain underwriting standards, as loan volumes increased, and the future of the housing market became more and more tenuous, the current benefit from originating questionable loans outweighed the future costs, and this led to a deterioration in issuers' incentives to properly underwrite loans. We should stress, however, that our results do not rule out the possibility that investors understood that such a deterioration in standards had taken place and that the prices or structures of the MBS reflected this.

We should point out that a recent paper by Jiang et al. (2009) also uses loan-level data on securitized and non-securitized loans to study the effect of loan sales on default. They study approximately 700,000 mortgages originated by a single lender specializing in Alt-A and, in particular, low-documentation loans. They find that broker-originated loans are more likely to default. However, they also find that low-doc loans that are securitized are *less* likely to default, after controlling for observable characteristics. Both of these findings are consistent with our results, as reported below. Since their data come from only a single lender, it is unclear how general their findings are. By contrast, because we work with a larger data set, we are able to establish a much broader set of results. In particular, our results highlight the different impact that securitization had in the prime and subprime markets.

Data

Introduction

We use loan-level data from the LPS Applied Analytics Inc, data set.[7] Other researchers have used this data set to study foreclosure outcomes; see Piskorski, Seru, and Vig (2009) and Foote et al. (2009); a more detailed description of the data may also be found in the latter paper. These data are provided by the servicers of the loans and include nine of the top 10 servicers.

As Table 1 demonstrates, coverage in LPS is approximately 75 percent of that in HMDA. However, subprime loans appear to be somewhat underrepresented, at least when compared to the Loan Performance data (Table 2).[8]

[7] This data set is also commonly known as the "McDash" data.
[8] The HMDA data do not break out loans by prime vs. subprime, so the HMDA shares in Table 2 are not directly comparable to LP and LPS.

Table 1: First Mortgage Originations: LPS vs. HMDA

	LPS	HMDA
2004	7.2m	10.2m
2005	7.4m	10.5m
2006	6.4m	8.6m
2007	5.1m	6.9m

Table 2: Subprime Share of Originations[9]

	LP	McDash	HMDA
2003	7%	2.7%	
2004	16%	8.5%	14%
2005	18%	9.2%	25%
2006	16%	17.3%	28%
2007		13.4%	18%

In order to make the analysis cleaner, we focus primarily on 30-year, owner-occupied first-lien mortgage loans.[10] We drop observations with missing data and obvious outliers. To reduce survival bias, we also restrict attention to loans that entered the LPS data set within 12 months of their origination date. Taken together, these restrictions eliminate 10-20 percent of the data we begin with. We also consider only loans originated from 2003-2007, since LPS coverage was more limited before 2003, and we want to have a sufficiently long time horizon following origination.[11] We then follow the loans through March 2009.

We consider the following products: fixed-rate mortgages (FRM), 5-year ARMs, 3-year ARMS, and 2/28 ARMs. These three classes of ARMs were chosen because, taken together, they make up over 60 percent of all the adjustable-rate mortgages originated in the LPS data set during this time period. For each class we break down the sample into prime and subprime loans (as reported by the servicers). Note that there is no separate category for Alt-A loans; depending on the issuer, they may be classed as either prime or subprime. We also consider prime low-doc

[9] Loan Performance statistics are from Mayer and Pence (2008). The HMDA share is the fraction of “higher-priced” loans in first-lien originations; pricing was first reported in HMDA in 2004.
[10] For the estimations reported here, we also dropped FHA and VA loans.
[11] We also repeated our analysis while restricting it to loans originated in 2005-2007; the results did not change significantly and are not reported here.

FRM and Jumbo FRM's separately. Except for prime FRM, where we draw a 25 percent random sample, we used all of the loans available in the LPS data set that met our criteria. Summary statistics for these different samples can be found in Table 6 of the Appendix.

The LPS data set is divided into a "static" file, whose values generally do not change over time, and a "dynamic" file. The static data set contains information obtained at the time of underwriting, such as the loan amount, house price, (origination) FICO score, documentation status, source of the loan (e.g., whether it was broker-originated), property location (zip code), type of loan (fixed-rate, ARM, prime, subprime, etc.), the prepayment penalty period (if any), and the termination date and termination status if the loan has indeed terminated. The termination types include "paid off," foreclosure (and other negative termination events such as REO sale), and the transfer of the loan to another servicer.

The dynamic file is updated monthly, and among other variables, it contains the status of the loan (current, 30 days delinquent, 60 days, etc.), the current interest rate (since this changes over time for ARMs), current balance, and investor type (private securitized, GNMA, FNMA, FHLMC, portfolio). The investor type variable is discussed in greater detail below.

We also generate several additional variables. First, we define a loan as being "in default" if it is 60+ days delinquent or if it experienced a negative termination event.[12] This is a relatively early definition of default, as opposed to a foreclosure, for example, which can occur many months later. We use this early definition for several reasons. First, state laws governing foreclosure differ widely, and this can have an effect on the length of time it takes to conclude a foreclosure.[13] Also, whether a delinquent loan is securitized or not may also affect the ease of modifying it and hence of avoiding foreclosure (Piskorski, Seru, and Vig, 2009 and Foote et al., 2009); thus, we choose to focus on the initial stages of distress.

In addition, we estimate the current house price by applying the FHFA house price index to the house price reported at origination[14] and use this to compute an estimate of the current loan-to-value ratio at quarterly frequency. We also calculate the house price appreciation in the

[12] We use the Mortgage Bankers Association (MBA) definition of delinquency: a loan increases its delinquency status if a monthly payment is not received by the end of the day immediately preceding the loan's next payment due date.

[13] Many papers have studied the effect of these state laws on foreclosure outcomes; for example, Ghent and Kudlyak (2009) use the LPS data to address laws that restrict deficiency judgments.

[14] For properties located within an MSA we use the MSA-level index, while for those not in an MSA we use the "rural" index (or the state-level index when the rural index is not available).

region over the four years *prior* to the origination date of the loan to capture the effect of a housing "boom" on lending standards (as in Dell'Ariccia, Igan, and Laeven, 2008).

The Investor Type

Since the investor type is a key variable in our analysis, we discuss it in more detail. First note that the investor type is dynamic: nearly half of all loans are initially recorded as "portfolio" loans and are only then subsequently securitized, typically within several months. So we must define the investor type carefully so as to capture the "intended" investor type at the time of origination. Roughly speaking, we adopted the most common investor type during the first year of the loan's life; we restricted attention to the first year in order to more closely capture the intended investor type when the loan was originated.[15] Table 3 compares this "final" investor type to the one reported at loan origination.

Another issue is that a loan may also end up in a lender's portfolio not by design but because the loan defaults before it can be securitized. In particular, investors are generally able to force lenders to take back any loans that experience "early default," that is, loans that default in the first three months.[16] This is a particular concern for subprime loans, especially in the 2006 vintage. As we discuss below, to address this possibility we repeat our analysis while dropping any loans that have defaulted within three months of origination.

[15] More precisely, for a given loan the investor type used was constructed as follows. We considered all investor types that occur during the first year of a loan's life. We then denoted an investor type to be "admissible" if it matches the modal investor type over the 12 months following the date on which it first appears. We then selected the admissible investor type that occurred first. Note that for almost all loans, there was only a single admissible investor type. On average, the investor type was determined within three months of the origination date.
[16] We are grateful to Amit Seru for highlighting the importance of this.

Table 3: Initial and "Final" Investor Type

	PRIME					SUBPRIME			
Year	**FHA**	**GSE**	**Private Securitized**	**Portfolio**		**FHA**	**GSE**	**Private Securitized**	**Portfolio**
Investor Type at Origination	2.81%	24.81%	22.84%	49.54%		0.05%	5.67%	52.84%	41.44%
"Final" Investor Type	9.02%	59.27%	22.43%	9.28%		0.09%	7.40%	84.12%	8.38%

Some Stylized Facts

Before we begin our formal analysis, it is useful to establish a few facts about the data. First, Table 4 reports the distribution of loans by investor type, for each product. We can see that prime ARMs represent an ideal laboratory for studying the effect of securitization, since issuers distributed their loans across all three investor types. Conversely, observe that the vast majority of subprime loans were privately securitized.

It is also useful to simply compare default rates across the different investor types (Table 5). Notice that private securitized loans do indeed seem to default at higher rates than other loans. However, this does not take account of the observable risk factors for these loans: for example, as we have already seen, private securitization was concentrated in the subprime market. Thus, a formal analysis is needed.

Table 4: Investor Type by Product

		GSE	Private Securitized	Portfolio	# Loans
FRM	Prime	78.6%	15.6%	5.7%	17.9 m
	Jumbo FRM	-	89.7%	10.3%	0.8 m
	Subprime	19.1%	74.1%	6.7%	0.8 m
	Lowdoc	80.7%	16.7%	2.5%	1.6 m
5 Yr ARMs	Prime	40.4%	35.2%	24.3%	2.2 m
	Subprime	17.9%	77.8%	4.2%	0.03 m
3 Yr ARMs	Prime	38.5%	35.5%	26.0%	0.6 m
	Subprime	0.0%	95.3%	4.7%	0.3m
2/28 ARMs	Prime	0.1%	95.4%	4.5%	0.4m
	Subprime	0.1%	91.0%	8.9%	0.8m

Table 5: Termination Status: by Investor Type (%)

	Entire Sample	FHA	GSE	Private Securitized	Portfolio
Paid Off or Did Not Terminate in Sample	93.47	92.4	97.31	85.38	89.62
Defaulted	1.7	2.59	0.59	3.78	1.62
Transferred to Other Servicer	4.83	5.37	2.1	10.84	9.13

Results

We estimate a Cox proportional hazard model, with default as the dependent variable, for each of our subsamples. The coefficients are reported in Tables 7-11, with standard errors in parentheses. Figure 1 plots the coefficients on securitized loans, by origination year, for the products we consider.

Turning first to the most standard product, prime FRM (Table 7), the coefficients in Panel A generally conform to our intuition. As others have found, broker-originated and low-doc loans are riskier. A borrower with a higher FICO score is less likely to default, while higher interest rates are riskier. We control for both current and origination LTV. A higher current LTV is has a strong positive association with delinquency, as expected. By contrast, loans with higher origination LTVs default at lower rates (this may reflect screening on unobservables).[17] Also note that origination LTV enters as a spline, with its effect permitted to differ based on whether the origination LTV is below, equal to, or above 80 percent. And indeed, loans originated at 80 percent LTV default at slightly higher rates.[18]

As others have found, high house price appreciation in the four years prior to the mortgage origination date is associated with a higher likelihood of default (for example, Dell'Ariccia, Igan, and Laeven, 2008; Nadauld and Sherlund, 2009), although when we include MSA dummies, this coefficient changes sign.

The interactions between the dummy variables for origination year and investor type are in Panel B. The baseline origination year is 2003, and the baseline investor type is a portfolio loan (we dropped FHA and VA loans). The coefficients for loans originated in subsequent years are positive; that is, loans originated in later vintages are riskier, after controlling for risk and house prices changes. This is consistent with other papers' findings: see for example, Demyanyk and Van Hemert (2009).

As for the marginal contribution of private securitization, observe that in 2003-2005, securitized loans are less risky. However, this coefficient attenuates over time, and, starting in 2006, the contribution of private securitization to default is positive. We also re-ran the estimation for the subset of loans originated in the top 25 MSAs, both with and without MSA fixed effects, and obtained similar results.

Finally, we also break up our sample into the broker-originated loans and those that were not originated through a broker. These results are reported in Table 13. This allows us to conclude that it is in fact the brokered-subsample that accounts for our earlier finding that prime securitized FRM are riskier. In addition, we also see that lowdoc loans are particularly risky when they are originated by brokers; this is also consistent with Jiang et al (2009).

Results for jumbo FRM are similar and are reported in Tables 8 and 13.

As discussed above, prime ARMs represent an ideal laboratory for studying the effect of securitization. In particular, 5-year and 3-year ARMs are split roughly evenly between all three

---

[17] Similar results have been observed in other contexts; for example, Berger and Udell (1990) find that riskier business loans tend to have more collateral.

[18] This may reflect the existence of unobserved "piggyback" (second) mortgages, which are more common for those loans originated at 80 percent LTV.

investor types, unlike prime FRM, where the GSEs dominated, and subprime ARMs, for which the vast majority of loans were in private securitized pools. For both of these, the coefficients on securitization are positive for every year (Tables 9 and 10). The results for 2/28 ARMs are similar, although they are negative for 2003 and 2004 (Table 10).

To help assess the economic significance of securitization on prime mortgage default rates, Table 12 reports the average two-year cumulative delinquency rate for loans originated in 2006, as well as the securitization coefficient for that year from the Cox regression. For example, for a typical prime 5-year ARM, private securitization would raise the delinquency rate by 23 percent,[19] from 14.6 percent to 18.4 percent.

By contrast, securitization seems to have a much smaller impact on subprime loans. For subprime ARMs, indeed, our baseline results have a negative coefficient in all years, which implies that securitized loans actually perform *better* (Table 11). However, it is important to note that a large part of this effect, particularly for the worst-performing vintages (such as 2006), is driven by early defaults.[20] This occurs because, as discussed above, early defaults can bias our definition of investor type; for example, loans that have defaulted can no longer be securitized. To control for this we re-run our estimations, but now excluding those loans that became delinquent within three months of origination.[21] This attenuates the coefficients, particularly for loans originated in 2005-2007; these coefficients are no longer statistically significant (and many become positive), once we make this restriction. By contrast, excluding these loans has little effect on the results for the prime sample (Tables 8-10), since it contains few early defaults.

The results for low-documentation FRM are similar (Table 8). Our baseline regression yields negative coefficients, but once we exclude those loans that become delinquent within three months of origination, they become positive or insignificant. This is also consistent with the findings of Jiang et al (2009) using their data set (from a single lender).

As discussed earlier, we conjecture that there are two reasons why riskier loans appear to have been securitized in prime markets, but less so in subprime. First, investors were aware of risks in the subprime and low-doc markets and may have scrutinized loans without the “prime” imprimatur more carefully. Along the same lines, “cherry picking” would have been riskier for

---

[19] That is, $0.26= e^{0.23}-1$.

[20] For example, 30 percent of all 2/28 subprime ARMs held in portfolio become delinquent within the first three months; for securitized loans the corresponding figure is only 10 percent.

[21] We also drop loans with small balances (<$50,000), since these are also less likely to be securitized. We thank Paul Calem for this suggestion.

subprime lenders who were dependent on securitized pools to hold their loans. There are two reasons to be cautious in interpreting these results on the subprime market, however. First, as we have shown, coverage of the LPS data of this segment of the market was less broad. Also, the fact that securitization was ubiquitous in the subprime market might imply that the few non-securitized loans may be special in some way.

Conclusion

Using a data set that covers approximately 75 percent of loan originations from the years 2003-2007, and that includes both private securitized, GSE, and mortgages held in portfolio, we have shown that prime (private) securitized loans originated at the peak of the bubble performed significantly worse than similar non-securitized loans, ceteris paribus. This is particularly striking for markets such as prime ARMs, in which issuers held non-negligible amounts of loans in portfolio, and for lenders who were less reliant on securitization. We argue that this is evidence that adverse selection was present in the prime mortgage market, and that this may have contributed to a deterioration in underwriting standards.

However, in contrast to some previous studies, we do not find that the small fraction of subprime loans that were held in portfolio performed better than securitized loans. We suggest that this may be the result of two factors: subprime lenders' reliance on securitization made cherry picking more risky for them, and investors were more careful in scrutinizing loans that did not have the “prime” imprimatur.

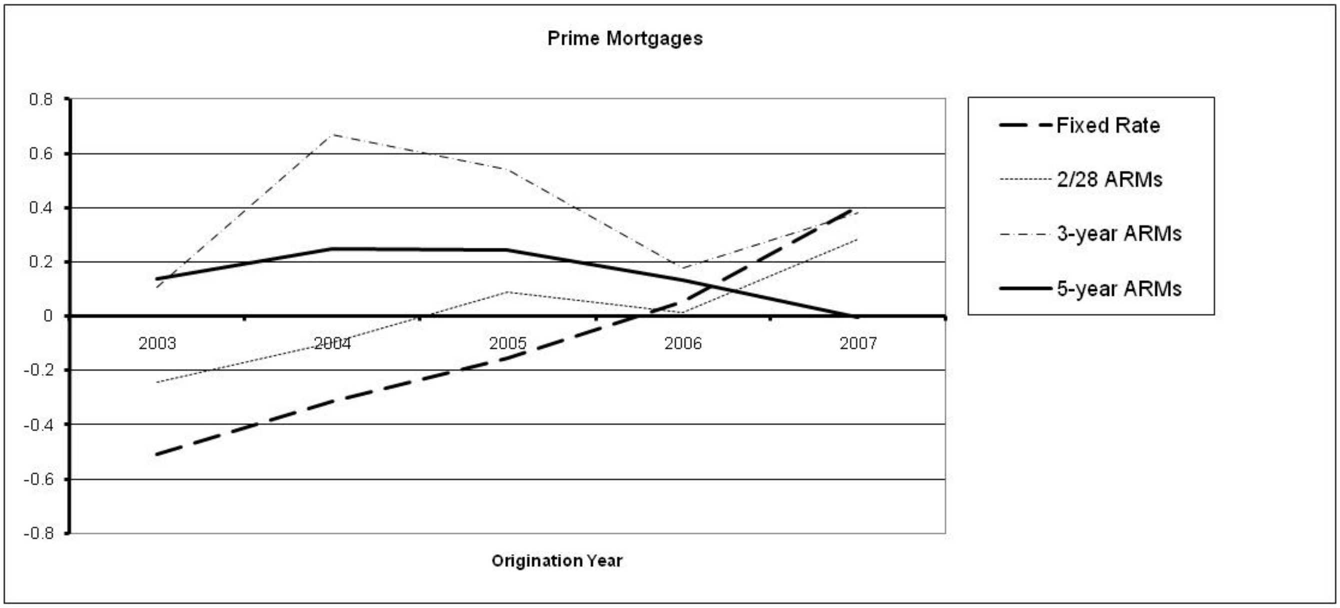
References

- Ambrose, Brent, Michael LaCour-Little, and Anthony Sanders (2005), "Does Regulatory Capital Arbitrage, Reputation, or Asymmetric Information Drive Securitization?," Journal of Financial Services Research, 28:1.
- Ashcraft, Adam, and Til Schuermann (2008), "Understanding the Securitization of Subprime Mortgage Credit," Federal Reserve Bank New York Staff Report #318.
- Berger, Allen N., and Gregory F. Udell (1990), "Collateral, Loan Quality, and Bank Risk," Journal of Monetary Economics, 25:1.
- Bhardwaj, Geetesh and Sengupta, Rajdeep (2009), "Where's the Smoking Gun? A Study of Underwriting Standards for US Subprime Mortgages," Federal Reserve Bank of St. Louis Working Paper No. 2008-036B.
- Dell'Ariccia, Giovanni, Deniz Igan, and Luc Laeven (2008), "Credit Booms and Lending Standards: Evidence from the Subprime Mortgage Market," IMF Working Paper WP/08/106.
- Demyanyk, Yuliya, and Otto Van Hemert (2009), "Understanding the Subprime Mortgage Crisis," Forthcoming, Review of Financial Studies.
- Foote, Christopher L., and Kristopher Gerardi, Lorenz Goette, and Paul S. Willen (2009), "Reducing Foreclosures," Federal Reserve Bank of Boston Public Policy Discussion Paper 09-2.
- Ghent, Andra, and Marianna Kudlyak (2009), "Recourse and Residential Mortgage Default: Theory and Evidence from US States," Federal Reserve Bank of Richmond Working Paper 09-10.
- Gorton, Gary (2008), "The Panic of 2007," Yale ICF Working Paper No. 08-24.
- Gorton, Gary, and Nicholas S. Souleles (2007), "Special Purpose Vehicles and Securitization," in Rene Stulz and Mark Carey (eds.), *The Risks of Financial Institutions.* Chicago: University of Chicago Press.
- Higgins, Eric, and Joseph Mason (2004), "What Is the Value of Recourse to Asset Backed Securities? A Study of Credit Card Bank ABS Rescues," Journal of Banking and Finance 28, 857-874.

- Jiang, Wei, Ashlyn Nelson, and Edward Vytlacil, (2009), “Liar’s Loan? Effects of Loan Origination Channel and Loan Sale on Delinquency,” manuscript, Columbia University.
- Keys, Benjamin, Tanmoy Mukherjee, Amit Seru, and Vikrant Vig (2009), “Did Securitization Lead to Lax Screening? Evidence from Subprime Loans,” forthcoming, Quarterly Journal of Economics.
- Krainer, John, and Elizabeth Laderman (2009), “Mortgage Loan Securitization and Relative Loan Performance,” manuscript, Federal Reserve Bank of San Francisco.
- Mayer, Christopher and Karen Pence (2008), “Subprime Mortgages: What, Where, and to Whom?,” FEDS Working Paper 2008-29.
- Mian, Atif R., and Amir Sufi (2009), “The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis,” forthcoming, Quarterly Journal of Economics.
- Nadauld, Taylor, and Shane Sherlund, (2009), “The Role of the Securitization Process in the Expansion of Subprime Credit,” Fisher College of Business Working Paper 2009-03-009.
- Piskorski, Tomasz, Amit Seru and Vikrant Vig (2009), “Securitization and Distressed Loan Renegotiation: Evidence from the Subprime Mortgage Crisis,” Chicago Booth School of Business Research Paper No. 09-02.

Appendix – Figures and Tables

Figure 1: Results: Coefficients on Private Securitization, by Origination Year



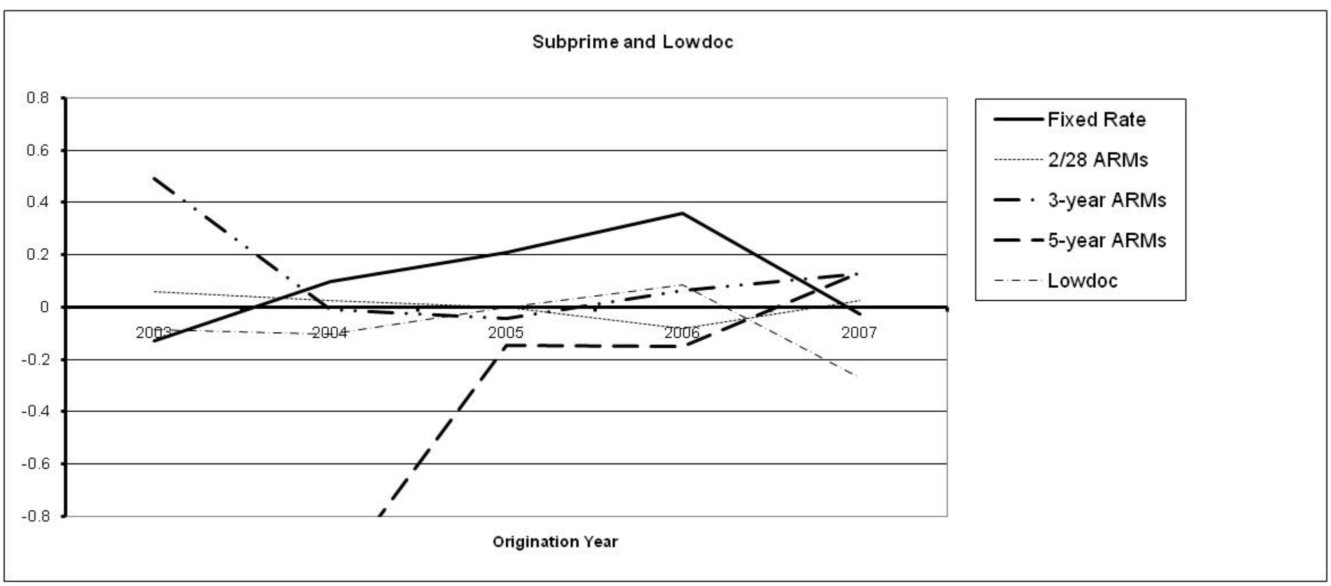


Table 6: Means of Selected Variables

	FRM			5-Yr ARMs		3-Yr ARMs		2-Yr ARMs	
	Prime	Lowdoc	Subprime	Prime	Subprime	Prime	Subprime	Prime	Subprime
Variable									
FICO (origination)	711	700	617	727	656	714	607	637	609
Current Interest Rate (%)	6.16	6.33	7.72	5.55	7.15	5.21	7.46	7.50	7.89
Margin (%; ARMs ONLY)				2.38	4.64	2.53	5.10	5.27	5.39
Jumbo Loan	0.05	0.04	0.06	0.26	0.16	0.19	0.08	0.14	0.09
Lowdoc Loan	0.10	1.00	0.06	0.22	0.15	0.23	0.03	0.14	0.09
Broker-originated	0.16	0.19	0.19	0.20	0.37	0.23	0.30	0.19	0.20
Correspondent-originated	0.29	0.42	0.17	0.20	0.05	0.25	0.17	0.02	0.12
Option-ARM Loan	0.00	0.01	0.00	0.14	0.41	0.09	0.05	0.20	0.16
Interest-Only Loan	0.02	0.03	0.03	0.50	0.39	0.35	0.23	0.27	0.16
Transferred from other Servicer	0.11	0.02	0.33	0.09	0.55	0.16	0.21	0.75	0.41
Prepayment Penalty Active	0.01	0.01	0.92	0.02	0.20	0.04	0.54	0.17	0.35
PMI	0.15	0.16	0.04	0.08	0.02	0.12	0.01	0.03	0.07
Refinancing	0.43	0.48	0.70	0.38	0.55	0.48	0.55	0.32	0.50
Cash-out Refinancing	0.11	0.17	0.47	0.10	0.37	0.10	0.47	0.06	0.32
LTV at Origination (%)	76	74	78	73	77	74	81	93	81
Current LTV Estimate (%)	66	67	72	65	66	64	74	76	75
County Unemployment Rate (%)	5.3	5.4	5.4	4.8	4.9	5.1	5.1	5.1	5.0
HPI Appreciation: 4 years before Orig.	0.39	0.42	0.44	0.54	0.52	0.49	0.48	0.58	0.51
Defaulted in-sample	0.09	0.10	0.34	0.09	0.27	0.12	0.35	0.43	0.39
Paid Off-in Sample	0.28	0.26	0.31	0.38	0.33	0.64	0.55	0.43	0.50

Table 7 (Panel A): Prime Fixed Rate Estimates

	Base Case	No Early Default	Top 25 MSA	
				Dummies
FICO at origination	-0.0124***	-0.0123***	-0.0121***	-0.0121***
	(5.74E-05)	*(6.02E-05)*	*(9.09E-05)*	*(0.000094)*
Interest Rate	0.511***	0.548***	0.600***	0.580***
	(0.00606)	*(0.00671)*	*(0.00879)*	*(0.00881)*
Loan Amount	0.000000464***	0.000000537***	0.000000551***	0.000000546***
	(3.19E-08)	*(3.39E-08)*	*(3.63E-08)*	*(3.85E-08)*
Jumbo loan	-0.334***	-0.360***	-0.392***	-0.361***
	(0.0195)	*(0.0206)*	*(0.0247)*	*(0.0249)*
Low-doc Loan	0.0381***	0.0347***	0.0143	0.0112
	(0.00713)	*(0.00759)*	*(0.01110)*	*(0.01110)*
Broker-originated	0.268***	0.240***	0.225***	0.226***
	(0.00734)	*(0.00790)*	*(0.01120)*	*(0.01130)*
Correspondent	0.0603***	0.0314***	0.0233*	0.0401***
	(0.00643)	*(0.00698)*	*(0.00956)*	*(0.00956)*
Prepayment Penalty Active	0.375***	0.381***	0.417***	0.318***
	(0.0316)	*(0.0332)*	*(0.0467)*	*(0.0451)*
PMI	0.521***	0.535***	0.599***	0.571***
	(0.0313)	*(0.0348)*	*(0.0387)*	*(0.0374)*
Refinancing	-0.0269***	0.00727	0.000655	0.016
	(0.00793)	*(0.00860)*	*(0.01260)*	*(0.01250)*
Cashout-refi	0.109***	0.103***	0.108***	0.105***
	(0.00852)	*(0.00902)*	*(0.01260)*	*(0.01270)*
LTV at Orig. (<80%)	-0.0188***	-0.0172***	-0.0229***	-0.0332***
	(0.00159)	*(0.00170)*	*(0.00133)*	*(0.00136)*
LTV at Orig. (=80%)	-0.01653***	-0.01478***	-0.02034***	-0.03058***
	(0.00154)	*(0.00165)*	*(0.00121)*	*(0.00125)*
LTV at Orig. (>80%)	-0.0194***	-0.0180***	-0.0243***	-0.0343***
	(0.00160)	*(0.00170)*	*(0.00143)*	*(0.00144)*
Current LTV	3.109***	3.006***	3.771***	4.863***
	(0.1310)	*(0.1430)*	*(0.0504)*	*(0.0696)*
County Unemp. Rate	0.0366***	0.0351***	0.0340***	0.0411***
	(0.00319)	*(0.00346)*	*(0.00249)*	*(0.00287)*
HPI appreciation	0.133***	0.178***	0.0848***	-0.425***
(4-years prior to orig)	*(0.0246)*	*(0.0277)*	*(0.0211)*	*(0.0404)*

Table 7 (Panel B): Prime Fixed Rate Estimates

	Base Case	No Early Default	Top 25 MSA	
				Dummies
Orig. 2004	0.280***	0.354***	0.398***	0.391***
	(0.0543)	*(0.0600)*	*(0.0793)*	*(0.0791)*
Orig. 2005	0.310***	0.387***	0.486***	0.457***
	(0.0515)	*(0.0570)*	*(0.0756)*	*(0.0758)*
Orig. 2006	0.351***	0.454***	0.382***	0.308***
	(0.0501)	*(0.0562)*	*(0.0740)*	*(0.0742)*
Orig. 2007	0.459***	0.661***	0.724***	0.608***
	(0.0506)	*(0.0568)*	*(0.0731)*	*(0.0730)*
GSE 2003	-0.428***	-0.354***	-0.373***	-0.362***
	(0.0429)	*(0.0480)*	*(0.0645)*	*(0.0643)*
GSE 2004	-0.517***	-0.467***	-0.536***	-0.509***
	(0.0363)	*(0.0391)*	*(0.0511)*	*(0.0507)*
GSE 2005	-0.388***	-0.350***	-0.447***	-0.429***
	(0.0299)	*(0.0320)*	*(0.0434)*	*(0.0433)*
GSE 2006	-0.360***	-0.328***	-0.247***	-0.236***
	(0.0260)	*(0.0283)*	*(0.0401)*	*(0.0397)*
GSE 2007	0.101***	0.103***	0.123***	0.111**
	(0.0252)	*(0.0274)*	*(0.0356)*	*(0.0347)*
Securitized-2003	-0.566***	-0.510***	-0.507***	-0.476***
	(0.0494)	*(0.0549)*	*(0.0714)*	*(0.0711)*
Securitized-2004	-0.398***	-0.314***	-0.359***	-0.315***
	(0.0402)	*(0.0425)*	*(0.0561)*	*(0.0558)*
Securitized-2005	-0.190***	-0.154***	-0.238***	-0.199***
	(0.0323)	*(0.0346)*	*(0.0463)*	*(0.0462)*
Securitized-2006	0.0663*	0.0555	0.117**	0.125**
	(0.0282)	*(0.0305)*	*(0.0433)*	*(0.0429)*
Securitized-2007	0.437***	0.405***	0.353***	0.327***
	(0.0309)	*(0.0331)*	*(0.0438)*	*(0.0429)*
No Early Default & Loan>$50,000	N	Y	Y	Y
MSA Subset	-	-	Y	Y
MSA Dummy	-	-	Y	Y
N	29128421	28257070	13701430	13701430

Table 8 (Panel A): Other Fixed-Rate Mortgages

	Jumbo FRM			Subprime FRM			Lowdoc Prime FRM		
	Base Case	No Early Default	Top 25 MSA	Base Case	No Early Default	Top 25 MSA	Base Case	No Early Default	Top 25 MSA
FICO at origination	-0.0108***	-0.0108***	-0.0109***	-0.00619***	-0.00583***	-0.00555***	-0.0113***	-0.0110***	-0.0108***
	(0.0000999)	*(0.000105)*	*(0.000130)*	*(0.0000497)*	*(0.0000543)*	*(0.0000774)*	*(0.0000584)*	*(0.0000618)*	*(0.0000895)*
Loan Amount	6.79e-08***	4.58e-08*	3.22e-09	0.00000117***	0.00000110***	0.000000957***	0.000000456***	0.000000509***	0.000000633***
	(1.76e-08)	*(1.93e-08)*	*(2.41e-08)*	*(5.34e-08)*	*(5.25e-08)*	*(6.29e-08)*	*(3.36e-08)*	*(3.59e-08)*	*(4.80e-08)*
Jumbo				-0.207***	-0.220***	-0.177***	-0.281***	-0.313***	-0.358***
				(0.0183)	*(0.0185)*	*(0.0206)*	*(0.0212)*	*(0.0224)*	*(0.0286)*
Low-doc	-0.0130	-0.0193	-0.00997	0.0568***	0.0578***	0.0475**			
	(0.0139)	*(0.0146)*	*(0.0174)*	*(0.0101)*	*(0.0113)*	*(0.0173)*			
Broker	0.284***	0.245***	0.231***	0.123***	0.109***	0.113***	0.245***	0.235***	0.228***
	(0.0113)	*(0.0119)*	*(0.0144)*	*(0.00681)*	*(0.00743)*	*(0.0108)*	*(0.00850)*	*(0.00896)*	*(0.0132)*
Correspondent	-0.0537***	-0.0740***	-0.0398**	0.114***	0.121***	0.0940***	0.0532***	0.0274***	0.0703***
	(0.0118)	*(0.0123)*	*(0.0150)*	*(0.00689)*	*(0.00755)*	*(0.0115)*	*(0.00739)*	*(0.00786)*	*(0.0119)*
Prepayment Penalty	0.240***	0.257***	0.224***	-0.126***	-0.143***	-0.0990***	0.141***	0.149***	0.0573
	(0.0147)	*(0.0157)*	*(0.0186)*	*(0.0102)*	*(0.0105)*	*(0.0143)*	*(0.0324)*	*(0.0353)*	*(0.0479)*
LTV at Orig. (<80%)	0.0247***	0.0275***	0.0286***	-0.00784***	-0.00590***	-0.00692***	-0.00998***	-0.00749***	-0.0139***
	(0.00131)	*(0.00126)*	*(0.00160)*	*(0.000427)*	*(0.000468)*	*(0.000714)*	*(0.000527)*	*(0.000556)*	*(0.000923)*
LTV at Orig. (=80%)	0.0258***	0.0283***	0.0295***	-0.00698***	-0.00514***	-0.00566***	-0.00946***	-0.00698***	-0.0133***
	(0.00128)	*(0.00123)*	*(0.00158)*	*(0.000403)*	*(0.00044)*	*(0.00068)*	*(0.000504)*	*(0.000528)*	*(0.000891)*
LTV at Orig. (>80%)	0.0191***	0.0219***	0.0233***	-0.00892***	-0.00718***	-0.00831***	-0.00736***	-0.00528***	-0.0120***
	(0.00130)	*(0.00125)*	*(0.00159)*	*(0.000374)*	*(0.000408)*	*(0.000643)*	*(0.000471)*	*(0.000492)*	*(0.000855)*
Current LTV	1.165***	1.179***	1.196***	2.484***	2.428***	2.639***	3.253***	3.200***	3.956***
	(0.101)	*(0.0927)*	*(0.124)*	*(0.0263)*	*(0.0274)*	*(0.0477)*	*(0.0323)*	*(0.0331)*	*(0.0682)*
Interest Rate	0.682***	0.695***	0.681***	0.231***	0.226***	0.242***	0.573***	0.606***	0.647***
	(0.00845)	*(0.00879)*	*(0.0107)*	*(0.00247)*	*(0.00274)*	*(0.00393)*	*(0.00965)*	*(0.00852)*	*(0.0103)*
HPI Appreciation	0.362***	0.386***	-0.166**	0.0716***	0.110***	0.153***	0.0966***	0.134***	-0.329***
(4-yrs prior)	*(0.0232)*	*(0.0240)*	*(0.0554)*	*(0.0107)*	*(0.0114)*	*(0.0418)*	*(0.0133)*	*(0.0141)*	*(0.0503)*
Refinance	0.0818***	0.151***	0.147***	-0.320***	-0.303***	-0.353***	0.00193	0.0326***	0.0798***
	(0.0106)	*(0.0111)*	*(0.0135)*	*(0.00777)*	*(0.00858)*	*(0.0129)*	*(0.00854)*	*(0.00877)*	*(0.0130)*
Cashout Refi	-0.127***	-0.122***	-0.117***	0.101***	0.124***	0.138***	0.160***	0.174***	0.203***
	(0.0122)	*(0.0128)*	*(0.0155)*	*(0.00747)*	*(0.00820)*	*(0.0121)*	*(0.00946)*	*(0.00988)*	*(0.0141)*
Unemployment Rate	0.0921***	0.0847***	0.104***	0.00811***	0.0116***	0.0291***	0.0158***	0.0146***	0.0164***
	(0.00366)	*(0.00352)*	*(0.00479)*	*(0.00125)*	*(0.00134)*	*(0.00302)*	*(0.00157)*	*(0.00163)*	*(0.00348)*
PMI	0.676***	0.609***	0.523***	-0.0582**	-0.0473*	-0.0509	0.0277*	0.0284*	0.0875***
	(0.0359)	*(0.0385)*	*(0.0456)*	*(0.0184)*	*(0.0203)*	*(0.0335)*	*(0.0110)*	*(0.0121)*	*(0.0189)*
Transfer	0.284***	0.214***	0.267***	0.179***	0.134***	0.120***	0.0701*	0.0253	0.0961**
	(0.0141)	*(0.0151)*	*(0.0180)*	*(0.00636)*	*(0.00699)*	*(0.0101)*	*(0.0274)*	*(0.0308)*	*(0.0367)*
Option-ARM	0.527***	0.604***	0.599***	0.726***	0.880***	0.834***			
	(0.0213)	*(0.0222)*	*(0.0262)*	*(0.0380)*	*(0.0408)*	*(0.0606)*			
Interest-Only	0.884***	0.908***	0.879***	0.414***	0.430***	0.445***			
	(0.0109)	*(0.0115)*	*(0.0138)*	*(0.00938)*	*(0.0102)*	*(0.0131)*			

Table 8 (Panel B): Other Fixed-Rate Mortgages

	Jumbo FRM			Subprime FRM			Lowdoc Prime FRM		
	Base Case	No Early Default	Top 25 MSA	Base Case	No Early Default	Top 25 MSA	Base Case	No Early Default	Top 25 MSA
Orig. 2004	0.330***	0.552***	0.566***	0.274***	0.234***	0.284**	0.714***	0.534***	0.462***
	(0.0738)	*(0.0850)*	*(0.101)*	*(0.0472)*	*(0.0573)*	*(0.0870)*	*(0.0786)*	*(0.0892)*	*(0.105)*
Orig. 2005	1.132***	1.102***	1.114***	0.522***	0.293***	0.355***	0.882***	0.586***	0.433***
	(0.0662)	*(0.0708)*	*(0.0867)*	*(0.0323)*	*(0.0370)*	*(0.0523)*	*(0.0776)*	*(0.0899)*	*(0.113)*
Orig. 2006	1.536***	1.566***	1.512***	0.768***	0.466***	0.482***	0.924***	0.651***	0.607***
	(0.0621)	*(0.0664)*	*(0.0850)*	*(0.0305)*	*(0.0359)*	*(0.0533)*	*(0.0784)*	*(0.0881)*	*(0.112)*
Orig. 2007	2.357***	2.557***	2.419***	1.162***	1.075***	1.118***	1.693***	1.555***	1.457***
	(0.0607)	*(0.0633)*	*(0.0785)*	*(0.0261)*	*(0.0294)*	*(0.0431)*	*(0.0658)*	*(0.0728)*	*(0.0843)*
Securitized-2003	-0.340***	-0.408***	-0.402***	-0.0395	-0.131***	-0.236***	-0.246**	-0.0866	-0.00332
	(0.0485)	*(0.0508)*	*(0.0598)*	*(0.0360)*	*(0.0390)*	*(0.0579)*	*(0.0813)*	*(0.0825)*	*(0.0825)*
Securitized-2004	-0.0461	-0.189*	-0.215*	0.0303	0.100	0.109	-0.398***	-0.104	-0.0158
	(0.0613)	*(0.0737)*	*(0.0876)*	*(0.0453)*	*(0.0554)*	*(0.0844)*	*(0.0572)*	*(0.0656)*	*(0.0839)*
Securitized-2005	-0.131**	-0.0296	-0.0381	0.0323	0.208***	0.225***	-0.350***	-0.00165	0.119
	(0.0466)	*(0.0507)*	*(0.0622)*	*(0.0287)*	*(0.0333)*	*(0.0463)*	*(0.0536)*	*(0.0654)*	*(0.0922)*
Securitized-2006	0.134**	0.221***	0.267***	0.107***	0.359***	0.397***	-0.217***	0.0836	0.0777
	(0.0429)	*(0.0469)*	*(0.0540)*	*(0.0258)*	*(0.0313)*	*(0.0461)*	*(0.0542)*	*(0.0624)*	*(0.0886)*
Securitized-2007	-0.163***	-0.166***	-0.116***	-0.0907***	-0.0249	-0.0158	-0.493***	-0.270***	-0.275***
	(0.0198)	*(0.0213)*	*(0.0259)*	*(0.0198)*	*(0.0228)*	*(0.0331)*	*(0.0344)*	*(0.0386)*	*(0.0497)*
GSE-2003				-0.554***	-0.602***	-0.451***	-0.278***	-0.218***	-0.249***
				(0.0846)	*(0.0918)*	*(0.114)*	*(0.0599)*	*(0.0655)*	*(0.0722)*
GSE-2004				0.335***	0.409***	0.399***	-0.508***	-0.253***	-0.253**
				(0.0478)	*(0.0578)*	*(0.0885)*	*(0.0539)*	*(0.0634)*	*(0.0804)*
GSE-2005				0.169***	0.328***	0.269***	-0.555***	-0.215***	-0.116
				(0.0313)	*(0.0360)*	*(0.0514)*	*(0.0518)*	*(0.0637)*	*(0.0897)*
GSE-2006				0.134***	0.365***	0.357***	-0.618***	-0.294***	-0.337***
				(0.0280)	*(0.0335)*	*(0.0499)*	*(0.0519)*	*(0.0600)*	*(0.0864)*
GSE-2007				0.0755***	0.144***	0.154***	-0.884***	-0.644***	-0.629***
				(0.0211)	*(0.0243)*	*(0.0352)*	*(0.0279)*	*(0.0321)*	*(0.0423)*
No Early Default	N	Y	Y	N	Y	Y	N	Y	Y
N	7850509	7682202	5387952	4724547	4405820	2093203	14438228	14001753	6995468

Table 9 (Panel A): 5-year Prime ARMs

	Base Case	No Early Default	No Option-ARM or IO		Top 25 MSA	
				No Early Default		Dummies
FICO at origination	-0.00974***	-0.00964***	-0.0122***	-0.0122***	-0.00967***	-0.00969***
	(0.0000605)	*(0.0000627)*	*(0.000138)*	*(0.000143)*	*(0.0000836)*	*(0.0000823)*
Loan Amount	0.000000239***	0.000000229***	-7.92e-09	-5.73e-08	0.000000168***	0.000000198***
	(1.18e-08)	*(1.25e-08)*	*(5.10e-08)*	*(5.65e-08)*	*(1.57e-08)*	*(1.53e-08)*
Jumbo	-0.0125	-0.0297***	-0.153***	-0.154***	-0.0486***	-0.0104
	(0.00865)	*(0.00891)*	*(0.0269)*	*(0.0288)*	*(0.0110)*	*(0.0113)*
Low-doc	-0.0422***	-0.0520***	-0.0591***	-0.0615***	-0.0539***	-0.0532***
	(0.00696)	*(0.00725)*	*(0.0154)*	*(0.0162)*	*(0.00931)*	*(0.00934)*
Broker	0.194***	0.186***	0.195***	0.202***	0.156***	0.170***
	(0.00724)	*(0.00756)*	*(0.0169)*	*(0.0177)*	*(0.00974)*	*(0.00978)*
Correspondent	-0.0183*	-0.0260**	-0.0385**	-0.0349*	-0.0147	-0.00982
	(0.00764)	*(0.00790)*	*(0.0149)*	*(0.0155)*	*(0.0103)*	*(0.0104)*
Prepayment Penalty	0.216***	0.190***	0.151***	0.116**	0.140***	0.0981***
	(0.0115)	*(0.0121)*	*(0.0335)*	*(0.0356)*	*(0.0160)*	*(0.0164)*
LTV at Orig. (<80%)	0.0222***	0.0212***	-0.00235*	0.00163	0.0224***	0.0228***
	(0.00136)	*(0.00122)*	*(0.000998)*	*(0.00105)*	*(0.00139)*	*(0.00139)*
LTV at Orig. (=80%)	0.02294***	0.02181***	-0.00134	0.002538*	0.0230***	0.0236***
	(0.00134)	*(0.00120)*	*(0.000948)*	*(0.000995)*	*(0.00136)*	*(0.00138)*
LTV at Orig. (>80%)	0.0171***	0.0164***	0.00610***	-0.00274**	0.0177***	0.0183***
	(0.00133)	*(0.00118)*	*(0.000929)*	*(0.000976)*	*(0.00139)*	*(0.00139)*
Current LTV	1.404***	1.688***	3.218***	3.117***	1.625***	1.543***
	(0.119)	*(0.103)*	*(0.0580)*	*(0.0594)*	*(0.117)*	*(0.119)*
Initial Interest Rate	0.651***	0.662***	0.412***	0.417***	0.701***	0.688***
	(0.00539)	*(0.00584)*	*(0.0112)*	*(0.0120)*	*(0.00804)*	*(0.00774)*
Margin	-0.178***	-0.167***	-0.195***	-0.197***	-0.164***	-0.159***
	(0.00570)	*(0.00569)*	*(0.00982)*	*(0.0104)*	*(0.00711)*	*(0.00684)*
HPI appreciation	0.633***	0.615***	0.462***	0.489***	0.604***	0.315***
(4-years prior to orig)	*(0.0194)*	*(0.0181)*	*(0.0230)*	*(0.0243)*	*(0.0186)*	*(0.0434)*
Refinancing	-0.0134	0.00242	0.0353*	0.0594***	-0.0123	-0.000275
	(0.00712)	*(0.00738)*	*(0.0148)*	*(0.0155)*	*(0.00974)*	*(0.00950)*
Cashout Refi	-0.0281**	-0.0241*	0.141***	0.151***	-0.0234	-0.0250
	(0.00980)	*(0.0102)*	*(0.0214)*	*(0.0223)*	*(0.0133)*	*(0.0131)*
Unemployment Rate	0.0810***	0.0703***	0.0568***	0.0570***	0.0896***	0.129***
	(0.00392)	*(0.00374)*	*(0.00379)*	*(0.00397)*	*(0.00528)*	*(0.00587)*
PMI	0.341***	0.281***	0.566***	0.580***	0.247***	0.206***
	(0.0187)	*(0.0205)*	*(0.0377)*	*(0.0405)*	*(0.0389)*	*(0.0347)*
Transfer	0.446***	0.465***	0.741***	0.819***	0.472***	0.461***
	(0.00825)	*(0.00860)*	*(0.0233)*	*(0.0241)*	*(0.0112)*	*(0.0112)*
Option-ARM	-0.128***	-0.143***			-0.122***	-0.109***
	(0.00931)	*(0.00974)*			*(0.0120)*	*(0.0120)*
IO	0.108***	0.121***			0.121***	0.109***
	(0.00768)	*(0.00790)*			*(0.0103)*	*(0.0101)*

Table 9 (Panel B): 5-year Prime ARMs

	Base Case	No Early Default	No Option-ARM or IO	No Option-ARM or IO No Early Default	Top 25 MSA	Top 25 MSA Dummies
Orig. 2004	0.291***	0.309***	0.206***	0.268***	0.283***	0.286***
	(0.0259)	*(0.0270)*	*(0.0375)*	*(0.0397)*	*(0.0338)*	*(0.0338)*
Orig. 2005	0.851***	0.842***	0.886***	0.945***	0.770***	0.791***
	(0.0291)	*(0.0294)*	*(0.0467)*	*(0.0495)*	*(0.0371)*	*(0.0367)*
Orig. 2006	1.309***	1.308***	1.082***	1.160***	1.284***	1.315***
	(0.0341)	*(0.0336)*	*(0.0525)*	*(0.0557)*	*(0.0419)*	*(0.0417)*
Orig. 2007	2.149***	2.116***	1.634***	1.743***	2.094***	2.036***
	(0.0436)	*(0.0420)*	*(0.0625)*	*(0.0679)*	*(0.0493)*	*(0.0501)*
Securitized-2003	0.210***	0.188***	0.543***	0.565***	0.133*	0.128*
	(0.0474)	*(0.0498)*	*(0.0572)*	*(0.0609)*	*(0.0640)*	*(0.0640)*
Securitized-2004	0.337***	0.344***	0.443***	0.450***	0.350***	0.352***
	(0.0181)	*(0.0186)*	*(0.0369)*	*(0.0384)*	*(0.0236)*	*(0.0236)*
Securitized-2005	0.264***	0.269***	0.199***	0.225***	0.306***	0.309***
	(0.0135)	*(0.0140)*	*(0.0399)*	*(0.0419)*	*(0.0187)*	*(0.0187)*
Securitized-2006	0.248***	0.235***	0.288***	0.317***	0.234***	0.231***
	(0.0140)	*(0.0142)*	*(0.0426)*	*(0.0448)*	*(0.0177)*	*(0.0175)*
Securitized-2007	0.0730***	0.108***	0.327***	0.351***	0.0932***	0.126***
	(0.0185)	*(0.0199)*	*(0.0660)*	*(0.0724)*	*(0.0245)*	*(0.0246)*
GSE 2003	0.363***	0.349***	0.327***	0.374***	0.294***	0.294***
	(0.0273)	*(0.0285)*	*(0.0347)*	*(0.0370)*	*(0.0368)*	*(0.0368)*
GSE 2004	0.409***	0.406***	0.421***	0.443***	0.392***	0.401***
	(0.0169)	*(0.0174)*	*(0.0268)*	*(0.0279)*	*(0.0226)*	*(0.0226)*
GSE 2005	0.147***	0.141***	-0.100**	-0.0827*	0.189***	0.191***
	(0.0144)	*(0.0150)*	*(0.0373)*	*(0.0392)*	*(0.0203)*	*(0.0203)*
GSE 2006	0.0883***	0.0746***	-0.105*	-0.0756	0.0553**	0.0468*
	(0.0151)	*(0.0153)*	*(0.0415)*	*(0.0436)*	*(0.0194)*	*(0.0192)*
GSE 2007	0.00832	0.0396*	-0.130*	-0.0911	0.0145	0.0422
	(0.0185)	*(0.0199)*	*(0.0647)*	*(0.0705)*	*(0.0252)*	*(0.0251)*
N	19628285	19238238	8275802	8135845	11514164	11514164
No Early Default	N	Y	N	Y	Y	Y
Top MSA Subsample	-	-	-	-	Y	Y
MSA Dummies	-	-	-	-	N	Y

Table 10 (Panel A): 3-year and 2/28 Prime ARMs

	3-yr ARMs			2/28 ARMs		
	Base Case	**No Early Default**	**Top 25 MSA Dummies**	**Base Case**	**No Early Default**	**Top 25 MSA Dummies**
FICO at origination	-0.00823***	-0.00830***	-0.00839***	-0.00276***	-0.00274***	-0.00234***
	(0.0000959)	*(0.0000995)*	*(0.000128)*	*(0.0000599)*	*(0.0000621)*	*(0.0000819)*
Loan Amount	0.000000251***	0.000000244***	0.000000240***	0.00000146***	0.00000144***	0.00000147***
	(2.99e-08)	*(2.98e-08)*	*(2.97e-08)*	*(3.58e-08)*	*(3.62e-08)*	*(4.93e-08)*
Jumbo	0.0507**	0.0451**	0.0808***	-0.102***	-0.102***	-0.0870***
	(0.0165)	*(0.0168)*	*(0.0197)*	*(0.0129)*	*(0.0132)*	*(0.0161)*
Low-doc	0.00626	-0.0159	-0.0308	0.185***	0.151***	0.167***
	(0.0118)	*(0.0123)*	*(0.0162)*	*(0.0102)*	*(0.0107)*	*(0.0138)*
Broker	0.194***	0.203***	0.185***	0.179***	0.246***	0.364***
	(0.0137)	*(0.0142)*	*(0.0190)*	*(0.0400)*	*(0.0427)*	*(0.0581)*
Correspondent	0.0202	0.0220	0.0398*	0.168***	0.277***	0.324***
	(0.0139)	*(0.0144)*	*(0.0193)*	*(0.0366)*	*(0.0389)*	*(0.0525)*
Prepayment Penalty	0.134***	0.105***	0.104***	-0.243***	-0.327***	-0.309***
	(0.0171)	*(0.0179)*	*(0.0239)*	*(0.0106)*	*(0.0112)*	*(0.0144)*
LTV at Orig. (<80%)	-0.0179***	-0.0178***	-0.0164***	-0.0186***	-0.0184***	-0.0227***
	(0.00115)	*(0.00104)*	*(0.00200)*	*(0.00121)*	*(0.00117)*	*(0.000921)*
LTV at Orig. (=80%)	-0.0145***	-0.0143***	-0.0125***	-0.0153***	-0.0152***	-0.0192***
	(0.00117)	*(0.00107)*	*(0.00201)*	*(0.00117)*	*(0.00113)*	*(0.000879)*
LTV at Orig. (>80%)	-0.0184***	-0.0182***	-0.0169***	-0.0173***	-0.0171***	-0.0212***
	(0.00113)	*(0.00103)*	*(0.00200)*	*(0.00112)*	*(0.00108)*	*(0.000812)*
Current LTV	1.934***	1.911***	1.774***	1.515***	1.496***	1.837***
	(0.152)	*(0.145)*	*(0.220)*	*(0.0854)*	*(0.0834)*	*(0.0546)*
Initial Interest Rate	0.243***	0.237***	0.223***	0.243***	0.247***	0.249***
	(0.00577)	*(0.00600)*	*(0.00782)*	*(0.00334)*	*(0.00353)*	*(0.00488)*
Margin	-0.00717	-0.00128	0.0105	0.00773**	0.00509	0.0121***
	(0.00476)	*(0.00497)*	*(0.00639)*	*(0.00257)*	*(0.00266)*	*(0.00356)*
HPI appreciation	0.357***	0.389***	0.601***	0.124***	0.154***	0.621***
(4-years prior to orig)	*(0.0210)*	*(0.0213)*	*(0.0949)*	*(0.0121)*	*(0.0125)*	*(0.0536)*
Refinancing	-0.160***	-0.163***	-0.195***	-0.190***	-0.184***	-0.199***
	(0.0112)	*(0.0117)*	*(0.0150)*	*(0.00749)*	*(0.00779)*	*(0.0101)*
Cashout Refi	0.00502	0.0112	0.0385	0.00171	0.0141	-0.00181
	(0.0161)	*(0.0166)*	*(0.0216)*	*(0.0151)*	*(0.0157)*	*(0.0201)*
Unemployment Rate	-0.0108*	-0.0114*	-0.0272***	0.0250***	0.0241***	0.0212***
	(0.00503)	*(0.00494)*	*(0.00711)*	*(0.00233)*	*(0.00235)*	*(0.00387)*
PMI	0.493***	0.484***	0.489***	0.0253	0.0347	0.0708*
	(0.0212)	*(0.0225)*	*(0.0290)*	*(0.0249)*	*(0.0261)*	*(0.0291)*
Transfer	0.267***	0.291***	0.287***	0.293***	0.426***	0.493***
	(0.0155)	*(0.0160)*	*(0.0213)*	*(0.0254)*	*(0.0273)*	*(0.0361)*

Table 10 (Panel B): 3-year and 2/28 Prime ARMs

	3-yr ARMs			2/28 ARMs		
	Base Case	**No Early Default**	**Top 25 MSA Dummies**	**Base Case**	**No Early Default**	**Top 25 MSA Dummies**
Option-ARM	0.0841***	0.0866***	0.145***	0.125***	0.0910*	0.0692
	(0.0202)	*(0.0211)*	*(0.0266)*	*(0.0362)*	*(0.0382)*	*(0.0521)*
IO	0.347***	0.365***	0.354***	0.143***	0.145***	0.0886***
	(0.0158)	*(0.0162)*	*(0.0180)*	*(0.00734)*	*(0.00756)*	*(0.00968)*
Orig. 2004	0.420***	0.405***	0.408***	0.668***	0.289***	0.411***
	(0.0333)	*(0.0344)*	*(0.0429)*	*(0.0485)*	*(0.0538)*	*(0.0788)*
Orig. 2005	0.872***	0.867***	0.857***	0.697***	0.533***	0.629***
	(0.0389)	*(0.0401)*	*(0.0514)*	*(0.0529)*	*(0.0546)*	*(0.0786)*
Orig. 2006	1.775***	1.783***	1.911***	0.811***	0.671***	0.825***
	(0.0468)	*(0.0481)*	*(0.0645)*	*(0.0431)*	*(0.0445)*	*(0.0647)*
Orig. 2007	2.120***	2.070***	2.308***	0.745***	0.547***	0.710***
	(0.0496)	*(0.0515)*	*(0.0730)*	*(0.0479)*	*(0.0502)*	*(0.0736)*
Securitized-2003	0.180**	0.106	0.150	-0.240***	-0.244***	-0.109
	(0.0651)	*(0.0692)*	*(0.0894)*	*(0.0425)*	*(0.0444)*	*(0.0687)*
Securitized-2004	0.633***	0.669***	0.691***	-0.407***	-0.0986*	-0.0723
	(0.0237)	*(0.0245)*	*(0.0309)*	*(0.0365)*	*(0.0424)*	*(0.0578)*
Securitized-2005	0.527***	0.543***	0.601***	0.0439	0.0898*	0.102
	(0.0248)	*(0.0256)*	*(0.0339)*	*(0.0390)*	*(0.0402)*	*(0.0539)*
Securitized-2006	0.165***	0.177***	0.183***	0.0254	0.0160	-0.00532
	(0.0232)	*(0.0245)*	*(0.0329)*	*(0.0202)*	*(0.0210)*	*(0.0282)*
Securitized-2007	0.303***	0.383***	0.350***	0.199***	0.285***	0.355***
	(0.0475)	*(0.0505)*	*(0.0613)*	*(0.0462)*	*(0.0484)*	*(0.0651)*
GSE 2003	0.288***	0.281***	0.320***			
	(0.0425)	*(0.0439)*	*(0.0562)*			
GSE 2004	0.427***	0.462***	0.475***			
	(0.0225)	*(0.0232)*	*(0.0298)*			
GSE 2005	0.211***	0.211***	0.242***			
	(0.0268)	*(0.0277)*	*(0.0374)*			
GSE 2006	-0.222***	-0.202***	-0.205***			
	(0.0309)	*(0.0323)*	*(0.0442)*			
GSE 2007	-0.386***	-0.268***	-0.378***			
	(0.0543)	*(0.0575)*	*(0.0791)*			
N	4668454	4560519	2579919	5052194	4833211	2652468
No Early Default	N	Y	Y	N	Y	Y

Table 11: Subprime ARMs (Panel A)

	5-yr ARMs		3-yr ARMs		2/28 ARMs	
	Base Case	MSA Subset - No Early Default	Base Case	MSA Subset - No Early Default	Base Case	MSA Subset - No Early Default
FICO (origination)	-0.00569***	-0.00551***	-0.00426***	-0.00360***	-0.00412***	-0.00369***
	(0.000296)	*(0.000419)*	*(0.0000715)*	*(0.000102)*	*(0.0000421)*	*(0.0000678)*
Loan Amount	0.00000107***	0.000000912***	0.00000126***	0.00000132***	0.00000123***	0.00000106***
	(1.28e-07)	*(1.44e-07)*	*(4.97e-08)*	*(6.39e-08)*	*(4.14e-08)*	*(3.64e-08)*
Jumbo	-0.105	-0.101	-0.163***	-0.174***	-0.104***	-0.0964***
	(0.0580)	*(0.0682)*	*(0.0177)*	*(0.0221)*	*(0.0133)*	*(0.0133)*
Low-doc	0.363***	0.330***	0.149***	0.0921***	0.183***	0.115***
	(0.0371)	*(0.0514)*	*(0.0184)*	*(0.0252)*	*(0.00666)*	*(0.00975)*
Broker	0.365***	0.295	0.141***	0.109***	0.429***	0.376***
	(0.105)	*(0.158)*	*(0.00913)*	*(0.0135)*	*(0.00750)*	*(0.0116)*
Correspondent	0.446***	0.454**	0.0424***	0.0162	0.293***	0.259***
	(0.110)	*(0.162)*	*(0.0103)*	*(0.0153)*	*(0.00736)*	*(0.0117)*
Prepayment Penalty	0.0388	-0.00935	0.0425***	0.0157	-0.115***	-0.113***
	(0.0536)	*(0.0731)*	*(0.00781)*	*(0.0115)*	*(0.00538)*	*(0.00802)*
LTV at Orig. (<80%)	-0.000607	-0.00324	-0.0332***	-0.0348***	-0.00263***	-0.0226***
	(0.00289)	*(0.00410)*	*(0.000520)*	*(0.000808)*	*(0.000154)*	*(0.000895)*
LTV at Origination (=80%)	0.0345	-0.120	-2.444***	-2.544***	-0.000958	-1.607***
	(0.218)	*(0.311)*	*(0.0404)*	*(0.0627)*	*(0.0114)*	*(0.0669)*
LTV at Orig. (>80%)	-0.00350	-0.00502	-0.0316***	-0.0332***	-0.00158***	-0.0225***
	(0.00252)	*(0.00362)*	*(0.000479)*	*(0.000749)*	*(0.000117)*	*(0.000766)*
Current LTV	2.781***	2.914***	2.422***	2.547***	0.115***	2.829***
	(0.148)	*(0.222)*	*(0.0367)*	*(0.0574)*	*(0.0128)*	*(0.0377)*
Initial Interest Rate	0.253***	0.273***	0.243***	0.271***	0.188***	0.201***
	(0.0145)	*(0.0216)*	*(0.00356)*	*(0.00547)*	*(0.00209)*	*(0.00322)*
Margin	-0.0386*	-0.0861***	0.0498***	0.0656***	0.0321***	0.0294***
	(0.0181)	*(0.0249)*	*(0.00471)*	*(0.00714)*	*(0.00269)*	*(0.00434)*
HPI Appreciation	-0.000969	-0.0169	-0.00207	0.0565**	0.185***	0.137***
(4-years prior to orig.)	*(0.0538)*	*(0.0718)*	*(0.0133)*	*(0.0184)*	*(0.00804)*	*(0.0109)*
Refinancing	-0.193***	-0.191**	-0.231***	-0.233***	-0.250***	-0.235***
	(0.0525)	*(0.0711)*	*(0.0138)*	*(0.0207)*	*(0.00620)*	*(0.0100)*
Cashout Refi	-0.0515	-0.0119	-0.0323*	-0.0268	0.0236***	0.0129
	(0.0532)	*(0.0724)*	*(0.0138)*	*(0.0206)*	*(0.00680)*	*(0.0105)*
Unemployment Rate	0.0161	0.0164	-0.00693***	-0.0145***	0.0600***	0.0102***
	(0.00946)	*(0.0158)*	*(0.00203)*	*(0.00343)*	*(0.00103)*	*(0.00223)*
PMI	0.259	0.326	-0.204***	-0.281***	-0.0607***	-0.0376*
	(0.151)	*(0.205)*	*(0.0376)*	*(0.0624)*	*(0.0102)*	*(0.0182)*
Transfer	0.226*	0.374**	0.135***	0.151***	0.370***	0.408***
	(0.0901)	*(0.135)*	*(0.0106)*	*(0.0156)*	*(0.00684)*	*(0.0107)*
Option-ARM	-0.239**	-0.0707	-0.0418*	0.0118	-0.175***	-0.0834***
	(0.0751)	*(0.107)*	*(0.0191)*	*(0.0265)*	*(0.00758)*	*(0.0112)*
IO	-0.173***	-0.121*	0.0408***	0.0436***	0.125***	0.0555***
	(0.0369)	*(0.0483)*	*(0.00849)*	*(0.0117)*	*(0.00547)*	*(0.00789)*

Table 11: Subprime ARMs (Panel B)

	5-yr ARMs		3-yr ARMs		2/28 ARMs	
	Base Case	MSA Subset - No Early Default	Base Case	MSA Subset - No Early Default	Base Case	MSA Subset - No Early Default
Orig. 2004	-0.355	0.0930	0.836***	0.883***	0.423***	0.264***
	(0.417)	*(0.484)*	*(0.109)*	*(0.169)*	*(0.0213)*	*(0.0347)*
Orig. 2005	-0.875***	-0.418	1.160***	0.989***	0.760***	0.376***
	(0.161)	*(0.231)*	*(0.0662)*	*(0.110)*	*(0.0184)*	*(0.0292)*
Orig. 2006			1.102***	0.983***	1.418***	0.562***
			(0.0598)	*(0.104)*	*(0.0234)*	*(0.0362)*
Orig. 2007	-0.589***	0.00523	1.063***	1.064***	1.280***	0.581***
	(0.156)	*(0.219)*	*(0.0711)*	*(0.118)*	*(0.0305)*	*(0.0516)*
GSE 2003	-3.170***	-2.649***	1.915*	2.237**		
	(0.182)	*(0.253)*	*(0.790)*	*(0.808)*		
GSE 2004	0.809	0.954	0.597	1.695***		
	(1.163)	*(1.207)*	*(0.544)*	*(0.372)*		
GSE 2005	-40.31***	-44.58	-0.254	0.270		
	(0.479)	*(0)*	*(0.470)*	*(0.207)*		
GSE 2006			4.711***			
			(0.0446)			
GSE 2007	-0.714**	-0.546	-0.413	-0.175		
	(0.259)	*(0.381)*	*(0.225)*	*(0.294)*		
Securitized-2003	-2.974***	-2.489***	0.347**	0.492**	0.301***	0.0582
	(0.176)	*(0.237)*	*(0.116)*	*(0.183)*	*(0.0385)*	*(0.0735)*
Securitized-2004	-1.140**	-1.055*	-0.0888	-0.0103	-0.0449**	0.0232
	(0.406)	*(0.460)*	*(0.0992)*	*(0.146)*	*(0.0172)*	*(0.0283)*
Securitized-2005	-0.226	-0.148	-0.239***	-0.0441	-0.0989***	0.000492
	(0.122)	*(0.173)*	*(0.0455)*	*(0.0636)*	*(0.0122)*	*(0.0198)*
Securitized-2006	-0.703***	-0.151	-0.0814*	0.0636	-0.500***	-0.0838**
	(0.111)	*(0.162)*	*(0.0341)*	*(0.0505)*	*(0.0195)*	*(0.0293)*
Securitized-2007	0.190	0.133	0.0262	0.126	-0.299***	0.0258
	(0.125)	*(0.173)*	*(0.0601)*	*(0.0858)*	*(0.0295)*	*(0.0499)*
N	187651	112985	2001750	1011367	12529196	5759399

Table 12: Two-year Delinquency Rates and Securitization Coefficients

	Two-Year Delinquency Rate (2006 Vintage)	Coefficient on Private Securitization (2006 Vintage)
Prime		
FRM	7.28%	0.055
Jumbo FRM	8.9%	0.221
Lowdoc FRM	9.75%	0.084
5-Year ARM	14.6%	0.235
3-Year ARM	28.6%	0.177
2/28 ARM	29.9%	0.016
Subprime		
FRM	41.4%	0.359
5-Year ARM	41.2%	-0.151
3-Year ARM	38.8%	0.064
2/28 ARMS	43.7%	-0.084

Table 13 (Panel A): Broker Originated Fixed-Rate Mortgages

	Prime FRM		Jumbo FRM		Low-doc FRM	
	Broker=1	Broker=0	Broker=1	Broker=0	Broker=1	Broker=0
FICO at origination	-0.0104***	-0.0127***	-0.00880***	-0.0116***	-0.00964***	-0.0115***
	(0.000125)	*(0.0000719)*	*(0.000201)*	*(0.000124)*	*(0.000131)*	*(0.0000692)*
Loan Amount	0.000000418	0.000000462***	-5.60E-08	0.000000146***	0.000000645***	0.000000394***
	(0.000000284)	*(3.25E-08)*	*(3.66E-08)*	*(2.24E-08)*	*(6.44E-08)*	*(4.46E-08)*
Jumbo	-0.220*	-0.374***			-0.229***	-0.380***
	(0.109)	*(0.0223)*			*(0.04)*	*(0.0274)*
Low-doc	0.0580***	0.0263**	-0.0279	-0.119***		
	(0.0161)	*(0.0086)*	*(0.0251)*	*(0.0174)*		
Correspondent		0.0386***		-0.0511***		0.0230**
		(0.00675)		*(0.0116)*		*(0.00787)*
Prepayment Penalty	0.215***	0.422***	0.334***	0.566***	0.0617	0.163***
	(0.0629)	*(0.0392)*	*(0.0277)*	*(0.0183)*	*(0.101)*	*(0.0334)*
LTV at Orig. (<80%)	0.00279	-0.0191***	0.0297***	0.0309***	-0.00648***	-0.00801***
	(0.00218)	*(0.000846)*	*(0.00165)*	*(0.00151)*	*(0.00118)*	*(0.000629)*
LTV at Orig. (=80%)	0.00307	-0.0166***	0.0309***	0.0318***	-0.00643***	-0.00728***
	(0.00204)	*(0.000763)*	*(0.00156)*	*(0.00146)*	*(0.00113)*	*(0.000595)*
LTV at Orig. (>80%)	-0.000688	-0.0195***	0.0210***	0.0235***	-0.00519***	-0.00548***
	(0.002)	*(0.000906)*	*(0.00171)*	*(0.00148)*	*(0.00104)*	*(0.000557)*
Current LTV	3.056***	3.117***	1.417***	1.214***	3.515***	3.151***
	(0.0868)	*(0.0439)*	*(0.052)*	*(0.11)*	*(0.0701)*	*(0.0375)*
Initial Interest Rate	0.574***	0.549***	0.750***	0.705***	0.637***	0.610***
	(0.0156)	*(0.00647)*	*(0.0177)*	*(0.0102)*	*(0.0189)*	*(0.0085)*
HPI Appreciation	0.185**	0.184***	0.274***	0.545***	0.0693*	0.158***
(4-yrs prior)	*(0.0574)*	*(0.0152)*	*(0.0377)*	*(0.0286)*	*(0.0297)*	*(0.0161)*
Refinancing	0.102***	0.00288	0.0364	0.188***	0.110***	0.0157
	(0.0174)	*(0.00912)*	*(0.0211)*	*(0.0133)*	*(0.0172)*	*(0.0102)*
Cashout refi	0.0583**	0.125***	-0.198***	-0.0531***	0.157***	0.175***
	(0.0184)	*(0.0101)*	*(0.0241)*	*(0.0152)*	*(0.0194)*	*(0.0114)*
Unemployment Rate	0.0357***	0.0315***	0.0702***	0.0944***	0.0191***	0.0128***
	(0.0031)	*(0.00161)*	*(0.00446)*	*(0.0044)*	*(0.00346)*	*(0.00186)*
PMI	0.372***	0.512***	0.613***	0.575***	0.0237	0.0176
	(0.0419)	*(0.0353)*	*(0.0981)*	*(0.0438)*	*(0.0236)*	*(0.0143)*

Table 13 (Panel B): Broker Originated Fixed-Rate Mortgages

	Prime FRM		Jumbo FRM		Low-doc FRM	
	Broker=1	Broker=0	Broker=1	Broker=0	Broker=1	Broker=0
Orig. 2004	0.217	0.363***	0.513*	0.489***	0.558***	0.408***
	(0.154)	*(0.0646)*	*(0.213)*	*(0.0929)*	*(0.162)*	*(0.112)*
Orig. 2005	0.0994	0.563***	0.497*	1.085***	0.667***	0.418***
	(0.134)	*(0.0632)*	*(0.196)*	*(0.0768)*	*(0.132)*	*(0.122)*
Orig. 2006	0.355**	0.473***	1.569***	1.462***	0.623***	0.533***
	(0.129)	*(0.0604)*	*(0.145)*	*(0.076)*	*(0.121)*	*(0.129)*
Orig. 2007	0.659***	0.608***	2.793***	2.590***	1.486***	1.449***
	(0.129)	*(0.0597)*	*(0.126)*	*(0.0726)*	*(0.107)*	*(0.101)*
Securitized 2003	-0.665***	-0.456***	-0.704***	-0.418***	-0.259*	0.0636
	(0.148)	*(0.0564)*	*(0.115)*	*(0.0566)*	*(0.12)*	*(0.0976)*
Securitized 2004	-0.214*	-0.278***	-0.25	-0.0878	-0.362*	-0.000313
	(0.105)	*(0.0463)*	*(0.191)*	*(0.0795)*	*(0.141)*	*(0.0723)*
Securitized 2005	0.248***	-0.383***	0.568***	-0.00844	-0.185	0.063
	(0.0682)	*(0.042)*	*(0.164)*	*(0.0532)*	*(0.103)*	*(0.0862)*
Securitized 2006	0.384***	-0.0567	0.577***	0.380***	0.157	0.0428
	(0.0563)	*(0.037)*	*(0.0883)*	*(0.0547)*	*(0.0801)*	*(0.0981)*
Securitized 2007	0.646***	0.356***	0.160***	-0.0685**	-0.0619	-0.410***
	(0.0558)	*(0.0401)*	*(0.0369)*	*(0.0262)*	*(0.0531)*	*(0.0565)*
GSE 2003	-0.133	-0.406***			-0.249*	-0.319***
	(0.122)	*(0.0517)*			*(0.0966)*	*(0.0893)*
GSE 2004	-0.0943	-0.544***			-0.236	-0.266***
	(0.0985)	*(0.0429)*			*(0.136)*	*(0.0703)*
GSE 2005	0.151*	-0.597***			-0.291**	-0.176*
	(0.0604)	*(0.0395)*			*(0.0959)*	*(0.0847)*
GSE 2006	-0.0507	-0.422***			-0.194*	-0.332***
	(0.0467)	*(0.0353)*			*(0.0758)*	*(0.0951)*
GSE 2007	0.300***	0.0764*			-0.437***	-0.717***
	(0.0438)	*(0.0324)*			*(0.0466)*	*(0.0476)*
N	3872913	24384157	1263913	6418289	2878852	11122901

Table 14 (Panel A): Broker Originated Prime ARMs

	5yr Prime		3yr Prime		2/28 Prime	
	Broker=1	Broker=0	Broker=1	Broker=0	Broker=1	Broker=0
FICO at origination	-0.00969***	-0.0100***	-0.00869***	-0.00810***	-0.00337***	-0.00268***
	(0.000144)	*(0.0000897)*	*(0.000233)*	*(0.00011)*	*(0.000225)*	*(0.0000649)*
Loan Amount	0.000000240***	0.000000179***	0.000000121	0.000000307***	0.00000111***	0.00000146***
	(2.50E-08)	*(1.96E-08)*	*(6.25E-08)*	*(3.68E-08)*	*(0.000000134)*	*(3.79E-08)*
Jumbo	-0.0886***	-0.0846***	0.0641	0.0353	-0.165**	-0.0902***
	(0.0199)	*(0.0137)*	*(0.0352)*	*(0.0196)*	*(0.0535)*	*(0.0136)*
Low-doc	0.0970***	-0.205***	-0.00755	-0.00174	0.0432	0.229***
	(0.0134)	*(0.0103)*	*(0.025)*	*(0.0145)*	*(0.0255)*	*(0.0121)*
Correspondent		-0.150***		0.0294*		0.338***
		(0.00984)		*(0.0146)*		*(0.0407)*
Prepayment Penalty	0.265***	0.159***	0.204***	0.0905***	0.861	-0.353***
	(0.0275)	*(0.0163)*	*(0.0569)*	*(0.019)*	*(1.073)*	*(0.0114)*
LTV at Orig. (<80%)	0.0171***	0.0111***	-0.00574**	-0.0173***	-0.0222***	-0.0175***
	(0.00121)	*(0.000746)*	*(0.00188)*	*(0.00111)*	*(0.00259)*	*(0.00117)*
LTV at Orig. (=80%)	0.0181***	0.0118***	-0.00399*	-0.0140***	-0.0210***	-0.0142***
	(0.00115)	*(0.000711)*	*(0.00179)*	*(0.00113)*	*(0.00247)*	*(0.00113)*
LTV at Orig. (>80%)	0.0124***	0.00676***	-0.00640***	-0.0173***	-0.0241***	-0.0162***
	(0.00125)	*(0.000675)*	*(0.00176)*	*(0.00111)*	*(0.00232)*	*(0.00108)*
Current LTV	2.332***	2.345***	3.417***	1.822***	4.172***	1.427***
	(0.0532)	*(0.0329)*	*(0.0998)*	*(0.147)*	*(0.177)*	*(0.0803)*
Initial Interest Rate	0.664***	0.609***	0.242***	0.243***	0.243***	0.239***
	(0.0133)	*(0.00922)*	*(0.0126)*	*(0.00673)*	*(0.0126)*	*(0.00371)*
Margin	-0.171***	-0.0589***	-0.218***	0.00703	0.0930***	0.00315
	(0.0124)	*(0.00886)*	*(0.0209)*	*(0.00515)*	*(0.0265)*	*(0.0027)*
HPI Appreciation	0.376***	0.576***	0.379***	0.394***	-0.09	0.185***
(4-yrs prior)	*(0.024)*	*(0.0161)*	*(0.0424)*	*(0.0233)*	*(0.054)*	*(0.013)*
Refinancing	0.150***	0.0182	0.0149	-0.192***	-0.232***	-0.174***
	(0.0151)	*(0.011)*	*(0.0255)*	*(0.0131)*	*(0.0228)*	*(0.00822)*
Cashout refi	-0.147***	-0.0853***	-0.0806*	0.0576**	1.223*	-0.0176
	(0.0183)	*(0.0143)*	*(0.0366)*	*(0.0189)*	*(0.497)*	*(0.016)*
Unemployment Rate	0.0623***	0.0460***	-0.0522***	-0.0118*	0.0267***	0.0230***
	(0.00393)	*(0.00242)*	*(0.00609)*	*(0.00502)*	*(0.00625)*	*(0.00233)*
PMI	0.287***	0.233***	0.135	0.436***	-37.46	0.0329
	(0.0655)	*(0.0247)*	*(0.072)*	*(0.0233)*	*(0)*	*(0.0251)*
Transfer		0.181***		0.286***		0.443***
		(0.0162)		*(0.0167)*		*(0.027)*
Option ARM	-0.0948***	-0.426***	0.250***	0.107***	21.02***	-0.0443
	(0.0165)	*(0.0156)*	*(0.0355)*	*(0.0293)*	*(0.668)*	*(0.0519)*
IO	0.0923***	0.157***	0.327***	0.348***	0.212	0.128***
	(0.0162)	*(0.0109)*	*(0.0279)*	*(0.0169)*	*(0.176)*	*(0.00755)*

Table 14 (Panel B): Broker Originated Prime ARMs

	5yr Prime		3yr Prime		2/28 Prime	
	Broker=1	Broker=0	Broker=1	Broker=0	Broker=1	Broker=0
Debt-to-Income Ratio		0.00966***				
		(0.000308)				
Orig. 2004	0.240***	0.367***	0.422***	0.384***	0.546***	0.315**
	(0.0504)	*(0.0412)*	*(0.0591)*	*(0.0431)*	*(0.0718)*	*(0.101)*
Orig. 2005	0.774***	0.831***	0.742***	0.841***	20.15	0.807***
	(0.056)	*(0.0416)*	*(0.068)*	*(0.0492)*	.	*(0.0909)*
Orig. 2006	1.181***	1.266***	1.549***	1.761***	20.75***	0.993***
	(0.0615)	*(0.0443)*	*(0.0793)*	*(0.0554)*	*(0.679)*	*(0.0857)*
Orig. 2007	2.194***	1.900***	2.180***	1.919***	21.08***	0.822***
	(0.0625)	*(0.0503)*	*(0.0823)*	*(0.0589)*	*(0.672)*	*(0.0904)*
Securitized 2003	-0.0253	-0.0474	0.355*	0.068	-0.243***	0.0274
	(0.126)	*(0.075)*	*(0.143)*	*(0.079)*	*(0.0487)*	*(0.155)*
Securitized 2004	0.414***	0.377***	0.707***	0.657***	-0.353***	0.169**
	(0.0382)	*(0.0265)*	*(0.0447)*	*(0.0305)*	*(0.0602)*	*(0.0618)*
Securitized 2005	0.197***	0.227***	0.280***	0.546***	-43.03	0.0981*
	(0.0365)	*(0.0198)*	*(0.054)*	*(0.0314)*	*(0)*	*(0.0403)*
Securitized 2006	0.133***	0.108***	-0.0864	0.161***	-0.436	-0.0138
	(0.0359)	*(0.0195)*	*(0.0741)*	*(0.0272)*	*(1.196)*	*(0.0213)*
Securitized 2007	-0.168***	0.221***	0.0654	0.447***	-42.69	0.309***
	(0.039)	*(0.0492)*	*(0.121)*	*(0.056)*	*(0)*	*(0.0508)*
GSE 2003	0.161*	-0.0529	0.249**	0.264***		
	(0.0731)	*(0.0492)*	*(0.0954)*	*(0.0507)*		
GSE 2004	0.426***	0.262***	0.430***	0.455***		
	(0.0325)	*(0.0237)*	*(0.0424)*	*(0.0286)*		
GSE 2005	0.128***	0.0516**	0.216***	0.203***		
	(0.0351)	*(0.0191)*	*(0.0491)*	*(0.0346)*		
GSE 2006	-0.0178	-0.0279	-0.177**	-0.210***		
	(0.0361)	*(0.019)*	*(0.0679)*	*(0.0375)*		
GSE 2007	-0.229***	0.0944*	-0.562***	-0.138		
	(0.033)	*(0.0377)*	*(0.0916)*	*(0.0752)*		
N	3613032	9296723	1053784	3506735	828741	4004470

Table 15 (Panel A): Broker Originated Subprime Mortgages

	5yr Subprime		3yr Subprime		2/28 Subprime		Subprime FRM	
	Broker=1	Broker=0	Broker=1	Broker=0	Broker=1	Broker=0	Broker=1	Broker=0
FICO at origination	-0.00254***	-0.00324***	-0.00380***	-0.00441***	-0.00228***	-0.00430***	-0.00515***	-0.00566***
	(0.000462)	*(0.000903)*	*(0.000134)*	*(0.0000927)*	*(0.000127)*	*(0.0000799)*	*(0.000125)*	*(0.00006)*
Loan Amount	0.00000165***	0.00000223***	0.00000119***	0.00000116***	0.00000114***	0.00000105***	0.00000143***	0.00000123***
	(0.000000211)	*(0.000000441)*	*(8.82E-08)*	*(6.48E-08)*	*(6.80E-08)*	*(4.32E-08)*	*(6.79E-08)*	*(6.03E-08)*
Jumbo	-0.298**	-0.38	-0.177***	-0.127***	-0.0997***	-0.0964***	-0.280***	-0.225***
	(0.0948)	*(0.198)*	*(0.031)*	*(0.0231)*	*(0.026)*	*(0.0155)*	*(0.0305)*	*(0.0212)*
Low-doc	0.280***	0.173	0.214***	0.0821***	0.147***	0.0683***	0.167***	0.0278*
	(0.0513)	*(0.117)*	*(0.0357)*	*(0.0243)*	*(0.0165)*	*(0.0129)*	*(0.0243)*	*(0.013)*
Correspondent		0.389*		0.0607***		0.252***		0.0730***
		(0.185)		*(0.0109)*		*(0.0121)*		*(0.00712)*
Prepayment Penalty	-0.177	-0.0685	-0.000312	0.0490***	-0.224***	-0.0903***	-0.197***	-0.195***
	(0.226)	*(0.125)*	*(0.0158)*	*(0.00982)*	*(0.0202)*	*(0.00881)*	*(0.0251)*	*(0.0113)*
LTV at Orig. (<80%)	0.00187	-0.00207	-0.0171***	-0.0326***	-0.0148***	-0.0221***	-0.00203	-0.00797***
	(0.00451)	*(0.00873)*	*(0.0014)*	*(0.000653)*	*(0.00139)*	*(0.000932)*	*(0.00115)*	*(0.000518)*
LTV at Orig. (=80%)	0.00267	-0.000847	-0.0156***	-0.0302***	-0.0124***	-0.0198***	-0.000553	-0.00717***
	(0.00427)	*(0.00817)*	*(0.00133)*	*(0.000634)*	*(0.00133)*	*(0.000869)*	*(0.00108)*	*(0.000487)*
LTV at Orig. (>80%)	-0.000435	-0.00535	-0.0198***	-0.0308***	-0.0157***	-0.0220***	-0.00432***	-0.00914***
	(0.00396)	*(0.00751)*	*(0.00121)*	*(0.0006)*	*(0.00123)*	*(0.0008)*	*(0.001)*	*(0.000453)*
Current LTV	2.408***	2.546***	3.030***	2.359***	2.996***	2.764***	2.560***	2.554***
	(0.213)	*(0.451)*	*(0.0717)*	*(0.046)*	*(0.0754)*	*(0.0436)*	*(0.0598)*	*(0.0315)*
Initial Interest Rate	0.261***	0.400***	0.255***	0.233***	0.193***	0.197***	0.248***	0.220***
	(0.0218)	*(0.0477)*	*(0.00756)*	*(0.00445)*	*(0.00666)*	*(0.00371)*	*(0.00575)*	*(0.00313)*
Margin	0.116*	0.0284	0.0913*	0.0632***	0.127***	0.0310***		
	(0.0586)	*(0.0627)*	*(0.0442)*	*(0.00525)*	*(0.022)*	*(0.00457)*		
HPI Appreciation	-0.0883	0.135	0.0875***	0.0439*	0.0880***	0.144***	0.137***	0.131***
(4-yrs prior)	*(0.0791)*	*(0.164)*	*(0.025)*	*(0.0173)*	*(0.0199)*	*(0.0129)*	*(0.0246)*	*(0.0128)*
Refinancing	-0.294***	-0.00298	-0.172***	-0.208***	-0.250***	-0.240***	-0.425***	-0.317***
	(0.0839)	*(0.18)*	*(0.0389)*	*(0.0159)*	*(0.0331)*	*(0.0106)*	*(0.0265)*	*(0.00915)*
Cashout refi	0.055	-0.201	-0.0426	-0.0175	-0.00804	0.0315**	0.132***	0.140***
	(0.0823)	*(0.173)*	*(0.0394)*	*(0.0158)*	*(0.0338)*	*(0.0113)*	*(0.0262)*	*(0.00866)*
Unemployment Rate	0.0273*	0.0329	-0.00632	-0.00895***	0.0275***	0.00346	0.00830**	0.0113***
	(0.0136)	*(0.0278)*	*(0.00407)*	*(0.00253)*	*(0.00439)*	*(0.00259)*	*(0.00301)*	*(0.0015)*
PMI	-0.12	-0.345	1.084	-0.204***		-0.0356	-0.373	-0.0836***
	(0)	*(0)*	*(0.727)*	*(0.0404)*		*(0.0186)*	*(0.21)*	*(0.0204)*
Transfer		-0.0853		0.181***		0.399***		
		(0.164)		*(0.0115)*		*(0.0112)*		
Option ARM	-0.0934	-0.408*	-0.0414	-0.0821	-0.0966***	-0.0708***		
	(0.151)	*(0.193)*	*(0.0314)*	*(0.0457)*	*(0.0189)*	*(0.0173)*		
IO	-0.0913	-0.126	0.0181	0.0576***	0.121***	0.0330***		
	(0.0558)	*(0.102)*	*(0.0152)*	*(0.0111)*	*(0.0166)*	*(0.00906)*		

Table 15 (Panel B): Broker Originated Subprime Mortgages

	5yr Subprime		3yr Subprime		2/28 Subprime		Subprime FRM	
	Broker=1	Broker=0	Broker=1	Broker=0	Broker=1	Broker=0	Broker=1	Broker=0
Debt-to-Income Ratio		0.00531						
		(0.00394)						
Orig. 2004			-0.568*	0.919***	0.404*	0.248***	0.0274	0.216***
			(0.27)	*(0.134)*	*(0.162)*	*(0.0398)*	*(0.176)*	*(0.0605)*
Orig. 2005	-0.558	18.29***	-0.378*	1.124***	0.454**	0.444***	0.161*	0.209***
	(0.511)	*(0.987)*	*(0.192)*	*(0.11)*	*(0.156)*	*(0.0429)*	*(0.0735)*	*(0.0516)*
Orig. 2006	0.00526	18.99***		1.085***	0.781***	0.580***	0.472***	0.532***
	(0.522)	*(0.965)*		*(0.0764)*	*(0.187)*	*(0.0391)*	*(0.126)*	*(0.0371)*
Orig. 2007	-0.0189	19.61	-0.306	0.971***	0.551**	0.692***	0.820***	1.103***
	(0.519)	.	*(0.213)*	*(0.088)*	*(0.168)*	*(0.0714)*	*(0.0716)*	*(0.0341)*
Securitized 2003			-41.26	0.422***		0.0631	-2.188*	-0.0714
			(0)	*(0.125)*		*(0.0744)*	*(1.024)*	*(0.039)*
Securitized 2004	-1.023	18.70***	0.111	-0.131	-0.117*	0.0459	-0.0389	0.153**
	(0.6)	*(0.819)*	*(0.198)*	*(0.125)*	*(0.0547)*	*(0.0344)*	*(0.168)*	*(0.0586)*
Securitized 2005	0.0924	0.713	0.00299	-0.285**	-0.0118	-0.051	0.0276	0.294***
	(0.154)	*(0.61)*	*(0.0606)*	*(0.097)*	*(0.023)*	*(0.0376)*	*(0.0477)*	*(0.0491)*
Securitized 2006	-0.0684	0.171	-0.224	-0.134*	-0.213*	-0.0870**	0.112	0.340***
	(0.162)	*(0.573)*	*(0.185)*	*(0.0541)*	*(0.105)*	*(0.0324)*	*(0.112)*	*(0.0323)*
Securitized 2007	0.0628	-0.0587	-0.00773	0.206*	0.0355	-0.041	0.0927*	-0.0211
	(0.166)	*(0.79)*	*(0.121)*	*(0.0803)*	*(0.0712)*	*(0.0713)*	*(0.0393)*	*(0.0283)*
GSE 2003		-20.94					-0.747*	-0.577***
		(0)					*(0.323)*	*(0.0982)*
GSE 2004	-40.76	-21.26					0.538**	0.412***
	(0)	*(0)*					*(0.175)*	*(0.0612)*
GSE 2005	-43.76	-43.66					0.415***	0.390***
	(0)	*(0)*					*(0.0817)*	*(0.0509)*
GSE 2006							0.284*	0.257***
							(0.12)	*(0.0345)*
GSE 2007	-43.05						0.184***	0.0672*
	(0)						*(0.0463)*	*(0.0294)*
N	49643	13434	588075	1341367	1297523	4461876	721616	3684204

Appendix E

Recourse and Residential Mortgage Default: Theory and Evidence from U.S. States*

Andra C. Ghent and Marianna Kudlyak†

Federal Reserve Bank of Richmond Working Paper No. 09-10

July 7th, 2009

Abstract

We analyze the impact of lender recourse on mortgage defaults theoretically and empirically across U.S. states. We study the effect of state laws regarding deficiency judgments in a model where lenders can use the threat of a deficiency judgment to deter default or to shorten the default process. Empirically, we find that recourse decreases the probability of default when there is a substantial likelihood that a borrower has negative home equity. We also find that, in states that allow deficiency judgments, defaults are more likely to occur through a lender-friendly procedure, such as a deed in lieu of foreclosure. (JEL: E44, G21, G28, K11, R20.)

Key Words: Deficiency Judgment. Foreclosure. Negative Equity. Residential Mortgage Default. Recourse.

---

*We thank Jay Weiser, Ned Prescott, Peter B. Ritz, and Paul Willen for helpful comments on an earlier draft. This paper has benefited from conversations with Patricia Antonelli, Roger Bear, Chris Cauble, Dale Cottam, Craig Doyle, Charles Just, Doug Hanoy, Bob Kroker, Lester Leu, Michael Manniello, Kenneth McFadyen, Richard Rothfuss, Sam Schulhofer-Wohl, Roman Sysuyev, Louis Vitti, Dave Weibel, and Randall Wharton. Anne Stilwell provided excellent research assistance. We also thank Mark A. Watson at the Federal Reserve Bank of Kansas City for his assistance with the LPS data. The views expressed here do not necessarily reflect those of the Federal Reserve Bank of Richmond or the Federal Reserve System. All errors and omissions are ours alone.

†Ghent: Dept. of Real Estate, Zicklin School of Business, Baruch College / CUNY; email andra.ghent@baruch.cuny.edu; phone 646-660-6929. Kudlyak: Research Department, Federal Reserve Bank of Richmond; email marianna.kudlyak@rich.frb.org; phone 804-697-8791.

I. Introduction

The recent surge in defaults on residential mortgages has renewed interest in understanding borrowers' decisions of whether to default and what factors influence that decision. One factor of interest is the recourse permitted to lenders. In some U.S. states, recourse in residential mortgages is limited to the value of the collateral securing the loan. In other U.S. states, the lender may be able to collect on debt not covered by the proceedings from a foreclosure sale by obtaining a deficiency judgment. Large increases in defaults in states that severely restrict lender recourse, such as California and Arizona, raise the question of whether allowing lenders more recourse substantially deters default.[1]

Existing literature usually models the default decision as a borrower exercising a default option when it is "in the money", i.e., when the borrower is in a negative equity situation.[2] Thus, if the lender has no recourse, even borrowers who do not experience a change in their income or mortgage payments, but who find themselves having substantial negative equity in their homes, will default on their mortgages. However, allowing the lender recourse to assets other than the mortgaged property lowers the value of the default option and thus reduces the borrower's incentive to default.

In this paper we explore the differences in recourse law across states to study the effect of recourse on residential mortgage default. We examine both how much recourse deters default and to what extent it changes how borrowers default. The effect of recourse on default is not clear a priori. Deficiency judgments may be rare in practice. This may be because it is often costly and time-consuming for a lender to pursue and collect on a deficiency judgment. Alternatively, the mere threat of a deficiency judgment may deter default implying few deficiency judgments in practice. Therefore, the number of deficiency judgments observed may not be a good indicator of their influence on borrowers' behavior.

We present a model in which lenders can use the threat of a deficiency judgment to get the borrower to agree to expedite the default process or to deter default altogether. The borrower first decides how to default and then, based on the expected payoff from default,

[1] As Martin Feldstein (2008) writes, "[t]he no-recourse mortgage is virtually unique to the United States. That's why falling house prices in Europe do not trigger defaults. The creditors' ability to go beyond the house to other assets or even future salary is a deterrent."

[2] See, for example, Kau, Keenan, Muller, and Epperson (1992) and Deng, Quigley, and Van Order (2000).

decides whether to default. In the subgame perfect equilibrium of the model, lenders rarely pursue deficiency judgments. However, allowing lenders recourse deters default in many situations. Further, recourse has an impact on how default happens: when the lender has recourse, defaults that do occur are likely to lead to smaller losses to lenders.

Following the solution of the theoretical model, we postulate empirical hypotheses and test them using a large sample of residential mortgages from the Lender Processing Services Inc. (LPS) Applied Analytics database. Our empirical findings are as follows:

1. Recourse has a negative effect on the probability of default when there is a substantial likelihood that a borrower has negative home equity (at high values of the default option). At the mean value of the default option at the time of default, the probability of default is 20% higher in states with no recourse as compared to states that allow recourse.

2. The magnitude of the deterrent effect of recourse on default depends on the borrower's wealth. The effect is significant only when the borrower is likely to have significant assets or income which we proxy for with the appraised value of the mortgaged property at origination. For borrowers with properties appraised at less than $200,000, there is no difference in the probability of default across recourse and non-recourse states. At the mean value of the default option at the time of default and for homes appraised at $300,000 to $500,000, borrowers in non-recourse states are 59% more likely to default than borrowers in recourse states. For homes appraised at $500,000 to $750,000, borrowers in non-recourse states are almost twice as likely to default as borrowers in recourse states while for homes appraised at $750,000 to $1 million, borrowers in non-recourse states are 66% more likely to default as borrowers in recourse states.

3. Recourse deters default on loans held privately. We cannot reject the hypothesis that recourse does not have an effect on loans held by government sponsored enterprises (GSEs).

4. Allowing the lender recourse increases the likelihood that default occurs by a more lender-friendly method, such as a deed in lieu, rather than foreclosure.

Our finding that recourse deters some borrowers from defaulting indicates that a non-negligible portion of U.S. mortgage default is in fact strategic rather than borrowers having no choice but to default because of liquidity constraints. This finding contrasts with the view that mortgage defaults are primarily driven by shocks to the borrower's ability to pay (see, for example, Foote, Gerardi, and Willen [2008]). Based on their analysis of Massachusetts data, Foote, Gerardi, and Willen (2008) conclude that negative equity is not a sufficient condition for default. However, Massachusetts is a recourse state and analyzing data only from recourse states gives an incomplete picture of the role of negative equity in the borrower's default decision. As our findings show, the borrower's decision to default in recourse states is substantially less sensitive to negative equity than in non-recourse states.

To our knowledge, ours is the first study looking at differences in how borrowers default. Earlier work by Clauretie (1987), Jones (1993), and Ambrose, Capone, and Deng (2001) has also looked empirically at differences in defaults across states.[3] Clauretie (1987) estimates a linear regression model of aggregate state default rates and finds that whether or not a state permits a deficiency judgment does not significantly affect the state's default rate. Jones (1993) looks at evidence from Alberta, which does not permit deficiency judgments, and British Columbia, which permits deficiency judgments, and finds that defaults in Alberta are more likely to be due to deliberate defaults, rather than trigger events in the borrower's life. Ambrose, Capone, and Deng (2001) include a dummy variable for whether a state allows a deficiency judgment in their study of the determinants of mortgage default in a sample of Federal Housing Administration (FHA) loans originated in 1989. Because the principal of FHA loans is guaranteed by the FHA, FHA lenders cannot seek a deficiency judgment such that FHA loans may be particularly poorly suited to studying the effect of recourse on default behavior.

Ambrose, Buttimer, and Capone (1997) study theoretically the effect of deficiency judgments on default and find that the probability of default is a decreasing function of the probability of obtaining a deficiency judgment. Our theoretical model builds on Ambrose, Buttimer, and Capone (1997) by exploring the interaction of recourse laws and the lengthiness of the foreclosure process but incorporates more fully the lender and borrower's incentives and the negotiation that goes on between them which determines how the borrower terminates the mortgage.

The remainder of the paper proceeds as follows: The next section describes how lender

[3] Pence (2006) does not directly study how recourse affects default rate; however, she looks at differences in average loan size in census tracts that span two states and finds that the average loan size is smaller in states with more defaulter-friendly foreclosure laws.

recourse varies across the US states. Section 3 presents a model of the negotiation between borrowers and lenders as a function of lender recourse, default costs, and homeowner equity. In section 4 we describe our data and variables. We present our empirical results in section 5. Section 6 concludes.

II. Foreclosure Law and Default

A. Foreclosure Law Across the U.S. States

States vary in the statutes governing how much recourse the lender has in the event the lender forecloses on the property and the proceeds from the foreclosure sale are not sufficient to cover the borrower's debt. States also differ markedly in how long it takes the lender to foreclose.

In most states, the lender may obtain a deficiency judgment to cover the difference between the balance owed and the value of the home in the event the lender must foreclose in a negative equity situation. However, in states that permit deficiency judgments, various restrictions often apply. Usually the lender must credit the borrower's account for fair market value of the property rather than the foreclosure sale price. The fair market value restriction is likely present because the lender is often the only bidder at the foreclosure sale (see, for example, Brueggeman and Fisher [2008]). In the absence of such a restriction, the lender could doubly profit from a foreclosure by bidding an artificially low price. In addition to lowering the likely recovery from a deficiency judgment, such restrictions sometimes imply that the lender must incur substantially higher legal costs and more time in pursuing a deficiency. The increase in costs and time depends on state statutes governing the determination of fair market value. In some states, a single appraiser determines fair market value. In other states, such as Minnesota, fair market value must be determined by a jury. Finally, states differ in how easy it is for the borrower to contest the fair market value of the property.

Lenders have less recourse in practice in states that require lenders to go through a lengthy judicial foreclosure process, rather than a quicker non-judicial foreclosure process, to obtain a deficiency judgment. In other states, such as Idaho and Nebraska, there is a relatively short time frame in which the lender can file. In practice, this can be onerous in states that also have a fair market value restriction since the lender does not immediately know the fair market value of the property and thus cannot determine whether there even is a deficiency to be sought. In some states, substantial personal property or wages are exempt from collection on the deficiency. For instance, Florida and Texas have nearly unlimited homestead exemptions such that the lender is very unlikely to collect on a deficiency judgment on an investment property or secondary residence since the borrower can easily shield his or her assets. Finally, in Ohio and Iowa, the lender has a relatively short period in which to collect on the deficiency after the foreclosure sale.

In states that allow deficiency judgments, a borrower retains the option to declare bankruptcy and have some portion or all of the deficiency judgment discharged. As White (1998) reports, prior to the 2005 bankruptcy reform, most unsecured debts were discharged in bankruptcy regardless of whether the borrower filed under chapter 7 or under chapter 13. Furthermore, filing for bankruptcy had a low pecuniary cost before the 2005 act such that the major cost to filing for bankruptcy was reduced availability of credit. In chapter 7 filings, it continues to be the case that deficiency judgments are completely discharged and, if the chapter 7 filing is concurrent with a foreclosure, the lender loses the right to a deficiency judgment. In chapter 13 filings, the lender may pursue a deficiency judgment. Following the 2005 bankruptcy reform, however, borrowers with incomes above the state median income must file under chapter 13, rather than chapter 7, which might make it more difficult to discharge a deficiency judgment for high income borrowers.

A few states explicitly forbid deficiency judgments on most homes (Arizona and Oregon) or on purchase mortgages. In other states, the restrictions on deficiency judgments are so onerous that it is highly impractical for the lender to pursue a judgment in the vast majority of cases which makes the state effectively non-recourse. Table 1 summarizes the extent of recourse the lender has in each state and the time it takes the lender to complete the foreclosure process if the borrower does not contest the foreclosure. We classify Alaska, Arizona, California, Iowa, Minnesota, Montana, North Carolina (purchase mortgages), North Dakota, Oregon, Washington, and Wisconsin as non-recourse states.[4]

[4] An appendix available from the authors describes the foreclosure and deficiency judgment procedures in the US states. Our time lines come from The National Mortgage Servicer's Reference Directory (2004) published by the USFN (America's Mortgage Banking Attorneys).

Our classification of states is similar to that of the USFN (2004): The states we classify as non-recourse are the same as those for which the USFN (2004, pp. 5-5 - 5-7) states that a deficiency judgment is either not available or for which getting one is impractical. However, we classify purchase mortgages in North Carolina as non-recourse since state law prohibits deficiency judgments on purchase mortgages and we treat South Dakota as a recourse state. We usually were able to speak with at least one foreclosure attorney in each state where the amount of recourse in practice was unclear or the statutes were difficult to understand.

B. Types of Default

In practice, lenders usually view litigiously foreclosing as a last resort in the event the borrower defaults and will usually try to recover a portion of principal through other means before resorting to foreclosure.[5] Furthermore, lenders have a strong interest in foreclosing quickly on the property even when the lender does choose to exercise the option to foreclose.[6]

Lenders prefer to avoid foreclosures, and, especially, contested foreclosures, for several reasons. First, properties depreciate substantially when the borrower is in default. Second, the property usually sells at a distressed value in a foreclosure sale. Third, lenders incur negative publicity and reputation costs among other prospective borrowers from forcibly removing a borrower from his or her home. For instance, Campbell, Giglio, and Pathak (2009) find that a foreclosure reduces the value of the home by approximately 28%. The depreciation rate is faster when a property is in default because the borrower has no incentive to adequately maintain the property and thus may deliberately accelerate the property's depreciation.

There are at least three ways by which a borrower can default: a short sale, a voluntary conveyance, or simply agreeing not to contest the foreclosure. In a short sale, the borrower finds a buyer for the property who pays a purchase price that is less than the full balance of the debt owed. Usually in a short sale the lender agrees to waive his right for a deficiency in exchange for the borrower selling the property and remitting the proceeds to the lender. Occasionally, the lender may only agree to waive his right to a deficiency if the borrower also

[5] See, for example, Larsen, Carey, and Carey (2007), Brueggeman and Fisher (2008), and Ling and Archer (2008).

[6] This view was also prevalent among the foreclosure attorneys to whom we spoke.

agrees to give the lender a lump sum payment in addition to the sale proceedings.

In a voluntary conveyance, the borrower hands over the deed to the property to the lender. In the most common voluntary conveyance, a deed in lieu, the lender forgives the debt owed in exchange for the deed. In addition to eliminating the risk of the lender pursuing a deficiency judgment, a deed-in-lieu affects a borrower's future access to credit less severely than if the lender must forcibly evict the borrower (Larsen, Carey, and Carey [2007]). The benefit to the lender is that, in addition to getting the property back more quickly, the lender's legal costs are lower and the deed in lieu of foreclosure "can be beneficial to the lender's public image and to the public perception of the property" (Ling and Archer [2008]).

However, a voluntary conveyance carries some risks to the lender. First, if the borrower declares bankruptcy within one year of a deed-in-lieu, the court may declare the conveyance improper. In such a case the lender's claim becomes an unsecured claim on the borrower's assets and, in the case of the borrower filing under chapter 13, future income which will generally give the lender a worse payoff. Second, a voluntary conveyance does not cut off any subordinate liens on the property the way a proper foreclosure does.

Finally, a borrower may simply agree to what is known as a "friendly foreclosure", i.e., to not contest the foreclosure and submit to the jurisdiction of the court regarding leaving the property and cooperating with the lender. The main benefit of this option is that the lender gets the property back more quickly relative to a contested foreclosure. This takes more time than a voluntary conveyance but is less time-consuming than a regular foreclosure (Brueggeman and Fisher [2008]). A friendly foreclosure may be preferable to the lender as it cuts off any subordinate interests that may exist in the property and protects the lender if the borrower subsequently declares bankruptcy (Ling and Archer [2008]). The benefits to the borrower from a friendly foreclosure relative to a more standard foreclosure are similar to those from a short sale and a deed in lieu: the lender may agree to waive his or her right to a deficiency judgment and the borrower's future credit availability suffers less.

Subsequent to a voluntary conveyance, the property becomes real estate owned (REO), i.e., the lender owns the property. A property can also become REO subsequent to a foreclosure sale if the lender acquires the property by virtue of being the only bidder.

III. A Model of the Default and Default Type Decisions

In this section, we present a static model to study the effect of recourse on default. The model predicts that allowing lenders recourse changes both default rates and the method of default even though lenders seldom actually pursue deficiency judgments in equilibrium. Finally, the model enables us to explore the interaction between recourse and laws that govern the rapidity of the foreclosure process.

A. The Economic Environment

The borrower makes two decisions regarding default: whether to terminate the mortgage and how to default if he defaults. The borrower's decision of whether to terminate the mortgage depends on transaction and search costs. Once the borrower decides to default, the borrower and lender must agree to a short sale, a friendly foreclosure / deed in lieu, or a contested foreclosure. We combine the friendly foreclosure and deed in lieu outcomes into one event in the model as we view the lender and borrower as having similar incentives in both situations and the outcomes are relatively similar. Herein, we refer to that outcome as a deed in lieu.

We first consider how default happens once the borrower has decided to default. We then use this analysis to explore the borrower's decision of whether to default conditional upon a given loan-value ratio in both recourse and non-recourse states. We further assume that 1) the borrower can contest foreclosure and, thus, slow down the foreclosure process; 2) the borrower receives free rent during the default period; 3) the borrower incurs search costs of finding a new home and moving costs once he agrees to a short sale, hands over the deed in lieu to the lender, or gets foreclosed upon; 4) the lender agrees to waive a deficiency judgment if the borrower agrees to a short sale or a deed in lieu; 5) at the foreclosure or REO sale, the lender recovers less than the fair market value of the property; 6) the lender recovers a greater fraction of the fair market value in an REO sale subsequent to a deed in lieu than at a foreclosure sale; 7) if the lender sues for a deficiency, the borrower receives credit for the fair market value of the home; and 8), the borrower incurs lower credit costs if he agrees to a short sale or a deed in lieu than if he is foreclosed upon.

Assumption (2) follows Ambrose, Buttimer, and Capone (1997). Assumption (5) is based on widespread evidence that properties depreciate substantially more rapidly during foreclosure and the lender is often the only bidder at a foreclosure sale (see, for example, Campbell, Giglio, and Pathak [2009]). Thus, the lender cannot recover the fair market value of the property in foreclosure. Assumption (7) is consistent with most states foreclosure laws requiring the borrower to receive credit for the fair market value of the home in any deficiency judgment. Assumption (8) is justified by the widespread observation that, in addition to a large drop in the FICO score in the event of foreclosure, the foreclosure itself goes on the borrower's credit record. In general, lenders are more willing to lend again to a borrower that has defaulted through a short sale or a deed in lieu than one that forced the lender to foreclose even if the FICO score drops by the same amount in a foreclosure as in a short sale.

We assume the lender will not agree to a loan modification. This assumption stems from empirical observation that significant principal write downs are rare prior to the end of our sample (December 2008). While loan modifications involving extending the amortization period, a temporary stay in payments, or a reduction in the interest rate, are somewhat more common, they will not alter the default decision of financially unconstrained borrowers.

Lenders may be unwilling to write down principal due to either unclear contractual obligations on securitized loans or borrower heterogeneity in the costs of defaulting that the lender cannot observe. While it is ex-post optimal in many circumstances for a lender to agree to a loan modification rather than have the borrower default, introducing the possibility of principal write-downs may ex-ante cause more borrowers to seek loan modifications than would in the absence of this policy. Because the lender cannot observe the cost of defaulting without the borrower actually defaulting, he would have to allow a loan modification on all loans with a given loan-to-value ratio that would reduce the value of his mortgage pool. See also Foote, Gerardi, Goette, and Willen (2009) for a model of why loan modifications are rare in practice.

B. The Model

We first consider the decision of how default happens in recourse and non-recourse states once the borrower has decided to default.

Recourse States.—Suppose that at time t the borrower decides to terminate his mortgage. The mortgage can be terminated in one of three ways: 1) the borrower and lender agree to a short sale (SHORT), 2) the borrower and lender agree to a deed in lieu or a friendly foreclosure without a deficiency judgment (DIL), or 3) the lender forecloses with a deficiency judgment (F). If the lender pursues a deficiency judgment, he cannot recover his collateral for $\tau+m$ periods. If the lender pursues foreclosure without a deficiency judgment, he cannot recover his collateral for τ periods. τ and m are exogenous to the model and determined by states' foreclosure laws. If default occurs, the house depreciates by a fraction δ^F each period between default and the foreclosure sale.

The lender's payoffs in these three scenarios are

1) $\mu^S H_t$ if short sale;

2) $\left(\frac{1}{1+r}\right)^{\tau}\left(1-\delta^F\right)^{\tau}\mu^{DIL}H_t$ if deed in lieu; or

3) $\left(\frac{1}{1+r}\right)^{\tau+m}\left(\left(1-\delta^F\right)^{\tau+m}\mu^F H_t+\phi\left[M_{t+\tau+m}-(1-\delta)^{\tau+m}H_t\right]\right)$ if foreclosure with deficiency judgment,

where ϕ is the expected present value the lender collects on the deficiency net of legal costs, H_t is the price of the house at the time the borrower announces he will default, $\mu^S H_t$, is the recovery amount in the event of a short sale, $\mu^{DIL}H_t$ is the discounted present value of the property to the lender at the time of a deed in lieu, $\mu^F H_t$ is the discounted present value of the property to the lender at the time of foreclosure, where $\mu^F<\mu^{DIL}<\mu^S<1$, $M_{t+\tau+m}$ is the unpaid mortgage balance at time $t+\tau+m$, δ^F is the fraction by which the house depreciates each period while the borrower is in default, and r is the discount rate.

The borrower receives free rent h in any period in which he is in default but has not been foreclosed upon or agreed to a short sale. Once the borrower either agrees to a short sale or gets foreclosed upon, he must pay the search and moving costs, s, of relocating. If the borrower defaults through a short sale or a deed in lieu, he incurs a cost c_0 that represents the cost of decreased availability of credit. If the borrower defaults by allowing himself to be foreclosed upon, he incurs a credit cost c_1. The borrower's payoffs from the three scenarios are thus

1) $-s-c_0$ if short sale;

2) $\sum\limits_{k=0}^{\tau}\left(\frac{1}{1+r}\right)^{k}h\left(1-\delta^F\right)^{k}H_t-\left(\frac{1}{1+r}\right)^{\tau}(c_0+s)$ if deed in lieu; or

3) $\sum_{k=0}^{\tau+m}\left(\frac{1}{1+r}\right)^k h\left(1-\delta^F\right)^k H_t - \left(\frac{1}{1+r}\right)^{\tau+m}\left\{\phi\left[M_{t+\tau_1}-\left(1-\delta^F\right)^{\tau+m}H_t\right]+c_1+s\right\}$ if foreclosure with deficiency judgment.

Non-Recourse States.—Since, conditional on opting for foreclosure, the borrower always prefers a lengthier foreclosure process, the lender's payoff in states without recourse are

1) $\mu^S H_t$ if short sale;

2) $\left(\frac{1}{1+r}\right)^\tau\left(1-\delta^F\right)^\tau \mu^{DIL}H_t$ if deed in lieu;

3) $\left(\frac{1}{1+r}\right)^{\tau+m}\left(1-\delta^F\right)^{\tau+m}\mu^F H_t$ if foreclosure.

Similarly, the borrower's payoffs become

1) $-s-c_0$ if short sale;

2) $\sum_{k=0}^{\tau}\left(\frac{1}{1+r}\right)^k h\left(1-\delta^F\right)^k H_t - \left(\frac{1}{1+r}\right)^\tau\left(c_0+s\right)$ if deed in lieu;

3) $\sum_{k=0}^{\tau+m}\left(\frac{1}{1+r}\right)^k h\left(1-\delta^F\right)^k H_t - \left(\frac{1}{1+r}\right)^{\tau+m}\left(c_1+s\right)$ if foreclosure with deficiency judgment.

The first conclusion is that, other things equal, borrowers should be more willing to agree to a deed in lieu or a short sale in states that allow deficiency judgments. To see this, note the following:

$$\begin{aligned}&\sum_{k=0}^{\tau+m}\left(\frac{1}{1+r}\right)^k h\left(1-\delta^F\right)^k H_t-\left(\frac{1}{1+r}\right)^{\tau+m}\left\{\phi\left[M_{t+\tau_1}-\left(1-\delta^F\right)^{\tau+m}H_t\right]+c_1+s\right\}\\ <\ &\sum_{k=0}^{\tau+m}\left(\frac{1}{1+r}\right)^k h\left(1-\delta^F\right)^k H_t-\left(\frac{1}{1+r}\right)^{\tau+m}\left(c_1+s\right).\end{aligned}$$

Then if the borrower's payoff from the short sale or deed in lieu exceeds the payoff from foreclosure in a non-recourse state, it will also exceed the payoff from the foreclosure in a recourse state.

The second conclusion is that lenders will be less willing to agree to a short sale or a deed in lieu in states that allow deficiency judgments. Clearly, the lender always prefers a short sale or deed in lieu to foreclosure in non-recourse states. In recourse states, there may be cases in which it is beneficial for the lender to pursue foreclosure if the deficiency is sufficiently large.

However, the threat of a deficiency judgment will often induce the borrower to agree to a short sale or a deed in lieu and the lender will often agree to this rather than actually pursue the deficiency judgment. The reason is that, when $\phi < 1$ or the borrower receives credit for the fair market value of the property, rather than the foreclosure sale price, the amount the lender can recover through a foreclosure sale may be smaller than what the lender can recover from foreclosing with a deficiency judgment.

The optimal default type is the equilibrium outcome of bargaining between lenders and borrowers. Lenders pursue the course of action that gives them the highest payoff conditional on what the borrower will agree to. The borrower and lender know each other's expected payoffs such that a threat to pursue a deficiency judgment in situations in which the payoff to the lender is higher without a deficiency judgment is not credible. Thus, the borrower knows before he decides whether to agree to a short sale whether the lender will be willing to agree to a deed in lieu such that, if the lender's payoff is higher with a deed in lieu than a contested foreclosure, the borrower can guarantee that his payoff will be that of the deed in lieu.

Conditional on the payoff from defaulting, the borrower decides whether to default conditional on the current loan-value ratio. Consider a borrower that currently owes M_t dollars on his mortgage used to finance a house that is now worth H_t, $M_t > H_t$. There are two possible outcomes:

1. Stay in the house and receive the payoff $H_t - M_t$.

2. Default and receive the payoff determined by the negotiation outlined above.

C. Model Solution

We solve a special parameterized case of the model to further explore its implications.

Optimal Type of Default.—Table 2 explores how recourse affects how the borrower chooses to terminate his mortgage as a function of how much the lender can recover and the home owner's LTV for a particular calibration. We normalize H to 1 and assume $\delta^F = 0.005$ per month, $m = 6$ months, $r = 0.05/12$, $s = 0.1$, $H = 0.1$, $h = 0.05/12$, $\mu^S = 0.9$, $\mu^{DIL} = 0.8$, and $\mu^F = 0.75$. We assume $\mu^S = 0.9$ because the borrower must find a buyer for the home quickly under distressed conditions such that the home sells for slightly less than its fair market value. Our choice for the rent-price ratio is equal to the U.S. average rent-price ratio from $1960Q1 - 2008Q4$ based on the calculations of Davis, Lehnert, and Martin (2008). We assume $c_0 = 0.08$ and $c_1 = 0.1$. The top panel, our benchmark case, assumes that it takes 6 months to complete an uncontested foreclosure. Finally, we assume the mortgage is a fully amortizing, 30 year constant-payment, 6% mortgage to get $M_{t+\tau+m}$.

The first column of table 2 illustrates the default method in non-recourse states for our benchmark case. In non-recourse states borrowers always do better by contesting the foreclosure process since they get an additional several months of rent and delay search and credit costs which, for our calibration, outweighs the lower credit costs they would incur by agreeing to a short sale or a deed in lieu.

In recourse states, for very low recovery rates ($\phi = 0.05$ or 0.1), the home owner never agrees to a short sale since the amount the lender can recover through a deficiency judgment is much lower than the present value of free rents the home owner receives during the foreclosure process. However, both the borrower and the lender do better with a deed in lieu in the case of low recovery rates. The lender has a higher payoff because the discount at which he sells the home is lower and it depreciates less. The homeowner has a higher payoff from deed in lieu because the amount he has to repay in the event of a deficiency judgment is sufficient to exceed the benefits of an additional six months of the free rent. The exception is when the initial LTV is 100% and the uncontested foreclosure time is 3 or 6 months. In this case there is only a slight deficiency – the amount of negative amortization that has accumulated since default. Then the borrower has such a small deficiency that the extra six months of free rent outweighs what he will eventually have to pay out on a deficiency judgment.

As the recovery rate and loan-to-value increase, short sales occur increasingly frequently and deeds in lieu become increasingly rare. The reason is that the lenders threat to pursue a deficiency judgment becomes credible as the lender receives a higher payoff from foreclosing with a deficiency judgment than agreeing to the deed in lieu. This ensures the borrower that he will receive the foreclosure payoff if he does not agree to a short sale and the borrower gets a higher payoff with a short sale than a foreclosure with a deficiency judgment.

We do not see lenders preferring foreclosing with a deficiency judgment rather than doing a short sale until ϕ rises to 35%. Starting at $\phi = 35\%$, if the initial LTV is sufficiently high, the lender can recover enough from a deficiency judgment such that he will always seek one and the borrower is guaranteed to have to pay back a portion of the deficit if he defaults.

Comparing across the panels in table 2 illustrates the effect of the length of the foreclosure process on the optimal default decision. The panels show the optimal default decision when the uncontested foreclosure time frame is 3, 6, and 12 months. Changing the foreclosure time frame does not usually result in a different foreclosure type. For very low LTVs and recovery rates, a faster foreclosure process benefits the borrower because the amount of amortization on the mortgage is smaller than the present value of the free rent he gets. Thus, for very low LTVs, we see a few more defaults completed by foreclosure in states with more rapid foreclosure processes. However, when the LTV and recovery rate are high, a faster foreclosure process works to the lenders advantage since the amount the lender can recover from a deficiency judgment is higher as the property has less time in which to depreciate. Thus, for high LTVs, we also expect to see more defaults completed by foreclosure in states with more rapid foreclosure processes.

Optimal Decision to Default.—Table 3 illustrates the effect of recourse on whether default occurs as a function of the recovery rate and home owner equity. When expected recovery is low, there is little difference in when the borrower defaults between recourse and non-recourse states. Overall, changing the time of default from 6 to 12 months or from 6 to 3 months changes the decision of whether to default in only a few cases and so we do not expect the length of the foreclosure process to substantially affect default rates.

It is only when the expected recovery is above 35% that recourse has a strong deterrent effect. This deterrent effect is stronger the lengthier is the foreclosure process. In some situations, the structure of the lender's and borrower's incentives imply that the borrowers know that the lender would prefer a short sale or a deed in lieu and the borrower's payoff from defaulting is thus that from a short sale such that allowing the lender recourse does not deter the borrower from defaulting.

Combining the information in the top panels of table 2 and table 3, we see that there are only a few combinations of recourse and the LTV ratio in which the outcome is actually that of the lender pursuing a deficiency judgment. When recourse is moderate, $35\% - 40\%$ and the LTV ratio is high, is it both worthwhile for the lender to pursue a deficiency judgment and the borrower to incur the fixed costs of defaulting. The lender wants to pursue a deficiency judgment because the payoff from a short sale or a deed in lieu is low relative to the balance owing but does not recover enough to deter default. Once the recovery rate becomes sufficiently high, the lender both pursues a deficiency judgment and the recovery rate is sufficiently large to deter default. In other circumstances in which default would have occurred through foreclosure with a deficiency judgment, the borrower earns a higher payoff by not defaulting because the balance owed is not high enough for it to be worthwhile for him to incur the credit and search costs of defaulting.

D. Discussion

To summarize, our model explains why lenders rarely pursue deficiency judgments. Furthermore, the model suggests that the presence of recourse will deter relatively well-off borrowers (i.e., those where the lender faces a high ϕ) from defaulting. It also suggests that borrowers will default differently, in ways that lead to lower losses for lenders, in states that allow lenders recourse. This is consistent with the results of Clauretie and Herzog (1990) and Crawford and Rosenblatt (1995) who find that, conditional upon foreclosure occurring, losses on foreclosures are lower in states that permit deficiency judgments. Our model suggests a reason why, even if lenders rarely actually pursue deficiency judgments, losses are lower in states that permit lenders recourse.

It is worth noting that recourse laws will affect the choice of how the borrower chooses to default both in the case of strategic defaulters (borrowers who can continue to make payments on their mortgage if they choose to) and non-strategic defaulters (borrowers who are insolvent and unable to make payments). In our model, any borrower that is insolvent defaults. In the case of non-strategic defaulters, lenders still can recover some portion of any deficiency in most states since the lender typically has 10 years to collect on a deficiency and can file for a 10 year extension on that recovery, ample time to see an improvement in a borrower's financial circumstances.

The results in tables 2 and 3 are stylized and assume that all borrowers face the same fixed default costs. In practice, these costs vary greatly across borrowers and so the type of default borrowers and lenders agree to is relevant. If the fixed default costs are lower for some borrowers, we should expect to see deficiency judgments primarily in the low-recovery regions.[7] When the recovery rate is higher, usually either the borrower is deterred from defaulting altogether by the threat of a deficiency or the lender and borrower agree to a short sale.

Which cases in tables 2 and 3 are most common, and thus how big a deterrent recourse is to default in practice, is an empirical question. In the next two sections, we empirically examine how big an effect recourse has on default rates. We then test whether defaults occur more frequently via deeds in lieu or short sales in states that permit lenders recourse.

IV. Data

The data used in the study is loan-level data from LPS Applied Analytics, Inc. The data contain information about loans on a monthly basis.[8]

A. Variable Definitions

Definition of Default.—We consider the loan as defaulted if the loan is terminated in one of the following ways: by REO sale, by short sale, by pay off out of foreclosure, by pay off out of bankruptcy and serious delinquency, or by liquidation to termination. In the analysis of the probability of default, the dependent variable takes a value of 1 in the month the loan defaults. We drop all observations on defaulted loans subsequent to the default month. Consequently, the dependent variable takes a value of 0 in months that we observe the defaulted loans prior to the default month and observations on loans that do not default, whether terminated or current.

Default Type.—In our analysis of whether recourse changes how default happens, we consider loans terminated by default. We divide defaults into defaults by foreclosure and other types of default such as short sales, deeds in lieu, and friendly foreclosures. We identify non-foreclosure defaults as loans which directly become REO loans or short sales. We also define any default where the lender received a payoff out of bankruptcy or serious delinquency

[7] This is consistent with foreclosure practice: Our conversations with foreclosure professionals indicated that deficiencies judgments typically sell for $5 - 10$ cents on the dollar.

[8] An appendix available from the authors provides details about the variables by LPS codes.

as a foreclosure since these are loans in which the borrower likely declared bankruptcy to halt foreclosure proceedings. Such a default is akin to a contested foreclosure process. Thus, default type takes on a value of 1 if the loan defaulted via a foreclosure and 0 otherwise.

Default Option Variables.—We define the value of the default option as the probability the borrower has negative equity in the house as in Deng, Quigley, and Van Order (2000) and Ambrose, Capone, and Deng (2001). Since we know the balance owed on the loan, we need only infer the distribution of individual house prices. The value of equity to market value k_i months after loan origination is

$$E_{i,t,k_i} = \frac{M_{i,t,k_i} - L_{i,t,k_i}}{M_{i,t,k_i}},$$

where M_{i,t,k_i} is a market value of the property purchased at time $t-k_i$, and L_{i,t,k_i} is a present value of the remaining loan balance. The market value of the property is

$$M_{i,t,k_i} = C_i \frac{HPI_{i,t}}{HPI_{i,t-k_i}},$$

where $C_{i,t-k_i}$ is a cost of a property at the time of a purchase, $HPI_{i,t}$ is house price index in the state where property i is located, and $\frac{HPI_{i,t}}{HPI_{i,t-k_i}}$ follows a lognormal distribution (see Case and Shiller [1987] and Deng, Quigley, and Van Order [2000] for details). The mean and variance of $\frac{HPI_{i,t}}{HPI_{i,t-k_i}}$ is obtained using the data available from the Office of Federal Housing Enterprise Oversight (OFHEO).[9]

The value of the default option for mortgage i k_i months after origination is the probability that equity is negative:

$$DEFAULT\_OPTION_{i,k_i} = \Pr(E_{i,t,k_i} < 0) = \Phi\left(\frac{\ln L_{i,k_i} - \ln M_{i,k_i}}{\sqrt{\sigma^2_{HPI_{i,k_i}}}}\right),$$

where $\Phi(\cdot)$ is the cumulative standard normal distribution and $\sigma^2_{HPI_{i,k_i}}$ is the variance of

[9]To calculate the standard deviation of $\frac{HPI_{i,t}}{HPI_{i,t-k_i}}$, $\sigma_{HPI_{i,k_i}}$, we use the volatility parameters A and B provided by OFHEO as follows:

$$\sigma_{HPI_{i,k_i}} = \sqrt{Ak_i + Bk_i{}^2}.$$

See Calhoun (1996) for the technical description of OFHEO index.

individual house prices in state i around the mean in state i.

We also include the default option squared as in Deng, Quigley, and Van Order (2000).[10]

Prepay Option Variables.—As a proxy for prepayment option, we use a spread between current market mortgage rate, r_t, and the mortgage rate on the contract, r_0. We use indicator variables, rather than a continuous variable, based on the results of Kau, Keenan, and Kim (1994) that the spread affects default rates in a nonlinear fashion. Following Ambrose, Capone, and Deng (2001), we define the following dummy variables: Rate1 $= 1$ if $r_0 + 2\% \leq r_t$, and 0 otherwise; Rate2 $= 1$ if $r_0 + 1\% \leq r_t < r_0 + 2\%$, and 0 otherwise; Rate3 $= 1$ if $r_0 - 1\% \leq r_t < r_0 + 1\%$, and 0 otherwise; Rate4 $= 1$ if $r_0 - 2\% \leq r_t < r_0 - 1\%$, and 0 otherwise; and Rate5 $= 1$ if $r_t < r_0 - 2\%$, and 0 otherwise, where r_t and r_0 are in percentages.

Foreclosure Timing and Recourse Variables.—We include the time it takes for uncontested foreclosure in the state in which the property is located since our model predicts that a lengthier foreclosure process will increase defaults. Table 1 contains the recourse classification of states and the foreclosure timelines. In one specification, we also include foreclosure timing as a dummy variable that takes on a value of 1 if the uncontested foreclosure time is less than 185 days, and zero otherwise.

Trigger Events.—We control for trigger events, as Capozza, Kazarian, and Thomson (1997) emphasize, by including the contemporaneous state divorce rate and the state unemployment rate. We use lagged monthly seasonally unadjusted unemployment rates from the BLS.[11]

Loan Level Variables and Borrower Characteristics.—We also use several loan level characteristics that other studies have found to be important in explaining defaults, such as the age of the loan (in months) and the LTV at origination, an indicator variable that takes a value of 1 if the loan is interest only at origination, an indicator variable that takes a value of 1 if the loan is an ARM, an indicator variable that takes a value of 1 if the loan is a jumbo, an indicator variable that takes a value of 1 if the loan is a first mortgage,

[10] As a robustness exercise, we verified that the coefficient on recourse is negative and significant (at slightly higher than the 5% level) on the linear term when we do not include the quadratic term.

[11] We do not use seasonally adjusted unemployment rates as there may be a seasonal pattern to defaults due to seasonal economic conditions.

an indicator variable that takes a value of 1 if the loan is not a purchase mortgage, and the borrower's FICO score at origination. We convert nominal appraisal amounts at origination into real 2005 dollars by deflating using the CPI excluding shelter.

B. Sample Description

We use information on loans originated between August 1997 and December 2008. August 1997 is the first month that the FICO score variable is available in the data. We drop all FHA and VA loans because deficiency judgments are prohibited on FHA loans and strongly discouraged on VA loans (Larsen, Carey, and Carey [2007]). We drop all loans with private mortgage insurance. We restrict our analysis to mortgages with constant principal and interest, ARMs, or Graduated Payment Mortgages (GPMs). We also drop mortgages for home improvement, debt consolidation, education, or medical expenses. We limit our analysis to first or second mortgages. Finally, we drop observations for loans on properties with more than one unit.

We then draw a 10% random sample from the LPS database. Our restrictions imply that we have 82,828,381 loan-month observations. 67% of our observations come from recourse states and on average there is a 1% probability that a home owner in our sample has negative equity. 7% of our observations are interest only at origination and 97% of our observations are on first mortgages. 20% of our observations are adjustable rate mortgages. In total, our sample includes $2,924,160$ loans and $38,440$ defaults. An appendix available from the authors lists the default rates in each state over the full sample, over the 1997-2004 period, and over the 2005-2008 period as well as summary statistics for the variables.

V. Empirical Results

We structure our empirical analysis in two parts: first, we examine the effect of recourse on the probability of default; second, we examine the effect of recourse on the way borrowers default.

A. The Impact of Recourse on Default

We assume that the borrower defaults if an unobserved variable $x = X\beta + \varepsilon$, falls below 0 where $\varepsilon \tilde{} N(0,1)$ and X is a vector of variables that controls for the borrower's prepay and default options, other loan-level characteristics, and trigger event variables.

As the theoretical model in section 3 shows, recourse affects the borrower's payoff from defaulting. Different payoffs from the default decision in recourse and non-recourse states may lead to different threshold values of the default option at which the borrower defaults in recourse and non-recourse states. Thus, to estimate the impact of the recourse on the probability of default, we model recourse in our empirical specification as an interaction term between the value of the default option and the recourse dummy variable. The recourse dummy variable takes a value of 1 if the mortgaged property is located in a state with a provision for recourse and 0 otherwise.

The first column of table 4 contains the results without recourse variables. The results in the column illustrate the effect of the prepay and default options, trigger events, and loan-level characteristics on default when we do not control for recourse. All of the coefficients have the expected sign. Having an interest-only loan, an ARM, or a second mortgage raises the probability of default. Borrowers with higher FICO scores at origination are less likely to default while loans with a high LTV at origination are more likely to default. Finally, younger loans are much more likely to default than older loans. The divorce rate has the expected sign but is significant only at the 10% level when we cluster the standard errors, likely because there is relatively little variation across time in the divorce rate within a state. The unemployment rate has the expected sign but becomes insignificant when the standard errors are clustered.

Following the theoretical model, we specify the hypotheses for our empirical analysis and test them against the alternatives that follow from the solution of the model.

Hypothesis 1.—*Hypothesis 1.0*: Recourse does not have an impact on the effect of the default option on the probability of default.

Hypothesis 1.1: Recourse decreases the probability of default for a given value of the default option.

Hypothesis 1.1 follows from the results in table 3 which depicts the decision of whether to default for a borrower who has negative home equity. There are many situations in table 3 in which recourse deters default for a borrower who has negative home equity. In particular, in all three panels default occurs at slightly higher values of the default option. At high values of the default option and when the recovery rate on the deficiency is high, recourse deters default altogether. The empirical analysis allows us to examine whether this effect is present and quantitatively important in the data. To test the effect of the recourse on the impact of negative equity on the probability of default, we include the interaction terms of the default option value and default option value squared with the recourse variable.

Column 2 of table 4 contains the results of our benchmark specification. The coefficient on the interaction term between recourse and negative equity is negative and statistically significant. The coefficient on the interaction between recourse and squared term of the probability of negative equity is positive and statistically significant. The negative coefficient on the linear term of default option value indicates that recourse decreases the impact of the negative equity on the probability of default. The positive coefficient on the square term indicates that the effect decreases as the default option value increases. Because of this nonlinear effect of default option value on the probability of default, the effect of recourse depends on a particular value of the default option.

To gauge the magnitude of the deterrent effect of recourse, we evaluate the probability of default in recourse and non-recourse states at different values of the default option. Table 5 contains the estimates of the probabilities. Column 1 shows the probabilities at the means of the continuous variables and the modes of the dummy variables at the time of default. At the mean of the default option at the time of default, borrowers in non-recourse states are 21% more likely to default than borrowers in recourse states.

In columns $2-4$ of table 5, we estimate default probabilities at the means of the continuous variables and the modes of the dummy variables for all observations at different values of the default option. At the mean of the default option, the probability of default is 6% higher in non-recourse states than in recourse states. At the 90th percentile of the value of the default option, 0.3%, the probability of default in non-recourse states is 2% higher. This difference increases to 13% at the 95th percentile, when the default option is

2.12%. The results in columns $2-4$ indicate that recourse has a deterrent effect on default at high values of the distribution of the default option value, which are precisely the values associated with default. Thus, the data reject Hypothesis 1.0 in favor of Hypothesis 1.1.

In columns 3 and 4 of table 4, we present the results for two of the specifications in which we verified the robustness of our result regarding recourse. In column 3, we include the difference between the contract rate and current mortgage rates in interactions with the probability of negative equity as in Ambrose, Capone, and Deng (2001). The results are quite similar to our benchmark specification although the log-likelihood is somewhat higher when rates are included in interactions suggesting that including rates in levels fits the data better.

In column 4, we explore whether our results regarding recourse are due to state-specific factors by including state dummy variables. We drop the divorce rate in this specification as our divorce rate data is only available at the annual frequency. Also, for some states we only have a few divorce rate observations over the entire sample, thus there is little variation remaining in the divorce rate after we control for state specific effects. When we control for the state specific fixed effects, the results on the effect of recourse carries through: the coefficient on the interaction between recourse and the default option value is statistically significant and are numerically similar to that in the benchmark specification. Thus, our results regarding the deterrent effect of recourse are not driven by unobserved differences between recourse and non-recourse states.[12]

Hypothesis 2.—*Hypothesis 2.0*: The probability of default is not affected by state foreclosure timelines.

Hypothesis 2.1: The probability of default is higher in states with longer foreclosure timelines.

In column 5 of table 4, we show the effect of the lengthiness of the uncontested foreclosure process, as stated in USFN (2004), on the probability of default. We model the lengthiness of the foreclosure process in two ways. In column 5 we include the length of the uncontested

[12]We also consider a specification in which we include year of origination dummies. The coefficient on the interaction between the default option value and recourse is very similar to that of our benchmark specification in column 2 and the results are in an appendix available from the authors. With 2003 as the omitted category, the coefficients on origination years prior to 2003 are negative.

foreclosure process in months for the state in which the property is located. When we do not cluster the standard errors, states with lengthier foreclosure processes appear to experience more defaults. However, the effect becomes insignificant when we cluster the standard errors. We also did not find that the lengthiness of the foreclosure process significantly affects default in other specifications in which we interacted the foreclosure time frame with recourse variables. While our model predicts that a lengthier foreclosure process will increase the default rate in a few cases, the empirical evidence in columns 5 suggests those cases are infrequent in practice. We also tried a specification in which we included foreclosure timing by using a dummy variable that takes on a value of 1 if the state's uncontested foreclosure process takes more than 6 months and 0 otherwise and obtained similar results. Thus we cannot reject Hypothesis 2.0 that the lengthiness of the foreclosure process has no effect on the probability of default.

Hypothesis 3.—The model in section 3 predicts that the deterrent effect of default on the probability of default depends on the amount of the deficiency judgment that a lender can actually recover. In particular, as the right hand side columns in table 3 show, the recourse has a substantial deterrent effect if the recovery rate is high. This leads to the following null hypothesis tested against the alternative that the model predicts:

Hypothesis 3.0: The effect of recourse on the probability of default does not depend on the lender's recovery rate.

Hypothesis 3.1: The effect of recourse on the probability of default is stronger when the lender's recovery rate is high.

In the empirical analysis we proxy for the lender's recovery rate with the appraised value of the mortgaged property. A higher appraisal amount likely indicates that the borrower has more assets that can be used by the lender to recover on the deficiency judgment. Additionally, a higher appraisal amount is more likely to be associated with higher income since the ratio of debt to income is a key ratio in the underwriting process. Higher income borrowers who declare bankruptcy also may have less chance to have their debt discharged during a bankruptcy proceeding. This is particularly true for borrowers considering default after the 2005 bankruptcy reform, which requires borrowers above the state median income to file under chapter 13 rather than under chapter 7. This implies that, unlike with poor borrowers, lenders have better recovery rates with richer borrowers.

Table 6 contains the results on estimating our benchmark specification separately for different values of the appraisal value of the mortgaged property. As the results in table 6 show, recourse does not deter default for all households in the same way. Recourse is a deterrent for default when the appraisal amount exceeds \$200,000: the coefficient on the recourse interaction with the default option value and its square are statistically insignificant when the appraisal amount is \$200,000 or less. The coefficient on the interaction of the recourse with a linear default option term is particularly large in the samples with appraisal amounts from \$300,000 to \$500,000, from \$500,000 to \$750,000, and from \$750,000 to \$1,000,000. The deterrent effect of recourse increases as the appraisal amount increases up until we reach property values that exceed \$1,000,000. For the sample with appraisal amounts of \$1,000,000 or higher, the coefficient has the expected sign and is similar in magnitude to our benchmark specification but loses its statistical significance.

The results of the estimation of the probability of default in the samples by appraisal amount indicate that the effect of recourse on the probability of default is mainly driven by the borrowers with mortgages on the properties appraised \$200,000 and higher. Thus we reject Hypothesis 3.0 in favor of the alternative that recourse has a substantial deterrent effect on default in case of high recovery rates on the deficiency judgment and does not have a statistically significant effect when the recovery on deficiency judgment is likely to be low.

To gauge the magnitude of the deterrent effect of recourse on the default probabilities, in table 5 we present estimates of the probabilities of default in recourse and non-recourse states. At the mean value of the default option at the time of default and for homes appraised at \$300,000 to \$500,000, borrowers in non-recourse states are 60% more likely to default than borrowers in recourse states. For homes appraised at \$500,000 to \$750,000, borrowers in non-recourse states are almost twice as likely to default as borrowers in recourse states. For homes appraised at \$750,000 to \$1 million, borrowers in non-recourse states are 66% more likely to default than borrowers in recourse states.

Hypothesis 4.—Some lenders may have a reputation for being more likely to pursue deficiency judgments. In this case, the borrower's decision to default has a lower expected payoff. This translates into a lower payoff from defaulting for a borrower with negative home equity in recourse states when the lender has a reputation of pursuing deficiency judgments relative to when the lender does not have a reputation of pursuing deficiency judgments. We postulate the following hypothesis against the alternative:

Hypothesis 4.0: The effect of recourse on the probability of default does not depend on the type of a lender.

Hypothesis 4.1: The effect of recourse on the probability of default depends on the type of a lender.

Under Hypothesis 4.1 we expect a stronger negative effect of the recourse interaction terms on the probability of default for some types of lenders. Table 7 presents the results from the probit regression estimated separately for loans held by Ginnie Mae (GNMA), loans held by Fannie Mae (FNMA), loans held by Freddie Mac (FHMLC), loans that are privately held and securitized, and loans held in a bank's portfolio.

As can be seen in table 7, the coefficient on the interaction of the recourse dummy with the default option value is negative, sizeable, and statistically significant for privately securitized and private portfolio loans. Table 5 presents estimates of the probabilities for recourse and non-recourse states. At the mean value of the default option at the time of default and for securitized privately held loans, borrowers in non-recourse states are 24% more likely to default than borrowers in recourse states while, for privately held portfolio loans, borrowers in non-recourse states are 31% more likely to default.

The estimation results in table 7 indicate that recourse does not have a significant deterrent effect on default for loans held by GNMA, FNMA or FHMLC. In particular, the coefficient on the interaction between the default option value and the recourse for the GNMA sample is negative but insignificant. The coefficients on the interaction between the default option value and recourse for the FNMA and FHMLC samples are two orders of magnitude smaller than the ones for privately securitized loans and statistically insignificant. This is true even when we consider only FNMA and FHMLC loans on properties appraised at $200,000 or more, the threshold above which we found recourse matters. We conclude that recourse has a statistically significant deterrent effect on default only for privately held loans.

Discussion.—Our empirical findings shed light on the ongoing discussions on whether there is strategic default (see, for example, Foote, Gerardi, and Willen [2008]) and whether the default decision depends on the borrower's income. The result that recourse deters default indicates that at least some of the defaults in the data are strategic rather than the borrower having no choice but to default because of liquidity constraints. Our results indicate that at least some borrowers choose not to default when the lender has recourse indicating that they are capable of continuing to make payments on their mortgage.

Additionally, our results on the differential effect of recourse by the appraisal amount of the mortgaged property indicate that at least some defaults on high and moderately priced homes are strategic. We cannot eliminate the possibility that some of the defaults on low priced homes are strategic as the appraisal amount proxies for both the amount of recourse the lender has and the borrower's financial means in general. Thus, recourse may not significantly affect default on low priced homes for one of two reasons. The first possibility is that most households with low priced homes are liquidity constrained, and thus default non-strategically. Alternatively, for households that buy low priced homes, the lender's recovery on a deficiency judgment may be low in practice such that the borrower's payoffs in recourse and non-recourse states are similar.

The finding that recourse has a differential effect on the probability of default depending on the appraisal amount of the mortgaged property also suggests that the default decision depends on the borrower's income in recourse states. This effect works via the expected deficiency judgment that allows the lender to claim a part of the borrower's assets. The fact that the default decision depends on income is relevant for policy discussions of the impact of default on welfare (see Hatchondo, Martinez, and Sanchez [2009]).

B. The Impact of Recourse on the Way a Borrower Defaults

We next turn to the question of how lender recourse affects the way in which the borrower defaults. We estimate a probit to determine which factors influence whether borrowers are more likely to default by foreclosure. The sample is restricted to the observations for which the default variable takes a value of 1. The dependent variable takes a value of 1 if the default is by a foreclosure and 0 otherwise.

Hypothesis 5.—From table 2, our model suggests that borrowers are less likely to default by litigious foreclosure in states with recourse. We are unable to empirically distinguish between friendly foreclosures and contested foreclosures although our model also predicts that we should see more defaults by friendly foreclosure than by contested foreclosure in recourse states. The model also predicts that the effect recourse has on the way a borrower defaults is influenced by how much recourse the lender has as well as the LTV at the time of default. In situations towards the upper right of table 2, the deterrent effect of a deficiency judgment is strong enough to deter default altogether (see table 3) so we expect foreclosures to be strictly decreasing in the amount of negative equity the borrower has in recourse states. To test the model's predictions we test the following hypothesis against the alternative predicted by the model:

Hypothesis 5.0. Recourse does not have an impact on the way the borrower defaults.

Hypothesis 5.1. Recourse reduces the probability of default by foreclosure. Recourse has a stronger negative effect on the probability of default by foreclosure at lower values of the default option.

To test the hypothesis, we first include only a recourse variable dummy as explanatory variable for the probability of default by foreclosure. Column 1 of table 8 contains the results of the estimation. As the results indicate, recourse lowers the probability of default by foreclosure. The estimated coefficient is negative and statistically significant. In particular, the probability of default by foreclosure in recourse states is 9% lower than the probability in non-recourse states.[13] The result carries through if we include additional explanatory variables. In particular, we include the borrower's FICO score at origination and the LTV at origination to control for any unobserved heterogeneity in the borrower's costs of decreased access to credit or search costs in columns $2-4$; for the specification in column 4, the probability of default by foreclosure is 11% lower in recourse states than in non-recourse states.

To test whether recourse has a stronger effect for higher values of the default option value, we add the default option value and default option value interacted with recourse dummy in addition to the recourse variable as the explanatory variables for the probability of default by foreclosure. If recourse has a stronger negative effect at higher values of the

[13] We calculate the partial effects at the mean of continuous variables and at the modes of dummy variables.

default option, we expect a negative coefficient on the interaction term between recourse and the default option value. As can be seen from the results in table 8, the negative effect of recourse on the probability to default by foreclosure is stronger for higher values of the default option.

Our model suggests the time it takes to foreclose on a home has an ambiguous effect on the share of short sales in defaults. On the one hand, a longer foreclosure process makes it more likely the lender will prefer a short sale to a foreclosure and is more likely to forgo a deficiency judgment in favor of a deed in lieu or a short sale. However, the borrower prefers foreclosure when he can delay the search and credit costs and receive a longer period of free rent as a result of a lengthier foreclosure process. A priori, it is unclear what effect foreclosure timing will have on the process. To examine the effect empirically, we include a dummy variable that takes a value of 1 if the uncontested foreclosure time is less than six months, and zero otherwise. As the results in columns 5 and 6 indicate, the foreclosure timing does not have a significant effect on the probability of default by the litigious foreclosure: the partial effect evaluated at the means implies an increase in probability of 1% and is far from statistically significant. The results were very similar when we included foreclosure timing as a continuous variable rather than as a dummy variable.

Default Types for Different Lenders and Appraisal Amounts.—Finally, we examine whether a lender's type and the appraisal amount affects the probability to default by litigious foreclosure. To examine the effect of a lender's type, we include a dummy variable that takes value 1 if a lender type is a GSE and 0 otherwise, i.e., when the investor type is "Private securitized" or "Private portfolio". As can be seen from the results in columns 7 and 8, mortgages held by a GSE are more likely to default by foreclosure than mortgages held by private lenders. This is consistent with our earlier findings that recourse does not have a significant impact on the probability of default for mortgages held by a GSE. In particular, for the specification in column 7 the probability of default by foreclosure increases by 7% for mortgages held by a GSE as compared to the mortgages held by private lenders. However, the effect decreases to 3% when we control for other variables.

To examine the effect of the appraisal amount of the property on the probability of default by foreclosure, we include the appraisal amount and the appraisal amount interacted with the recourse dummy as explanatory variables. We present the estimation results in columns 9 and 10 of table 8. The coefficient on the appraisal amount is positive and statistically significant. The effect on the interaction term is negative but statistically insignificant.

VI. Conclusions

Our model predicts that we do not need to actually observe lenders frequently pursuing deficiency judgments to conclude that recourse alters borrowers' behavior. The threat of a deficiency judgment deters would-be strategic defaulters under many combinations of negative equity and the degree of lender's recourse. In other situations, if the borrower does default, allowing lenders to pursue a deficiency judgment changes how borrowers default. In particular, in states that allow lenders recourse, default occurs more frequently by deeds in lieu and short sales as recourse gives lenders a better negotiating position.

Empirically, we find that, at the mean value of the default option at the time of default, the probability of default is 20% higher in non-recourse states than in recourse states. The deterrent effect on default is significant only for borrowers with appraised property values of $200,000 or more. At the mean value of the default option at the time of default and for homes appraised at $300,000 to $500,000, borrowers in non-recourse states are 59% more likely to default than borrowers in recourse states. For homes appraised at $500,000 to $750,000, borrowers in non-recourse states are almost twice as likely to default as borrowers in recourse states while for homes appraised at $750,000 to $1 million, borrowers in non-recourse states are 66% more likely to default. We also find that recourse deters default on loans held privately; we cannot reject the hypothesis that recourse does not have an effect on loans held by the Government Sponsored Enterprises. Finally, we find that allowing lenders recourse increases the likelihood that default occurs by a more lender-friendly method, such as a deed in lieu of foreclosure.

REFERENCES

Ambrose, Brent W., Richard J. Buttimer Jr., and Charles A. Capone. 1997. "Pricing Mortgage Default and Foreclosure Delay." *Journal of Money, Credit, and Banking*, 29(3): 314-25.

Ambrose, Brent W., Charles A. Capone, and Yongheng Deng. 2001. "Optimal Put Exercise: An Empirical Examination of Conditions for Mortgage Foreclosure." *Journal of Real Estate Finance and Economics*, 23(2): 213-34.

Brueggeman, William B. and Jeffrey D. Fisher. 2008. *Real Estate Finance and Investments*. 13th Ed. New York: McGraw-Hill Irwin.

Calhoun, Charles A. 1996. "OFHEO House Price Indexes: HPI Technical Description." Manuscript, OFHEO.

Campbell, John Y., Stefano Giglio, and Parag Pathak. 2009. "Forced Sales and House Prices." NBER Working Paper 14866.

Capozza, Dennis R., Dick Kazarian, and Thomas A. Thomson. 1997. "Mortgage Default in Local Markets." *Real Estate Economics*, 25(4): 631-55.

Case, Karl E. and Robert J. Shiller. 1987. "Prices of Single-Family Homes since 1970: New Indexes for Four Cities." *New England Economic Review*, Sept./Oct.: 45-56.

Clauretie, Terrence M. 1987. "The Impact of Interstate Foreclosure Cost Differences and the Value of Mortgages on Default Rates." *American Real Estate and Urban Economics Association Journal*, 15(3): 152-67.

Clauretie, Terrence M. and Thomas Herzog. 1990. "The Effect of State Foreclosure Laws on Loan Losses: Evidence from the Mortgage Industry." *Journal of Money, Credit, and Banking*, 22(2): 221-33.

Crawford, Gordon W. and Eric Rosenblatt. 1995. "Efficient Mortgage Default Option Exercise: Evidence from Loss Severity." *Journal of Real Estate Research,* 10(5): 543-55.

Davis, Morris A., Andreas Lehnert, and Robert F. Martin. 2008. "The Rent-Price Ratio for the Aggregate Stock of Owner-Occupied Housing." *Review of Income and Wealth*, 54(2): 279-84.

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

Feldstein, Martin. 2008. "How to Help People Whose Home Values Are Underwater." *Wall Street Journal*, November 18th, A21.

Foote, Christopher L., Kristopher S. Gerardi, Paul S. Willen. 2008. "Negative Equity and Foreclosure: Theory and Evidence." *Journal of Urban Economics*, 64(2): 234-45.

Foote, Christopher L., Kristopher S. Gerardi, Lorenz Goette, and Paul S. Willen. 2009. "Reducing Foreclosures." Working Paper No. 09-2, Federal Reserve Bank of Boston.

Hatchondo, Juan C., Leonardo Martinez, and Juan M. Sanchez. 2009. "Mortgage Defaults and Welfare." Manuscript, Federal Reserve Bank of Richmond.

Jones, Lawrence D. 1993. "Deficiency Judgments and the Exercise of the Default Option in Home Mortgage Loans." *Journal of Law and Economics*, 36(1): 115-38.

Kau, James B., Donald C. Keenan, and Taewon Kim. 1994. "Default Probabilities for Mortgages." *Journal of Urban Economics*, 35(3): 278-296.

Kau, James B., Donald C. Keenan, Walter J. Muller, and James F. Epperson. 1992. "A Generalized Valuation Model for Fixed-Rate Residential Mortgages." *Journal of Money, Credit, and Banking*, 24(3): 279-99.

Larsen, James E. with Bill Carey and Chantal Howell Carey. 2007. *Real Estate*. Hoboken: Wiley.

Ling, David C. and Wayne R. Archer. 2008. *Real Estate Principles: A Value Approach*. 2nd Ed. New York: McGraw-Hill Irwin.

Pence, Karen M. 2006. "Foreclosing on Opportunity: State Laws and Mortgage Credit." *Review of Economics and Statistics*, 88(1): 177-82.

USFN. 2004. *The National Mortgage Servicer's Reference Directory*. USFN: Tustin, CA.

White, Michelle J. 1998. "Why it Pays to File for Bankruptcy: A Critical Look at the Incentives Under US Personal Bankruptcy Law and a Proposal for Change." *The University of Chicago Law Review*, 65(3): 685-732.

Table 1: State Foreclosure Laws

State	Judicial or Non-Judicial Foreclosure	Optimum Timeline*	Recourse Classification	State	Judicial or Non-Judicial Foreclosure	Optimum Timeline*	Recourse Classification
Alabama	NJ	49-74	Recourse	Nebraska	NJ	121	Recourse
Alaska	NJ	108-111	Non-Recourse	Nebraska	J	176	
Arizona	NJ	115	Non-Recourse	Nevada	Nj	116	Recourse
Arkansas	NJ	90	Recourse	New Hampshire	NJ	75	Recourse
California	NJ	120	Non-Recourse	New Jersey	J	295	Recourse
Colorado	NJ	173	Recourse	New Mexico	J	225	Recourse
Connecticut	J, strict	160	Recourse	New York (NYC)	J	445	Recourse
Connecticut	J, by decree of sale	235		New York (Outside NYC)	J	299	
DC	NJ	48	Recourse	New York (Outside NYC)	NJ	355	
Delaware	J	200-300	Recourse	North Carolina Purchase Mortgages	NJ	120	Non-Recourse
Florida	J	150	Recourse	North Carolina Other Mortgages	NJ	120	Recourse
Georgia	NJ	48	Recourse	North Dakota	J	150	Non-Recourse
Hawaii	NJ	195	Recourse	Ohio	J	217	Recourse
Hawaii	J	320		Oklahoma	NJ	201	Recourse
Idaho	NJ	150	Recourse	Oregon	NJ	160	Non-Recourse
Illinois	J	345	Recourse	Pennsylvania	J	300	Recourse
Indiana	J	266	Recourse	Rhode Island	NJ	74	Recourse
Iowa	J	180	Non-Recourse	South Carolina	J	180	Recourse
Kansas	J	230	Recourse	South Dakota	J	340	Recourse
Kentucky	J	198	Recourse	Tennessee	NJ	50-55	Recourse
Louisiana	J, executory process	209	Recourse	Texas	NJ	35-60	Recourse
Louisiana	J, non-executory	269		Texas	J	80-180	
Maine	J	270	Recourse	Utah	NJ	139	Recourse
Maryland	J	46	Recourse	Vermont	J	275	Recourse
Massachusetts	J	75	Recourse	Virginia	NJ	60	Recourse
Michigan	NJ	360**	Recourse	Washington	NJ	140-150	Non-Recourse
Minnesota	NJ	270-280***	Non-Recourse	West Virginia	NJ	120	Recourse
Missouri	NJ	61-65	Recourse	Wisconsin	J	315	Non-Recourse
Montana	NJ	163	Non-Recourse	Wyoming	NJ	180	Recourse
Mississippi	NJ	90	Recourse				

Notes: * These are optimum timelines from The National Mortgage Servicer's Reference Directory, 21st edition (2004). The optimum timelines assume no delays and are based on uncontested foreclosure actions. ** The non-judicial foreclosure optimally takes 60 days; however, after that the redemption period begins to run, typically for 6 months. Estimated time for completion for uncontested foreclosure without eviction action is 12 months. ***The sale in non-judicial foreclosure can generally be held within 90 days; however, there are substantial redemption rights in Minnesota. Thus, including the redemption period the optimum timeframe for non-judicial foreclosure is 270-280 days.

Table 2: Recourse and the Default Method

Initial LTV	Recovery Percent: Non-recourse	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%
	Uncontested Default Time = 6 months (Benchmark)														
150%	F	DIL	DIL	SH	SH	SH	SH	F	F	F	F	F	F	F	F
145%	F	DIL	DIL	SH	SH	SH	SH	SH	F	F	F	F	F	F	F
140%	F	DIL	DIL	DIL	SH	SH	SH	SH	SH	F	F	F	F	F	F
135%	F	DIL	DIL	DIL	SH	SH	SH	SH	SH	F	F	F	F	F	F
130%	F	DIL	DIL	DIL	SH	SH	SH	SH	SH	SH	SH	F	F	F	F
125%	F	DIL	DIL	DIL	DIL	SH	SH	SH	SH	SH	SH	SH	F	F	F
120%	F	DIL	DIL	DIL	DIL	SH	SH	SH	SH	SH	SH	SH	SH	SH	F
115%	F	DIL	DIL	DIL	DIL	DIL	SH	SH	SH	SH	SH	SH	SH	SH	SH
110%	F	DIL	DIL	DIL	DIL	DIL	DIL	DIL	SH	SH	SH	SH	SH	SH	SH
105%	F	DIL	DIL	DIL	DIL	DIL	DIL	DIL	DIL	DIL	DIL	SH	SH	SH	SH
100%	F	F	DIL	DIL	DIL	DIL	DIL	DIL	DIL	DIL	DIL	DIL	DIL	DIL	SH
	Uncontested Default Time = 3 months														
150%	F	DIL	DIL	SH	SH	SH	SH	F	F	F	F	F	F	F	F
145%	F	DIL	DIL	DIL	SH	SH	SH	F	F	F	F	F	F	F	F
140%	F	DIL	DIL	DIL	SH	SH	SH	SH	F	F	F	F	F	F	F
135%	F	DIL	DIL	DIL	SH	SH	SH	SH	SH	F	F	F	F	F	F
130%	F	DIL	DIL	DIL	DIL	SH	SH	SH	SH	SH	F	F	F	F	F
125%	F	DIL	DIL	DIL	DIL	SH	SH	SH	SH	SH	SH	F	F	F	F
120%	F	DIL	DIL	DIL	DIL	DIL	DIL	SH	SH	SH	SH	SH	SH	SH	F
115%	F	DIL	DIL	DIL	DIL	DIL	DIL	SH	SH	SH	SH	SH	SH	SH	SH
110%	F	DIL	DIL	DIL	DIL	DIL	DIL	DIL	SH	SH	SH	SH	SH	SH	SH
105%	F	DIL	DIL	DIL	DIL	DIL	DIL	DIL	DIL	DIL	DIL	SH	SH	SH	SH
100%	F	F	F	DIL	DIL	DIL	DIL	DIL	DIL	DIL	DIL	DIL	DIL	DIL	SH
	Uncontested Default Time = 12 months														
150%	F	DIL	DIL	SH	SH	SH	SH	SH	F	F	F	F	F	F	F
145%	F	DIL	DIL	SH	SH	SH	SH	SH	F	F	F	F	F	F	F
140%	F	DIL	DIL	SH	SH	SH	SH	SH	SH	F	F	F	F	F	F
135%	F	DIL	DIL	SH	SH	SH	SH	SH	SH	SH	F	F	F	F	F
130%	F	DIL	DIL	DIL	SH	SH	SH	SH	SH	SH	SH	F	F	F	F
125%	F	DIL	DIL	DIL	SH	SH	SH	SH	SH	SH	SH	SH	F	F	F
120%	F	DIL	DIL	DIL	SH	SH	SH	SH	SH	SH	SH	SH	SH	F	F
115%	F	DIL	DIL	DIL	DIL	DIL	SH	SH	SH	SH	SH	SH	SH	SH	SH
110%	F	DIL	DIL	DIL	DIL	DIL	DIL	SH	SH	SH	SH	SH	SH	SH	SH
105%	F	DIL	DIL	DIL	DIL	DIL	DIL	SH	SH	SH	SH	SH	SH	SH	SH
100%	F	DIL	DIL	DIL	DIL	DIL	DIL	DIL	DIL	SH	SH	SH	SH	SH	SH

Notes: F = contested foreclosure with a deficiency judgment, DIL = deed in lieu or friendly foreclosure without a deficiency judgment, SH = short sale. Initial LTV refers to the LTV at the time of default.

Table 3: Recourse and the Decision to Default

Initial LTV	Recovery Percent														
	Non-recourse	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%	55%	60%	65%	70%
	Uncontested Default Time = 6 months (Benchmark)														
150%	D	D	D	D	D	D	D	D	D	D	D	D	ND	ND	ND
145%	D	D	D	D	D	D	D	D	D	D	D	ND	ND	ND	ND
140%	D	D	D	D	D	D	D	D	D	D	ND	ND	ND	ND	ND
135%	D	D	D	D	D	D	D	D	D	ND	ND	ND	ND	ND	ND
130%	D	D	D	D	D	D	D	D	D	D	D	ND	ND	ND	ND
125%	D	D	D	D	D	D	D	D	D	D	D	D	ND	ND	ND
120%	D	D	D	D	D	D	D	D	D	D	D	D	D	D	ND
115%	D	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
110%	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
105%	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
100%	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
	Uncontested Default Time = 3 months														
150%	D	D	D	D	D	D	D	D	D	D	D	D	ND	ND	ND
145%	D	D	D	D	D	D	D	D	D	D	D	ND	ND	ND	ND
140%	D	D	D	D	D	D	D	D	D	D	ND	ND	ND	ND	ND
135%	D	D	D	D	D	D	D	D	D	ND	ND	ND	ND	ND	ND
130%	D	D	D	D	D	D	D	D	D	D	ND	ND	ND	ND	ND
125%	D	D	D	D	D	D	D	D	D	D	D	ND	ND	ND	ND
120%	D	D	D	D	D	D	D	D	D	D	D	D	D	D	ND
115%	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
110%	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
105%	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
100%	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
	Uncontested Default Time = 12 months														
150%	D	D	D	D	D	D	D	D	D	D	D	ND	ND	ND	ND
145%	D	D	D	D	D	D	D	D	D	D	D	ND	ND	ND	ND
140%	D	D	D	D	D	D	D	D	D	D	ND	ND	ND	ND	ND
135%	D	D	D	D	D	D	D	D	D	D	ND	ND	ND	ND	ND
130%	D	D	D	D	D	D	D	D	D	D	D	ND	ND	ND	ND
125%	D	D	D	D	D	D	D	D	D	D	D	D	ND	ND	ND
120%	D	D	D	D	D	D	D	D	D	D	D	D	D	ND	ND
115%	D	D	D	D	D	D	ND	ND	ND	ND	ND	ND	ND	ND	ND
110%	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
105%	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
100%	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND

Notes: D = default, ND = no default. Initial LTV refers to the LTV at the time of default.

Table 4: Recourse and the Probability of Default

	(1) No Recourse Dummies	(2) Benchmark	(3) Rates in Interactions	(4) State Dummies	(5) Foreclosure Timing
Default Option	**0.61**	**1.44**	**1.59**	**1.43**	**1.46**
	(0.26)	(0.22)	(0.24)	(0.22)	(0.20)
Default Option Squared	**-1.22**	**-2.27**	**-3.08**	**-2.15**	**-2.29**
	(0.33)	(0.33)	(0.43)	(0.34)	(0.33)
Default Option * Recourse	-	**-1.46**	**-1.44**	**-1.62**	**-1.50**
		(0.49)	(0.49)	(0.40)	(0.45)
Default Option Sq. * Recourse	-	**1.74**	**1.41**	**2.03**	**1.78**
		(0.64)	(0.63)	(0.58)	(0.62)
Rate 1	**-0.326**	**-0.318**	-	**-0.327**	**-0.318**
	(0.0288)	(0.027)		(0.025)	(0.026)
Rate 2	**-0.295**	**-0.290**	-	**-0.304**	**-0.291**
	(0.040)	(0.038)		(0.036)	(0.037)
Rate 4	**0.302**	**0.302**	-	**0.314**	**0.302**
	(0.011)	(0.011)		(0.0089)	(0.011)
Rate 5	**0.436**	**0.440**	-	**0.456**	**0.441**
	(0.015)	(0.0157)		(0.014)	(0.015)
Default Option * Rate 1	-	-	**-3.43**	-	-
			(1.14)		
Default Option * Rate 2	-	-	**-2.81**	-	-
			(0.65)		
Default Option * Rate 4	-	-	**0.76**	-	-
			(0.06)		
Default Option * Rate 5	-	-	**0.92**	-	-
			(0.12)		
Divorce Rate	0.026	0.025	0.027	-	**0.029**
	(0.015)	(0.015)	(0.014)		(0.014)
Lagged Unemp Rate	0.013	0.012	0.020	**-0.057**	0.010
	(0.018)	(0.020)	(0.019)	(0.013)	(0.016)
Fico Score at Origination	**-0.274**	**-0.275**	**-0.352**	**-0.274**	**-0.275**
	(0.015)	(0.016)	(0.022)	(0.016)	(0.016)
Interest Only Dummy	**0.214**	**0.206**	**0.125**	**0.206**	**0.207**
	(0.039)	(0.037)	(0.039)	(0.035)	(0.036)
Jumbo Dummy	0.0513	0.0367	-0.0197	**0.0237**	0.0379
	(0.0297)	(0.0253)	(0.0292)	(0.0116)	(0.0254)
Mortgage Type Dummy	**-0.757**	**-0.781**	**-1.179**	**-0.765**	**-0.782**
	(0.071)	(0.084)	(0.086)	(0.085)	(0.084)
ARM Dummy	**0.274**	**0.271**	**0.339**	**0.247**	**0.271**
	(0.010)	(0.010)	(0.011)	(0.008)	(0.009)
LTV Ratio at Origination	**0.0104**	**0.0110**	**0.0130**	**0.0110**	**0.0110**
	(0.0008)	(0.0012)	(0.0011)	(0.0012)	(0.0011)
Ln Loan Age	**0.0847**	**0.0836**	**0.0710**	**0.0950**	**0.0837**
	(0.0053)	(0.0049)	(0.0051)	(0.0072)	(0.0048)
Purpose Type Dummy	**-0.0967**	**-0.0983**	**-0.1056**	**-0.1113**	**-0.0994**
	(0.0199)	(0.0207)	(0.0207)	(0.0199)	(0.0204)

Table 4 (Continued) : Recourse and the Probability of Default

	(1) No Recourse Dummies	(2) Benchmark	(3) Rates in Interactions	(4) State Dummies	(5) Foreclosure Timing
Foreclosure Timing	-	-	-	-	0.00378 (0.00758)
Constant	**-2.06** (0.14)	**-2.05** (0.15)	**-1.20** (0.18)	**-1.53** (0.14)	**-2.08** (0.17)
% Defaults	0.0464%	0.0464%	0.0464%	0.0464%	0.0464%
Log ps. likelihood	-283,742	-283,476	-289,682	-280,881	-283,450
Pseudo R-squared	14.9%	15.0%	13.1%	15.8%	15.0%
Number of obs.	82,828,381	82,828,381	82,828,381	82,828,381	82,828,381

Notes: The dependent variable in the probit is a binary variable that takes a value of 1 if the loan defaults in that month, 0 otherwise. Default Option is the probability that the borrower has negative home equity. Recourse is a dummy variable that takes a value of 1 if the property is in a recourse state, 0 otherwise; for North Carolina, recourse takes a value of 1 if the loan is not a purchase mortgage, 0 otherwise. The rate variables control for the difference between the current mortgage rate and the contract rate. Mortgage Type Dummy takes a value of 1 if the mortgage is a first mortgage, 0 otherwise. Purpose Type Dummy takes a value of 1 if the loan is not a purchase mortgage, 0 otherwise. Standard errors are in parentheses. The coefficients and standard errors for Fico Score at Origination show the effect of a 100 point increase in the FICO score. Standard errors are clustered by state. Coefficients in bold font are significant at the 5% level. % Defaults is the percentage of observations that are defaults. The standard error on Foreclosure Timing when the standard errors are not clustered is 0.00053.

Table 5: Estimated Default Probabilities in Recourse and Non-Recourse States

	At Time of Default	All Loans Value of Default Option		
	At Mean of Default Option	At Mean	At 90th percentile	At 95th percentile
	(1)	(2)	(3)	(4)
		Benchmark Specification		
Default option value	4.02%	1.02%	0.30%	2.12%
Non-Recourse Def. Prob.	0.1191%	0.0083%	0.0079%	0.0088%
Recourse Def. Prob.	0.0988%	0.0078%	0.0078%	0.0078%
Ratio NR/R	121%	106%	102%	113%
		By Appraisal Amount		
		$200,000 to $300,000		
Default option value	3.90%	0.89%	0.23%	1.41%
Non-Recourse Def. Prob.	0.1028%	0.0058%	0.0056%	0.0059%
Recourse Def. Prob.	0.0823%	0.0054%	0.0055%	0.0054%
Ratio NR/R	125%	106%	102%	110%
		$300,000 to $500,000		
Default option value	5.57%	1.26%	0.26%	2.40%
Non-Recourse Def. Prob.	0.1136%	0.0042%	0.0040%	0.0045%
Recourse Def. Prob.	0.0713%	0.0037%	0.0039%	0.0035%
Ratio NR/R	159%	115%	103%	131%
		$500,000 to $750,000		
Default option value	6.20%	1.62%	0.34%	4.25%
Non-Recourse Def. Prob.	0.0887%	0.0036%	0.0034%	0.0042%
Recourse Def. Prob.	0.0461%	0.0029%	0.0032%	0.0024%
Ratio NR/R	192%	125%	105%	175%
		$750,000 to $1,000,000		
Default option value	3.70%	1.21%	0.12%	2.05%
Non-Recourse Def. Prob.	0.0373%	0.0034%	0.0032%	0.0035%
Recourse Def. Prob.	0.0225%	0.0028%	0.0032%	0.0025%
Ratio NR/R	166%	122%	102%	139%
		$1,000,000 <		
Default option value	3.06%	0.93%	0.04%	0.99%
Non-Recourse Def. Prob.	0.0205%	0.0051%	0.0050%	0.0052%
Recourse Def. Prob.	0.0200%	0.0051%	0.0050%	0.0051%
Ratio NR/R	103%	101%	100%	101%
		By Investor Type		
		Private Securitized		
Default option value	3.98%	2.21%	2.40%	11.32%
Non-Recourse Def. Prob.	0.5539%	0.0214%	0.0216%	0.0295%
Recourse Def. Prob.	0.4463%	0.0182%	0.0181%	0.0143%
Ratio NR/R	124%	118%	119%	206%
		Private Portfolio		
Default option value	6.95%	2.24%	1.49%	12.17%
Non-Recourse Def. Prob.	0.0537%	0.0068%	0.0065%	0.0118%
Recourse Def. Prob.	0.0405%	0.0061%	0.0060%	0.0070%
Ratio NR/R	133%	112%	108%	168%

Note: The benchmark specification is specification (2) from table 5. The probabilities are estimated at the modes for dummy variables and means for the variables other than the default option value and default option value squared. In column 1, we estimate the probabilities at the modes of dummy variables and the means of all variables at the time of default for defaulted loans. Ratio is the ratio of the probabilities in non-recourse and recourse states.

Table 6: Recourse and the Probability of Default by Appraisal Amount

	All	< $100,000	$100,000 to $200,000	$200,000 to $300,000	$300,000 to $500,000	$500,000 to $750,000	$750,000 to $1,000,000	> $1,000,000
Default Option	**1.44**	0.65	0.32	**1.41**	**1.50**	**1.42**	**1.19**	**1.01**
	(0.22)	(0.39)	(0.26)	(0.20)	(0.25)	(0.23)	(0.23)	(0.27)
Default Option	**-2.27**	**-1.46**	-0.49	**-2.14**	**-2.46**	**-2.55**	**-2.11**	**-1.61**
Squared	(0.33)	(0.62)	(0.38)	(0.34)	(0.25)	(0.28)	(0.22)	(0.28)
Default Option *	**-1.46**	0.24	-0.01	**-1.75**	**-2.75**	**-3.30**	**-3.88**	-0.16
Recourse	(0.49)	(0.39)	(0.33)	(0.44)	(0.46)	(0.69)	(1.24)	(1.23)
Default Option Sq. *	**1.74**	-0.22	-0.45	**1.92**	**3.37**	**4.52**	**5.02**	-1.98
Recourse	(0.64)	(0.69)	(0.39)	(0.57)	(0.57)	(0.88)	(1.39)	(2.68)
Rate 1	**-0.318**	-0.125	**-0.268**	**-0.279**	**-0.368**	**-0.449**	**-0.250**	-0.089
	(0.027)	(0.096)	(0.046)	(0.033)	(0.048)	(0.055)	(0.090)	(0.073)
Rate 2	**-0.290**	**-0.147**	**-0.232**	**-0.300**	**-0.327**	**-0.463**	**-0.337**	**-0.279**
	(0.038)	(0.028)	(0.028)	(0.031)	(0.066)	(0.046)	(0.124)	(0.069)
Rate 4	**0.302**	**0.229**	**0.237**	**0.298**	**0.379**	**0.381**	**0.525**	**0.440**
	(0.011)	(0.016)	(0.013)	(0.019)	(0.024)	(0.017)	(0.052)	(0.053)
Rate 5	**0.440**	**0.356**	**0.364**	**0.455**	**0.556**	**0.594**	**0.713**	**0.719**
	(0.016)	(0.025)	(0.020)	(0.021)	(0.028)	(0.033)	(0.069)	(0.072)
Divorce Rate	0.025	0.018	-0.009	0.031	**0.069**	**0.109**	**0.065**	0.012
	(0.015)	(0.022)	(0.019)	(0.016)	(0.024)	(0.029)	(0.030)	(0.026)
Lagged Unemp Rate	0.012	**0.043**	0.015	-0.005	-0.014	-0.015	-0.048	-0.029
	(0.020)	0.021	0.022	0.015	0.013	0.022	0.030	0.027
Fico Score at	**-0.275**	**-0.218**	**-0.296**	**-0.309**	**-0.295**	**-0.300**	**-0.285**	**-0.296**
Origination	(0.016)	(0.020)	(0.015)	(0.012)	(0.015)	(0.025)	(0.020)	(0.029)
Interest Only	**0.206**	-0.025	**0.160**	**0.231**	**0.247**	**0.246**	**0.244**	**0.183**
Dummy	(0.037)	(0.029)	(0.026)	(0.030)	(0.023)	(0.025)	(0.025)	(0.071)
Mortgage Type	**-0.781**	**-0.484**	**-0.746**	**-0.851**	**-0.999**	**-1.080**	**-0.767**	**-0.552**
Dummy	(0.085)	(0.059)	(0.056)	(0.075)	(0.135)	(0.203)	(0.189)	(0.084)
ARM Dummy	**0.271**	**0.266**	**0.294**	**0.306**	**0.272**	**0.138**	**0.083**	0.021
	(0.010)	(0.014)	(0.017)	(0.020)	(0.013)	(0.020)	(0.032)	(0.062)
LTV Ratio at	**0.0110**	**0.0064**	**0.0096**	**0.0117**	**0.0158**	**0.0198**	**0.0156**	**0.0114**
Origination	(0.0012)	(0.0007)	(0.0006)	(0.0009)	(0.0021)	(0.0037)	(0.0031)	(0.0012)
Ln Loan Age	**0.0836**	**0.0339**	**0.0868**	**0.1120**	**0.1278**	**0.1097**	**0.0870**	**0.1073**
	(0.0049)	(0.0061)	(0.0062)	(0.0086)	(0.0142)	(0.0226)	(0.0142)	(0.0236)
Purpose Type	**-0.0983**	**-0.0523**	**-0.0833**	**-0.1102**	**-0.1536**	**-0.1504**	**-0.0920**	**-0.0956**
Dummy	(0.0207)	(0.0155)	(0.0171)	(0.0201)	(0.0296)	(0.0157)	(0.0385)	(0.0245)
Constant	**-2.05**	**-2.25**	**-1.75**	**-1.87**	**-2.20**	**-2.37**	**-2.15**	**-1.86**
	(0.15)	(0.16)	(0.18)	(0.14)	(0.12)	(0.17)	(0.28)	(0.22)
% Defaults	0.046%	0.102%	0.044%	0.035%	0.042%	0.043%	0.027%	0.021%
Log ps. likelihood	-283,476	-58,922.51	-91,326.18	-50,894.29	-53,570.32	-20,252.92	-3,938.02	-2,778.57
Pseudo R-squared	15.0%	11%	14%	16%	18%	18%	16%	13%
Number of obs.	82,828,381	8,231,808	27,580,379	19,146,465	17,840,518	6,567,914	1,858,369	1,602,928

Notes: The dependent variable in the probit is a binary variable that takes a value of 1 if the loan defaults in that month, 0 otherwise. Default Option is the probability that the borrower has negative home equity. Recourse is a dummy variable that takes a value of 1 if the property is in a recourse state, 0 otherwise; for North Carolina, recourse takes a value of 1 if the loan is not a purchase mortgage, 0 otherwise. The benchmark specification is specification (2) from table 5. The coefficients and standard errors for Fico Score at Origination show the effect of a 100 point increase in the FICO score. The rate variables control for the difference between the current mortgage rate and the contract rate. Mortgage Type Dummy takes a value of 1 if the mortgage is a first mortgage, 0 otherwise. Purpose Type Dummy takes a value of 1 if the loan is not a purchase mortgage, 0 otherwise. % Defaults is the percentage of observations that are defaults. Standard errors are in parentheses. Standard errors are clustered by state. Coefficients in bold font are significant at the 5% level. Appraisal amounts are in 2005 dollars.

Table 7: Recourse and the Probability of Default by Investor Type

	All	GNMA	Fannie Mae (FNMA) All	Fannie Mae (FNMA) Appraisal >$200,000	Freddie Mac (FHMLC) All	Freddie Mac (FHMLC) Appraisal >$200,000	Private Securitized	Private Portfolio
Default Option	**1.44**	4.32	**1.41**	**1.97**	**1.54**	**1.94**	**1.27**	**1.69**
	(0.22)	(2.39)	(0.33)	(0.28)	(0.39)	(0.27)	(0.22)	(0.18)
Default Option	**-2.27**	-13.18	**-1.56**	**-2.39**	**-2.27**	**-2.75**	**-2.34**	**-2.09**
Squared	(0.33)	(9.33)	(0.47)	(0.34)	(0.62)	(0.30)	(0.32)	(0.21)
Default Option *	**-1.46**	-1.07	0.07	-0.69	0.11	0.51	**-1.98**	**-1.23**
Recourse	(0.49)	(2.41)	(0.32)	(0.57)	(0.52)	(0.71)	(0.61)	(0.41)
Default Option Sq. *	**1.74**	5.78	-0.69	-0.02	-0.60	-2.19	**2.53**	**1.29**
Recourse	(0.64)	(9.61)	(0.62)	(0.69)	(0.70)	(1.76)	(0.79)	(0.50)
Rate 1	**-0.318**	-	**-0.128**	**-0.138**	**-0.143**	-0.211	**-0.416**	**-0.256**
	(0.027)		(0.039)	(0.067)	(0.066)	(0.144)	(0.028)	(0.059)
Rate 2	**-0.290**	**-0.206**	**-0.163**	**-0.159**	**-0.142**	**-0.133**	**-0.393**	**-0.298**
	(0.038)	(0.063)	(0.020)	(0.021)	(0.027)	(0.038)	(0.039)	(0.053)
Rate 4	**0.302**	0.073	**0.266**	**0.317**	**0.275**	**0.343**	**0.249**	**0.348**
	(0.011)	(0.090)	(0.019)	(0.020)	(0.0185)	(0.043)	(0.013)	(0.019)
Rate 5	**0.440**	0.223	**0.444**	**0.501**	**0.526**	**0.608**	**0.326**	**0.566**
	(0.016)	(0.132)	(0.024)	(0.058)	(0.032)	(0.068)	(0.012)	(0.028)
Divorce Rate	0.025	**0.064**	0.005	0.002	0.012	0.017	0.029	**0.044**
	(0.015)	(0.032)	(0.023)	(0.016)	(0.020)	(0.017)	(0.017)	(0.019)
	0.012	**-0.080**	0.021	-0.002	0.011	-0.016	0.017	0.011
Lagged Unemp Rate	(0.020)	(0.029)	(0.026)	(0.025)	(0.022)	(0.024)	(0.018)	(0.022)
Fico Score at	**-0.275**	**-0.138**	**-0.302**	**-0.324**	**-0.270**	**-0.284**	**-0.259**	**-0.179**
Origination	(0.016)	(0.038)	(0.014)	(0.011)	(0.017)	(0.018)	(0.010)	(0.028)
Interest Only	**0.206**	-	**0.195**	**0.279**	**0.291**	**0.363**	**0.150**	**0.240**
Dummy	(0.037)		(0.0645)	(0.067)	(0.076)	(0.085)	(0.037)	(0.042)
Mortgage Type	**-0.781**	-	-	-	-	-	**-0.850**	**-0.249**
Dummy	(0.085)						(0.097)	(0.101)
ARM Dummy	**0.271**	0.134	**0.149**	**0.173**	0.044	0.041	**0.240**	**0.200**
	(0.010)	(0.324)	(0.017)	(0.027)	(0.034)	(0.052)	(0.012)	(0.018)
LTV Ratio at	**0.0110**	0.0026	**0.0123**	**0.0128**	**0.0103**	**0.0102**	**0.0115**	**0.0086**
Origination	(0.0012)	(0.0023)	(0.0009)	(0.0011)	(0.0009)	(0.0011)	(0.0016)	(0.0014)
Ln Loan Age	**0.0836**	**0.1881**	**0.0827**	**0.1067**	**0.1152**	**0.1198**	**0.0846**	**0.1238**
	(0.0049)	(0.0385)	(0.0068)	(0.0106)	(0.0076)	(0.0103)	(0.0079)	(0.0078)
Purpose Type	**-0.0983**	-0.4019	0.0282	0.0109	**0.0751**	**0.0600**	**-0.1439**	**-0.0873**
Dummy	(0.0207)	(0.211)	(0.0152)	(0.0266)	(0.0244)	(0.0418)	(0.0172)	(0.0158)
% Defaults	0.046%	0.1167%	0.0179%	0.010%	0.0119%	0.0072%	0.1472%	0.0478%
Log ps. likelihood	-283,476	-985	-54,559	-16,861	-23,555	-7,772	-172,033	-28,420
Pseudo R-squared	15.0%	5.1%	10.0%	10.1%	7.9%	8.3%	10.9%	11.4%
Number of obs.	82,828,381	114,351	35,258,604	18,385,916	21,389,732	11,226,747	17,465,885	7,755,789

Notes: The dependent variable in the probit is a binary variable that takes a value of 1 if the loan defaults in that month, 0 otherwise. Default Option is the probability that the borrower has negative home equity. Recourse is a dummy variable that takes a value of 1 if the property is in a recourse state, 0 otherwise; for North Carolina, recourse takes a value of 1 if the loan is not a purchase mortgage, 0 otherwise. The benchmark specification is specification (2) from table 5. The coefficients and standard errors for Fico Score at Origination show the effect of a 100 point increase in the FICO score. The rate variables control for the difference between the current mortgage rate and the contract rate. Mortgage Type Dummy takes a value of 1 if the mortgage is a first mortgage, 0 otherwise. Purpose Type Dummy takes a value of 1 if the loan is not a purchase mortgage, 0 otherwise. % Defaults is the percentage of observations that are defaults. Standard errors are in parentheses. Standard errors are clustered by state. Coefficients in bold font are significant at the 5% level. All regressions include a constant.

Table 8: Recourse and the Type of Default

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Recourse	**-0.347**	**-0.460**	**-0.313**	**-0.449**	**-0.333**	**-0.436**	**-0.383**	**-0.450**	**-0.266**	**-0.371**
	(0.119)	(0.142)	(0.105)	(0.133)	(0.090)	(0.118)	(0.116)	(0.121)	(0.136)	(0.125)
Default Option	-	-	0.365	**-0.942**	-	**-0.950**	-	**-0.899**	-	**-1.001**
			(0.188)	(0.214)		(0.213)		(0.229)		(0.190)
Default Option * Recourse	-	-	-0.80	**-1.70**	-	**-1.68**	-	**-1.68**	-	**-1.63**
			(0.56)	(0.47)		(0.42)		(0.41)		(0.41)
Fico Score at Origination	-	0.00937	-	0.01596	-	0.014890	-	-0.001760	-	0.001060
		(0.02750)		(0.03280)		(0.035000)		(0.04040)		(0.03230)
LTV Ratio at Origination	-	**0.0255**	-	**0.0285**	-	**0.0285**	-	**0.0283**	-	**0.0287**
		(0.0026)		(0.0023)		(0.0024)		(0.0024)		(0.0023)
Foreclosure Timing Dummy	-	-	-	-	0.0469	0.0450	-	0.0674	-	0.0297
					(0.2226)	(0.2555)		(0.2534)		(0.2547)
Investor Type 1	-	-	-	-	-	-	**0.263**	**0.131**	-	
							(0.076)	(0.050)		
Appraisal Amount	-	-	-	-	-	-	-	-	**0.0217**	**0.0341**
									(0.0048)	(0.0075)
Appraisal Amount * Recourse	-	-	-	-	-	-	-	-	-0.0241	-0.0107
									(0.0150)	(0.0225)
Constant	**1.05**	**-0.74**	**1.03**	**-0.93**	**1.01**	**-0.96**	**1.03**	**-0.87**	**0.98**	**-0.99**
	(0.01)	(0.32)	(0.02)	(0.31)	(0.20)	(0.33)	(0.03)	(0.37)	(0.03)	(0.31)
% Foreclosures	79.2%	79.2%	79.2%	79.2%	79.2%	79.2%	79.3%	79.3%	79.2%	79.2%
Log ps. likelihood	-19,437.70	-16,619	-19,420	-16,254	-19,433	-16,251	-19,245	-16,136	-19,431	-16,222
Pseudo R-squared	1.2%	15.5%	1.3%	17.4%	1.2%	17.4%	1.8%	17.7%	1.3%	17.6%
Number of obs.	38,444	38,444	38,444	38,444	38,444	38,444	38,390	38,390	38,444	38,444

Notes: The dependent variable in the probit is a binary variable that takes a value of 1 if default is by foreclosure, 0 otherwise. Default Option is the probability that the borrower has negative home equity. Recourse is a dummy variable that takes a value of 1 if the property is in a recourse state, 0 otherwise. The coefficients and standard errors for Fico Score at Origination show the effect of a 100 point increase in the FICO score.
Foreclosure timing dummy is a dummy variable that takes a value of 1 if the uncontested foreclosure time is less than six months, and zero otherwise. Investor type 1 is a dummy variable that takes a value of 1 if the lender type is not "Private Portfolio" or "Private Securitized", 0 otherwise. Appraisal amount is the appraisal amount of the property at origination; coefficients and standard errors shown are for the effect of a $100,000 increase in the appraisal amount. The number of observations in columns 7 and 8 differs from the number of observations in other columns because we exclude observations with investor type "Unknown" for these specifications. In all specifications standard errors are clustered by state the property is located in. Coefficients in bold font are significant at the 5% level. Appraisal amounts are in 2005 dollars.

A Foreclosure Laws State by State

Alabama: Lenders may foreclose through either a judicial or a non-judicial procedure. State law permits deficiency judgments without significant restrictions. We classify Alabama as a RECOURSE state. The borrower retains a right of redemption for one year after foreclosure. The relevant statutes are in section 35-10 of the Alabama Code.

Alaska: Lenders may foreclose through either a judicial or a non-judicial procedure. The usual financing instrument is a deed of trust and non-judicial foreclosure is the usual foreclosure process. State law permits deficiency judgments only if the lender pursues judicial foreclosure under the promissory note, no separate "deficiency judgment" is entered. The property sold at a judicial sale is subject to a right of redemption, and the redemption period is 12 months. As judicial foreclosure is substantially more time consuming and cumbersome, we classify Alaska as a NON-RECOURSE state. The relevant statutes are in Title 34, Ch. 20, Section 100 of the Alaska Statutes.

Arizona: Lenders may foreclose through either a judicial or a non-judicial procedure. The usual financing instrument is a deed of trust and non-judicial foreclosure is the usual foreclosure process. Deficiency judgments are not permitted if the property is residential and on 2.5 acres or less and its intended use was for a one-family dwelling or two-family dwelling. We classify Arizona as a NON-RECOURSE state. The relevant statute is Article 33 of the Arizona State Code.

Arkansas: Lenders may foreclose through either a judicial or a non-judicial procedure. Lenders usually foreclose on a deed of trust through a non-judicial procedure. State law permits deficiency judgments with the restriction that borrowers must receive credit for the greater of the foreclosure sales price or the fair market value of the property. We classify Arkansas as a RECOURSE state. The relevant statutes are in sections 18-50-212 and 18-50-216 of the Arkansas Code.

California: Lenders may foreclose through either a judicial or a non-judicial procedure. Non-judicial foreclosure is the usual foreclosure process. The borrower has five days to reinstate in a non-judicial foreclosure process. State law prohibits deficiency judgments on purchase mortgages. On other residential mortgages, state law permits deficiency judgments only if the lender pursues the more expensive and time-consuming judicial foreclosure process rather than the non-judicial foreclosure process. The lender may only file for a payment of the difference between the debt owed and the fair market value of the property. A deficiency suit also gives the borrower a right to redemption. We classify California as a NON-RECOURSE state. The relevant statutes are in sections 2920-2944.5 of the California Code.

Colorado: Lenders may foreclose through either a judicial or a non-judicial procedure. Non-judicial foreclosure is the norm. State law permits deficiency judgments. However, judges require lenders to bid fair market value on the property in the event that total debt owed exceeds the property value less reasonable expenses; if the borrower can show that lenders bid less than fair market value, the borrower can avoid a deficiency judgment. After the sale there is a redemption period of 75 days. There are no unreasonably burdensome statutory limitations on either filing or collecting on a deficiency or collection. We classify Colorado as a RECOURSE state. The relevant statutes are Title 38, Articles 37-39 of the Colorado Revised Statutes.

Connecticut: Lenders may foreclose only through one of two judicial procedures. The two procedures are a strict foreclosure and a decree of sale foreclosure. State law permits deficiency judgments under both procedures; however, if the lender pursues decree of sale foreclosure the lender must first credit the borrower with one-half the difference between the debt and the appraised value if the property is sold pursuant to a court-order and the property sells for less than the appraised value. In strict foreclosure, the judge determines the fair market value of the property for which the borrower receives credit; a motion for deficiency judgment must be filed within 29 days of title vesting. There is no statutory deadline to file the motion for deficiency judgment after foreclosure-by-sale. We classify Connecticut as a RECOURSE state. The relevant statutes are sections 49-14 and 49-28 of the General Statutes of Connecticut.

Delaware: Lenders may foreclose only through a judicial procedure. State law permits deficiency judgments without significant restrictions. We classify Delaware as a RECOURSE state. The relevant statute is Title 10, Ch. 49:XI of the Delaware Code.

District of Columbia: Lenders may only foreclose through a non-judicial procedure. At any time within thirty days after the time limit for redemption has expired, any party to a mortgage foreclosure may file a motion seeking a deficiency judgment. We classify the District of Columbia as a RECOURSE district. The relevant statute is Title 42, Ch. 8 of the District of Columbia Code.

Florida: Lenders may foreclose only through judicial foreclosure. State law permits deficiency judgments subject to the borrower receiving credit for the greater of fair market value of the property or the foreclosure sale price. A deficiency judgment can be pursued against the original makers of a note even if they were not a party to the foreclosure action. However, Florida has an extremely generous homestead exemption such that if the property is an investment property, rather than a primary residence, the borrower can partially shield his or her assets from collection on the deficiency. We classify Florida as a RECOURSE state. The relevant statutes are Title 40, Ch. 702 of the Florida Statutes.

Georgia: Lenders may foreclose through either a judicial or a non-judicial procedure. Non-judicial foreclosure is the usual process. A prerequisite to a deficiency judgment is that the court has confirmed and approved the sale which in turn requires that the sale price was equal to at least the fair market value of the property. The lender must receive such confirmation and approval within 30 days of the foreclosure sale. There is no right of redemption. We classify Georgia as a RECOURSE state. The relevant statutes are in Title 44, Ch. 14 of the Official Code of Georgia.

Hawaii: Lenders may foreclose through either a judicial or a non-judicial procedure. A judicial foreclosure takes 320 days; non-judicial takes 195 days if uncontested. State law permits deficiency judgments if the lender pursues judicial foreclosure. The deficiency judgment process, if not contested, is fairly inexpensive. We classify Hawaii as a RECOURSE state. The relevant statutes are Ch. 667-5 and Ch. 667-38 of the Hawaii Revised Statutes.

Idaho: Lenders may foreclose through either a judicial or a non-judicial procedure although judicial foreclosure is exceptionally rare. State law permits a deficiency judgment provided one is filed within 90 days of the foreclosure sale. The deficiency is limited to the difference between the balance owed and the fair market value of the property. The deficiency judgment process is onerous in practice since the lender must prove fair market value and the borrower can contest the fair market value of the property. We classify Idaho as a RECOURSE state. The relevant statutes are in Idaho Statutes, Title 45, Ch. 15, section 45.12.

Illinois: Lenders may foreclose only through judicial foreclosure. State law permits deficiency judgments provided the borrower is personally served with the deficiency suit. Furthermore, a judge must confirm the sale and, according to chapter 735, article XV, section 15-1508, the judge may opt to not confirm the sale on the grounds that "justice was not otherwise done". In practice, this means that is at the discretion of the judge whether to grant a deficiency judgment and judges rarely grant deficiency judgments on residential property. We decided to classify Illinois as a RECOURSE state as the possibility of personal recourse may be sufficient to deter some strategic defaulters even if deficiency judgments are rarely granted. The relevant statutes are in chapter 735, article XV of the Illinois Compiled Statutes.

Indiana: Lenders may foreclose only through judicial foreclosure which optimally takes 266 days if uncontested. State law permits deficiency judgments on residential properties without significant restrictions. The borrower must be served in person which is not a significant restriction in practice. We classify Indiana as a RECOURSE state. The relevant statutes are in Article 29, chapter 7 of the Indiana State code.

Iowa: Lenders may foreclose only through judicial foreclosure. State law permits deficiency judgments on non-agricultural residential properties. However, seeking a deficiency judgment significantly delays the foreclosure process. Furthermore, there is a two year statute of limitations on collecting on the deficiency judgment and generous limits on garnishment of wages. The law makes it much faster to foreclosure on property if the lender waives the right to a deficiency judgment. Because deficiencies are hard to collect in Iowa, lenders may even compensate the borrower who agrees to vacate the property fast by paying the first month of rent on new housing. We classify Iowa as a NON-RECOURSE state. The relevant statute is Ch. 654.6 of the Iowa code. There was a bill pending that may change the foreclosure laws significantly as of March 2009.

Kansas: Lenders may foreclose only through judicial foreclosure. Following a foreclosure sale, a deficiency judgment is automatically entered if the sale proceeds less expenses are not sufficient to cover the debt owed. The borrower may contest the deficiency if the foreclosure sales price was less than the fair market value of the property. Kansas is unusual as redemption rights can be sold to third parties such that if the lender bids substantially less for the property than its fair market value, the holder of the redemption rights may obtain the property at significantly below market value. Further, second lien holders lose the right to a deficiency if they do not ask for a foreclosure themselves. We classify Kansas as a RECOURSE state. The relevant statute is Ch. 60, 2417 of the Kansas Statutes.

Kentucky: Lenders may foreclose only through judicial foreclosure. Following a foreclosure sale, a deficiency judgment is automatically entered if the sale proceeds less expenses are not sufficient to cover the debt owed. There are no significant restrictions. We classify Kentucky as a RECOURSE state. The relevant statutes are in Ch. 426 of the Kentucky Revised Statutes.

Louisiana: Lenders may foreclose only through judicial foreclosure. State law permits deficiency judgments on residential properties without significant restrictions. We classify Louisiana as a RECOURSE state. The relevant statutes are in Title 10:9-629 of the Louisiana Code.

Maine: Lenders may foreclose only through judicial foreclosure. State law permits deficiency judgments on residential properties provided the lender credits the borrower's account for fair market value of the property. We classify Maine as a RECOURSE state. The relevant statutes are in Title 14, part 4, Ch. 403 of the Revised Maine Statutes.

Maryland: Lenders may foreclose through either a judicial or a non-judicial procedure. State law permits deficiency judgments on residential properties without significant restrictions. We classify Maryland as a RECOURSE state. The relevant statutes are in the Maryland Rules, Title 14, Ch. 200.

Massachusetts: Lenders may foreclose through either a judicial or a non-judicial procedure. State law permits a deficiency judgment provided that the lender gives the borrower notice in writing prior to the foreclosure sale that he or she intends to pursue a deficiency. We classify Massachusetts as a RECOURSE state. The relevant statutes are in Ch. 244 of the General Laws of Massachusetts.

Michigan: Lenders may foreclose through either a judicial or a non-judicial procedure. There is typically a six month redemption period after the completion of a non-judicial foreclosure. State law permits a deficiency judgment without significant restrictions in the case of judicial foreclosure; in the case of non-judicial foreclosure, the borrower can contest the deficiency if the property sold for substantially less than the fair market value of the property. We classify Michigan as a RECOURSE state. Michigan Compiled Laws, Ch. 451; EPIC Act 236, Sections 600 and 700.

Minnesota: Lenders may foreclose through either a judicial or a non-judicial procedure although in the vast majority of cases lenders foreclose through a non-judicial process. There are substantial redemption rights in Minnesota. In particular, the mortgagor is entitled to a six- or twelve-month period after the foreclosure sale. The mortgagor is entitled to possession of the property and the lender has limited right to enter the property. The redemption period can be shortened to 6 months if certain conditions are met. A separate court procedure is required to shorten the redemption period to 5 weeks if the residential property is deemed “abandoned” and of less than 5 units and is on less than 10 acres. Thus, including the redemption period the optimum time-frame for non-judicial foreclosure is 270-280 days. In the event the lender forecloses by advertisement, state law prohibits deficiency judgments. In judicial foreclosure, the lender may obtain a deficiency judgment subject to the borrower receiving credit for the fair market value of the property. The fair market value of the property is determined by a jury. Because judicial foreclosure is substantially more onerous than the non-judicial procedure, lenders pursue non-judicial foreclosure in the vast majority of cases. We classify Minnesota as a NON-RECOURSE state. The relevant statutes are in 580 and 582 of the 2008 Minnesota Statutes and, particularly, 582.2, subdivision 2.

Mississippi: Lenders may foreclose on deeds of trusts or mortgages in default using either a judicial or non-judicial foreclosure process. State law permits a deficiency judgment provided the lender files for one within one year of the foreclosure sale date. If a mortgagee participates in foreclosure sale auction, his bid must pass a judicial standard of reasonableness. We classify Mississippi as a RECOURSE state. The relevant statutes are in section 89-1-305 of the Mississippi State Code.

Missouri: Lenders may foreclose through either a judicial or a non-judicial procedure. The state has a statutory right of redemption, but a burden on the borrower is prohibitively heavy and this right can be rarely exercised. In the case of non-judicial foreclosure sale a separate court action must be filed to obtain a deficiency judgment but there are no other significant restrictions on obtaining a deficiency judgment. We classify Missouri as a RECOURSE state. The relevant statutes are in the Missouri Revised Statutes, Chapter 141 sections 400-590.

Montana: Lenders may foreclose through either a judicial or a non-judicial procedure. Deficiency judgments are prohibited on purchase mortgages by title 71, chapter 1-232 of the Montana Code Annotated. Deficiency judgments are permitted on other types of residential mortgages only if the lender pursues judicial foreclosure; however, judicial foreclosure is often impractical because the grantor is entitled to a one year right of redemption. The non-judicial foreclosure process is also substantially less complicated and costly. We classify Montana as a NON-RECOURSE state. The relevant statutes are in title 71, chapter 1 of the Montana Code Annotated.

Nebraska: Lenders may foreclose through either a judicial or a non-judicial procedure. Lenders may obtain a deficiency judgment; however, the borrower must receive credit for the fair market value of the property and the deficiency must be filed for within 90 days of the foreclosure sale by non-judicial foreclosure and within 5 years in case of judicial foreclosure. We classify Nebraska as a RECOURSE state. The relevant statutes are in the Nebraska Revised Statutes Chapter 76-1013.

Nevada: Lenders may foreclose through either a judicial or a non-judicial procedure. Usually properties are foreclosed through a non-judicial procedure. A deficiency judgment can be obtained; however, the borrower must receive credit for the greater of the fair market value of the property, as determined through a hearing, or the foreclosure sale price. The lender must file for a deficiency judgment with 90 days of the foreclosure sale. We classify Nevada as a RECOURSE state. The relevant statutes are in the Nevada Revised Statutes, chapters 40, 106, and 107.

New Hampshire: Lenders may foreclose through either a judicial or a non-judicial procedure. Almost all properties are foreclosed non-judicially. There are no significant restrictions on deficiency judgments. We classify New Hampshire as a RECOURSE state. The relevant statutes are in Title 38, chapter 479 of the New Hampshire Revised Statutes.

New Jersey: Lenders foreclose through a judicial process. State law permits deficiency judgments but the borrower must be given credit for the fair market value of the property and must be brought within three months of the foreclosure sale. The pursuit of a deficiency judgment extends the redemption period from 10 days to 6 months. We classify New Jersey as a RECOURSE state. The relevant statutes are in the New Jersey Permanent Statutes Title 2A, section 50.

New Mexico: Lenders foreclose on residential properties through a judicial process. Deficiency judgments on mortgages and deeds of trust other than those used to finance low-income housing can be obtained and there are no significant restrictions. We classify New Mexico as a RECOURSE state. The relevant statutes are in Ch. 48, Articles 48-7-1 to 48-7-24 and Articles 48-10-1 to 48-10-21 of the New Mexico Statutes Annotated.

New York: Lenders may foreclose through either a judicial or a non-judicial procedure, although non-judicial foreclosure is exceptionally rare. State law permits a deficiency judgment provided that the lender submits a request for a deficiency judgment within 90 days of filing the foreclosure suit. However, the borrower receives credit for the greater of the foreclosure sale price or the fair market value of the property. The judge usually sides with the borrower regarding the fair market value of the property. A typical deficiency judgment is relatively expensive. We classify New York as a RECOURSE state. The relevant statutes are in Article 13 of the New York State Consolidated Laws.

North Carolina: Lenders may foreclose through either a judicial or a non-judicial process. Ch. 45, Article 2B, section 21.38 of the North Carolina General Statutes prohibits deficiency judgments on purchase mortgages. We classify purchase mortgages in North Carolina as NON-RECOURSE. Deficiency judgments are permitted on other types of residential mortgages but the borrower has the right to contest the deficiency judgment such that he or she receives credit for the fair market value of the property. The deficiency judgment must be filed within 1 year. North Carolina law does not permit garnishment of wages to collect debt. We classify non-purchase mortgages in North Carolina as RECOURSE. The relevant statutes are sections 21.36 and 21.38 of Article 2B in Ch. 45 of the North Carolina General Statutes.

North Dakota: Lenders foreclose through a judicial process. Chapter 32-19-01 of the North Dakota Century Code prohibits deficiency judgments on residential properties. There is a provision for so called deficiency mortgages but the value must be determined by a juror trial and is not pursued in practice. We classify North Dakota as a NON-RECOURSE state.

Ohio: Lenders may foreclose only through judicial foreclosure. If the debt is greater than the foreclosure sales price plus reasonable expenses, a deficiency judgment is automatic. However, lenders have only two years to collect on the deficiency. We classify Ohio as a RECOURSE state. The relevant statutes are in the Ohio Revised Code, section 2329.08.

Oklahoma: Lenders may foreclose through either judicial or non-judicial foreclosure. The optimum time-frame for non-judicial foreclosure is 201 days. Lenders may only receive a deficiency judgment if they pursue non-judicial foreclosure and the borrower must receive credit for the greater of the fair market value or the foreclosure sale price. The lender must file for a deficiency judgment within 90 days of the foreclosure sale. We classify Oklahoma as a RECOURSE state. The relevant statute is Title 12, Chapter 12, section 686 of the Oklahoma Statutes Citationized.

Oregon: Lenders may foreclose through either a judicial or a non-judicial procedure. Lenders can generally not obtain a deficiency judgment on a residential property. We classify Oregon as a NON-RECOURSE state.

Pennsylvania: Lenders foreclose through a judicial procedure. Pennsylvania Law permits the lender to file for a deficiency judgment through a separate suit from the foreclosure but the borrower must receive credit for the fair market value of the property. The deficiency suit must be brought within six months of the foreclosure sale. We classify Pennsylvania as a RECOURSE state. The relevant statute is the Pennsylvania Deficiency Judgment Act, Chapter 81 Section 8103 of the Pennsylvania Consolidated Statutes.

Rhode Island: Lenders may foreclose through either a judicial or a non-judicial procedure. Deficiency judgments can be obtained and there are no significant restrictions. We classify Rhode Island as a RECOURSE state. The relevant statutes are in Ch. 34-27 of the Rhode Island General Laws.

South Carolina: Lenders foreclose through a judicial procedure. State law permits deficiency judgments subject to the restriction that the borrower receive may present a motion to receive credit for the fair market value of the property. In such a circumstance, the borrower, judge, and lender all hire appraisers to determine the fair market value of the property. We classify South Carolina as a RECOURSE state. The relevant statutes are in Title 29, Ch. 3, Article 7 of the South Carolina Code of Laws.

South Dakota: Lenders may foreclose through either a judicial or a non-judicial procedure. State law permits deficiency judgments provided the borrower is credited for the fair market value of the property. We classify South Dakota as a RECOURSE state. The relevant statutes are in ch. 21-47 of the South Dakota Codified Laws.

Tennessee: Lenders may foreclose through either a judicial or a non-judicial procedure although lenders seldom use the judicial foreclosure process. State law permits deficiency judgments without significant restrictions. We classify Tennessee as a RECOURSE state. The relevant statutes for non-judicial foreclosure are Title 21, Ch. 1, Section 803 of the Tennessee Code.

Texas: Lenders may foreclose through either a judicial or a non-judicial procedure. The lender must foreclose on a home equity loan through a judicial foreclosure process, however. State law permits deficiency judgments subject to the borrower receiving credit for the fair market value of the property. However, Texas has a nearly unlimited homestead exemption such that lenders have less recourse on mortgages backed by investment properties if the borrower's primary residence is also in Texas. We classify Texas as a RECOURSE state. The relevant statutes are in Title 5, Section 51 of Texas Statutes.

Utah: Lenders may foreclose through either a judicial or a non-judicial procedure. State law permits deficiency judgments without significant restrictions. We classify Utah as a RECOURSE state. The relevant statutes are in Title 38, Ch.1-16 and Title 57, Ch. 1 of the Utah Code.

Vermont: Lenders may foreclose through either a judicial or, if the mortgage contains a power of sale clause, a non-judicial procedure. The norm, however, is judicial foreclosure. State law permits deficiency judgments with no significant restrictions. We classify Vermont as a RECOURSE state. The relevant Vermont Statutes are in Title 12, Chapter 163.

Virginia: Lenders may foreclose through either a judicial or non-judicial process. State law permits deficiency judgments with no significant restrictions. We classify Virginia as a RECOURSE state. The relevant statutes are in Title 8.9A Part 6 and Title 55, Ch. 4 of the Code of Virginia.

Washington: Lenders may foreclose through either a judicial or non-judicial process. If the lender wishes to pursue a deficiency judgment, however, it must pursue judicial foreclosure and pursuit of a deficiency judgment triggers a 12 month right of redemption. Furthermore, the judicial foreclosure process is substantially more time-consuming than the non-judicial process. Deficiency judgments can also not be obtained if the property has been abandoned for six months or more which we view as one way a strategic defaulter could evade a deficiency judgment relatively easily. We classify Washington as a NON-RECOURSE state. The relevant statutes are in Title 61, Ch. 61-12 of the Revised Code of Washington.

West Virginia: Lenders may foreclose through either a judicial or non-judicial process. West Virginia permits deficiency judgments without significant restrictions. We classify West Virginia as a RECOURSE state. The relevant statutes are in Articles 1 and 16 of Ch. 38 of the West Virginia Code.

Wisconsin: Lenders foreclose through a non-judicial process. A deficiency judgment must be filed at the time the foreclosure action starts. A waiver of a deficiency judgment may reduce a redemption period of 12 months to 6 months, and a redemption period of 6 months to 3 months. The redemption period depends on a number of characteristics including parcel size. We classify Wisconsin as a NON-RECOURSE state. The relevant statutes can be found in Wisconsin Statutes and Annotations, Ch. 846.

Wyoming: Lenders may foreclose through either a judicial or non-judicial process. The lender generally bids the lesser of the debt owed or the fair market value for the property at a foreclosure sale. State law permits deficiency judgments without significant restrictions. We classify Wyoming as a RECOURSE state. The relevant statutes are in Title 34, Ch. 4 of the Wyoming Statutes.

B Data Description

A. Sample Restrictions

We restrict our analysis to mortgages with constant principal and interest, ARMs, or Graduated Payment Mortgages (GPM) (variable INT_TYPE takes values 1, 2, 5, respectively). Also, we restrict the analysis to mortgages taken for purchase or refinance (PURPOSE_TYPE_MCDASH variable takes values 1 = Purchase, 2 = Refinance (Cash out), 3 = Refinance (No cash out), 5 = Refinance (unknown cash). Mortgages for home improvement, debt consolidation, education, medical, or other were dropped. The analysis is limited to first or second mortgages (Variable MORT_TYPE takes values 1 = First mortgage, 2 = Second mortgage, 4 = First mortgage, grade "B" or "C", 5 = Second Mortgage (Home Equity), Grade "B" or "C"). We also drop all observations for loans on properties with more than one unit (Variable is UNITS_NO).

B. Variable Definitions

Definition of Default.—We consider the loan as defaulted if the loan is terminated in one of the following ways: by REO sale, by short sale, by pay off out of foreclosure, pay off out of bankruptcy and serious delinquency or by liquidation to termination. We do not count terminations by voluntary pay off, by a loan transfer from a servicer, or by a third party sale as defaults. The default month is determined as the first month the loan that defaulted was reported as being in foreclosure, in REO proceedings or under liquidation, whichever comes first (MBA_STAT variables takes values F, R, L, respectively). In addition, if the loan is terminated by default without loan status reported as any of the three mentioned above, the default month is the month when the loan is reported as paid off. Finally, if TERMINATION_TYPE=8, we count the loan as defaulted since the FORECLOSURE_TYPE for these variables is non-zero, indicating that there was a foreclosure although less than 0.1% of loans are terminated in this fashion.

In the analysis of the probability of default the dependent variable takes value 1 if it corresponds to the default month of the loan that defaulted. Thus, all subsequent to default month observations on the defaulted loans are dropped. Consequently, the dependent variable takes value 0 for all months that we observe the defaulted loans prior to the default month and all observations on loans that did not default, whether terminated or current. Observations on current mortgages or mortgages terminated not be default for which the value of the principal balance amount is 0, i.e. the balance is paid off, are dropped.

Default Type.—If a loan goes from being in foreclosure to being an REO loan, we treat that as a foreclosure. That is, we define a foreclosure as any loan for which MBA_STAT=F prior to it being any other MBA_STAT.

Default Option.—For the current principal balance amount we use variable PRIN_BAL_AMT (the balance the borrower owns on the loan); for the cost of a purchase we use variable ORIG_AMT (original loan amount). Loans for which the principal balance amount at the time of default (which is described bellow) is 0 or missing and cannot be imputed from up to two previous months are dropped from the analysis. To calculate k_i we use the loan closing date (CLOSE_DT; as is used by McDash).

The OFHEO provides a quarterly (not seasonally adjusted) measure of the House Price Index by state (http://www.ofheo.gov/hpi_download.aspx). The OFHEO provides A and B in quarters and so we convert months since origination into quarters since origination. We also construct monthly values of $HPI_{i,t}$ by linearly interpolating from the quarterly values attributing the quarterly value to the second month of the quarter.

Prepay Option Variables.—The ongoing contract rate on the mortgage is contained in variable CUR_INIT_RATE. (Variable ARM_INT_RATE contains initial interest rate on the loan; however, it is sparsely populated). The market mortgage rate is a contract rate on the composite of all conventional mortgage loans (fixed- and adjustable-rate) from the Finance Board's Monthly Survey of Rates and Terms on Conventional Single-Family Non-farm Mortgage Loans. The survey collects information on fully amortized conventional mortgage loans used to purchase single-family non-farm homes; mortgage loans insured by the Federal Housing Administration or guaranteed by the Veterans Administration are excluded. Also loans used to refinance houses and non-amortized and balloon loans are excluded. The data are available in Table 17, http://www.fhfb.gov/Default.aspx?Page=8&Top=4.

Trigger Events.—State divorce rates are available on an annual basis for most years in our sample from the Division of Vital Statistics, National Center for Health Statistics, CDC. The data are available at http://www.cdc.gov/nchs/data/nvss/Divorce%20Rates%2090%2095%20and%2099-07.pdf. We interpolate the values for 1997 and 1998 from the 1995 and 1999 and use the 2007 value for 2008.

Loan Level Characteristics.—

- ln loan age in months from the closing date to the contemporaneous month,
- LTV (LTV_RATIO),
- an indicator variable if the loan was interest only at origination (IO_FLAG),
- an indicator variable if the loan was an option ARM (INT_TYPE),
- an indicator variable if the loan is a jumbo (JUMBO_FLG),
- an indicator variable if the loan is a first mortgage (MORT_TYPE),
- the borrower's FICO score at origination (FICO_ORIG - original FICO score, available from 8/1997).

Table A1: Default Rates by State

State	*1997-2008*	*1997-2004*	*2005-2008*	*State*	*1997-2008*	*1997-2004*	*2005-2008*
AK	0.78%	0.19%	0.83%	MT	0.72%	0.52%	0.62%
AL	1.68%	0.82%	1.66%	NC	1.18%	0.47%	1.21%
AR	1.19%	0.30%	1.34%	ND	0.33%	0.14%	0.34%
AZ	1.40%	0.27%	1.59%	NE	1.48%	0.54%	1.64%
CA	1.51%	0.07%	1.78%	NH	0.83%	0.16%	0.95%
CO	1.79%	0.39%	2.13%	NJ	0.52%	0.19%	0.53%
CT	0.65%	0.15%	0.73%	NM	0.80%	0.39%	0.79%
DC	0.44%	0.18%	0.44%	NV	2.13%	0.26%	2.42%
DE	0.55%	0.18%	0.57%	NY	0.59%	0.19%	0.60%
FL	1.17%	0.21%	1.28%	OH	2.34%	0.77%	2.48%
GA	2.13%	0.41%	2.43%	OK	1.99%	0.74%	2.02%
HI	0.34%	0.06%	0.35%	OR	0.72%	0.49%	0.63%
IA	1.09%	0.36%	1.21%	PA	0.80%	0.39%	0.77%
ID	0.80%	0.44%	0.77%	RI	1.08%	0.12%	1.30%
IL	1.03%	0.28%	1.12%	SC	1.42%	0.74%	1.35%
IN	2.62%	1.03%	2.68%	SD	0.65%	0.42%	0.60%
KS	1.30%	0.41%	1.39%	TN	2.14%	0.50%	2.40%
KY	1.60%	0.44%	1.74%	TX	1.56%	0.43%	1.65%
LA	1.11%	0.54%	1.09%	UT	1.02%	0.60%	0.97%
MA	0.81%	0.09%	0.98%	VA	0.94%	0.09%	1.10%
MD	0.57%	0.18%	0.60%	VT	0.21%	0.05%	0.22%
ME	0.68%	0.20%	0.74%	WA	0.70%	0.31%	0.68%
MI	3.23%	0.52%	3.67%	WI	0.93%	0.31%	1.11%
MN	1.22%	0.29%	1.52%	WV	1.43%	0.63%	1.41%
MO	1.88%	0.43%	2.11%	WY	0.67%	0.38%	0.65%
MS	2.13%	0.60%	2.31%				

Table A2: Summary Statistics

	Mean	Std. Dev.	5th Percentile	95th Percentile
Recourse	0.67	0.47	0	1
Default Option (Probability of Negative Equity)	0.010	0.063	0	0.0212
Rate 1	0.014	0.118	0	1
Rate 2	0.14	0.35	0	1
Rate 4	0.087	0.281	0	1
Rate 5	0.051	0.221	0	1
Divorce Rate	3.80	0.88	2.6	5.1
Lagged Unemployment Rate	5.1	1.1	3.3	7.1
Fico Score at Origination	721	61	609	798
Interest Only (at Origination) Dummy	0.068	0.253	0	1
Jumbo Dummy	0.090	0.286	0	1
Mortgage Type Dummy	0.97	0.17	1	1
ARM Dummy	0.20	0.40	0	1
LTV Ratio at Origination	65	18	26.4	80
Natural Log of Loan Age	3.06	0.88	1.39	4.20
Purpose Type Dummy	0.63	0.48	0	1
Foreclosure Timing (in months)	6.37	3.28	2	12
Appraisal Amount (at Origination)	309,667	350,761	83,000	750,000
Number of Loans	2,924,160			
Number of Defaults	38,440			

Notes: Recourse is a dummy variable that takes a value of 1 if the property is in a recourse state, 0 otherwise; for North Carolina, recourse takes a value of 1 if the loan is not a purchase mortgage, 0 otherwise. The rate variables control for the difference between the current mortgage rate and the contract rate. Mortgage Type Dummy takes a value of 1 if the mortgage is a first mortgage, 0 otherwise. Purpose Type Dummy takes a value of 1 if the loan is not a purchase mortgage, 0 otherwise.

Appendix F

NBER WORKING PAPER SERIES

WHY DON'T LENDERS RENEGOTIATE MORE HOME MORTGAGES? REDEFAULTS, SELF-CURES AND SECURITIZATION

Manuel Adelino
Kristopher Gerardi
Paul S. Willen

Working Paper 15159
http://www.nber.org/papers/w15159

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2009

We thank John Campbell, Chris Foote, Scott Frame, Anil Kashyap, Chris Mayer, Eric Rosengren, Julio Rotemberg, and seminar audiences at Freddie Mac, Harvard Business School, University of Connecticut, LSE, the SAET meetings in Ischia, ESSET-Gerzensee, and the MIT Finance Lunch for thoughtful comments, Kyle Carlson and Suzanne Lorant for editorial help, and the Studienzentrum Gerzensee, where the authors completed this paper, for their generous hospitality. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research, the Federal Reserve Banks of Atlanta or Boston or the Federal Reserve System.

Why Don't Lenders Renegotiate More Home Mortgages?
Redefaults, Self-Cures and Securitization
Manuel Adelino, Kristopher Gerardi, and Paul S. Willen
NBER Working Paper No. 15159
July 2009
JEL No. D11,D12,G21

ABSTRACT

We document the fact that servicers have been reluctant to renegotiate mortgages since the foreclosure crisis started in 2007, having performed payment reducing modifications on only about 3 percent of seriously delinquent loans. We show that this reluctance does not result from securization: servicers renegotiate similarly small fractions of loans that they hold in their portfolios. Our results are robust to different definitions of renegotiation, including the one most likely to be affected by securitization, and to different definitions of delinquency. Our results are strongest in subsamples in which unobserved heterogeneity between portfolio and securitized loans is likely to be small and for subprime loans. We use a theoretical model to show that redefault risk, the possibility that a borrower will still default despite costly renegotiation, and self-cure risk, the possibility that a seriously delinquent borrower will become current without renegotiation, make renegotiation unattractive to investors.

Manuel Adelino
MIT Sloan
50 Memorial Dr, E52-458
Cambridge, MA 02138
madelino@mit.edu

Kristopher Gerardi
Federal Reserve Bank of Atlanta
1000 Peachtree St. NE
Atlanta, GA 30309
kristopher.gerardi@atl.frb.org

Paul S. Willen
Research Department
Federal Reserve Bank of Boston
P.O. Box 55882
Boston, MA 02210
and NBER
willen968@yahoo.com

1 Introduction

Many commentators have attributed the severity of the foreclosure crisis in the United States in the 2007–2009 period to the unwillingness of lenders to renegotiate mortgages, and, as a consequence, have placed renegotiation at the heart of the policy debate. Every major policy action to date has involved encouraging lenders, in one way or another, to renegotiate loan terms in order to reduce borrower debt loads. According to the Treasury-sponsored HopeNow initiative, in December of 2007 lenders were expected to prevent adjustable-rate mortgages from increasing to higher rates at the first reset of the mortgage.[1] "Hope For Homeowners," enacted by Congress in July of 2008, envisioned that lenders would write off a substantial portion of the principal balance of mortgages for financially distressed households.[2] The Obama Adminstration's Making Home Affordable Plan, announced in February of 2009, provided financial incentives to servicers to renegotiate loans on the condition that the lenders reduce the interest rate for a significant period of time.[3]

The appeal of renegotiation to policy makers is simple to understand. If a lender makes a concession to a borrower by, for example, reducing the principal balance on the loan, it can prevent a foreclosure. This is clearly a good outcome for the borrower, and possibly good for society as well. But the key to the appeal of renegotiation is the belief that it can also benefit the lender, as the lender loses money only if the reduction in the value of the loan exceeds the loss the lender would sustain in a foreclosure. In short, according to proponents, renegotiation of home mortgages is a type of public policy holy grail, in that it helps both borrowers and lenders at little or no cost to the government.[4]

In this paper, we explore the renegotiation of home mortgages using a dataset from Lender Processing Services (LPS), a large, detailed sample of residential mortgages. Our primary empirical analysis involves following borrowers over the year subsequent to their first serious delinquency and counting the frequency of renegotiation.[5] Measuring renegotiation in the LPS data is a challenge because there is no field in the data that identifies whether or not a servicer has changed the terms of, or "modified," the loan. We overcome this difficulty by developing an algorithm to identify modifications that we validate on an unrelated dataset that includes a modification flag.

We explore several different definitions of renegotiation in the data. Our first definition of "renegotiation" is concessionary modifications that serve to reduce a borrower's monthly

[1] Edmund L. Andrews, In Mortgage Plan, Lenders Set Terms, *New York Times*, Dec. 7, 2007.
[2] "Bush Signs Wide-Ranging Housing Bill Into Law," *Wall Street Journal*, July 31, 2008.
[3] See "$275 Billion Plan Seeks To Address Crisis In Housing," *New York Times*, Feb. 18, 2009.
[4] See this discussion in Congressional Oversight Panel (2009), Zingales (2008), and Geanakoplos and Koniak (2008), as examples.
[5] Until 2008, the dataset was known as McDash.

payment. These may be reductions in the principal balance or interest rate, extensions of the term, or combinations of all three. This definition of renegotiation is a key focus of our analysis because there is a consensus among many market observers that concessionary modifications are the most, or possibly the only, effective way of preventing foreclosures. As the Congressional Oversight Panel (COP) for the Troubled Asset Recovery Program (TARP) has written, "Any foreclosure mitigation plan must be based on a method of modifying or refinancing distressed mortgages into affordable ones. Clear and sustainable affordability targets achieved through interest rate reductions, principal write-downs, and/or term extensions should be a central component of foreclosure mitigation."[6]

Because the pooling and servicing agreements (PSAs), which govern the conduct of servicers when loans are securitized, often place limits on the number of modifications a servicer can perform, we broaden our definition of renegotiation to include any modification, regardless of whether it lowers the borrower's payment. Modifications are often thought to always involve concessions to the borrower, but many, and in some subsets most, modifications involve the capitalization of arrears into the balance of the loan, and thus lead to increased payments.

Finally, we attempt to include in our definition of renegotiation the transactions whereby lenders allow borrowers to extinguish their liabilities by repaying less than the outstanding balance of the loan. These transactions are known as short payoffs, short sales, or deeds-in-lieu of foreclosure, depending on the structure. We measure this component of renegotiation by counting the number of seriously delinquent loans that the servicer reports as "paid off."

No matter which definition of renegotiation we use, one message is quite clear: lenders rarely renegotiate. Fewer than 3 percent of the seriously delinquent borrowers in our sample received a concessionary modification in the year following the first serious delinquency. More borrowers received modifications under our broader definition, but the total still accounted for fewer than 8 percent of the seriously delinquent borrowers. And finally, fewer than 5 percent of all of our troubled borrowers repaid their mortgages, putting an upper bound on the number who could have repaid less than the principal balance of the loan. These numbers are small both in absolute terms, and relative to the approximately half of the sample for whom foreclosure proceedings were initiated, and the nearly 30 percent for whom they were also completed.

We next turn to the question of why renegotiation is so rare. If the logic described in the second paragraph is correct, lenders should find renegotiation attractive, even in the

[6]See the Congressional Oversight Panel (2009). This view is widely held and is the main focus of the Administration's Making Home Affordable foreclosure prevention plan was to encourage servicers to modify loans to reduce monthly payments to 31 percent of income.

absence of government prodding. Yet, we observe very little renegotiation in the data. We address this apparent paradox.

The leading explanation attributes the reluctance of lenders to renegotiate to the process of securitization.

> The complex webs that securitization weaves can be a trap and leave no one, not even those who own the loans, able effectively to save borrowers from foreclosure. With the loan sliced and tranched into so many separate interests, the different claimants with their antagonistic rights may find it difficult to provide borrowers with the necessary loan modifications, whether they want to or not. In the tranche warfare of securitization, unnecessary foreclosures are the collateral damage. (Eggert 2007)

More precise institutional evidence appears to confirm the role of securitization in impeding renegotiation. As mentioned in more detail below, PSAs do sometimes place global limits on the number of modifications a servicer can perform for a particular pool of mortgages. In addition, the rules by which servicers are reimbursed for expenses may provide a perverse incentive to foreclose rather than modify. Furthermore, because servicers do not internalize the losses on a securitized loan, they may not behave optimally. Another issue is the possibility that those investors whose claims are adversely affected by modification will take legal action. Finally, historically, SEC rules have stated that contacting a borrower who is fewer than 60-days delinquent constitutes an ongoing relationship with the borrower and jeopardizes the off-balance sheet status of the loan.

But some market observers express doubts about the renegotiation-limiting role of securitization. Hunt (2009) conducted an exhaustive review of a sample of PSAs and concluded, "it appears that large-scale modification programs may be undertaken without violating the plain terms of PSAs in most cases." Although some servicers have expressed concern about lawsuits, of the more than 800 lawsuits filed by investors in subprime mortgages through the end of 2008, not one involved the right of a servicer to modify a loan.[7] Even the Congressional Oversight Panel (2009), which did view securitization as a problem in general, conceded, "The specific dynamics of servicer incentives are not well understood." Finally, the SEC ruled in 2008 that if default was "reasonably forseeable," then contact with a borrower prior to 60-day delinquency would not affect the accounting status of the loan.

Our empirical analysis provides strong evidence against the role of securitization in preventing renegotiation. The LPS dataset includes loans that are serviced for private securitization trusts that are not sponsored by any of the government sponsored enterprises

[7] Navigant report, Congressional Oversight Panel (2009).

(GSEs), so-called "private-label" loans, which are subject to all of the contract frictions described above. It also includes loans owned by servicers, so-called "portfolio" loans, which are immune to such problems. We compare renegotiation rates, controlling for observable characteristics of the loans. For our narrowest definition of renegotiation, payment-reducing modification, we find that the differences in the likelihood of renegotiation in the 12 months subsequent to the first 60-day delinquency between the two types of loans is neither economically nor statistically significant. When we consider the broader definition that includes any modification at all, which, as we mentioned above, we would expect to be most affected by securitization, the data even more strongly reject the role of securitization in preventing renegotiation. We also find that servicers are *more* likely to peform modifications, broadly defined, and to allow the borrower to prepay on a private-label loan than on a portfolio loan.

Our results are highly robust. One potential problem with the data is that there is unobserved heterogeneity in the characteristics of portfolio and private-label loans. To address this, we exploit subsets of the LPS data, in which servicers provide an exceptional amount of information about borrowers. When we exclude observations where the servicer failed to report whether the borrower fully documented income at origination, or what the debt-to-income ratio was at origination, our results become even stronger. When we focus only on loans for which the borrower fully documented income, we obtain results that are broadly consistent or, in some cases, stronger than the results for the full sample. Finally, we limit our sample to only subprime loans (as defined in LPS). These loans comprise only 7 percent of the LPS data, but they account for more than 40 percent of all serious delinquencies and almost 50 percent of the modifications that we identify in the data. The results that we obtain for the subprime sample are also consistent with our results for the full sample.

Another potential issue with our focus on 60-day delinquent loans is that portfolio lenders can contact borrowers at any time, whereas some securitization agreements forbid lenders from contacting borrowers until they are at least 60 days delinquent (two missed payments). When we shift our focus to 30-day delinquent borrowers (one missed payment), our results continue to show no meaningful difference between renegotiation of private-label and portfolio loans.

One other possibility is that our algorithm for identifying modifications is somehow missing a class of loss-mitigation actions taken by servicers. Forbearance agreements and repayment plans, for example, would not necessarily show up in our data. However, neither of these actions constitutes renegotiation in any classic sense, because the lender still expects the borrower to repay in full, including interest on any delayed payment. In addition, unlike modifications, PSAs never place any limits on the use of forbearance agreements or repayment plans, so, *a priori*, we would have less reason to expect a difference in their use across private-label and portfolio loans. Finally, most successful forbearance agreements conclude with a modification to allow the borrower to repay the arrears incurred in forbearance. With all of that said, we test the proposition that servicers engage in other loss mitigation actions by looking at the "cure rate." This is the percentage of loans that transition to current status after becoming 60-days delinquent. We find that in the full sample, private-label loans are less likely to cure, but that the gap, although statistically significant, is small — correcting for observable characteristics, we estimate a cure rate of around 30 percent for the typical portfolio loan and a cure rate of about 2 percentage points less for an otherwise equivalent private-label loan. However, for the subprime subsample, the subsample with information about documentation and debt-to-income (DTI) status, and the sample of fully documented loans, we find that private-label loans are significantly *more* likely to cure.

The policy debate has focused exclusively on the ways securitization impedes renegotiation and implicitly assumes that portfolio lenders face no institutional impediments, but this is not realistic. Portfolio lenders complain about accounting rules, including the need to identify modifications, even when the borrowers are current prior to the modification, as "troubled debt restructurings," which leads to reduction of the amount of Tier II capital and increased scrutiny from investors and cumbersome accounting requirements. The shortage of qualified staff, an oft-heard complaint from borrowers seeking renegotiation, affects servicers of portfolio loans and private label loans equally. Finally, the interests of the managers of a loan portfolio are not necessarily any more likely to be aligned with their investors than are the interests of the trustees of a mortgage pool; many have attributed the catastrophic failures of financial institutions like AIG in 2008 to misaligned incentives of managers and shareholders.

Our results are consistent with the hypothesis that securitization does impede renegotiation but that a different set of impediments leads to similar problems with portfolio loans and generates our finding that there is no difference. However, the small differences would represent a remarkable coincidence.[8] More importantly, the low overall levels of renegotiation mean that even if contract frictions cut the overall number of concessionary modifications in half, 94 percent of seriously delinquent borrowers would still fail to receive a concessionary modification. So the puzzle remains why so few loans are renegotiated.

If contract frictions are not a significant problem, then what is the explanation for

[8]Yet another possible explanation is that equal treatment provisions in PSAs force servicers to modify similar numbers of portfolio and private-label loans and that servicers are reluctant to modify portfolio loans in spite of the fact that they internalize the benefits because they must then modify private label loans for which they don't.

why lenders do not renegotiate with delinquent borrowers more often? We argue for a very mundane explanation: lenders expect to recover more from foreclosure than from a modified loan. This may seem surprising, given the large losses lenders typically incur in foreclosure, which include both the difference between the value of the loan and the collateral, and the substantial legal expenses associated with the conveyance. The problem is that renegotiation exposes lenders to two types of risks that can dramatically increase its cost. The first is what we will call "self-cure" risk. As we mentioned above, more than 30 percent of seriously delinquent borrowers "cure" without receiving a modification; if taken at face value, this means that, in expectation, 30 percent of the money spent on a given modification is wasted. The second cost comes from borrowers who redefault; our results show that a large fraction of borrowers who receive modifications end up back in serious delinquency within six months. For them, the lender has simply postponed foreclosure; in a world with rapidly falling house prices, the lender will now recover even less in foreclosure. In addition, a borrower who faces a high likelihood of eventually losing the home will do little or nothing to maintain the house or may even contribute to its deterioration, again reducing the expected recovery by the lender.

In Section 4 of the paper, we formalize the basic intuition of the investor renegotiation decision, with a simple model. We show that higher cure rates, higher redefault rates, higher expectations of house price depreciation, and a higher discount rate all make renegotiation less attractive to the investor. Thus, one cannot evaluate a modification by simply comparing the reduction in the interest rate on the loan or in the principal balance with the expected loss in foreclosure. One must take into account both the redefault *and* the self-cure risks, something that most proponents of modification fail to do.[9]

To our knowledge, this paper is the first to estimate directly the likelihood of renegotiation of private-label and portfolio-held loans. Piskorski, Seru, and Vig (2009) address the question of the effects of securitization on renegotiation, but rather than directly identifying renegotiation, they run "black-box" foreclosure regressions using LPS data and argue that observed differences in foreclosure rates *imply* differences in renegotiation activity. Our results contradict this interpretation. For renegotiation to explain the differences in foreclosure rates, there would have to be large errors in our algorithm for identifying renegotiation, and those errors would have to be significantly biased toward portfolio loans, a possibility that is particularly problematic given that the renegotiations we focus on are precisely the type that PSAs supposedly prevent. In addition, most of the loan histories in the LPS

[9]Many proponents of aggressive modification take into account redefault risk, and the MHA plan did address it by providing some insurance against further house price declines to investors who modified loans. However, none of the main proponents ever mentions self-cure risk, even though it is well-known in the industry, see: http://www.calculatedriskblog.com/2008/09/loan-modifications-anecdotes-and-data.html.

sample are right-censored, meaning that the borrowers have neither lost their homes nor paid off their mortgages when the data end, making it impossible to equate the absence of a foreclosure with successful renegotiation. By contrast, a "cure" is a necessary condition for renegotiation, and thus the differences we report in cure rates across portfolio and private-label loans that are neither large nor of consistent sign contradict the claim that securitization is a major obstacle to renegotiation.

The implications of our research for policy are three-fold. First, "safe harbor" provisions, which shelter servicers from investor lawsuits, are unlikely to affect the number of modifications and should have little effect. Second, and more broadly, the number of "preventable foreclosures" may be far fewer than many believe.

Finally, we point out that while our model shows why investors may not want to perform modifications, that does not necessarily imply that modifications may not be socially optimal. One key input to our theoretical model is the discount rate, and it is possible that investors, especially in a time when liquidity is highly valued, may be less patient than society as a whole, and therefore foreclose when society would prefer renegotiation. Large financial incentives to investors or even to borrowers to continue payment could mitigate this problem.

1.1 Related Literature and Existing Evidence

Our research draws on existing literature in several different fields. First, there has been substantial interest in the question of renegotiation of home mortgages among real estate economists, both prior to, and as a result of the current crisis. Riddiough and Wyatt (1994a), Riddiough and Wyatt (1994b), and Ambrose and Capone (1996) addressed informational issues that inhibit efficient renegotiation. We draw extensively on this research in Section 4. Springer and Waller (1993), in an early example, explores patterns in the use of forbearance as a loss mitigation tool. Capone (1996) and Cutts and Green (2005) both discuss the institutional issues, with the former study providing historical evidence and focusing on issues in the mid-1990s, and the latter study discussing innovations since then.

The start of the subprime crisis in 2007 led to a resurgence of interest in the topic among real estate economists and aroused new interest from other fields, in particular, the field of law. In real estate, Quercia, Ding, and Ratcliffe (2009), Cutts and Merrill (2008), Stegman, Quercia, Ratcliffe, Ding, Davis, Li, Ernst, Aurand, and Van Zandt (2007), and Mason (2007), all discuss issues with contemporary loss mitigation approaches. Legal researchers, White (2008) and White (2009), for example, have addressed empirical questions about the frequency and characteristics of loan modifications, closely related to the analysis in this paper. In addition, they have also looked at issues related to the restrictions imposed by contracts (Hunt 2009 and Gelpern and Levitin 2009) and the interactions among foreclosure, renegotiation, and personal bankruptcy (Levitin 2009a and Levitin 2009b).

More broadly, real estate economists have explored the factors that lead delinquent mortgages to transition to foreclosure or to cure, one of which is renegotiation. Pre-crisis papers include Ambrose and Capone (1998), Ambrose, Buttimer Jr, and Capone (1997), Ambrose and Capone (2000), Lauria, Baxter, and Bordelon (2004), Danis and Pennington-Cross (2005), Pennington-Cross (2009), and Pennington-Cross and Ho (2006). Mulherin and Muller (1987) discusses conflicts between mortgage insurers and owners that may lead servicers to induce or postpone foreclosure inefficiently. In light of the crisis, Piskorski, Seru, and Vig (2009) and Cordell, Dynan, Lehnert, Liang, and Mauskopf (2008a) have revisited the question.

The issue of dispersed ownership and debt renegotiation has received a fair amount of attention in the corporate finance literature. Gan and Mayer (2006), for example, focus on commercial mortgages, and find that servicers delay liquidation of delinquent mortgages when they are also the holders of the equity tranche of the deal. This suggests that participating in the losses due to liquidation may alleviate some of the agency problems posed by the separation of ownership and servicing pointed out before. However, it may also lead to conflicts of interest between holders of different tranches. In their setting, Gan and Mayer (2006) find that the servicers' behavior is consistent with asset substitution, as servicers seek to benefit from the option-like payoff of their position. Also, the contractual restrictions imposed by PSAs (discussed above) and standard economic arguments on the effects of dispersed ownership of debt (as in Bolton and Scharfstein 1996 and Asquith, Gertner, and Scharfstein 1994) further reduce the incentives of servicers to modify mortgages.

2 Data

We use a dataset constructed by LPS. This is a loan-level dataset that covers approximately 60 percent of the U.S. mortgage market and contains detailed information on the characteristics of both purchase-money mortgages and mortgages used to refinance existing debt.[10] This dataset is especially useful in the context of this paper, as it includes both securitized mortgages and loans held in portfolio.[11] The LPS data specifically denote whether a mort-

[10] We use a 10 percent random sample of the LPS data when estimating all of our empirical models. The dataset is simply too big to use in its entirety from a computational standpoint. However, we have checked the robustness of our results to using different sample sizes, and we do not find substantial differences.

[11] For a more detailed discussion of the LPS data, we direct the reader to Foote, Gerardi, Goette, and Willen (2009).

gage is held in portfolio, or securitized by a non-agency, private institution.[12] If institutional constraints are restricting the modification process for private-label, securitized loans, we would expect to see relatively few modifications among them, as compared to portfolio loans. Unfortunately, our LPS sample does not include direct information regarding loan modifications.[13] However, LPS does provide monthly updates to loan terms, so it is possible to identify loan modifications indirectly (and imperfectly). Table 1 shows two examples of modifications in the data. In the first example, the servicer cuts the interest rate, capitalizes arrears into the balance of the loan, and extends the term of the loan to 40 years. In the second example, the servicer just capitalizes arrears into the balance of the loan. In both cases the loan is reported as "current" after the modification, whereas before it was 90+ days delinquent.

We denote a loan as being modified if there is a change in its terms that was not stipulated by the initial terms of the contract. Such modifications include interest-rate reductions, principal-balance reductions, and term extensions. We can also identify principal-balance and mortgage-payment *increases* that reflect the addition of arrears into the balance of a loan.[14] We spell out our algorithm for identifying modifications in more detail in Appendix A.

There are two potential mistakes we can make in this exercise. First, we may falsely identify modifications ("false positives") because of measurement error in the data (for example, a mistake in the updated balance or interest rate) or some endogenous behavior on the part of the borrower (for example, a borrower making extra principal payments). Second, we could miss modifications ("false negatives") because our algorithm for finding modifications is incomplete. In order to test our algorithm, we use data from the Columbia files put together by Wells Fargo's CTSLink service. This dataset includes a similar set of variables to those in the LPS dataset (on performace of the loans and characteristics of the borrower at origination) but is limited to private-label loans. These files do include,

[12]The LPS data also denote when a loan is securitized by a GSE (Government Sponsored Enterprise) such as Freddie Mac or Fannie Mae. We eliminate this class of loans, since the GSEs hold all credit risk, and thus are not subject to any modification restrictions.

[13]In a recent report, the Office of Thrift Supervision (OTS), in collaboration with the Office of the Comptroller of Currency (OCC), used data from LPS to analyze the outcomes of recent mortgage modification programs (OCC and OTS Mortgage Metrics Report, Third Quarter 2008). In this report, they had access to supplementary data from servicers that include the identification of loans in the LPS data that had been modified. We have not been able to obtain access to this data.

[14]One of the major types of loan modifications that we are largely unable to identify are interest rate freezes for subprime ARMs, which reset after two or three years. However, the reason that we cannot identify those freezes is because many are not binding; the fully-indexed rate is lower than the initial rate. These modifications will have no major effect on the current terms of the mortgage, so we do not view this as a major drawback.

however, explicit flags for modifications. This allows us to use the same algorithm described in Appendix A and compare the modifications we identify to the "true" modifications. Results are reported in Table 2. Overall our algorithm performs well, with 17 percent false negatives (that is, we do not identify around 17 percent of the "true" modifications) and around the same percentage of false positives (that is, approximately 17 percent of the modifications we identify are not flagged as modifications on the CTSLink data). By type of modification, our algorithm performs best for principal reductions, term increases, and fixed-rate mortgage reductions, and comparatively worse for ARM rate reductions and for principal increases.

2.1 Summary Statistics from the Data

Table 3 reports the number of modifications performed each quarter from the first quarter of 2007 through the final quarter of 2008, disaggregated by the type of modification. Each of the numbers is a multiple of 10 because we used a 10 percent random sample and scaled up the numbers we found. The first column of Table 3 simply reports the total number of loan modifications made. Not surprisingly, modifications have become more common as the housing market has weakened. There appear to be more than 7–8 times as many modifications performed in the fourth quarter of 2008 as in the first quarter of 2007. In addition to the rapid growth in loan modifications, the composition of modifications has changed over time. This can be seen in the remaining columns of Table 3, which list the incidence of modifications of different types.[15]

An interesting finding is that most modifications entailed *increases* in the principal balance of a mortgage. Such increases are likely due to the addition of arrears to the outstanding mortgage balance for delinquent borrowers, and these often increase the monthly mortgage payment by a nontrivial amount. While the absolute numbers of balance-increasing modifications are still rising, they are falling as a percentage of total modifications. In the last few quarters, interest-rate reductions, which necessarily involve a decrease in the monthly mortgage payment, have become more frequent, rising to more than 26 percent of all modifications performed in 2008:Q4. Table 3 provides further information regarding the behavior of monthly mortgage payments for loans that have undergone a modification. There are several notable patterns in this table. First, as of 2008:Q4, modifications that involved payment decreases were more common than those that involved payment increases. Furthermore, the

[15] In many cases a mortgage will experience multiple types of modifications at the same time. For example, we see cases in the data in which the interest rate is decreased and at the same time the term of the loan is extended. Thus, the percentages in Table 3 are not calculated with respect to the number of loans modified, but rather with respect to the number of modifications performed.

average and median magnitude of payment decreases has recently increased in our sample. From 2007:Q1 to 2008:Q2, the median payment decrease ranged from approximately 10 percent to 14 percent, but then increased to approximately 20 percent and 22 percent in 2008:Q3 and 2008:Q4, respectively. Based on the logic from our simple framework above, it is likely that these will have more success than modifications involving increases in the payment and/or balance.

Another interesting observation from Table 3 is that the incidence of principal reductions is quite low in our data. This is likely due to two factors. First, the LPS dataset under-represents the subprime mortgage market.[16] A few servicers that focus almost exclusively on subprime mortgages have recently begun modification programs that involve principal reduction.[17] In addition, from a theoretical perspective, principal reduction plans suffer from the severe incomplete-information problem noted earlier. Balance reductions are appealing to both borrowers in danger of default and those who are not. In a recent paper, we argued that to avoid such moral hazard concerns, lenders have a strong incentive to only provide modifications to those borrowers who are most likely to default.[18] Table 3 contains summary statistics regarding the characteristics at origination of both the sample of modified mortgages and the sample of all loans in the LPS dataset. The patterns that emerge from the table are consistent with such an argument. We discuss this point in more detail below. The sample of modified mortgages is characterized by substantially lower credit scores, higher loan-to-value (ltv) ratios, and slightly higher debt-to-income ratios. The discrepancy in ltv ratios may be underestimated, as the percentage of mortgages with an ltv ratio of exactly 80 percent is significantly higher in the modification sample than in the full sample. As we argued above, this likely implies a larger fraction of highly leveraged loans, for which the second liens are not observable in the data. In addition, the modification sample includes a higher fraction of mortgages with non-traditional amortization schedules, such as interest-only loans, option ARMS, hybrid ARMs, and subprime loans.

In Table 4 we compare the size of payment decrease and payment increase modifications for loans held in private-label trusts and loans held in portfolio. The results are somewhat mixed, as the size (as a percentage of the original payment) of the median payment decrease due to modification is larger for private-label loans in the first three quarters of 2008, but smaller in the final quarter. We see a similar pattern for the median payment increase due

[16]The majority of subprime mortgages are securitized by non-agency firms, and the LPS dataset includes approximately 35 percent of mortgages securitized by non-agency corporations.

[17]According to an October report by Credit Suisse, Ocwen Loan Servicing, LLC and Litton Loan Servicing LP were the only subprime servicers that had performed a nontrivial number of principal reduction modifications. Neither of these servicers contributes to the LPS dataset.

[18]See Foote, Gerardi, and Willen (2008) for a more detailed discussion.

to modification, while the differences are small for the mean and median payment increase.

3 Differences in Modification Behavior

In this section, we directly address the question of whether the incidence of modification is impeded by the process of securitization. We show evidence that private-label loans and portfolio loans perform similarly, both unconditionally and when observable differences between securitized and portfolio-held loans are controlled for, using both a logit model with a 12-month horizon and a Cox proportional hazard model that takes into account the problem of right censoring in the data.

To make sure that our results are robust to the type of modification performed, we use several different definitions of modification in this section. Our first measure is the number of concessionary modifications, which we define as reductions in the interest rate, reductions in the principal balance, extensions of the term, or combinations of all three. Any or a combination of these serves to reduce a borrower's monthly mortgage payment. We use this as our primary definition of modification in our analysis, as there is a consensus among most market observers that concessionary modifications are the most, or perhaps the only, effective way of preventing foreclosures. Because pooling and servicing agreements, which govern the conduct of servicers when loans are securitized, often limit modifications that change *any* of the contract terms (not just those that result in payment decreases), we broaden our definition of renegotiation to include any modification, regardless of whether it lowers the borrower's payment. As we discussed above, many, and in some subsets, most modifications, involve the capitalization of arrears into the balance of the loan and thus lead to increased payments. Finally, we attempt to include in our measure of renegotiation the number of times that lenders allow borrowers to extinguish their liabilities by repaying less than the outstanding balance of the loan. These transactions are known as short payoffs, short sales, or deeds-in-lieu of foreclosure, depending on the structure. We do this by counting the number of seriously delinquent loans that the servicer reports as paid off, and including these observations in our definition of modification.

Before turning to the regressions, however, it is instructive to look at the unconditional frequencies of modifications in the data. Panel A of Table 5 shows the unconditional frequencies for each type of investor. The first takeaway from the table is the extremely low percentages of modifications for *both* types of mortgages. Only 3 percent of 60-day delinquent loans received concessionary modifications in the 12 months following the first serious delinquency, and only 8.5 percent of the delinquent loans received *any* type of modification in the same period. These are extremely low levels of modifications, and they suggest that even if there are contract frictions that are preventing modifications in securitized trusts, the economic effects are small. The second takeaway from the table is that the unconditional differences between portfolio loans and private-label loans are very small in absolute terms. There is a difference of approximately 0.6 percentage points and 0.3 percentage points for concessionary modifications and all modifications, respectively. These are very small differences, and they suggest that contract frictions do not play an important role in inhibiting the renegotiation process for loans in securitized trusts. However, these are unconditional statistics, and it is possible that once observable differences in the characteristics of each type of loan and borrower are accounted for, the results may change.[19] Thus, we now estimate differences in modification behavior while controlling for observable loan and borrower characteristics. These characteristics include the contract interest rate at origination; the credit score of the borrower at origination; the loan-to-value ratio of the mortgage (not including second or third liens) at origination[20]; the logarithm of the nominal dollar amount of loan; an indicator of whether the purpose of the loan was a refinance of a previous mortgage or a home purchase; an indicator of whether the loan was considered to be subprime[21]; a measure of the amount of equity in the property at the time of delinquency, specified as a percentage of the original loan balance and updated by state-level house price indexes calculated by the Federal Housing Finance Agency (FHFA)[22] (and an indicator for a borrower who is in a position of negative equity at the time of delinquency, where the value of the mortgage exceeds the value of the home); and the unemployment rate of the county in which the borrower resides, calculated by the Burea of Labor Services (BLS).[23] We also include, but do not show because of space considerations, a set of cohort dummies that control for the quarter when the mortgage was originated, information regarding the amortization schedule of the mortgage (interest-only or negative amortization, including mortgages commonly referred to as option ARMs), an indicator for whether the size of the mortgage is greater than the GSE conforming loan limits, an indicator for whether the

[19] For example, if private-label loans are significantly riskier, and thus better candidates for modification on average, then the unconditional difference will significantly understate things.

[20] Because of the lack of information on second liens in the LPS data and the prevalence of second mortgages as a way to avoid paying mortgage insurance, we include an indicator variable if the ltv ratio is exactly equal to 80 percent. These are the borrowers who likely took out second mortgages, as the requirement for mortgage insurance occurs at ltv ratios above 80 percent. Our experience with other, more complete datasets also confirms that many of these borrowers are likely to have second mortgages that bring the cumulative ltv ratio up to 100 percent.

[21] This definition of subprime comes from the mortgage servicers that contribute to the LPS dataset.

[22] House prices are measured at the state level using the FHFA index. We also tried using Case-Shiller metropolitan area house price indexes and found no substantive differences. We chose to use the OFHEO prices for our primary specifications because of their greater sample coverage.

[23] Equity and periods of unemployment are very important determinants of a borrower's decision to default, and thus should also be important factors in the modification decision.

house is a primary residence, an indicator for adjustable rate mortgages that contain a reset provision (so-called "hybrid ARMs"), and, finally, an indicator for a borrower who does not use the corresponding property as a principal residence (this includes both properties used strictly for investment purposes, and vacation homes).

3.1 Canonical Specification Results

Panel B of Table 5 displays the estimated marginal effects from a set of logit models for the three different types of modification definitions. The dependent variable is 1 if a 60-day deliquent loan is modified at any point in the 12 months following the first delinquency. The first column considers payment-reducing (concessionary) modifications, the second column includes both payment-reducing and payment-increasing modifications, and the third column contains all modifications considered before, as well as prepayments. In all regressions, the group of portfolio-held loans is omitted from the estimation and is thus assumed to be the reference group. We cluster the standard errors at the zip code level to account for the fact that loans in the same geographical area are likely to suffer correlated (unobserved) shocks.

According to the estimates in the first column, private-label loans were approximately 0.3 percentage points less likely to receive concessionary modifications than loans held in portfolio. This estimate is economically small but statistically significant at the 10 percent level. When we consider all modifications the point estimate flips sign and becomes 0.2 percentage points (statistically insignificant), while for the third specification, private-label loans were actually 0.9 percentage points more likely to receive concessionary modifications (statistically significant). As discussed above, all of these specifications include a number of additional loan characteristics that are important in the underwriting process and, thus, likely to play an important role in the modification decision. The first observation to make regarding the results reported in Panel B is that the difference between the incidence of modification for portfolio-held loans and private-label loans becomes even smaller when these variables are controlled for in the estimation. The results also imply that loans with higher credit scores were modified less, loans with higher ltv ratios were modified less, larger loans were modified more, and loans with more equity at the time of delinquency were modified less. We find a sizeable difference in terms of the frequency of modification for both refinances and subprime loans. Conditional on being 60-days delinquent, subprime loans were modified about 2 percentage points more than prime loans. We estimate a model separately for subprime loans in Table 6.

Censoring is an important issue for any sample of mortgages, as there are currently many delinquent loans that are, or will soon be, good candidates for modification, as the housing market continues to decline. For this reason, we estimate a Cox proportional hazard model of the transition from serious delinquency to modification. The Cox model is very common in the survival analysis literature, and it has the advantage of being both flexible in terms of functional form considerations, as the baseline hazard function can be treated as an incidental parameter, and easy to estimate in terms of computational considerations. The results, expressed as hazard ratios, are reported in Panel C. A hazard ratio less than 1 indicates that private-label loans were less likely to receive a modification compared to portfolio loans, while a ratio greater than 1 signifies the opposite. The estimates are consistent with what we report for the logits in the previous panel. Private-label loans were less likely to receive concessionary modifications, but this coefficient estimate is statistically insignificant. For the our other two modification definitions the sign flips, but again the result is not statistically significant. All three specifications include the same covariates that were included in the logit models.

3.2 Subsample Results

Table 6 contains further logit estimation results for various subsamples of interest to see if there are different probabilities than in the full sample. Since the subprime indicator seems to be such a powerful predictor of modification conditional on serious delinquency in Table 5, we report the estimated marginal effects for only the sample of subprime loans in the second column of Table 6. The subprime sample also has the advantage that the agencies (Fannie Mae and Freddie Mac) were unlikely to be the marginal investor for this type of loans, so it is less likely that the portfolio and private-label samples differ significantly on unobservable characteristics. In the third column, we report results from the sample of LPS mortgages for which the borrower had a FICO score of less than 620, since automated underwriting systems generally instruct lenders to engage in increased scrutiny for such loans because of increased default risk. In the fourth and fifth columns, we focus on samples of loans that we believe contain the most information regarding the borrowers, in order to try to minimize the amount of unobservable heterogeneity that could potentially be biasing the results. In the fourth column, we focus on the sample of loans for which both the DTI ratio and the documentation status contain non-missing values, while the fifth column contains results for only the loans that were fully documented (in terms of income and assets) at origination. Panel A contains both unconditional means and estimated marginal effects for concessionary modifications, while Panel B contains results for the broader definition that also includes non-concessionary modifications.

The results are largely consistent with those contained in Table 5. We redisplay the results from the full sample in the first column of Table 6 for ease of comparison. The difference in modification frequency between private-label and portfolio-held, subprime mortgages for 60-day delinquent loans is small, and not statistically different from zero for both definitions of modification. Using a FICO cutoff of 620 as an alternative definition of subprime does not seem to make much difference. The unconditional means are smaller (for both types of loans) compared to the LPS subprime sample, as the LPS definition includes most of the loans with a FICO less than 620, but also some loans with higher associated FICOs. However, the marginal effects of private-label loans estimated from the logit models are quite similar to those from the LPS subprime sample, as they are economically small, and not statistically significant. Finally, we also find small and largely insignificant results for the last two subsamples, displayed in the fourth and fifth columns of Table 6. Although, it is worth pointing out that we do find a statistically significant, positive estimate of private-label loans for the broad definition of modification (Panel B).

3.3 Alternative Delinquency Definition

As an additional robustness check, we broaden our definition of delinquency and focus on modifications performed on loans subsequent to their first 30-day delinquency, which corresponds to one missed mortgage payment. While waiting until a borrower becomes seriously delinquent (defined as 60-days) to renegotiate is common practice in the servicing industry, there are no direct contractual stipulations (to our knowledge) that restrict a servicer from modifying the loan of a borrower who is 30-days delinquent. Thus, in Table 7 we repeat our analysis of Tables 5 and 6, but condition on 30-days delinquency rather than 60-days. The table contains three panels of estimation results, one for each of our modification definitions, and all of the subsamples described considered in Table 6. The unconditional means, logit marginal effects, and Cox hazard ratios are all reported for each combination of subsample and modification definition.

The results are very similar to those from the analysis of 60-day delinquent loans. According to the full sample and subprime sample logit models, portfolio loans received slightly more concessionary modifications, and the differences (0.3 and 0.5 percentage points respectively) are statistically significant at conventional levels. However, according to the subprime sample and full documentation sample Cox models, private-label loans actually received more concessionary modifications, although those differences are also small.[24] The results

[24]The logit marginal effects correspond to percentage point differences, while the Cox hazard ratios correspond to percent differences. If one expresses the logit marginal effects as a percent change of the unconditional means, those percent changes are very similar in magnitude to the Cox results.

for our second modification definition are similar, although we find more evidence of statistically significant, positive differences between the incidence of portfolio and private-level modifications. The samples of portfolio loans with non-missing information for DTI and documentation status were modified more often than the corresponding sample of private-label loans, but the magnitudes are still relatively small (10 to 20 percent difference from the unconditional mean). Finally, in Panel C, we see strong evidence for both the logit and Cox specifications, that delinquent private-label loans prepayed more often than portfolio loans. The differences are statistically significant for every one of the subsamples.

3.4 Redefault Probabilities and Cure Rates

In the previous subsections, we showed that there is little difference in the frequency of mortgage loan modifications between servicers of loans held in a private trust versus loans held in portfolio. There are two potential reasons that may explain the failure of those exercises to pick up important differences in servicer behavior that may truly exist. First, it may be that contract frictions in securitization trusts do not result in substantial differences in the frequency of modifications (the extensive margin) but do result in significant differences in the intensive margin, with respect to the types of modifications performed, the extent to which contract terms are modified, and, more broadly, the care or effort expended in each modification by private-label servicers compared to that expended by portfolio servicers. Second, there may be a type of renegotiation that our algorithm does not identify, but that is used to a large extent in loss mitigation efforts and used differently by servicers of private-label loans than by servicers of portfolio loans. For example, forms of forbearance, which are often called repayment plans in the industry, would not be picked up by our algorithm.[25] In this subsection, we use the LPS data to attempt to address these possibilities.

We perform two separate empirical exercises to address each of these concerns in turn. First, we compare redefault rates of private-label modified loans with those of portfolio modified loans. We define redefault as a loan that is 60 days delinquent or more, in foreclosure process or already foreclosed and now owned by the lender (REO for "real-estate-owned") six months after the time of the modification. If there are important differences in the manner by which servicers of private-label loans modify mortgages relative to the foreclosure procedures of servicers of portfolio loans, then we would expect to see significant differences in the subsequent performance of modified loans.

Second, to address the possibility that our algorithm misses an important aspect of

[25] However, as we argued above, PSAs do not contain restrictions on repayment plans, because they do not involve changing the terms of the mortgage. Thus, we would argue that differences in forbearance behavior that might exist could not be the result of contract frictions in securitization trusts.

renegotiation, we compare the cure rates of seriously delinquent, private-label loans to those of seriously delinquent portfolio loans. The idea behind this exercise is that any appreciable difference in servicer renegotiation behavior will manifest itself in differences in cure rates. It is important to stress however, that differences in servicer renegotiation behavior are only one potential explanation for differences that may exist in cure rates. To put this idea in the terms of logical reasoning, differences in cure rates are a necessary condition for significant differences in renegotiation behavior, but they are not a sufficient condition.

Table 8 contains the results of the redefault analysis. The first observation to note from the table is that the unconditional probability that a modified mortgage redefaults in this six-month period is very large, at about 20–40 percent for payment-reducing modifications (Panel A), and 40–50 percent for all modifications (Panel B). We argue below that the high level of redefault rates could explain why we observe so few modifications — very often they do not lead to successful outcomes even as little as six months after the modification. The second observation to note is that there is no statistically significant difference between the redefault rates of private-label loans and those of portfolio loans, once the observable characteristics of the mortgages are taken into account (this is valid for all of the subsamples). These results, combined with the statistics displayed in Table 4 suggest that there are no substantial differences in either the type of modification employed or in the care/effort expended by the two types of servicers.

Table 9 shows the results of logit models for the probability that a seriously delinquent loan subsequently cures. Our definition of a cure is that the loan is either current, 30-days delinquent, or prepaid after 12 months following the first 60-day delinquency. The first important point to make is that the unconditional cure probabilities are large (around 30 percent). Given that the unconditional modification probability is about 8 percent, this means that many loans cure without any intervention on the part of servicers. The second important observation to note in this table is that the cure probabilities for portfolio loans and private-label loans are quite similar. The unconditional cure probability is smaller by about 4.4 percentage points for private-label loans in the whole sample, but that is reduced to only 2.2 percentage points (statistically significant) when we control for observable characteristics of the loans and borrowers. We also include results for the subsamples of interest in columns 2–5. For each of the subsamples the sign of the difference actually reverses, as private-label loans were *more* likely to cure (the marginal effects are statistically significant, with the exception of the $FICO < 620$ sample). This is an important robustness check, as we argued above that unobserved heterogeneity is likely to be less of a problem in the subsamples (especially for the non-missing documentation status and DTI ratios sample and the full documentation sample). Thus, the change in the sign of the differences in cure rates between private-label servicers and portfolio servicers suggests that unobserved heterogeneity between the two loan types plays an important role.

4 Understanding the Empirical Results

If securitization does not block renegotiation, then why is it so rare? In this section, we build a simple model of the renegotiation decision, which, in a stylized way, mirrors the net present value (NPV) calculation that servicers are supposed to perform when deciding whether to offer a borrower a modification. We show that servicer uncertainty about whether the borrower will redefault even after successful renegotiation or uncertainty about whether the borrower will cure without renegotiation can dramatically affect the NPV calculation, ruining what a naive observer might think of as a "win-win" deal for the borrower and lender. While many proponents of modification are aware of the former problem, "redefault risk," none seem to be aware of the latter problem, which we call "self-cure risk."

In addition to the model, we also provide institutional evidence in this section that supports our arguments and findings above. This includes evidence of low modification frequencies in previous housing busts, well before the advent of securitization trusts; the equal treatment provision statements contained in the PSAs, which direct the servicer to behave as if it was in fact the investor of the mortgage-backed security and thus the owner of the mortgages; and finally, the absence of lawsuits to date directed at servicers by investors in mortgage-backed securities, which one would expect to find if modifications were unambiguously better than foreclosures from an NPV calculation.

4.1 A Simple Model of Loss Mitigation

We consider a simple model of a lender's decision to modify a delinquent loan.[26] There are three periods: $t = 0, 1, 2$. The borrower owes a mortgage payment of size m at time 1 and is due to repay the loan balance M in period 2. The mortgage is collateralized by a house, which is worth P_1 and P_2 in periods 1 and 2, respectively. In period 0, the lender has to make a decision to either modify the loan, or do nothing. If the lender fails to modify the loan, then, with probability α_0, the borrower will default in period 1, and the lender will foreclose and recover $P_1 - \lambda$, where λ is the cost of foreclosing on the property. If the borrower does not default next period, then the lender receives the periodic payment m in period 1, and the borrower repays the loan in full in period 2. The value to the lender of

[26] Our model shares some basic similarities with the approach in Ambrose and Capone (1996), who also identify a role for self-cure risk in assessing the profitability of a loss mitigation action.

the loan without modification equals the present discounted value of the cash flow:

$$\alpha_0 * \min[(P_1 - \lambda), M] + (1 - \alpha_0)[m + (1/R)M], \tag{1}$$

where we ignore discounting for the first period because there is no income in period 0. If the lender modifies the loan, then we assume that the borrower makes a reduced periodic payment m^* in period 1 with certainty, but then either defaults with probability α_1 or repays a modified amount M^* in period 2. The value to the lender of the modified loan is:

$$m^* + (1/R)\alpha_1 * \min[(P_2 - \lambda), M^*] + (1 - \alpha_1)(1/R)M^*. \tag{2}$$

Taking the difference between expressions (2) and (1) yields the following proposition:

Proposition 1 *Modification makes sense if:*

$$\begin{aligned}(\alpha_0 - \alpha_1)[m^* + \tfrac{1}{R}M^* - \min[(P_1 - \lambda), M]]& \\ - (1 - \alpha_0)[m + \tfrac{1}{R}M - (m^* + \tfrac{1}{R}M^*)]& \\ + \alpha_1[m^* + \tfrac{1}{R}\min[(P_2 - \lambda), M^*] - \min[(P_1 - \lambda), M]] &> 0. \end{aligned} \tag{3}$$

To interpret equation (3), divide the population of borrowers into three groups. The first group, with mass of $\alpha_0 - \alpha_1$ are borrowers who will repay in full with a modification but who will default otherwise. For this group, the investor gains the difference between the present value of the modified repayment $m^* + \frac{1}{R}M^*$ and the recovery given foreclosure, $\min[(P_1 - \lambda), M]$. The second group, with mass $1 - \alpha_0$, includes borrowers who will repay whether or not they receive a modification. For this group, the investor loses the difference between full repayment and the modified repayment. Gerardi and Willen (2009) refer to the first two terms as Type I error and Type II error, respectively, in analogy with the statistical concepts. In this context, Type I error corresponds to the cost of not renegotiating loans that need modifying, while Type II error corresponds to the cost of modifying loans that would be repaid in the absence of assistance. The third term, with mass α_1, includes borrowers who will default regardless of whether they receive a modification. For these borrowers, modification yields a periodic payment, but postpones foreclosure. Whether this is good or bad for the lender depends on the evolution of house prices and the rate at which the lender discounts the cash flow.

To illustrate the implications of the model, we compute some simple comparative statics. All else being equal, an increase in α_0 makes modification more attractive to the investor, while an increase in α_1 makes modification less attractive. Intuitively, a higher α_0 means higher Type I error and lower Type II error, and a higher α_1 implies higher Type II error. Since, in general, one would think that α_0 and α_1 would move in the same direction across borrowers, it is useful to note that an increase the gap, $\alpha_0 - \alpha_1$, makes modification more attractive.

We make three points about the model. First, when looking at the data, it is not sufficient to show that one would recover more from a modified loan than from foreclosure *ex post*, to prove that modification is *ex ante* optimal. To prove that a modification makes sense from the perspective of the lender, one must show that the Type I error, the value of the modified loans that would have defaulted, exceeds the Type II error, the value of the modified loans that would have paid off in the absence of modification. White (2009), among many others, focuses entirely on Type I error:

> The average loss for the 21,000 first mortgages liquidated in November was \$145,000, representing an average loss of 55 percent of the amount due. Losses on second lien mortgages were close to 100 percent. In comparison, for the modified loans with some amount of principal or interest written off, the average loss recognized was \$23,610. This seven-to-one difference between foreclosure losses and modification write-offs is striking, and lies at the heart of the failure of the voluntary mortgage modification program. At a minimum, there is room for servicers to be more generous in writing down debt for the loans they are modifying, while still recovering far more than from foreclosures in the depressed real estate market of late 2008. I will consider some of the reasons for this apparently irrational behavior in a later section.[27]

To see why this is wrong, take an extreme example with $\alpha_1 = 0$. In that case, the gain to modifications equals

$$\alpha_0[m^* + \tfrac{1}{R}M^* - \min[(P_1 - \lambda), M]] - (1 - \alpha_0)[m + \tfrac{1}{R}M - (m^* + \tfrac{1}{R}M^*)]. \tag{4}$$

With α_0 sufficiently low, modification will not make sense. To be clear, our criticism of White (2009) and others has nothing to do with the possibility that the modified loan will default, as we have assumed here that the modified loan will pay off in full.

The second point here is that both the rate at which lenders discount future payoffs and the evolution of prices affect the gains to modification. For mass $(1 - \alpha_1)$ of the borrowers, modification will simply delay foreclosure. In that case, the lender will get some extra income from any mortgage payments the borrower makes before redefaulting, but the lender has to wait longer to obtain the final payout and will get less if prices continue to fall.

[27] White (2009), p. 14–15

The third point is that the lender's information set plays a crucial role here, and one could argue that it should only contain information outside the control of the borrower. This would limit the set to the origination characteristics of the loan, prices, and interest rates. Employment status, income, and marital status all present problems, although they can be partially overcome—as in the case of unemployment insurance. Delinquency status, which seems a natural candidate, is a difficult issue. On one hand, a borrower has virtually complete control over it. On the other hand, it is a costly signal, as a 60-day delinquency does adversely affect one's credit history and future access to credit markets. Thus, when considering ways to design a profitable modification program, which implies attempting to maximize α_0 and minimize α_1, a lender must restrict its information set to a relatively small set of variables that are contemporaneously exogenous to the borrower.

4.2 Institutional Evidence

While the results from Section 3 may be surprising to market commentators who believe that contract frictions inherent in securitization trusts are preventing large-scale modification efforts in mortgage markets, we argue in this section that both historical evidence and evidence from securitization contracts actually support our findings.

First, we look at history. If securitization, or more precisely private-label securitization, inhibits renegotiation, then we would expect that renegotiation would have been common in the 1990s, when there was little private-label securitization, or in the 1970s, when securitization itself was rare. But, the historical evidence we have does not bear that out. In 1975, Touche Ross surveyed loss mitigation activities at savings and loans and found, "Lenders... were unwilling to either modify loans through extended terms or refinancing to a lower rate."[28] In the 1990s, a report commissioned by Congress to study foreclosure alternatives, said, "Along with loan modifications, long-term forbearance/repayment plans are the most under utilized foreclosure avoidance tools currently available to the industry."[29]

Second, many observers have focused on institutional factors that inhibit loan modification when the loan is securitized, but other factors may play a similar role for portfolio lenders as well. In particular, accounting rules force lenders to take writedowns at the time of the modification (reducing Tier II capital), to identify modified loans as troubled debt restructurings (under FAS 15), and also to impose burdensome reporting requirements on modified loans including loan-specific allowances for potential losses (under FAS 114). Additionally, payments made by borrowers for loans that are subject to "troubled debt re-

[28]Capone (1996), p. 20–21.
[29]Capone (1996), p. x.

structurings" are recognized only as principal repayments and generate to interest income until the bank can demonstrate that a borrower is "performing." All of the above accounting requirements potentially make modifications costly for a bank. Downey Financial, for example, attempted to refinance current borrowers out of risky option ARMs into safer, fixed-rate instruments and argued that the change should not affect their balance sheet because the borrowers had never missed payments. However, their accountants viewed the refinancings as "troubled debt restructurings," and forced the firm to restate the share of nonperforming assets for November 2007 to 5.77 percent from 3.65 percent.[30]

If modifications were truly in the best financial interest of investors in mortgage-backed-securities (MBS) as many commentators have alleged, we would expect to see concern on their part regarding the low levels of modifications performed to date. But, according to Cordell, Dynan, Lehnert, Liang, and Mauskopf (2008b), who interviewed a number of MBS investors, they (the investors) are not concerned that servicers are foreclosing on many more mortgages than they are modifying. Thus, there does not seem to be much concern by market participants that either incentives or contract frictions are inhibiting servicers from performing loan modifications. The evidence in the literature seems to suggest a small role for contract frictions in the context of renegotiation. In a 2007 study of a small sample of PSAs, Credit Suisse found that fewer than 10 percent of the contracts ruled out modifications completely, while approximately 40 percent allowed modifications, but with quantity restrictions,[31] and the rest, about half, contained no restrictions on renegotiation behavior. Hunt (2009) also analyzed a sample of subprime PSAs and concluded that outright modification bans were extremely rare. A 2008 report by the COP analyzed a number of securitized mortgage pools with quantity restrictions and concluded that none of the restrictions were binding. In terms of incentive issues, Hunt (2009) found that most of the contracts in his sample explicitly instructed the mortgage servicer to behave as if it were the owner of the pool of the loans:

> The most common rules [in making modifications] are that the servicer must follow generally applicable servicing standards, service the loans in the interest of the certificate holders and/or the trust, and service the loans as it would service loans held for its own portfolio. Notably, these conditions taken together can be read as attempting to cause the loans to be serviced as if they had not been securitized. (p. 8, insertion added)

[30] http://www.housingwire.com/2008/01/14/downey-financial-accounting-rules-suck/

[31] The quantity restrictions often took the form of a limit (usually 5 percent) on the percentage of mortgages in the pool that could be modified without requesting permission from the trustee.

5 Conclusion

There is widespread concern that an inefficiently low number of mortgages have been modified during the current crisis, and that this has led to excessive foreclosure levels, leaving both families and investors worse off. We use a large dataset that accounts for approximately 60 percent of mortgages in the United States originated between 2005 and 2007, to shed more light on the determinants of mortgage modification, with a special focus on the claim that delinquent loans have different probabilities of renegotiation depending on whether they are securitized by private institutions or held in a servicer's portfolio. By comparing the relative frequency of renegotiation between private-label and portfolio mortgages, we are able to shed light on the question of whether institutional frictions in the secondary mortgage market are inhibiting the modification process from taking place.

Our first finding is that renegotiation in mortgage markets during this period was indeed rare. In our full sample of data, approximately 3 percent of the seriously delinquent borrowers received a concessionary modification in the year following their first serious delinquency, while fewer than 8 percent received any type of modification. These numbers are extremely low, considering that foreclosure proceedings were initiated on approximately half of the loans in the sample and completed for almost 30 percent of the sample. Our second finding is that a comparison of renegotiation rates for private-label loans and portfolio loans, while controlling for observable characteristics of loans and borrowers, yields economically small, and for the most part, statistically insignificant differences. This finding holds for a battery of robustness tests we consider, including various definitions of modification, numerous subsamples of the data, including subsamples for which we believe unobserved heterogeneity to be less of an issue, and consideration of potential differences along the intensive margin of renegotiation.

Since we conclude that contract frictions in securitization trusts are not a significant problem, we attempt to reconcile the conventional wisdom held by market commentators, that modifications are a win-win proposition from the standpoint of both borrowers and lenders, with the extraordinarily low levels of renegotiation that we find in the data. We argue that the data are not inconsistent with a situation in which, on average, lenders expect to recover more from foreclosure than from a modified loan. At face value, this assertion may seem implausible, since there are many estimates that suggest the average loss given foreclosure is much greater than the loss in value of a modified loan. However, we point out that renegotiation exposes lenders to two types of risks that are often overlooked by market observers and that can dramatically increase its cost. The first is "self-cure risk," which refers to the situation in which a lender renegotiates with a delinquent borrower who does not need assistance. This group of borrowers is non-trivial according to our data, as we find that approximately 30 percent of seriously delinquent borrowers "cure" in our data without receiving a modification. The second cost comes from borrowers who default again after receiving a loan modification. We refer to this group as "redefaulters," and our results show that a large fraction (between 30 and 45 percent) of borrowers who receive modifications, end up back in serious delinquency within six months. For this group, the lender has simply postponed foreclosure, and, if the housing market continues to decline, the lender will recover even less in foreclosure in the future.

We believe that our analysis has some important implications for policy. First, "safe harbor provisions," which are designed to shelter servicers from investor lawsuits, are unlikely to have a material impact on the number of modifications and thus will not significantly decrease foreclosures. Second, and more generally, if the presence of self-cure risk and redefault risk do make renegotiation less appealing to investors, the number of easily "preventable" foreclosures may be far smaller than many commentators believe.

A Appendix: Identifying Modifications in the LPS Dataset

In this section we discuss in detail the assumptions that we used to identify modified loans in the LPS dataset. The LPS dataset is updated on a monthly basis, and the updated data include both new mortgages originated and a snapshot of the current terms and delinquency status of outstanding mortgages. Essentially, for a given mortgage, we compare the updated terms to the terms at origination, as well as the change in terms from the proceeding month, and if there is a material change over and above the changes stipulated in the mortgage contract, then we assume that the contract terms of the mortgage have been modified.

A.1 Interest Rate Reductions

We use a different set of rules to identify reduced interest rates for fixed-rate mortgages (FRM) and adjustable-rate mortgages (ARM). In principle, identifying a rate change for an FRM should be easy, since by definition the rate is fixed for the term of the mortgage. However, after a detailed inspection of the LPS data, it became apparent that some of the smaller rate fluctuations were likely due to measurement error rather than to an explicit modification. Thus, we adopt a slightly more complex criterion: The difference between the rate at origination and the current rate must be greater than 50 basis points; *and* the difference between the rate in the previous month and the current rate must be greater than 50 basis points; *and* either the mortgage must be 30-days delinquent with the loan currently in loss mitigation proceedings (as reported by the servicer) or the difference between the rate in the previous month and the current rate must be greater than 300 basis points (which allows for the possibility that a loan that is current could feasibly qualify for a modification).

Identifying interest rate reductions for ARMs is slightly more complicated, since by definition the interest rate is variable and can move both up and down. The LPS data contain the information necessary to figure out how much the interest rate should move from month to month. This rate is often referred to as the fully indexed rate, as it is normally specified as a fixed spread above a common nominal interest rate. The LPS dataset contains information regarding the initial rate, the appropriate index rate, and the spread between the index and the mortgage rate. In addition, the majority of ARMs are characterized by a period at the beginning of the contract in which the interest rate is held constant (these mortgages are often referred to as hybrid ARMs). At the end of this period, the interest rate adjusts (or resets) to a certain spread above an index rate and then subsequently adjusts at a specific frequency. The LPS dataset also contains information regarding the length of the initial fixed period, enabling us to identify this period in the data and determine the point at which the interest rate should begin to adjust (we refer to this period as the reset date). Our criterion for identifying an interest rate reduction for an ARM is as follows: The difference between the rate at origination and the current rate must be greater than 50 basis points; *and* the difference between the rate in the previous month and the current rate must be greater than 50 basis points; *and* if the reset date has passed, then the difference between the fully-indexed rate and the current rate must be at least 100 basis points ; *and* either the mortgage must be 30-days delinquent with the loan currently in loss mitigation proceedings (as reported by the servicer) or the difference between the rate in the previous month and the current rate must be greater than 300 basis points (which allows for the possibility that a loan that is current could feasibly qualify for a modification). In addition, we allow for more modest month-to-month decreases in the interest rate (200 to 300 basis points) as long as there is also a positive change in the delinquency status of the loan (that is, the loan is reported to be less delinquent). Our inspection of the data suggests that the majority of modifications involve a resetting of the delinquency status back to current, or a minor delinquency, so conditioning on this change likely eliminates many false positives.

A.2 Term Extensions

In theory, it should be straightforward to identify term extensions in the LPS data, but it can be tricky to do so because of possible measurement error in the variable that measures the remaining maturity of each loan. We defined a term extension in the LPS dataset to be a case in which the loan was at least 30-days delinquent at some point and the number of years remaining increases by at least 20 months *or* the change in number of years remaining is greater than the difference between the original term of the loan and the remaining term (for example, if the original maturity is 360 months, and the loan has 350 months remaining, then the increase in length must be at least 10 months) and, finally, either the monthly payment decreases *or* the principal balance increases *or* the loan is in loss mitigation.

A.3 Principal Balance Reductions

A reduction in the remaining balance of a mortgage is perhaps the most difficult type of modification to identify because of the prevalence of "curtailment" or partial prepayment among mortgage borrowers. For example, it is common for borrowers to submit extra mortgage payments in order to pay down the loan at a faster rate. For this reason, we were forced to adopt strict criteria to limit the number of false positives. Our criterion for identifying a principal balance reduction is as follows: The month-to-month decrease in the remaining principal balance must be at least -10 percent and cannot be more than -30 percent (the upper bound does not matter as much as the lower bound—we experimented with -40 percent and -50 percent, but did not find a substantial difference); the principal balance recorded in the previous month must be greater than $25,000 (since we throw second liens out, and look only at mortgages originated after 2004, this cutoff does not bind often); the month-to-month payment change must be negative (there are only a few cases in which the principal balance is reduced without a corresponding decrease in the payment, but in these cases the term is extended, and thus is picked up in our code for identifying term extensions); and, finally, the mortgage must be either 30-days delinquent or currently in loss mitigation proceedings (as reported by the servicer).

A.4 Principal Balance Increases

For interest-only and fully-amortizing mortgages, identifying an increase in the principal balance due to the addition of arrears is relatively straightforward. It becomes trickier for mortgages that allow for negative amortization, as the principal balance is allowed to increase over the course of the contract, by definition. For interest-only and fully-amortizing mortgages our criterion is: The month-to-month principal balance must increase by at least 0.5 percent (to rule out measurement error in the data); the loan must have been at least 30-days delinquent at the time of the balance increase; and, finally, the month-to-month payment change must be positive unless there is also a corresponding increase in the term of the loan. For mortgages that allow for negative amortization, the criterion is similar, except that the balance increase must be at least 1 percent and there must be a positive change in the delinquency status of the loan.

References

Ambrose, B., R. Buttimer Jr, and C. Capone (1997). Pricing Mortgage Default and Foreclosure Delay. *Journal of Money, Credit & Banking 29*(3).

Ambrose, B. and C. Capone (1996). Cost-benefit analysis of single-family foreclosure alternatives. *The Journal of Real Estate Finance and Economics 13*(2), 105–120.

Ambrose, B. and C. Capone (1998). Modeling the conditional probability of foreclosure in the context of single-family mortgage default resolutions. *Real Estate Economics 26*(3), 391–429.

Ambrose, B. and C. Capone (2000). The hazard rates of first and second defaults. *The Journal of Real Estate Finance and Economics 20*(3), 275–293.

Asquith, P., R. Gertner, and D. Scharfstein (1994). Anatomy of financial distress: An examination of junk-bond issuers. *Quarterly Journal of Economics 109*(3), 625–658.

Bolton, P. and D. S. Scharfstein (1996). Optimal debt structure and the number of creditors. *Journal of Political Economy 104*(1), 1–25.

Capone, C. A. (1996). Providing Alternatives to Mortgage Foreclosure: A Report to Congress. US Department of Housing and Urban Development.

Congressional Oversight Panel (2009). Foreclosure Crisis: Working Toward a Solution. March Oversight Report.

Cordell, L., K. Dynan, A. Lehnert, N. Liang, and E. Mauskopf (2008a). The incentives of mortgage servicers: Myths and realities. *Federal Reserve Board Finance and Economics Discussion Series* 2008-46.

Cordell, L., K. Dynan, A. Lehnert, N. Liang, and E. Mauskopf (2008b). The Incentives of Mortgage Servicers: Myths and Realities. *Finance and Economics Discussion Series, Federal Reserve Board 46*.

Cutts, A. and R. Green (2005). Innovative servicing technology: smart enough to keep people in their houses? *Building Assets, Building Credit: Creating Wealth In Low-Income Communities*, 348–377.

Cutts, A. and W. Merrill (2008). Interventions in Mortgage Default: Policies and Practices to Prevent Home Loss and Lower Costs. Technical report, Working Paper.

Danis, M. A. and A. N. Pennington-Cross (2005). The Delinquency of Subprime Mortgages. *SSRN eLibrary*.

Eggert, K. (2007). Comment on michael a. stegman et al.: What prevents loan modifications? *Housing Policy Debate 18*(2).

Foote, C., K. Gerardi, L. Goette, and P. Willen (2009, June). Reducing foreclosures: No easy answers. Working Paper 15063, National Bureau of Economic Research.

Foote, C., K. Gerardi, and P. Willen (2008). Negative equity and foreclosure: Theory and evidence. *Journal of Urban Economics 6*(2), 234–245.

Gan, Y. H. and C. Mayer (2006). Agency conflicts, asset substitution, and securitization. *NBER Working Paper No. 12359*.

Geanakoplos, J. and S. Koniak (2008). The best way to help homeowners – and the economy. Available at http://cowles.econ.yale.edu/ gean/crisis/index.htm.

Gelpern, A. and A. Levitin (2009). Rewriting frankenstein contracts: Workout prohibitions in mortgage-backed securities. http://ssrn.com/abstract =1323546.

Gerardi, K. and P. Willen (2009). Suprime Mortgages, Urban Neighborhoods and Foreclosure. *The B.E. Journal of Economic Analysis & Policy 9*(3 (Symposium).

Hunt, J. P. (2009). What do subprime securitization contracts actually say about loan modification? *Berkeley Center for Law, Business and the Economy working paper*.

Lauria, M., V. Baxter, and B. Bordelon (2004). An investigation of the time between mortgage default and foreclosure. *Housing Studies 19*(4), 581–600.

Levitin, A. (2009a). Helping homeowners: Modification of mortgages in bankruptcy. *Harvard Law & Policy Review Online 3*.

Levitin, A. (2009b). Resolving the foreclosure crisis: Modification of mortgages in bankruptcy. *Wisconsin Law Review*.

Mason, J. (2007). Mortgage Loan Modification: Promises and Pitfalls.

Mulherin, J. and W. Muller (1987). Volatile interest rates and the divergence of incentives in mortgage contracts. *Journal of Law, Economics, and Organization 3*(1), 99–115.

Pennington-Cross, A. (2009). The duration of foreclosures in the subprime mortgage market: a competing risks model with mixing. *The Journal of Real Estate Finance and Economics*, 1–21.

Pennington-Cross, A. and G. Ho (2006). Loan Servicer Heterogeneity and the Termination of Subprime Mortgages.

Piskorski, T., A. Seru, and V. Vig (2009). Securitization and distressed loan renegotiation: Evidence from the subprime mortgage crisis. *Chicago Booth School of Business Research Paper No. 09-02*.

Quercia, R., L. Ding, and J. Ratcliffe (2009). Loan Modifications and Redefault Risk.

Riddiough, T. and S. Wyatt (1994a). Strategic default, workout, and commercial mortgage valuation. *The Journal of Real Estate Finance and Economics 9*(1), 5–22.

Riddiough, T. and S. Wyatt (1994b). Wimp or tough guy: Sequential default risk and signaling with mortgages. *The Journal of Real Estate Finance and Economics 9*(3), 299–321.

Springer, T. and N. Waller (1993). Lender forbearance: evidence from mortgage delinquency patterns. *Real Estate Economics 21*(1), 27–46.

Stegman, M., R. Quercia, J. Ratcliffe, L. Ding, W. Davis, W. Li, K. Ernst, A. Aurand, and S. Van Zandt (2007). Preventive Servicing Is Good for Business and Affordable Homeownership Policy. *Housing Policy Debate 18*(2).

White, A. M. (2008). Rewriting Contracts, Wholesale: Data on Voluntary Mortgage Modifications from 2007 and 2008 Remittance Reports. *Fordham Urban Law Journal forthcoming*.

White, A. M. (2009). Deleveraging the American Homeowner: The Failure of 2008 Voluntary Mortgage Contract Modifications. *Connecticut Law Review forthcoming*.

Zingales, L. (2008). Plan b. *Economists' Voice 5*(6), 4.

Table 1: Examples of modifications in the data.

Example 1: Servicer cuts interest rate, capitalizes arrears in the balance of the loan and extends term to 40 years.

Date	MBA Delinq. Stat.	Interest Rate	Monthly Payment	Outstanding Balance	Remaining Term in Months
2008m10	9	6.5	907	141,323	340
2008m11	9	6.5	907	141,323	339
2008m12	9	6.5	907	141,323	338
2009m1	C	4.5	660	146,686	479

Example 2: Servicer capitalizes arrears into the balance of the loan but otherwise leaves the loan unchanged.

Date	MBA Delinq. Stat.	Interest Rate	Monthly Payment	Outstanding Balance	Remaining Term in Months
2008m5	6	9.25	1,726	208,192	346
2008m6	9	9.25	1,726	208,192	346
2008m7	9	9.25	1,726	208,192	346
2008m8	C	9.25	1,815	218,316	341
2008m9	C	9.25	1,815	218,184	340

Table 2: Robustness of the modifications algorithm

False positives by type of modifications

	# of Modifications Using WF CTS Data	False Positives
FRM Rate Reduction	5,381	8.0%
ARM Rate Reduction	8,951	22.0%
Principal Reductions	470	1.9%
Principal Increases	13,010	12.8%
Term Increases	394	2.3%

Overall success of algorithm

	No Mod Using Our Algorithm	Mod Using Our Algorithm	Total
No Mod in WF Data	2,329,187	3,559	2,332,746
Mod in WF Data	3,627	17,514	21,141
Total	2,332,814	21,073	2,353,887

Notes: We test our algorithm on a dataset of securitized mortgages in which the trustee has identified modifications (data is from Wells Fargo Trustee Services). The lower panel shows that about 17.2% of our modifications are false positives, meaning that we identify modifications but the trustee does not and about 16.9% are false negatives, meaning that the trustee identifies a modification but we do not.

Table 3: Modification Statistics

(1) By Type of Modification: 2007:Q1–2008:Q4

	# Loans Modified	Interest Rate Reductions		Principal Balance Reductions		Principal Balance Increases		Term Extensions	
		#	(% total)	#	(% total)	#	(% total)	#	(% total)
2007:Q1	10,940	600	5.3	700	6.2	8,660	76.4	1,380	12.2
2007:Q2	14,600	820	5.4	550	3.7	11,630	77.3	2,050	13.6
2007:Q3	17,720	770	4.1	810	4.3	15,170	81.2	1,940	10.4
2007:Q4	27,150	2,990	9.7	700	2.3	22,520	72.8	4,740	15.3
2008:Q1	36,230	6,010	13.8	900	2.1	32,100	73.8	4,500	10.3
2008:Q2	44,750	9,050	16.4	1,300	2.4	39,750	72.1	5,030	9.1
2008:Q3	62,190	16,280	20.3	940	1.2	56,940	70.9	6,110	7.6
2008:Q4	74,800	28,630	26.7	1,450	1.4	65,960	61.5	11,230	10.5

(2) By Payment Change

	Payment Decreases					Payment Increases				
	#	mean Δ		median Δ		#	mean Δ		median Δ	
		$	%	$	%		$	%	$	%
2007:Q1	2,080	-492	-13.2	-157	-10.0	5,020	106	6.7	62	4.4
2007:Q2	2,060	-464	-12.7	-141	-9.6	7,710	120	7.0	63	4.4
2007:Q3	2,470	-290	-12.9	-125	-9.7	10,380	110	6.7	60	4.3
2007:Q4	5,600	-367	-15.3	-159	-11.7	14,540	100	5.9	59	3.9
2008:Q1	11,500	-358	-14.0	-210	-13.2	18,720	108	6.5	62	4.3
2008:Q2	18,660	-425	-16.1	-239	-14.1	20,770	124	7.4	69	4.1
2008:Q3	31,770	-562	-21.5	-365	-20.2	26,400	124	6.3	63	3.6
2008:Q4	48,000	-503	-22.9	-315	-21.7	22,520	104	6.0	53	3.6

(3) Loan Characteristics of Modified Mortgages

	All Loans					Modifications				
	#	mean	p25	p50	p75	#	mean	p25	p50	p75
FICO (at origination)	1,892,777	706	660	713	762	17,533	622	580	621	662
LTV (at origination)	2,250,162	75	67	79	85	21,675	82	78	80	90
DTI (at origination)	1,346,093	37	28	38	45	13,945	41	35	41	47
Mortgage balance (at origination)	2,267,497	231K	121K	185K	288K	21K	234K	121K	186K	294K
% characterized as										
LTV = 80	14.4					21.7				
Subprime	6.8					47.4				
Fixed	71.2					39.7				
Hybrid ARM	7.7					26.2				
IO-ARM	11.3					13.1				
IO-Fixed	2.1					2.7				
Option-ARM	5.1					12.0				
Option-Fixed	0.3					1.4				
Owner	89.3					96.0				
Investor	7.1					2.6				
Vacation Home	3.7					1.1				
Purchase	51.9					49.0				
Low/no documentation	29.2					20.4				

Notes: These statistics were computed using a 10% random sample of the LPS data. Quantities obtained from the data are multiplied by a factor of 10. The percentages in panels (1) and (2) are taken with respect to the total number of modifications, and *not* loans modified. Thus, there is double-counting in the sense that some loans received multiple types of modifications in a given quarter.

Table 4: Modification Comparison by Payment Change

Private-label Modifications										
	Payment Decreases					**Payment Increases**				
	#	**mean**		**median**		#	**mean**		**median**	
		$	%	**$**	%		**$**	%	**$**	%
2007:Q1	106	-614	-14.42	-162	-10.85	239	121	6.02	76	3.37
2007:Q2	110	-505	-12.02	-222	-9.30	364	168	7.96	76	3.49
2007:Q3	128	-261	-11.82	-131	-8.42	558	145	7.52	75	3.65
2007:Q4	288	-313	-13.38	-163	-12.36	741	125	6.24	74	3.52
2008:Q1	634	-393	-16.12	-261	-15.65	938	133	6.76	79	4.08
2008:Q2	1,014	-540	-18.94	-334	-17.89	1,241	152	8.14	83	4.08
2008:Q3	1,778	-641	-22.01	-423	-19.95	1,805	137	6.22	70	3.31
2008:Q4	1,993	-565	-21.73	-367	-20.13	1,398	118	5.91	61	3.23
Portfolio Modifications										
	Payment Decreases					**Payment Increases**				
	#	**mean**		**median**		#	**mean**		**median**	
		$	%	**$**	%		**$**	%	**$**	%
2007:Q1	28	-759	-20.90	-428	-17.19	128	106	7.78	52	5.46
2007:Q2	19	-1172	-25.17	-656	-28.07	222	81	6.11	55	5.28
2007:Q3	31	-395	-17.13	-168	-15.29	255	71	6.13	43	5.37
2007:Q4	90	-474	-11.11	-90	-2.48	292	70	5.50	37	4.29
2008:Q1	187	-369	-10.00	-183	-8.08	331	80	6.59	33	3.97
2008:Q2	309	-304	-10.90	-117	-6.64	405	63	5.59	34	3.56
2008:Q3	376	-585	-25.19	-295	-17.85	359	105	7.04	39	4.26
2008:Q4	616	-794	-31.91	-384	-25.04	389	59	5.48	35	3.51

Table 5: Modifications (Main Sample)

Panel A: Unconditional Percentages

	Concessionary Mods	All Mods	All Mods + Prepayments
Portfolio	0.032	0.087	0.147
Private-label	0.026	0.084	0.155

Panel B: Logit Regressions (12 month horizon)

	Concessionary Mods	All Mods	All Mods + Prepayments
Private-label	-0.003	0.002	0.009
	-1.69	0.58	1.95
Initial Rate	0.001	-0.004	-0.007
	1.45	-5.7	-7.25
LTV Ratio	0	0	-0.002
	-0.24	-1.68	-11.14
$LTV = 80$	0	-0.014	-0.034
	-0.18	-6.25	-11.7
$FICO$	0	0	-0.002
	-0.02	-0.43	-4.62
$FICO^2$	0	0	0
	-0.39	-0.08	3.95
$FICO < 620$	0.002	0.029	0.034
	0.53	3.43	3.42
$620 \leq FICO < 680$	0.005	0.017	0.024
	1.46	2.95	3.41
Log Original Amount	0.004	0.007	0.022
	3.12	2.96	7.47
Equity at Delinquency	-0.001	-0.003	0
	-0.4	-1.09	0
Negative Equity	-0.006	-0.022	-0.022
	-1.6	-3.17	-1.77
Unemployment	0	-0.002	-0.005
	-0.37	-3.13	-4.37
Refi	0.006	0.015	0.04
	4.14	5.98	11.67
Subprime	0.02	0.037	0.042
	9.32	11.71	10.87
Other Controls	Y	Y	Y
# Mortgages	66,541	66,541	66,541

Panel C: Duration Model

	Concessionary Mods	All Mods	All Mods + Prepayments
Private-label	0.921	1.002	1.018
	-1.41	0.07	0.68
# Mortgages	87,343	87,343	87,343

Notes: Other controls include indicator variables for Jumbo, Option, Hybrid and Interest-Only mortgages, as well as for condos and multifamily homes. Panel B shows the marginal effects of logit regressions with a 12-month horizon, t-statistics shown below the coefficients. Standard errors are clustered at the zip code level. Panel C shows hazard ratio estimates from a Cox proportional hazards model.

Table 6: Modifications (Robustness tests with alternative samples)

Panel A: Concessionary Modifications

	All Loans	Subprime	$FICO < 620$	Non-missing Documentation and DTI	Fully Documented
Portfolio Mean	0.032	0.047	0.034	0.028	0.023
Private-label Mean	0.026	0.037	0.031	0.033	0.037
Marginal Effect	-0.003	-0.004	-0.003	0	0.007
(private-label)	-1.69	-0.94	-0.77	-0.14	1.46
# Mortgages	66,541	33,719	27,639	25,543	18,097

Panel B: All Modifications

	All Loans	Subprime	$FICO < 620$	Non-missing Documentation and DTI	Fully Documented
Portfolio Mean	0.087	0.111	0.097	0.092	0.077
Private-label Mean	0.084	0.103	0.109	0.107	0.124
Marginal Effect	0.002	0.004	0.007	0.006	0.025
(private-label)	0.58	0.61	1.06	0.97	2.94
# Mortgages	66,541	33,719	27,639	25,543	18,097

Notes: Portfolio and private-label means are unconditional probabilities of modification in each sample. Marginal effects are computed from logit models with a 12-month horizon that include all the controls in Table 5. Standard errors are clustered at the zip code level. t-statistics are reported below the marginal effects.

Table 7: Modifications Conditional on 30 Days Delinquency (Logits)

Panel A: Concessionary Mods

	All Loans	Subprime	$FICO < 620$	Non-missing Documentation and DTI	Fully Documented
Portfolio Mean	0.014	0.025	0.016	0.014	0.012
Private-label Mean	0.014	0.021	0.016	0.017	0.019
Marginal Effect	-0.003	-0.005	-0.001	-0.002	0.001
(Logit)	-2.72	-2.31	-0.55	-1.57	0.37
Hazard Ratio	1.03	1.147	1.027	0.969	1.237
(Cox)	0.59	1.83	0.31	-0.42	2.34
# Mortgages	120,558	51,285	43,550	47,993	34,403

Panel B: All Mods

	All Loans	Subprime	$FICO < 620$	Non-missing Documentation and DTI	Fully Documented
Portfolio Mean	0.038	0.056	0.051	0.042	0.052
Private-label Mean	0.042	0.055	0.051	0.047	0.035
Marginal effect	-0.004	-0.007	-0.004	-0.008	-0.001
(Logit)	-2.39	-1.79	-1.22	-3.16	-0.2
Hazard Ratio	1.043	0.951	1.008	0.909	1.065
(Cox)	1.42	-1.05	0.17	-2.23	1.21
# Mortgages	120,558	51,285	43,550	47,993	34,403

Panel C: All Mods + Prepayment

	All Loans	Subprime	$FICO < 620$	Non-missing Documentation and DTI	Fully Documented
Portfolio Mean	0.145	0.195	0.152	0.147	0.13
Private-label Mean	0.174	0.211	0.218	0.185	0.198
Marginal effect	0.023	0.021	0.044	0.016	0.029
(Logit)	7.31	2.98	6.46	3.47	4.54
Hazard Ratio	1.158	1.05	1.181	1.098	1.202
(Cox)	9.09	1.69	5.72	3.88	6.56
# Mortgages	120,558	51,285	43,550	47,993	34,403

Notes: Portfolio and private-label means are unconditional probabilities of modification in each sample. Marginal effects are computed from logit models with a 12-month horizon that include all the controls in Table 5. Hazard ratios are computed from Cox proportional hazard models with the same controls as in Table 5. z-statistics are shown below the coefficients, and t-statistics are reported below the marginal effects. Standard errors are clustered at the zip code level. Sample sizes refer to the logit regressions. The sample sizes for the Cox models are slightly larger.

Table 8: redefault Conditional on Modification

Panel A: Payment Reducing Mods

	All Loans	Subprime	$FICO < 620$	Non-missing Documentation and DTI	Fully Documented
Portfolio Mean	0.308	0.386	0.332	0.228	0.249
Private-label Mean	0.358	0.392	0.371	0.362	0.359
Marginal effect	0.016	-0.001	-0.015	0.03	-0.004
(Logit)	0.66	-0.03	-0.35	0.81	-0.1
# Mortgages	4,626	2,514	1,562	1,475	1,135

Panel B: All Mods

	All Loans	Subprime	$FICO < 620$	Non-missing Documentation and DTI	Fully Documented
Portfolio Mean	0.393	0.53	0.444	0.404	0.403
Private-label Mean	0.449	0.5	0.501	0.482	0.482
Marginal effect	0.008	-0.023	-0.009	-0.021	-0.033
(Logit)	0.58	-0.84	-0.38	-0.97	-1.24
# Mortgages	14,796	7,073	5,344	4,594	3,620

Notes: redefault is defined as loans that are 60 days delinquent, 90 days delinquent, in the process of foreclosure or in REO 6 months after the modification. Marginal Effects refer to the marginal effects of a logit model with a horizon of 6 months. t-statistics shown below the marginal effects. Standard errors are clustered at the zip code level.

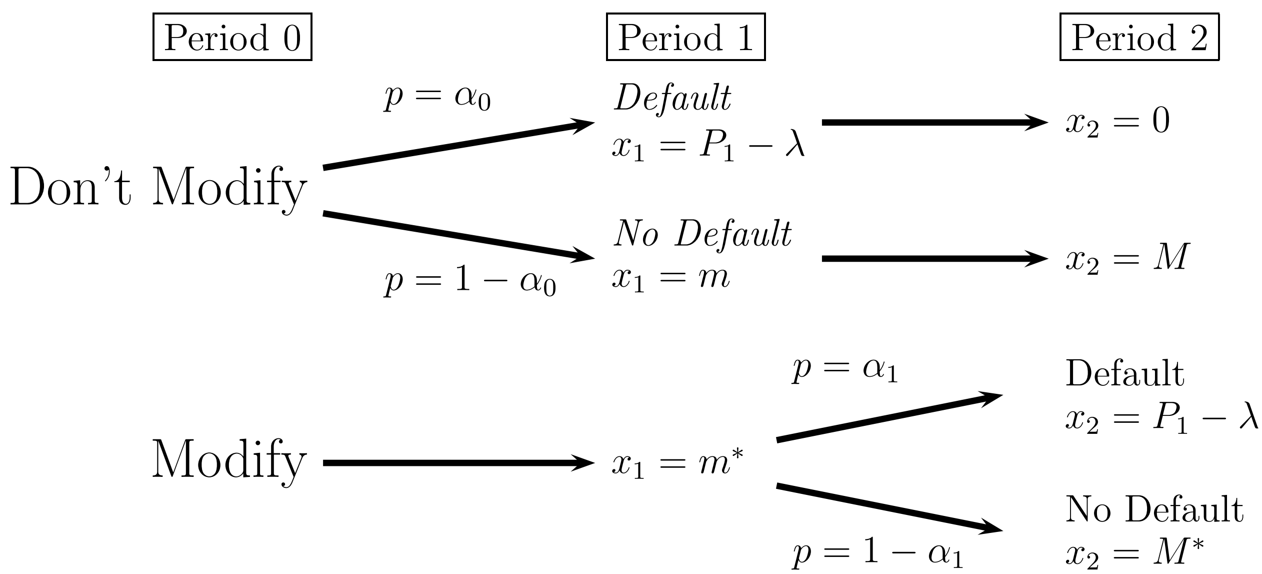
Table 9: Cure Conditional on 60 Days Delinquency

	All Loans	Subprime	$FICO < 620$	Non-missing Documentation and DTI	Fully Documented
Portfolio Mean	0.300	0.257	0.320	0.280	0.299
Private-label Mean	0.256	0.289	0.328	0.289	0.324
Marginal effect	-0.022	0.043	0.004	0.022	0.025
(Logit)	-4.32	4.31	0.44	2.8	2.43
# Mortgages	66,451	33,719	27,639	25,543	18,097

Notes: The dependent variable ("Cure") is defined as a loan that is either current, 30 days delinquent, or prepaid 12 months after the first 60-day delinquency. Portfolio and Private-label means are unconditional probabilities of modification in each sample. Marginal effects are computed from logit models with a 12-month horizon that include all the controls in Table 5. Standard errors are clustered at the zip code level. t-statistics are reported below the marginal effects.

Figure 1:

(1) Model of loan modification



(2) Understanding the lender's gains from modification

Share of borrowers	$1 - \alpha_0$	$\alpha_0 - \alpha_1$	α_1
Description	Borrower always repays Lender loses because borrower would have paid in full	Modification effective Lender gains because modified payments worth more than foreclosure	Borrower never repays Foreclosure is delayed May or may not help lender
Net gain	$m^* + \frac{1}{R}M^* - (m + \frac{1}{R}M)$	$m^* + \frac{1}{R}M^* - (P_1 - \lambda)$	$m^* + \frac{1}{R}(P_2 - \lambda) - (P_1 - \lambda)$
Error	"Type II error" Costly assistance to borrowers who can pay	"Type I error" Don't help borrowers who would have defaulted	"Type III error" Lender loses if R is large or if $P_1 - P_2$ is big

Appendix G

The Contagion Effect of Foreclosed Properties

John P. Harding*
Professor of Finance and Real Estate
University of Connecticut
2100 Hillside Road
Storrs, CT 06269-1041
860-486-3229
johnh@business.uconn.edu

Eric Rosenblatt
Fannie Mae
3900 Wisconsin Ave NW
Washington, D.C. 20016
202-752-3254
eric_rosenblatt@fanniemae.com

Vincent W. Yao
Fannie Mae
3900 Wisconsin Ave NW
Washington, D.C. 20016
202-752-1044
vincent_w_yao@fanniemae.com

July 13, 2009

*This paper has benefited from the helpful comments and suggestions of two anonymous referees. We are especially grateful to the editor, William Strange for his guidance and suggestions for improving the paper. We also thank the participants at the January 2009 AREUEA meetings, the participants at the University of San Diego research seminar and those attending the January 2009 session of the Weimer School for Advanced Studies in Real Estate and Land Economics for helpful comments and suggestions.
The views expressed in this article are those of the authors and should not be interpreted as those of either Fannie Mae or the University of Connecticut.

The Contagion Effect of Foreclosed Properties

Abstract

Although previous research shows that prices of homes in neighborhoods with foreclosures are lower than those in neighborhoods without foreclosures, it remains unclear whether the lower prices are the result of a general decline in neighborhood values or whether foreclosures reduce the prices of nearby non-distressed sales through a contagion effect. We provide robust evidence of a contagion discount by simultaneously estimating the local price trend and the incremental price impact of nearby foreclosures. At its peak, the discount is roughly one percent per nearby foreclosed property. The discount diminishes rapidly as the distance to the distressed property increases. The contagion discount grows from the onset of distress through the foreclosure sale and then stabilizes. This pattern is consistent with the contagion effect being the visual externality associated with deferred maintenance and neglect.

Keywords: Foreclosure, Contagion
JEL Classification: G12, G21, R31

Introduction

After an extended period of house price appreciation, in 2008 the United States experienced declining house prices and rapidly increasing foreclosures. It is widely accepted that home foreclosures have significant negative externalities.[1] Recent studies (e.g., Immergluck and Smith, 2006 and Lin et al., 2009) have reported that the presence of foreclosed properties is associated with lower sales prices for nearby non-distressed properties. These results are often interpreted as supporting the idea that nearby foreclosed properties lower the prices of neighboring houses (e.g., Center for Responsible Lending, 2008 and 2009). There are several possible mechanisms through which a foreclosed property can affect the values of nearby properties. The first is through a negative visual externality as the appearance of the neglected property deteriorates. In addition to normal depreciation, many properties undergoing foreclosure experience gross neglect, abandonment and vandalism which significantly alter their exterior appearance. A second mechanism, social interaction, is described by Ioannides (2002) who shows that individuals' valuations of their own homes are influenced by those of their immediate neighbors. As a result, a decline in value of a nearby foreclosed property can result in lower seller reservation prices and lower sales prices for nearby non-distressed properties. Foreclosed properties also increase the supply of homes and the sellers of foreclosed properties are highly motivated to sell quickly putting downward pressure on local prices. Finally, the prospect of imminent foreclosure reduces the incentive of homeowners to invest in socially desirable individual and community activities which can reduce the attractiveness of the neighborhood to potential buyers. Concern about such negative externalities has helped shape public policy since the 1930's. In 2008, the Bush Administration, Congress and various regulators introduced new programs to aid troubled homeowners, reduce foreclosures and fund the acquisition of distressed properties by local governments. In 2009, the

[1] For example, in a May 5, 2008 speech, Federal Reserve Board Chairman, Ben Bernanke, stated "High rates of foreclosure can have substantial spillover effects on the housing market, the financial market and the broader economy."

Obama Administration introduced a $275 billion program to shore up the housing markets and aid homeowners deemed to be “at risk” of foreclosure. The justification for the use of public funds for such efforts is based on the negative externalities of foreclosures.

Three recent papers (Lin et al., 2009, Immergluck and Smith, 2006, and Shlay and Whitman, 2004) have documented a negative relationship between non-distressed sales prices and the number of nearby foreclosures.[2] Significantly, none of these studies have focused on whether foreclosed properties cause a decline in the value of nearby homes or whether all homes in the neighborhood experienced a decline in value which, in turn, led to more local foreclosures. A general decline in home values would trigger defaults at some houses in the neighborhood, but not all, making it difficult to distinguish cause and effect.[3] The critical question is whether the presence of nearby foreclosures causes an incremental decline in values of nearby properties or is just a symptom of a general decline in house prices.

All three recent papers derive their estimates using hedonic models. The contagion effect is estimated by including measures and/or indicators of nearby distressed properties as additional independent variables. A general problem with the hedonic specification is that it is impossible to observe all house, location and local market characteristics and thus the coefficient estimates of the included variables are subject to omitted variable bias. The omitted variable problem arises very frequently in the study of urban externalities, including the impact on home values of environmental problems, schools, and commercial development. A recent example is the analysis by Pope (2008) of the impact of the presence of a registered sex offender on the prices of nearby homes. Pope finds that sex offenders tend to locate in lower valued areas and shows that estimating the impact of a nearby sex

[2] See Schwartz et al. (2003) for a summary of earlier literature on housing spillover effects.
[3]Theory suggests that although a decline in house value that creates negative equity is a necessary condition for a foreclosure, the foreclosure is triggered by a cash flow or income problem that reduces the ability of the household to make the mortgage payments. Thus, the incidence of foreclosure in a neighborhood undergoing a general decline in value depends on the distribution of loan-to-value ratios in the area and the incidence of income disruption (e.g., illness, divorce, loss of job). It is reasonable to assume that the distribution of original loan to value ratios and the incidence of trigger events are independent of the local price changes. See Avery et al. (1996) for more discussion.

offender using an indicator variable significantly overstates the effect of the sex offender. Measures of the number of nearby foreclosures in a cross-sectional hedonic model are especially vulnerable to the omitted variable problem because it is likely that the number of nearby foreclosures is correlated with unobserved property and location characteristics and especially the local trend in market prices.[4]

The repeat sales approach provides an alternative estimation procedure that substantially reduces the omitted variable problem of hedonic models and is well-suited to identify the separate effects of the overall price trend and the contagion effect of nearby foreclosures. We use an extension of the repeat sales model (first suggested by Bailey et al., 1963, and later utilized by Schwartz, et al., 2003, and Harding et al., 2007) that explicitly controls for property characteristics that are expected to change between sales and thus do not difference out of the repeat sales regression. We treat the number of nearby foreclosures as such a variable.

We use a sample of approximately 400,000 repeat sales transactions to study this issue. At each sale, we collect information about the number and distance of distressed properties in the vicinity of the subject property. Because the foreclosure process includes three distinct phases (a period of delinquency leading to a foreclosure sale, the period after the lender takes title through foreclosure and the period after the REO sale to a new permanent owner), for each foreclosed property, we collected information about the foreclosure sale date and the subsequent REO sale date. This information enables us to identify the phase of the foreclosure process the nearby property was in when the subject property sold and thereby estimate whether the contagion effect differs with the phase of foreclosure. The results provide insights into the mechanism by which foreclosed properties affect nearby values.

Our results confirm that nearby distressed properties have significant negative contagion effects over and above the local trend in house prices. We estimate a peak contagion effect from the closest

[4] Immergluck and Smith (2006) discuss this problem and try to address it by including a very large set of neighborhood characteristics. However, they do not control for the overall market trend in prices.

foreclosures of approximately one percent. The estimated contagion effect varies from one MSA to another but in all cases it diminishes quickly with the distance between the subject property and the foreclosed property. We also find that the strongest contagion effect generally increases quickly during the year preceding the foreclosure sale. In general, after the foreclosure sale, the trend slows or stops, but the negative effect remains well after the lender resells the property. We find that the maximum negative effect of a single foreclosure on a nearby sale occurs around the time of the foreclosure sale. At that point, a single foreclosure reduces the value of a house located within 300 feet of the foreclosure by approximately one percent. We estimate the model for seven different MSAs with different rates of price appreciation in the 2000's and find a significant negative contagion effect in all seven, although the magnitude of the effect varies by MSA.

Our estimate of the peak contagion effect of a single nearby foreclosed property is smaller than the previous hedonic estimates. We also find a sharper decline with distance than did earlier studies. Based on our estimates, a foreclosed property 1/8th of a mile away would have a peak negative effect of less than .5% or roughly half that estimated by Immergluck and Smith (2006). The cumulative effect of all nearby foreclosures can be significantly higher. We find that if there are three or more foreclosed properties within 300 feet of a non-distressed sale, the non-distressed property sells at a price approximately three percent below market.

Historically, lender forbearance and foreclosure moratoria have been popular policy responses to high levels of mortgage delinquency and foreclosures. The current housing and financial crisis has renewed interest in these responses.[5] However, because our results suggest that the root cause of the contagion effect may be reduced maintenance, neglect and vandalism, the most efficient way to reduce the negative externality of foreclosures would be to avoid extended periods of reduced maintenance and

[5] For information on the use of foreclosure moratoria in the 1930s, see Wheelock (2008). In early 2009, many large mortgage lenders implemented voluntary foreclosure moratoria and in June, 2009, the state of California imposed a 90-day moratorium on housing foreclosures.

neglect. One way to achieve this would be to negotiate a permanent loan modification assuring that the homeowner has both the resources and incentive to maintain and protect the property. Failing that, the best solution is a quick resolution of the problem through foreclosure and subsequent transfer of the property to a new owner who has the capacity and incentive to properly maintain and protect the home.

The remainder of the paper is organized as follows. Section 2 presents the methodology used to estimate the contagion effect. The next section presents the data and discusses the base estimation results while Section 4 explores the cumulative effect of multiple nearby foreclosures. Section 5 discusses robustness tests and Section 6 concludes the paper.

2. Methodology

We begin with the standard log-linear hedonic specification for modeling house prices based on the premise that the price of a bundled good such as a specific house can be expressed as a function of the inner product of a vector of characteristics and the market-determined shadow prices of those characteristics (see Griliches, 1971, Rosen, 1974 and Epple, 1987).

$$P_t = e^{[sC]} \quad \text{or} \quad \ln(\mathrm{P_t}) = s'C \tag{1}$$

In equation (1), P_t represents the price of a house at time t. The house and its locational attributes are fully described by the vector of characteristics, C. The important insight of Rosen (1974) was that under the assumption of sufficient variation in the traded bundles, the vector of shadow prices is revealed to agents in the economy through trades that differ in a single characteristic. If the presence of nearby distressed properties affects the value of the house, in theory, we should simply include the presence of distressed properties in the vector of attributes.

In practice, the vector of shadow prices, s, is estimated by regressing observed house prices on the vector of observed characteristics. A problem arises, however, because we do not observe all characteristics of the subject house and its neighborhood and market. To describe this problem, consider partitioning the vector C into two components -- $C = \begin{pmatrix} C_1 \\ C_2 \end{pmatrix}$, where C_1 denotes the vector of observed characteristics (including the presence of a nearby distressed property) and C_2 the vector of unobserved characteristics, including the local trend in house prices and other unobserved property and location attributes. If C_1 and C_2 are independent, then estimating equation (1) using only C_1 provides unbiased estimates of the corresponding shadow prices, s_1. However, if an element of C_1 is correlated with elements of C_2, then the estimated coefficient on that characteristic will be a combination of the effects of the unobserved characteristics on price and the direct effect (i.e., shadow price) of the observed characteristic. Specifically, consider including the number of nearby distressed properties in C_1. It is well established that the likelihood of foreclosure of a given property increases with the contemporaneous loan-to-value (LTV) ratio. If the overall trend in prices is downward, that will directly lower the price of the subject home and also trigger nearby foreclosures.[6] If the empirical model does not control for the overall price level, a large number of nearby foreclosures may proxy for an overall decline in home prices and thus a negative coefficient on the measure of foreclosures could actually be an estimate of the local decline in prices and not a true contagion effect.

The repeat sales model derived by Bailey et al. (1963) and Case and Shiller (1989) provides a way to jointly estimate the overall trend in prices and the direct contagion effect of nearby distressed properties. The basic idea of a repeat sales specification is that most elements of the characteristic vector (both observed and unobserved) remain constant between the two sales and consequently difference out when we model the rate of price appreciation instead of the price. This greatly simplifies the estimation of the model because those characteristics can be deleted from the model as shown below.

[6] Nearby foreclosures would increase because the current LTV ratio depends on past financing decisions as well as the current price. Different households will have taken out loans with different LTV ratios and made different use of secondary financing. Consequently, the same house price decline will leave some properties with prices in excess of the current debt while other households are "under water" because they owe more than their home is currently worth. For "under water" borrowers, random trigger events (e.g., divorce, job loss, sickness etc.) are more likely to result in default and foreclosure.

We begin with a slightly expanded version of equation (1), describing the price at time t for property i:

$$P_t^i = e^{\gamma_t} e^{\left[s_1 C_1^i + s_2 C_2^i\right]} e^{aN_t^i} e^{\eta_t^i} \quad \text{or } \ln(P_t^i) = \gamma_t + s_1 C_1^i + s_2 C_2^i + aN_t^i + \eta_t^i \tag{2}$$

C_1^i and C_2^i include observed and unobserved explanatory variables related to the price of an individual property. The error term, η_t^i, is assumed to be independent and identically distributed and captures pure random shocks to the transaction price. There are two new elements in equation (2)—γ_t and aN^i_t. The first term, e^{γ_t}, represents the overall market price level and adjusts the base value of the bundle of house attributes to current market price. The second term, $e^{aN_t^i}$, adjusts the price of the particular house for the effects of nearby distressed properties; N^i_t equals the number of nearby (to property i) distressed properties.[7] While N^i_t is technically a location characteristic, we separate it from the other elements of C because we anticipate that it will change between sales and is thus different from the other elements of C. Using the same notation, equation (3) describes the price at the time of the next sale of the same property. The second sale occurs at time t +τ.

$$P_{t+\tau}^i = e^{\gamma_{t+\tau}} e^{\left[s_1 C_1^i + s_2 C_2^i\right]} e^{aN_{t+\tau}^i} e^{\eta_{t+\tau}^i} \quad \text{or } \ln(P_{t+\tau}^i) = \gamma_{t+\tau} + s_1 C_1^i + s_2 C_2^i + aN_{t+\tau}^i + \eta_{t+\tau}^i \tag{3}$$

Following the standard derivation of the repeat sales equation, we difference the log versions of the two equations, assuming s_1C_1 and s_2C_2 are unchanged between t and t+τ for a particular property *i*:

$$\ln\left(P_{t+\tau}^i \big/ P_t^i\right) = \left(\gamma_{t+\tau} - \gamma_t\right) + a(N_{t+\tau}^i - N_t^i) + \varepsilon_{t,t+\tau}^i \tag{4}$$

[7] To simplify the presentation of the basic approach, we use the imprecise notion of "nearby" at this point. The reader can think of nearby being defined as a distance of less than three hundred feet. When we implement the approach, we use several different classes of distance.

The assumption that s_1C_1 and s_2C_2 are unchanged between t and $t+\tau$ is essential to the repeat sales model. While this assumption is common to all repeat sales applications (including research papers such as Harding et al., 2007, and the widely cited Case-Shiller and OFHEO price indices), it is also common to screen the repeat sales pairs to eliminate those where the assumption is questionable. Our standard filter screens out repeat sales pairs that have a high rate of appreciation (ten percent per quarter) combined with a short holding period (less than two years) as well as pairs with an appreciation rate of eight percent per quarter and a holding period longer than two years. Nevertheless, to the extent that property characteristics change between observations, the effect of such changes will be included in the error term and the resulting coefficients would be biased if the variables on the right hand side of equation (4) are correlated with the error term.[8]

The estimates of γ_t and $\gamma_{t+\tau}$ provide an estimate of the change in the local price level between t and $t+\tau$. The contagion effect of distressed properties is measured by the coefficient a, which in this specification, is assumed to be constant over time.[9] The ability of the repeat sales specification to jointly estimate the overall price level change and the contagion effect is a significant advantage over the hedonic specification for this application.

The estimated contagion effect will be unbiased as long as $(N^i_{t+\tau}-N^i_t)$ is uncorrelated with the error term. Given the assumptions underlying equations (2) and (3), N^it is independent of the random idiosyncratic variation in individual house prices, η^i_t. The random shock to the value of house i at time t cannot influence the number of nearby distressed properties. Therefore, if the error term in equation (4), $\varepsilon^i_{t,t+\tau}$ is the just the difference in the two price equation error terms, then the change in the number of

[8] The repeat sales methodology also assumes no change in attribute prices between sales. If the attribute vector is fixed, then this assumption reduces to assuming that the inner product of shadow prices and attributes remains constant. The model allows for an overall shift in price levels and so changes in tastes that lower one price and raise another can have offsetting effects which do not alter the inner product. Even if the effects do not offset, it is unlikely that such shifts in attribute prices relative to the overall price level will be correlated with the change in the number of nearby foreclosures.

[9] The specification also implies that the contagion effect is linear in the number of foreclosures. Although the linearity assumption proves to be reasonable, we later estimate models with different specifications and discuss the impact of multiple nearby foreclosures in Section 4 later in the paper.

nearby foreclosed properties will be independent of the error term. However, it is possible that the error term, $\varepsilon^{i}_{t,t+\tau}$ includes a component attributable to the changes in other elements of the vector, C, that are not explicitly included in the model. If these omitted variables were correlated with the change in the number of foreclosed properties, the estimate of the contagion effect could be biased. As mentioned above, one defense against this potential problem is to screen out repeat sales pairs where it is likely that there has been some unobserved change in characteristics. A second solution is to use instrumental variables to estimate the model. In Section 5 we discuss the application of different filters for screening unusual observations and the use of instrumental variables to address the possible endogeneity issue. Briefly summarizing our findings, the results presented in Section 3 are robust to changes in data filters and further we do not find evidence of significant endogeneity.

3. Data and Results

3.1 Data

In order to estimate equation (4), we need to identify local markets where we have both a large sample of repeat sales and the ability to identify all nearby foreclosures that could potentially influence the transaction price. It is critical that the data include a complete inventory of all foreclosures near each sale with enough information about the foreclosure to precisely locate the foreclosed property in both space and time. At a minimum, one needs the latitude and longitude of the property, the foreclosure date and the REO sale date.

We begin with a large proprietary mortgage database, containing approximately half of all national mortgage transactions over the period from 1989 through 2007. From this mortgage data, we can identify home purchase and sale transactions and the outcome of the mortgage: prepaid, foreclosed or still active at the end of the sample period. For the housing transactions in this database, it is known with near certainty whether the ownership of the underlying property was transferred through a foreclosure sale. For these foreclosed properties, the data include both the foreclosure sale date and the

REO sale date. However, since the database does not contain all mortgages, and particularly since it does not contain many subprime mortgages, it does not provide a complete record of foreclosures in any given area. To augment that record, additional data were acquired from vendors of housing transactions data. This purchased data includes information on foreclosure sales and REO sales as well as normal purchase transactions. Unfortunately, even the purchased transactions data have gaps in coverage of foreclosures and as a result we must restrict our analysis to those geographic areas where we are highly confident that we have complete coverage of foreclosure activity.

We use the proprietary data in conjunction with the purchased data to identify zip codes where the purchased data provides close to complete coverage of foreclosures by finding those zips where the purchased transaction data correctly identifies at least eighty percent of the foreclosure sales from the proprietary database. We assume that if the purchased data can identify at least eighty percent of the known foreclosure sales in the mortgage database for a particular zip code, its coverage of foreclosures in that location resulting from other mortgages is similarly good. For the 296 zip codes that met this criterion, we combine all foreclosures from the proprietary mortgage database with those from the purchased transaction database to create a local inventory of foreclosures which we believe provides a coverage rate well above eighty percent.

To estimate equation (4), we need information on repeat sales pairs as well as the foreclosure information discussed above. The basic source for the repeat sales pairs is the GSE loan database used to generate the Federal Housing Finance authority (FHFA) home price indices augmented with the previously discussed purchased transactions data covering non-GSE home purchase transactions. The repeat sales pairs were restricted to single-family detached houses and include only true market transactions. Refinancings and all foreclosure-related sales were excluded.[10] In addition, we filter the

[10] Leventis (2009) provides evidence that the inclusion of distressed sales has a small but significant negative effect on repeat sales indexes. In the analysis presented here, we exclude all properties that are included in the foreclosure sample (i.e., experience a foreclosure at any time during the sample period) from the repeat sales sample.

repeat sales pairs to eliminate outliers and records where the holding period returns appear abnormal and suggest that the underlying assumption of no change in property and neighborhood characteristics is questionable. Our primary screen eliminates any observation with price appreciation greater than eight percent per quarter – although the return threshold is somewhat higher (ten percent) when the holding period is less than two years.[11] Finally, imposing the requirement that each zip code has sufficient transactions (post screening) to compute a robust repeat sales index for the period from 1990 through 2007 reduced the number of zip codes to 140. For each repeat sales observation in these zip codes we collected the property address, the initial purchase date and price, and the second sale date and price.

The final step in building the data base was to geocode all records in both files so that we could identify nearby foreclosures for each repeat sales transaction using the geocoded locations. We categorized the distance between the subject property in the repeat sales pair and a foreclosure using four concentric rings with different radii around each subject property.[12] The rings we selected were: 1) 0 to 300 feet (Ring 1); 2) 300 feet to 500 feet (Ring 2); 3) 500 feet to 1000 feet (Ring 3); 4) 1000 feet to 2000 feet (Ring 4). The innermost ring can be thought of as including the two to three nearest neighboring properties in each direction (which are probably visible from the sale property), while the second ring can be thought of as having a foreclosure on the same block as the subject property. Properties in this ring might be seen by potential buyers visiting the home for sale but are not likely to be visible from the property. Foreclosures in the two outer rings would not be visible from the subject property but could influence the subject price by altering a potential buyer's perception of the neighborhood and/or by providing competition for buyers. For each sale in the repeat sales sample, we

[11] As a robustness check we also estimated the model using other more restrictive screens. The results of these robustness checks are discussed in Section 5.
[12] For computational reasons, the count of foreclosures in a ring is restricted to include only those in the same zip code as the subject property. This truncation of the rings creates measurement error in the count of nearby foreclosures for some properties. We check the robustness of our results to this potential problem using instrumental variables and also by estimating a model for a single zip code area where we are able to calculate the estimates with and without this restriction. The results are very similar.

searched the foreclosure file to identify all foreclosed properties that were somewhere in the foreclosure process[13] on the date of the sale and also were located in one of the concentric rings.

In addition to identifying the distance of the foreclosed property from the subject property, we categorized each foreclosed property by the phase of the foreclosure process it was in at the date of the repeat sale transaction. To do this, we considered thirteen windows of time linked to either the foreclosure sale date, F, or the REO sale date, S. Figure 1 shows how we identified the different phases of the foreclosure process. We break the time before F into four quarterly periods or "windows" and the time after F and before S into at most five windows. Because the time between F and S varies from loan to loan, not all loans pass through all the indicated post-foreclosure windows. In our classification scheme, as soon as S occurs (which can be shortly after F, but on average occurs approximately ten months after F) we assign the date to one of the four post-S categories and not the post-F categories.[14] For example, consider a foreclosure sale, F, that occurred on July 1 with an REO sale, S, to a third party on August 15. If we observed a nearby repeat sales transaction on August 30, we would classify the foreclosed property as falling into the {S to (S+3months)} window as opposed to the {F to (F+3 months)} category. If the repeat sales date had been August 1 (falling before S), the foreclosed property would be classified as falling into the {F to (F+3)} window.[15]

The 104 zip codes with good foreclosure coverage represent thirty-seven MSAs and thirteen states. The distribution of observations across states is provided in Table 1. Because we want to control as well as possible for local market conditions, we further limit our sample to the seven MSAs with at least 7,500 repeat sales pairs. We estimate equation (4) for each of the seven selected MSAs. These

---

[13] The foreclosure process is defined as the period beginning twelve months before the foreclosure sale and ending twelve months after the REO sale.

[14] As a result of this convention, the number of foreclosures observed in the windows after F declines with time. For this reason, we define the last post-foreclosure window to include all properties where more than twelve months have passed without an REO sale.

[15] To simplify notation in future discussions of the phase windows, we describe a window using only the later date, as long as the meaning is clear.

MSAs represent different economic and housing market conditions ranging from those that are often cited as examples of "boom" areas (e.g., Las Vegas) to those that experienced more modest price changes and less robust economic growth (e.g., Memphis and Columbus). These seven local markets include 72 zip codes and more than sixty-five percent of the total repeat sales transactions in our data. In most cases, the zip codes in these MSAs that passed the earlier screens for good foreclosure coverage represent a small fraction of the total zip codes in the whole MSA. Further, the selected zip codes are not necessarily contiguous.

Table 2 presents summary statistics for the foreclosure sample. The first column provides data on the combined sample of foreclosures for all seven MSAs while the next seven columns report the data for each of the selected MSAs. The table shows a dramatic increase in foreclosures over the sample period but also shows significant variations across MSAs.[16] In the Los Angeles MSA, foreclosures were quite high in the local housing recession of the mid -1990s and although they are increasing at the end of the sample, they remain below the peak of 1997. In Atlanta, Charlotte, Columbus and Las Vegas, the number of foreclosures in the 2005-2007 period is much higher than at any other period in the sample. The data for Memphis and St. Louis show a history of chronic foreclosure problems related to local economic conditions. The bottom portion of the table provides data from the Mortgage Bankers Association (MBA) for the rate of new foreclosure filings in the nation (column one) and the state corresponding to each of the seven MSAs. These numbers are provided to give an overall perspective on foreclosure rates but it is important to keep in mind that in addition to the geographic differences between state and MSA, the numbers in the top portion of the table represent the stock of homes

[16] The apparent decline in foreclosures in 2007 is the result of lags in recording data related to the foreclosure process — especially the key foreclosure and REO sale dates. We do not include foreclosures for which we do not have both the foreclosure date and the REO sale date. Because the foreclosure sample was drawn in the summer of 2008, some foreclosures from the second half of 2007 still had incomplete information and were excluded from the sample. We tested for sensitivity of our findings to the exclusion of these foreclosures by estimating our models using repeat sales through 2006 instead of 2007. Our parameter estimates were essentially the same.

somewhere in the foreclosure process while those in the bottom panel report the flow of new foreclosure filings.

Table 3 presents descriptive statistics for the repeat sales sample. The purchase price ($140,600) reported in the table represents the average price at the time of the initial purchase in the repeat sales pair, while the sale price ($171,377) reflects the second sale of the pair. The typical holding period (a measure of local mobility) was just under five years. Homeowners in the sample earned an average nominal annual return from price appreciation of approximately four percent. The data for the individual MSAs confirms that these seven MSAs represent a diverse sample of housing markets. The average purchase price ranges from $93,752 in St. Louis to $194,537 in Los Angeles. The average holding period ranges from 3.4 years in Las Vegas to 5.4 years in St. Louis and annual holding period returns averaged 2.6% per year in Charlotte and 9.2% per year in Las Vegas. The repeat sales were generated by 281,088 distinct properties. More than two-thirds of the properties had a single repeat sale and the properties with at most two repeat sales account for more than ninety percent of the sample. This pattern is typical for large repeat sales samples such as those used by FHFA. Panel B of Table 3 provides information about the timing of the sales in the repeat sales pairs.

3.2 Model Specification

We follow the standard repeat sales methodology (see Case and Shiller, 1989), to implement the OLS estimation of equation (4). As discussed in the previous section, we measure the contagion discount using a total of fifty-two "buckets"; where each bucket contains the difference in the number of nearby foreclosures that are within the specified ring and are at the specified phase of the foreclosure process (see Figure 1). The resulting equation to be estimated by OLS is:

$$\ln\left({P^i}_{t+\tau} \Big/ P^i_t \right) = \sum_{j=1}^{72} \gamma_j D_{i,j} + \sum_{d=1}^{4} \sum_{p=1}^{13} a_{dp} {B^i}_{dp} + {\varepsilon^i}_{t,t+\tau} \tag{5}$$

where ${B^i}_{dp} \equiv (N^{idp}_{t+\tau} - N^{idp}_t)$ and $N^{idp}_t \equiv$ the number of foreclosures at distance d from property i in phase p at t.

In equation (5), D is the standard matrix of indicators that identify sales dates. In estimating equation (5), we use the Case and Shiller (1989) GLS approach which allows for an increase in the variance of the disturbance term, $\varepsilon_{t,t+\tau}$, with increases in the holding period, τ.

3.3 Results

We estimated the parameters of equation (5) separately for each of the seven selected MSAs. We present the estimated indices in graphical form where all indices are normalized to a value of 100 in the first quarter of the plot. Figure 2 shows the indices for each MSA, estimated with and without controlling for nearby foreclosures (i.e., dropping the terms in the double summation of equation (5)). Because the scale would be significantly compressed by plotting the index from 1990 through 2007 (typical prices have more than doubled), the plots in Figure 2 show the indices for the last five years (2003 through 2007). The figures confirm the previously discussed variation in housing market conditions represented by the seven MSAs. Four markets show steady but moderate increases in prices through mid-2007 (Atlanta, Charlotte, Memphis and St. Louis). Las Vegas and Los Angeles exhibit much steeper rates of price appreciation prior to 2006 followed by steep declines.

In six of the seven MSAs (the exception being Los Angeles), the index estimated without controlling for the nearby foreclosures shows less market-based appreciation than the indices estimated with such controls. This is consistent with the recent sharp increase in foreclosures in those MSAs because when nearby foreclosures are more common for the second sale of the repeat sale pair, the observed holding period returns are lower and the unadjusted index underestimates the true appreciation rate because it is "pulled down" by the large number of foreclosure discounts at the second sale. The model with controls for nearby foreclosures correctly reflects a higher market appreciation rate and an offsetting negative contagion effect. When the estimation does not control for contagion effects, the negative contagion effect is erroneously attributed to the general price level, γ_t. When foreclosures affect the first and second sales with roughly the same frequency (as in Los Angeles), the contagion effect on the price index is minimal.

We turn next to the estimates of the contagion effects in the seven models. Because we have seven MSAs, four different rings and thirteen time windows, we need to review the results by selected categories. We focus first on the estimated contagion effect for the nearest foreclosed properties – those in Ring 1. Table 4 presents the estimated Ring 1 contagion effects for all thirteen different phases of foreclosure for each MSA. At the bottom of the Table, we aggregate the thirteen different windows of the foreclosure process into three categories: 1) the twelve months before foreclosure, 2) the time between foreclosure sale and REO sale and 3) the twelve months following the REO sale. Nineteen of the twenty-one aggregated contagion parameters are negative and, of those, sixteen are significantly different from zero at the five percent significance level or better. The estimated effects for properties in the post-foreclosure, pre-REO sale phase are uniformly negative (and significant) and generally larger in absolute value than the estimated pre-foreclosure effects. The average of the seven different MSA estimates is reported in the far right column and shows that the average effect for the post-foreclosure phase is approximately twice as large as the pre-foreclosure effect. The average effects for properties in the post REO sale period are generally negative and somewhat larger in magnitude than the estimated effects for the post foreclosure phase. This shows that the stigma effect is persistent and not easily reversed, even after the property is sold by the lender.

The individual contagion effects reported in the top portion of the table show significant variation across MSAs and across the different phase windows within an MSA. Furthermore, given the limited sample sizes for certain MSA models, the individual parameters are not always estimated precisely enough to achieve statistical significance. The sample size issue is especially significant for the properties that have spent more than twelve months without a lender sale and we place little confidence in those estimates – only one of which is statistically significant at normal confidence levels.

Turning to the average over the seven MSAs for each phase window, all but one of the estimated phase effects are negative. Figure 3 plots this average effect and shows that the estimated contagion effect is negligible a year before the foreclosure sale.[17] At that time, most properties are occupied by their owners and most owners still have an expectation of retaining ownership, even if they are facing financial challenges. Hence, they still have an incentive to maintain the property, albeit at a somewhat lower intensity than owners with more certainty of long-term ownership.[18] The negative effect grows in magnitude as the delinquency extends and foreclosure nears. During this period, the hope of retaining ownership diminishes as the borrower falls further behind in payments and the foreclosure process begins (on average around {F-9}). A homeowner with little prospect of curing the default has no incentive to maintain the property and the lender does not yet have control of the property or the authority to undertake needed repairs. The sharp increase in discount related to foreclosures less than three hundred feet away over the period just preceding the foreclosure sale is consistent with a growing negative externality arising from serious neglect and an increased frequency of abandonment and/or vandalism. By the time of the foreclosure sale, the estimated contagion effects for the two windows that bracket F suggest an approximate one percent discount in price for nearby non-distressed sales. After the foreclosure sale, the property is under the control of the lender who has an incentive to maximize the net proceeds from the disposition of the property. In some cases, the best strategy for the lender is to repair the property, but in other situations the best strategy is to sell the property in "As Is" condition. Generally speaking, the property is left vacant during the marketing period and may be at risk of vandalism. Figure 3 shows that the average contagion effect generally stabilizes while the property is under the control of the lender. Even when the property is resold to a new owner there is little immediate improvement. This could be because it takes the new owner some time to repair the home

[17] In Figure 3 we exclude the window for properties that have been in the post-foreclosure phase for more than a year. The average for that window is dominated by the large positive but insignificant estimate for Los Angeles.
[18] See Harding et al. (2000) for a discussion of how current LTV influences an owner's maintenance decisions.

and offset the previous owner's neglect, but it could also reflect a tendency for purchasers of foreclosed properties to either rent the home or invest less in maintenance and renovation than would the purchaser of a non-distressed property.

Table 5 presents similar results for Ring 2 where the foreclosed property is between three hundred and five hundred feet of the non-distressed sale. Looking at the results for the aggregated windows (near the bottom of the table), we again see a preponderance of negative contagion effects. However, now five of the twenty-one estimates are positive. More than half of the significant negative effects are from the Charlotte and Columbus MSAs. In the other MSA models, most of the significant estimated coefficients are negative, but many of the estimated effects are quite small. The average effect for all seven MSAs is negative for all three aggregate windows, but less than half the magnitude of the average effects in Ring 1. Shifting attention to the top portion of the table, only Charlotte and Columbus consistently show statistically significant negative contagion effects. Taken as a whole, the estimated coefficients reported in top portion of Table 5 suggest a weak negative effect from these more distant foreclosures.

Figure 4 provides a graph comparing the average phase effects from the seven MSAs (shown in the far right column) for all four Rings. The figure shows that the largest negative effect in Ring 2 occurs around the time of the REO sale by the lender, not the foreclosure sale date as was the case for Ring 1. The figure also shows that the estimated contagion coefficients for Rings 3 & 4 are much smaller in magnitude than those for the inner rings.

In summary, our results show that foreclosed properties within 300 feet of the subject property create a significant negative externality effect which is approximately one percent per distressed property at its peak. This contagion discount diminishes rapidly with distance and falls to approximately .5% for properties that are between 300 feet and 500 feet from the non-distressed sale. Beyond five hundred feet (.1 mile), we find very small negative effects. Our results with respect to the phase of foreclosure show that the contagion discount is negligible a year before the foreclosure sale but increases sharply and peaks in Ring 1 around the time of the foreclosure sale. In Ring 2, the effect is small until close to the lender's REO sale date.

The different time patterns for Ring 1 and Ring 2 suggest different transmission mechanisms for the negative externality. It is reasonable to assume that owner neglect of a property undergoing foreclosure peaks just before the foreclosure sale and eviction. At about the same time, many foreclosed properties become vacant and subject to vandalism. As a result, the visual negative externality of having a neighboring property in foreclosure peaks at about the same time. Because of the close proximity, potential buyers of the non-distressed property will observe the effects of neglect and also face the uncertainty about the future owner and whether the property will be repaired and reasonably maintained in the future. This suggests that the transmission mechanism for how a foreclosed property influences the value of its immediate neighbors is largely visual – visitors to the non-distressed property are confronted with the problem each time they visit the non-distressed property. More distant properties, even those on the same block, have a less direct visual impact on potential buyers. However, such properties can still affect the value of non-distressed properties that are being sold through increased competition for buyers as highly motivated sellers try to sell REO as quickly as possible—often with ready financing. Thus, our finding that the effect of these more distant properties peaks during the lender's REO marketing time is consistent with the hypothesis that the primary transmission mechanism for these more distant properties is through increased competition for a limited number of buyers.[19]

Our results confirm that the presence of a nearby distressed property has a significant, negative effect on the prices of nearby homes over and above the overall trend in market prices. The finding that

[19] Turnbull and Dombrow (2006) argue that greater concentrations of sellers in a local market have two potentially offsetting effects on a given seller: a negative competition effect and a positive shopping externality. They find empirical support for both effects and also find that a higher concentration of vacant houses has a consistent negative effect on the prices (and a generally positive effect on marketing time) of nearby homes. This suggests that the competition effect from nearby vacant homes generally dominates the shopping externality. This is consistent with the results reported here.

the contagion effect diminishes rapidly with distance is intuitive and consistent with the estimated effect being truly a contagion effect. An effect that exhibited persistence with distance would be more consistent with an unobserved trend in local house prices.

4. **Multiple Foreclosures**

The model specification of equations 2-5 and the results discussed in the preceding section assume that the contagion discount increases linearly with the number of nearby foreclosures. In this section, we relax that restriction in two different ways. First, we estimate a model with indicators for exactly one, exactly two and three or more nearby foreclosures in each concentric ring around the subject property. Second, we use an alternative specification that allows for a quadratic effect in the number of nearby foreclosures. To better focus the results and discussion, we reduce the number of contagion parameters estimated in each model by combining the thirteen phase buckets used earlier into a single bucket spanning the period from twelve months before the foreclosure to twelve months after the REO sale.

Indicator Models

We first establish a baseline by estimating a model with a single indicator for each concentric ring; the indicator flags the presence of one or more distressed properties in the ring. The initial specification of the underlying price hedonic (the equivalent of equation (2) without the disturbance term) using these four indicators is given below:

$$P_t = e^{\gamma_t} e^{\left[s_1 C_1 + s_2 C_2\right]} e^{\left[\sum_{d=1}^{4} a_d I_t^d\right]} \tag{6}$$

In equation (6), d indicates the specific ring and I^d_t takes on a value of one if there is one or more distressed properties at time t in Ring d. A distressed property is defined as any property that falls in the range from {F-12} to {S+12} on the sale date, t. In this specification, a_d represents the cumulative effect of nearby foreclosures and is influenced by the distribution of the observed number of nearby foreclosures. For example, if there are typically two nearby foreclosures when I^d_t equals one, then a_d will reflect the average effect corresponding to two nearby foreclosures. In this specification, the effect of nearby foreclosures is summarized by four coefficients not fifty-two. As in the previous section, to derive the final OLS model to be estimated, we write an equation similar to equation (6) for time t + τ, take logs of both equations and difference the two. The resulting OLS specification is:

$$\ln\left(P^i{}_{t+\tau} / P^i{}_t\right) = \sum_{j=1}^{72} \gamma_j D_{ij} + \sum_{d=1}^{4} a_d (I^{id}_{t+\tau} - I^{id}_t) + \varepsilon^i{}_{t,t+\tau} \tag{7}$$

where $I^{id}_t = 1$ if there are one or more distressed properties in Ring d around property i at t.

The results of estimating a_d using equation (7) are presented in the top panel of Table 6. We focus our discussion on the rightmost column that reports the average of the estimates for the seven MSAs. As expected, the contagion coefficients in Table 6 are larger in absolute value than those in Tables 4 and 5. The effect of one or more nearby foreclosures in Ring 1 is -1.5%. The estimated contagion effect in this specification is larger than the earlier estimated coefficients because it represents the cumulative effect summed over all thirteen phases as well as the average number of foreclosures in each phase bucket. Second, the reported contagion effect appears to be more persistent as the distance from the subject property increases. This is because each of the outer rings contains a larger total area and consequently the frequency of foreclosures in each ring increases. The coefficients estimated with equation (7) reflect the combination of a declining marginal effect per foreclosure that is partially offset by an increase in the average number of foreclosures in the bucket.

To further study the effect of multiple foreclosures, we extend equations (6) and (7) to include three different indicators for each ring: one that indicates the presence of exactly one foreclosure in the specified ring, a second to identify cases where there are exactly two foreclosures in the indicated ring and a third to flag the cases with three or more foreclosures. The estimates of these twelve discount effects are reported in Panel B of Table 6. The results are generally consistent with a linear effect over the range from zero to two. For example, the effect of exactly one nearby foreclosure in Charlotte is -1.64% while the effect of having exactly two is -3.6%. The table also shows that several of the individual MSA effects are not significantly different from zero. This is likely the result of a limited number of occurrences in an MSA where only one sale in the repeat sales pair had exactly two nearby foreclosures.[20] As conventional wisdom predicts, the effect of having three or more nearby foreclosures in Ring 1 is quite severe -- roughly three percent based on the average of all MSAs and as high as six percent in Charlotte. Significantly, however, even this effect diminishes quickly with distance. For example, looking at the average of the seven MSAs, the effect declines to -1.3% in Ring 2, despite the fact that the area of Ring 2 is much larger than that of Ring 1 and the average number of foreclosures increases with the area of the ring.

Quadratic Specification.

The indicator specifications discussed above suggest that a linear model for the effect of nearby foreclosures may be appropriate, but does not shed much information on the "tail" of the distribution or the effect of a large number of nearby foreclosures. Therefore, to further explore this issue, we modified equation (2) to include a quadratic effect (while still using a single phase bucket).

$$P_t = e^{\gamma_t} e^{\left[s_1 C_1 + s_2 C_2\right]} e^{\sum_{d=1}^{4} a_d N_t^d + b_d (N_t^d)^2} \tag{8}$$

Following the same procedure of taking logs and differencing, generates equation (9):

$$\ln\left(P^i{}_{t+\tau} \big/ P^i{}_t \right) = \sum_{j=1}^{72} \gamma_j D_{ij} + \sum_{d=1}^{4} a_d (N_{t+\tau}^{id} - N_t^{id}) + b_d ((N_{t+\tau}^{id})^2 - (N_t^{id})^2) + \varepsilon_{t,t+\tau} \tag{9}$$

where N_t^{id} = the number of distressed properties in Ring d around property i at t.

[20] The parameter a_d is identified by the cases where only one pair in the sale has an indicator equal to one because if both indicators are one or both are zero, the term differences to zero.

The estimates of a_d and b_d for each of the four rings are presented in Panel C of Table 6. Figure 5 plots the resulting average effect as a function of N_t^{id}. To avoid extrapolating a quadratic function beyond the range of the estimating data, the lines for the various rings are truncated at four, five, seven and ten foreclosures, respectively. The table and figure provide further confirmation of the fact that foreclosures in the innermost ring have a very significant negative externality and that the effect is roughly linear over the observed range of the number of nearby foreclosures. The quadratic term has a significant positive coefficient in all four rings, suggesting that the marginal effect of each new foreclosure is somewhat smaller, but the offsetting quadratic effect is small relative to the linear effect for all rings.

In summary, these alternative specifications suggest that the linear assumption for the contagion effect used in equations (2) to (5) is reasonable. The finding of a rapid decline in the contagion effect with distance is also robust to these different model specifications.

5. Robustness

Although the repeat sales methodology for estimating the contagion effect is preferable to estimating the effect in a hedonic model without controls for local price trends, there is a price associated with that advantage -- the assumption that there is no change in property and neighborhood characteristics between the paired sales. One way to test for robustness to this assumption is to apply different filters to the repeat sales pairs used to estimate the model. The results discussed in the previous section were based on a sample that excluded observations with unusually high holding period returns. This filter serves to screen out property "flips" where an owner/investor buys a property, renovates it and resells the renovated property. To test the sensitivity of our findings to the specific terms of this screen, we re-estimated the models for all seven MSAs using alternative screens. Table 7 describes the various filters used in the estimation. The Base filter is the one described earlier in the paper and is used for all of the previously discussed results. The rows of the table describe the alternative filters applied to the data and the resulting additional exclusions relative to the base sample. While most of the alternative screens reduce the sample size by two to three percent, the most stringent (Screen 6) reduces the total sample by approximately eight percent.

The results of estimating the seven MSA models are robust to varying the filter applied to the repeat sales pairs. Figure 6 shows the estimated Ring 1 phase coefficients (averaged over all seven MSAs) for the various screens. The figure clearly shows that the coefficients are not sensitive to the various screens. We also reviewed the individual MSA estimated coefficients and found them to be similarly robust. The largest single change in a Ring 1 coefficient (from -.0108% to -.0068%) was observed in Los Angeles.

Another possible concern is that the change in the number of foreclosures between the two sales dates of a repeat sales pair is correlated with changes in other variables not explicitly included in the model specification. If this were the case, then the change in the number of foreclosures in equation (5) would be correlated with the error term.[21] As a robustness check for this possible endogeneity, we used the instrumental variables (IV) technique. IV estimation can address both the possibility of endogeneity and the problem of measurement error.[22] To implement this approach we need to predict the number of nearby foreclosures using variables that are uncorrelated with the disturbance term. Previous mortgage research and traditional underwriting rules have shown that borrower credit histories, original loan-to-value ratios and income are important predictors of foreclosure. Consequently, we developed estimates of the distribution of FICO scores, loan-to-value ratios and homeowner incomes in the four rings surrounding each property at each transaction date. We created these estimates using data from loan originations in each ring for each date. We selected the Los Angeles MSA because it had the smallest

[21] Using the change in foreclosures for fifty-two different phase/distance buckets already provides some degree of control for this problem because changes in other house or location characteristics are less likely to be correlated with these very fine measures of foreclosure change.

[22] Although our inventory of nearby foreclosures is quite good, it is nevertheless not perfect. To the extent that our inventory of nearby foreclosed properties is incomplete, our measure of the change in the number of nearby foreclosures in each phase bucket will have measurement error. Another source of measurement error arises because of the boundaries of the zip codes used as the basic geographic unit. This latter problem is discussed below.

number of repeat sales transactions which kept the effort required to create the instruments manageable. We used the 90th percentile of FICO score and loan-to-value ratio, the median income level along with the housing stock in each ring and the subject property size as instruments in a first stage regression to predict the change in the number of foreclosures in each ring for each repeat sales pair (equation 5). We are able to use the property size as an instrument because the property characteristics difference out in the derivation of the repeat sales OLS specification. The property size proxies for average neighborhood house characteristics.

The first stage regression uses the change in the number of foreclosures as the dependent variable and the set of all exogenous variables, including the five instruments described above as the independent variables. In the second stage regression, we used the predicted change in the number of foreclosures in place of the actual number of foreclosures to estimate equation (5). The estimated contagion parameters and index values were qualitatively similar to the original estimates. We tested for the presence of significant endogeneity using the Hausmann endogeneity test.[23] The test rejects the presence of significant endogeneity in the Los Angeles model and provides support for using OLS to estimate the models. We did not repeat this robustness check for the other MSAs because gathering the necessary data on census tract credit scores and loan-to-value ratios is difficult and the results from Los Angeles suggest that endogeneity is not a significant problem.

Another potential problem with our data is that we measure the number of nearby foreclosures looking at only those foreclosures that fall in the same zip code as the non-distressed sale. Thus for a sale that is near the border of its zip code, the concentric circles around the property will cross the zip code boundary leading to an underestimate of the number of nearby foreclosures and therefore introduce measurement error in the change in nearby foreclosures variable. To test whether this influences our

[23]The Hausman endogeneity test provides a test of whether a variable in an OLS specification is endogenous. The test entails projecting the variable to be tested (in our case the change in the number of foreclosures) on the set of all exogenous variables, including the instruments. The residuals from this first stage regression are then added to the original OLS specification and if the estimated coefficient is significant, the test rejects the null hypothesis of no endogeneity.

results,[24] we identified a cluster of zip codes in the Atlanta MSA where there was a central core zip code (30044) that had passed our screens and was fully surrounded by zip codes that had also passed our criteria for having full foreclosure inventories. We re-estimated the model for the central zip code only, treating it as a stand-alone zip code and therefore excluding all foreclosures outside the zip code in the count of nearby foreclosures. We then re-estimated the model for the same zip code (30044) including foreclosures from the surrounding zip codes in the count of nearby foreclosures. We found that the measurement error in the count of nearby foreclosures was quite small for this zip code and that the resulting contagion effect estimates were essentially the same as those estimated using the zip code as a stand-alone location.

6. **Conclusion**

We use the repeat sales methodology to provide joint estimates of the local trend in house prices and the contagion discount associated with selling a home with a foreclosed property nearby. Using the repeat sales approach reduces the omitted variable bias that is likely a problem when using hedonic models for this purpose. We find that having a neighboring property in the process of foreclosure can result in a discount to market value of up to one percent per nearby distressed property. We find that for small numbers of foreclosures the discount is approximately linear in the number of nearby distressed properties.

The estimated contagion effect declines rapidly with distance between the foreclosed property and the non-distressed sale. The effect of a foreclosed property in a ring 300 to 500 feet from the subject property is roughly half that of an immediate neighbor. The size of the discount continues to fall as the distance increases; beyond five hundred feet we find almost no statistically significant contagion

[24] Note that instrumental variables estimation can also be used to address measurement error and as discussed previously we do not find significant differences in Los Angeles using instrumental variables.

effect. These results showing a rapid decline of the contagion effect with distance are different than those reported by previous researchers.

Because of the size of our database, we are also able to study how the contagion effect varies with the phase of the foreclosure process which, in turn, sheds light on the transmission mechanism. For properties within three hundred feet of a foreclosed property, we find that the contagion discount is negligible a year before the foreclosure sale but increases rapidly as the delinquency becomes more serious. The peak negative externality occurs near the time of the foreclosure sale. Between the foreclosure sale and the REO sale, the discount stabilizes as the lender resumes maintenance of the property and markets the property. Despite some improvement after the REO sale, the contagion discount resulting from immediate neighbors lingers for at least a year after the REO sale. The pattern of effects from foreclosed properties between three hundred and five hundred feet (Ring 2) is different in several respects. First, it does not increase as rapidly during the delinquency phase and second it peaks near the time of the REO sale by the lender. We interpret these different patterns as suggesting that the negative externality from immediate neighbors is attributable to property neglect and uncertainty about the future owner. Properties located further away affect the sale prices of non-distressed properties largely through a competition effect.

From a policy perspective, our results confirm the existence of a significant negative externality associated with foreclosed properties and support publicly funded efforts to reduce the problem. However, our estimates of that externality controlling for the local trend in house prices are generally smaller than previous estimates in the literature and we provide evidence that the most significant externalities are attributable to immediate neighbors. As a result, a million additional foreclosures would significantly affect three to five million homes not the forty million that has been estimated using earlier estimates of contagion effects. Finally, our analysis of the impact by phase shows that the externality grows rapidly during the period required for the lender to take control of the property. This suggests that when foreclosure is inevitable, efforts to speed the foreclosure process would be effective at reducing the costs associated with the contagion effect.

References

Avery, R., Bostic, R., Calem, P., Canner, G., 1996. Credit Risk, Credit Scoring, and the Performance of Home Mortgages, *Federal Reserve Bulletin* 82: 621-648.

Bailey, M. J.; Muth, R. F., Nourse, H. O., 1963. A Regression Model for Real Estate Price Index Construction, *Journal of the American Statistical Association* 58(304): 933-942.

Case, K., Shiller, R., 1989. The Efficiency of the Market for Single-Family Homes, *The American Economic Review* 79(1): 125-137.

Center for Responsible Lending, 2008. Subprime Spillover: Foreclosures Cost Neighborhoods $202 Billion; 40.6 Million Homes Lose $5,000 on Average, CRL Issue Paper.

Center for Responsible Lending, 2009. Soaring Spillover: Accelerating Foreclosures to Cost Neighbors $502 billion in 2009 alone; 69.5 million homes lose $7,200 on average, CRL Issue Paper.

Epple, D., 1987. Hedonic Prices and Implicit Markets: Estimating Demand and Supply Functions for Differentiated Products, *Journal of Political Economy,* 95(1):59-80.

Griliches, Z., ed., 1971. *Price Indexes and Quality Change,: Studies in New Methods of Measurement.* Cambridge, MA: Harvard University Press.

Harding, J., Knight, J., Sirmans, C. F., 2003. Estimating Bargaining Effects in Hedonic Models: Evidence from the Housing Market, *Real Estate Economics* 31(4): 601-622.

Harding, J., Rosenthal, S., Sirmans, C. F., 2007. Depreciation of Housing Capital, Maintenance and House Price Inflation: Evidence from a Repeat Sales Model, *Journal of Urban Economics* 61(2): 193-217.

Harding, J., Miceli, T., Sirmans, C.F., 2000. Do Owners Take Better Care of Their Housing Than Renters? *Real Estate Economics* 28(4): 663-681.

Immergluck, D., Smith, G., 2006. The External Costs of Foreclosure: The Impact of Single-Family Mortgage Foreclosures on Property Values, *Housing Policy Debate* 17(1):57-79.

Ioannides, Y., 2002. Interactive Property Valuations. *Journal of Urban Economics.* 53: 145-170.

Leventis, A., 2009. The Impact of Distressed Sales on Repeat-Transaction House Price Indexes, Federal Housing Finance Agency, Research Paper**.**

Lin, Z., Rosenblatt, E., Yao V. W., 2009. Spillover Effects of Foreclosure on Neighborhood Property Values, *Journal of Real Estate Finance and Economics* 38(4), *f*orthcoming.

Pope, J. C., 2008. Fear of Crime and Housing Prices: Household Reactions to Sex Offender Registries, *Journal of Urban Economics* 64:601-614.

Rosen, S., 1974. Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition, *Journal of Political Economy* 82: 34-55.

Schwartz, A., Ellen, I., Voicu, I., Schill, M.,. 2003. Estimated External Effects of Subsidized Housing Investment on Property Values. Lincoln Institute of Land Policy. Working Paper WP03AS1.

Shlay, A., Whitman, G., 2004. Research for Democracy: Linking Community Organizing and Research to Leverage Blight Policy, *City and Community* 5(2): entire monograph.

Turnbull, G., Dombrow, J., 2006. Spatial Competition and Shopping Externalities: Evidence from the Housing Market, *Journal of Real Estate Finance and Economics* 32:391-408.

Wheelock, D., 2008. The Federal Response to Home Mortgage Distress: Lessons from the Great Depression, *Federal Reserve Bank of St. Louis Review.* 90(3) 133-148.

Table 1
Repeat Sales Observations: Distribution by State and MSA

	State Totals				7-MSA Samples		
State	**Number of Repeat Sale**	**Percent of Total Sampl**	**Number of Zip Codes**	**# of MSAs**	**MSA**	**Number of Repeat Sale**	**Number of Zip Codes**
CA	42,521	6.80%	5	5	Los Angeles	7,767	1
GA	190,514	30.30%	30	2	Atlanta	186,655	29
IN	25,111	4.00%	14	7	--	--	--
MO	25,065	4.00%	7	2	--	--	--
NC	64,022	10.20%	15	6	St. Louis	19,594	4
NV	27,895	4.40%	4	1	Charlotte	46,219	10
OH	192,852	30.70%	51	8	Las Vegas	27,895	4
SC	5,194	0.80%	3	3	Columbus	84,751	18
TN	49,331	7.90%	8	3	--	--	--
UT	2,023	0.30%	1	1	Memphis	32,802	6
WA	2,237	0.40%	1	1	--	--	--
WY	1,766	0.30%	1	1	--	--	--
Total	628,531	100%	140	40		405,683	72

The repeat sales sample was drawn from the FHFA (previously OFHEO) joint GSE mortgage loan files, augmented with transaction data purchased from private vendors describing transactions not financed by a GSE.
The repeat sales were drawn from 140 zip codes for which we have a nearly complete record of foreclosures and a sufficent volume of regular sales transactions large enough to estimate an accurate repeat sales index.

Table 2
Summary Statistics for Foreclousre Sample
(Foreclosures from 1989 through 2007)

	Total Sample	Atlanta	Charlotte	Columbus	Las Veg.	LA	Memphis	St. Louis
Number (total)	60,708	24,334	8,711	11,858	3,303	2,887	6,087	3,528
% (of all foreclsures)	100.0%	40.1%	14.3%	19.5%	5.4%	4.8%	10.0%	5.8%
Average Time from F to S (months)	9.89	9.86	9.89	10.01	9.43	9.46	10.15	9.97
Foreclosures by Year								
1989	79	61	1	5	-	1	9	2
1990	127	87	1	4	-	22	9	4
1991	252	119	-	10	-	106	10	7
1992	287	83	4	6	1	173	14	6
1993	355	58	4	12	11	248	20	2
1994	369	50	8	13	7	251	29	11
1995	325	45	8	10	19	221	16	6
1996	813	189	7	28	31	352	136	70
1997	2,126	987	133	217	56	385	178	170
1998	1,961	645	136	297	157	275	221	230
1999	1,865	501	220	291	138	173	258	284
2000	2,172	535	312	473	159	110	287	296
2001	3,221	1,059	502	556	197	74	466	367
2002	4,700	1,688	859	937	283	59	541	333
2003	7,009	2,943	1,319	1,525	233	21	648	320
2004	7,548	3,144	1,470	1,865	33	10	695	331
2005	8,271	3,915	1,374	1,859	68	7	735	313
2006	10,881	4,927	1,423	2,438	615	93	970	415
2007	8,156	3,142	926	1,302	1,295	305	834	352

Mortgage Foreclosure Started (%): US Total and 7 States

	U.S.	GA	NC	OH	NV	CA	TN	MO
2000	0.62	1.48	1.43	2.18	2.18	1.28	1.63	1.14
2003	0.79	2.25	2.33	2.87	2.87	0.74	2.31	1.78
2005	0.72	2.27	2.02	3.36	3.36	0.59	2.20	1.84
2007	1.30	3.23	2.12	4.37	4.37	3.26	2.62	2.61

Notes:
1. The foreclosures described in Table 2 represent the combination of foreclosures identified in the GSE mortgage database and foreclosures identified from purchased transaction data
2. All foreclosures reported in Table 2 are drawn from the 72 zip codes identified in table 1.
Each selected zip code met the criteria needed to assure better than eighty percent coverage of foreclosures described in the text.
3. The foreclosaure data for 2007 are incomplete because the complete information on sale date and REO date are ofetn reported with a lag and as a result were excluded from the inventory of foreclosures because of incomplete information.
3. The state foreclosure rates are from Mortgage Banker Association (MBA) and may not be indicative of the foreclosure rate in the corresponding MSA.

Table 3
Descriptive Statistics for the Repeat sales Observations
(standard deviations in parentheses)

Panel A	**All Seven MSAs**	**Atlanta**	**Charlotte**	**Col-umbus**	**Las Vegas**	**Los Angeles**	**Memphis**	**St. Louis**
Number of Repeat Sales Pairs	405,631	186,626	46,205	84,743	27,895	7,767	32,801	19,594
% of Total	100	46.01	11.39	20.89	6.88	1.91	8.09	4.83
Purchase Price ($)	140,606	142,097	142,759	130,929	180,224	194,537	135,612	93,752
	(84,120)	(83,329)	(85,673)	(75,356)	(102,691)	(113,378)	(80,062)	(39,849)
Sale Price ($)	171,383	171,829	163,185	154,460	258,605	261,675	156,733	124,225
	(102,143)	(97,375)	(104,730)	(84,970)	(139,027)	(129,458)	(91,153)	(49,735)
Holding Period (yrs)	4.64	4.59	4.78	4.90	3.40	4.39	4.70	5.39
	(3.2)	(3.3)	(3.1)	(3.2)	(2.7)	(3.8)	(3.2)	(3.6)
Holding Period Return								
Total (%)	19.8	19.7	12.9	17.7	35.1	30.7	15.2	28.5
	(19.8)	(18.6)	(14.7)	(16.4)	(27.8)	(37.7)	(15.5)	(22.5)
Per Year (%)	4.0	4.0	2.6	3.4	9.2	6.3	3.1	4.8
	(5.7)	(5.3)	(4.5)	(4.9)	(9.6)	(8.8)	(4.7)	(5.9)
House Price Appreciation (OFHEO)								
1990 -2007	4.47%	4.14%	4.06%	4.19%	5.82%	5.46%	3.29%	4.36%
Number of Distinct Properties	281,063	127,930	32,236	59,024	19,569	5,504	22,076	14,724
Percent of Properties with one or more repeat sales								
1	67.76	65.79	67.73	66.87	67.66	68.7	64.56	73.04
2	23.94	24.95	23.48	24.67	24.24	23.4	25.2	21.62
3	6.58	7.25	6.89	6.78	6.35	6.3	7.85	4.61
4	1.42	1.67	1.56	1.41	1.43	1.34	1.92	0.63
5	0.25	0.29	0.3	0.23	0.25	0.18	0.41	0.07
6	0.04	0.05	0.03	0.03	0.06	0.04	0.04	0.01
7 or more	< .01	< .01	< .01	< .01	< .01	< .01	< .01	< .01

Panel B

Timing of Repeat Sales Transactions

	Initial Purchase		**Resale**	
Year	**N**	**%**	**N**	**%**
1990	22,721	5.60	383	0.09
1991	23,845	5.88	1,419	0.35
1992	28,044	6.91	4,095	1.01
1993	29,700	7.32	6,982	1.72
1994	28,179	6.95	9,241	2.28
1995	27,151	6.69	11,539	2.84
1996	31,524	7.77	15,999	3.94
1997	29,439	7.26	18,085	4.46
1998	32,773	8.08	24,440	6.03
1999	32,342	7.97	28,657	7.06
2000	28,061	6.92	28,140	6.94
2001	27,408	6.76	33,338	8.22
2002	21,615	5.33	33,482	8.25
2003	18,499	4.56	37,214	9.17
2004	13,669	3.37	39,903	9.84
2005	7,379	1.82	43,788	10.80
2006	2,787	0.69	39,515	9.74
2007	495	0.12	29,411	7.25
Sum	405,631	100	405,631	100

Table 4
Estimated Contagion Effects --Ring 1
(coefficients above, t-statistics below in parentheses)

	Atlanta	Charlotte	Columbus	Las Vegas	Los Angeles	Memphis	St. Louis	Avg of All MSAs
Phase of Foreclosure								
{F-12 to F-9}	0.04	-0.68	-1.11	-0.23	1.68	-0.74	0.00	-0.15
	(0.32)	-(3.05)	-(5.68)	-(1.23)	(3.79)	-(2.81)	(0.01)	-(1.39)
{F-9 to F-6}	0.15	-0.86	-0.77	0.58	0.22	0.00	-0.67	-0.19
	(1.07)	-(3.86)	-(3.89)	(2.88)	(0.48)	-(0.01)	-(1.44)	-(1.78)
{F-6 to F-3}	-0.27	-0.83	-1.22	0.17	0.23	-0.59	-0.50	-0.43
	-(1.82)	-(3.62)	-(6.10)	(0.84)	(0.44)	-(2.08)	-(1.09)	-(3.55)
{F-3 to F}	-0.01	-1.06	-1.30	0.25	-3.47	-0.93	-1.02	-1.08
	-(0.09)	-(4.68)	-(6.40)	(1.22)	-(6.84)	-(3.32)	-(2.17)	-(8.99)
{F to F+3}	-0.67	-1.67	-1.12	0.33	-1.08	-0.74	-0.86	-0.83
	-(4.34)	-(7.41)	-(5.43)	(1.70)	-(1.94)	-(2.58)	-(1.76)	-(6.55)
{F+ 3 to F+6}	-0.79	-1.75	-1.08	-0.52	-1.56	-0.93	-0.12	-0.96
	-(5.23)	-(7.83)	-(5.16)	-(2.85)	-(2.99)	-(3.17)	-(0.26)	-(7.95)
{F+6 to F+9}	-0.32	-1.21	-0.81	-0.24	-0.73	-0.15	-1.40	-0.69
	-(2.05)	-(5.23)	-(3.84)	-(1.19)	-(1.37)	-(0.51)	-(2.97)	-(5.59)
{F+9 to F+12}	-0.77	-1.73	-0.96	-0.40	-1.57	-0.05	-0.21	-0.81
	-(3.43)	-(5.47)	-(3.30)	-(1.17)	-(1.77)	-(0.14)	-(0.34)	-(4.43)
{> F+12}	-0.75	-0.51	-0.33	-0.17	5.06	-0.17	2.65	0.83
	-(1.81)	-(0.75)	-(0.51)	-(0.13)	(1.52)	-(0.20)	(1.77)	(1.41)
{S to S+3}	-0.62	-2.41	-1.08	-0.24	-1.03	-0.77	-0.64	-0.97
	-(3.66)	-(10.20)	-(4.97)	-(1.08)	-(1.68)	-(2.44)	-(1.26)	-(7.10)
{S+3 to S+6}	-0.96	-1.99	-1.46	-0.23	-0.58	-1.34	-0.24	-0.97
	-(5.50)	-(8.12)	-(6.41)	-(0.97)	-(0.97)	-(4.11)	-(0.47)	-(7.08)
{S+6 to S+9}	-0.56	-1.60	-1.37	-0.30	-1.22	-1.32	0.54	-0.83
	-(3.18)	-(6.59)	-(5.83)	-(1.23)	-(2.09)	-(3.92)	(1.06)	-(6.08)
{S+9 to S+12}	-0.67	-2.29	-1.54	-1.12	-1.93	-1.02	1.23	-1.05
	-(3.56)	-(9.17)	-(6.34)	-(4.81)	-(3.51)	-(3.00)	(2.29)	-(7.66)
Average Discounts								
{F-12 through F}	-0.02	-0.86	-1.10	0.19	-0.34	-0.57	-0.55	-0.46
Joint F-Test	0.09	64.4***	131.98***	4.14**	1.96	16.79***	5.52**	
{F through F+12}	-0.64	-1.59	-0.99	-0.21	-1.24	-0.47	-0.65	-0.83
Joint F-Test	54.14***	171.21***	76.87***	3.12*	15.11***	8.79***	6.47**	
{S through S+12}	-0.70	-2.07	-1.36	-0.47	-1.19	-1.11	0.22	-0.96
Joint F-Test	64.81***	323.85***	149.04***	17.51***	18.01***	46.27***	0.78	

Notes: 1. Table 5 presents estimates of the contagion effect for a single nearby distressed property. The estimates reported here are for the effect of distressed properties less than 300 feet from the subject.
2. The rows of the table show how the estimated effect varies with the phase of foreclosure.
3. "F" denotes the foreclosure sale and "S" the REO sale by the lender.
4. As soon as a property is sold by the lender (REO sale), the phase is defined as being post REO sale. Consequently, not all properties proceed through each time bucket following F and the number of observations in each post F bucket declines.
5. The F-statistics reported for the Average Discounts are based test that the sum of four coefficients equals zero.
6. The t-statistics reported for the Average of All MSAs are calculated assuming each of the seven MSA estimates is independent

Table 5
Estimated Contagion Effects --Ring 2
(coefficients above, t-statistics below in parentheses)

	Atlanta	Charlotte	Columbus	Las Vegas	Los Angeles	Memphis	St. Louis	Avg of All MSAs
Phase of Foreclosure								
{F-12 to F-9}	0.62	-0.32	-0.94	0.36	0.87	-0.27	-1.30	-0.14
	(4.76)	-(1.59)	-(5.63)	(2.33)	(2.71)	-(1.22)	-(3.22)	-(1.49)
{F-9 to F-6}	0.40	-0.46	-0.52	0.31	-0.93	0.34	-0.50	-0.19
	(2.96)	-(2.31)	-(3.07)	(1.94)	-(2.42)	(1.48)	-(1.27)	-(1.99)
{F-6 to F-3}	0.33	-0.44	-0.40	0.16	0.19	-0.27	-0.71	-0.16
	(2.46)	-(2.20)	-(2.35)	(0.95)	(0.47)	-(1.11)	-(1.81)	-(1.60)
{F-3 to F}	0.14	-0.69	-0.23	0.11	0.16	-0.22	-0.48	-0.17
	(0.99)	-(3.41)	-(1.34)	(0.61)	(0.38)	-(1.00)	-(1.19)	-(1.69)
{F to F+3}	-0.08	-0.90	-0.37	0.40	0.71	-0.51	-0.26	-0.15
	-(0.60)	-(4.38)	-(2.10)	(2.36)	(1.74)	-(2.11)	-(0.62)	-(1.42)
{F+ 3 to F+6}	-0.04	-0.78	-0.98	0.23	0.63	-0.45	0.21	-0.17
	-(0.29)	-(3.92)	-(5.52)	(1.33)	(1.48)	-(1.78)	(0.53)	-(1.63)
{F+6 to F+9}	0.06	-0.77	-0.89	0.06	-0.96	-0.70	-0.46	-0.52
	(0.38)	-(3.71)	-(5.07)	(0.35)	-(2.08)	-(2.73)	-(1.12)	-(4.83)
{F+9 to F+12}	-0.25	-0.28	-0.46	-0.43	-0.59	-0.57	0.42	-0.31
	-(1.16)	-(1.00)	-(1.90)	-(1.67)	-(0.80)	-(1.75)	(0.76)	-(1.97)
{> F+12}	-0.71	0.77	-0.12	-1.26	-10.05	-1.75	1.14	-1.71
	-(1.85)	(1.17)	-(0.22)	-(1.10)	-(3.18)	-(2.65)	(0.95)	-(3.19)
{S to S+3}	-0.18	-1.40	-0.66	-0.46	-1.80	-0.68	-0.51	-0.81
	-(1.15)	-(6.55)	-(3.55)	-(2.52)	-(3.67)	-(2.58)	-(1.19)	-(7.15)
{S+3 to S+6}	-0.22	-1.23	-0.88	-0.44	-0.69	-0.23	0.13	-0.51
	-(1.41)	-(5.54)	-(4.56)	-(2.24)	-(1.51)	-(0.82)	(0.29)	-(4.51)
{S+6 to S+9}	-0.05	-1.32	-0.68	-0.31	-0.64	-0.17	-0.17	-0.48
	-(0.29)	-(5.97)	-(3.41)	-(1.61)	-(1.26)	-(0.61)	-(0.39)	-(4.04)
{S+9 to S+12}	-0.12	-1.12	-0.88	-0.29	-0.45	0.32	1.53	-0.14
	-(0.71)	-(4.92)	-(4.27)	-(1.45)	-(0.94)	(1.21)	(3.34)	-(1.22)
Average Discounts								
{F-12 through F}	0.37	-0.48	-0.52	0.23	0.08	-0.10	-0.75	-0.17
Joint F-Test	31.19***	25.27***	40.96***	8.53***	0.16	0.83	14.16***	
{F through F+12}	-0.08	-0.69	-0.67	0.07	-0.05	-0.56	-0.02	-0.29
Joint F-Test	0.95	40.16***	50.85***	0.49	0.04	16.96***	0.01	
{S through S+12}	-0.14	-1.27	-0.78	-0.38	-0.90	-0.19	0.24	-0.49
Joint F-Test	3.2*	147.61***	68.14***	15.47***	14.24***	2.00	1.22	

Notes: 1. Table 5 presents estimates of the contagion effect for a single nearby distressed property. The estimates reported here are for the effect of distressed properties less than 300 feet from the subject.
2. The rows of the table show how the estimated effect varies with the phase of foreclosure.
3. "F" denotes the foreclosure sale and "S" the REO sale by the lender.
4. As soon as a property is sold by the lender (REO sale), the phase is defined as being post REO sale. Consequently, not all properties proceed through each time bucket following F and the number of observations in each post F bucket declines.
5. The F-statistics reported for the Average Discounts are based test that the sum of four coefficients equals zero.
6. The t-statistics reported for the Average of All MSAs are calculated assuming each of the seven MSA estimates is independent

Table 6

The Effect of Multiple Foreclosures

Estimated Coefficients (with T-statistics in parentheses below)

Panel A One Indicator per Ring -- Identifies the Presence of One or More Foreclosures in the Ring

	Indicator	Atlanta	Charlotte	Columbus	Las Vegas	Los Angeles	Memphis	St. Louis	Average
Ring 1	1 or more	-0.77	-3.04	-2.38	-0.48	-1.01	-2.21	-0.40	-1.47
		-(10.08)	-(24.74)	-(21.96)	-(3.85)	-(3.26)	-(14.48)	-(1.63)	-(21.59)
Ring 2	1 or more	-0.24	-2.57	-1.79	-0.23	0.04	-1.65	-0.23	-0.95
		-(3.23)	-(21.19)	-(17.47)	-(1.89)	(0.12)	-(11.54)	-(0.98)	-(14.50)
Ring 3	1 or more	-0.69	-1.98	-1.32	0.25	-1.71	-2.15	-0.90	-1.21
		-(11.07)	-(17.80)	-(14.30)	(1.81)	-(5.50)	-(16.06)	-(4.13)	-(18.76)
Ring 4	1 or more	-1.04	-0.63	-1.23	2.30	-0.13	-1.42	-2.03	-0.60
		-(16.59)	-(5.44)	-(13.12)	(13.08)	-(0.35)	-(9.47)	-(7.93)	-(7.84)

Panel B Indicators for Exactly One, Exactly Two and Three or More Foreclosures in the Ring

		Atlanta	Charlotte	Columbus	Las Vegas	Los Angeles	Memphis	St. Louis	Average
	Exactly 1	-0.77	-1.64	-1.49	-0.41	-0.34	-1.43	-0.49	-0.94
		(5.99)	-(2.17)	-(0.52)	(0.72)	(0.99)	(1.50)	(4.40)	-(12.65)
Ring 1	Exactly 2	-0.71	-3.56	-2.58	-0.70	-1.12	-1.79	0.23	-1.46
		-(4.64)	-(15.97)	-(12.20)	-(3.32)	-(2.09)	-(6.44)	(0.49)	-(11.78)
	3 or more	-0.89	-5.99	-4.84	-0.45	-4.97	-2.96	-0.53	-2.95
		-(4.14)	-(22.22)	-(18.52)	-(1.77)	-(7.65)	-(7.96)	-(0.79)	-(18.41)
	Exactly 1	-0.29	-1.40	-0.95	-0.30	0.24	-0.72	-0.51	-0.56
		-(3.59)	-(10.47)	-(8.26)	-(2.23)	(0.72)	-(4.64)	-(1.97)	-(7.75)
Ring 2	Exactly 2	-0.05	-2.65	-2.00	0.25	0.17	-1.89	0.78	-0.77
		-(0.33)	-(12.75)	-(10.61)	(1.29)	(0.33)	-(7.58)	(1.86)	-(6.21)
	3 or more	0.36	-4.12	-3.81	-0.20	-0.34	-2.33	1.02	-1.34
		(1.88)	-(17.01)	-(17.21)	-(0.94)	-(0.62)	-(7.75)	(1.91)	-(10.11)
	Exactly 1	-0.60	-0.95	-1.05	0.28	-1.30	-1.09	-0.51	-0.75
		-(8.77)	-(7.70)	-(10.10)	(1.76)	-(3.75)	-(7.45)	-(2.08)	-(10.28)
Ring 3	Exactly 2	-0.82	-1.96	-1.02	0.44	-3.23	-2.18	-1.55	-1.47
		-(8.30)	-(11.81)	-(7.14)	(2.30)	-(7.23)	-(11.26)	-(4.92)	-(15.71)
	3 or more	-0.52	-3.54	-1.99	-0.25	-1.96	-4.31	-0.05	-1.80
		-(4.92)	-(21.54)	-(13.83)	-(1.42)	-(4.63)	-(21.63)	-(0.15)	-(19.71)
	Exactly 1	-0.85	-0.17	-1.10	2.18	-0.81	-0.68	-1.51	-0.42
		-(12.31)	-(1.28)	-(10.30)	(10.48)	-(1.86)	-(4.11)	-(5.27)	-(4.79)
Ring 4	Exactly 2	-1.12	-0.55	-1.46	2.05	0.19	-1.36	-1.70	-0.56
		-(12.63)	-(3.36)	-(10.73)	(8.67)	(0.41)	-(6.82)	-(4.94)	-(5.70)
	3 or more	-1.56	-2.06	-1.50	2.62	0.95	-3.85	-3.27	-1.24
		-(17.45)	-(13.13)	-(11.77)	(12.73)	(2.04)	-(19.84)	-(10.17)	-(12.90)

Panel C Quadratic Specification for the Number of Forecl soures in the Ring

		Atlanta	Charlotte	Columbus	Las Vegas	Los Angeles	Memphis	St. Louis	Average
Ring 1	a_d	-0.71	-1.57	-1.29	-0.19	-0.89	-1.16	-0.88	-0.96
		-(10.68)	-(17.88)	-(15.36)	-(2.50)	-(5.65)	-(9.26)	-(4.47)	-(20.72)
	b_d	0.10	-0.03	-0.01	0.01	0.03	0.05	0.27	0.06
		(5.99)	-(2.17)	-(0.52)	(0.72)	(0.99)	(1.50)	(4.40)	(5.07)
Ring 2	a_d	-0.36	-1.25	-0.97	0.10	-0.25	-1.09	-0.50	-0.62
		-(6.07)	-(15.26)	-(13.60)	(1.45)	-(1.99)	-(10.25)	-(2.92)	-(15.62)
	b_d	0.11	0.02	0.03	-0.02	0.01	0.11	0.19	0.06
		(9.19)	(1.78)	(2.28)	-(1.89)	(0.30)	(5.15)	(4.22)	(7.41)
Ring 3	a_d	-0.28	-0.86	-0.67	0.10	-0.20	-1.05	-0.53	-0.50
		-(8.91)	-(17.20)	-(19.11)	(2.27)	-(3.01)	-(17.00)	-(4.85)	-(21.41)
	b_d	0.02	0.04	0.02	-0.01	0.00	0.05	0.06	0.03
		(7.15)	(8.54)	(10.68)	-(3.37)	(1.18)	(10.22)	(5.86)	(13.25)
Ring 4	a_d	-0.43	-0.25	-0.13	0.13	-0.03	-0.58	-0.63	-0.28
		-(23.40)	-(8.69)	-(7.76)	(4.66)	-(0.72)	-(16.45)	-(11.52)	-(21.38)
	b_d	0.02	0.01	0.00	0.00	0.00	0.01	0.01	0.01
		(21.23)	(4.56)	(6.86)	-(3.96)	-(0.28)	(10.34)	(8.29)	(16.02)

Table 7 Alternative Screens

	Short Term Holding Period		Long Term Holding Period		Total Sample Size	
Screen	**Years**	**Quarterly Return Threshold**	**Years**	**Quarterly Return Threshold**	**Number of Observations**	% of Base
Base	≤ 2	>10%	>2	>8%	405,631	--
Screen 2	≤ 2	>7.5%	>2	>6%	395,535	97.51%
Screen 3	≤ 1	>10%	>1	>8%	398,727	98.30%
Screen 4	≤ 1	>7.5%	>1	>6%	394,273	97.20%
Screen 5	≤ 1.5	>7.5%	>1.5	>5%	392,456	96.75%
Screen 6	<=.5	all pairs with holding period < .5 year excluded	>.5	>8%	374,558	92.34%

Notes:

1. Table 7 defines the Base Filter and Alternative Filters used to screen the data to exclude unusual observations where it is likely that the property or neighborhood characteristics have changed
2. All screens (except Screen 6) apply different limits based on the holding period: a higher threshold for shorter holding periods and a lower threshold for longer holding periods. See text discussion for the rationale.
3. The column labelled "Years" defines the holding period for each screen. For example the Base screen excludes all observations with a holding period of two years or less and a holding period return that exceeds 10% per quarter. The Base Long Term screen excludes any obseravtion with a holding period greater than two years and a quarterly holding period return in excess of 8%.
4. The alternative screens are made more restrictive (i.e., exlcude more observations) either by lowering the thresholds (e.g., Screen 2) or by loweirng the definition of the short term holding period.
5. Screen 6 excludes all repeat sales pairs where the holding period is less than .5 year and retains the same return threshold for the pairs with holding periods greater than .5 year.

Figure 1
Foreclosure Process

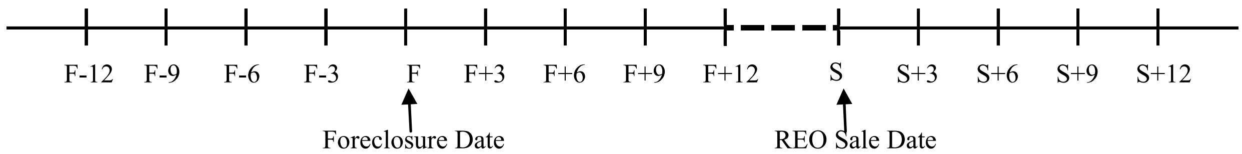
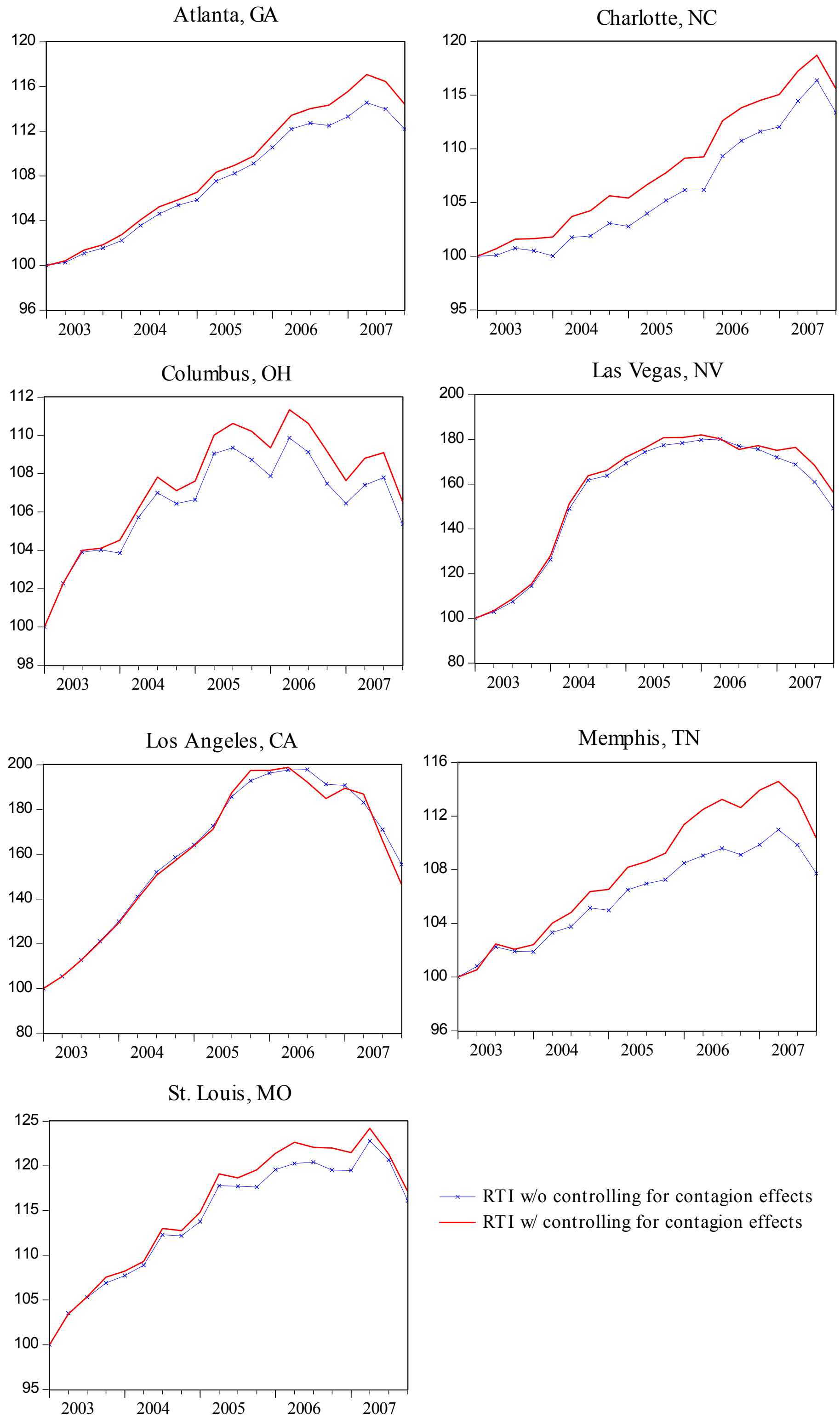


Figure 2 shows the thirteen different phases of the foreclosure process used in the paper. The two key reference dates are the foreclosure date (F) and the REO sale date (S). We classify a property as being in the foreclosure process from twelve months before the foreclosure sale date until twelve months after the REO sale date. Under our classification scheme, as soon as a property is sold by the lender, the property is classified as being post-REO sale. Therefore, not all properties pass through all five post-foreclosure sale windows and the number of observations in each post-foreclosure window declines.

Figure 2
Repeat Transaction Indices w/ and w/o Controlling Nearby Foreclosures



Note. Figure 2 compares the House Price Index estimated with and without controlling for the contagion effect The scales for each MSA vary because the house price appreciation rates vary.

Figure 3

Average Contagion Effect

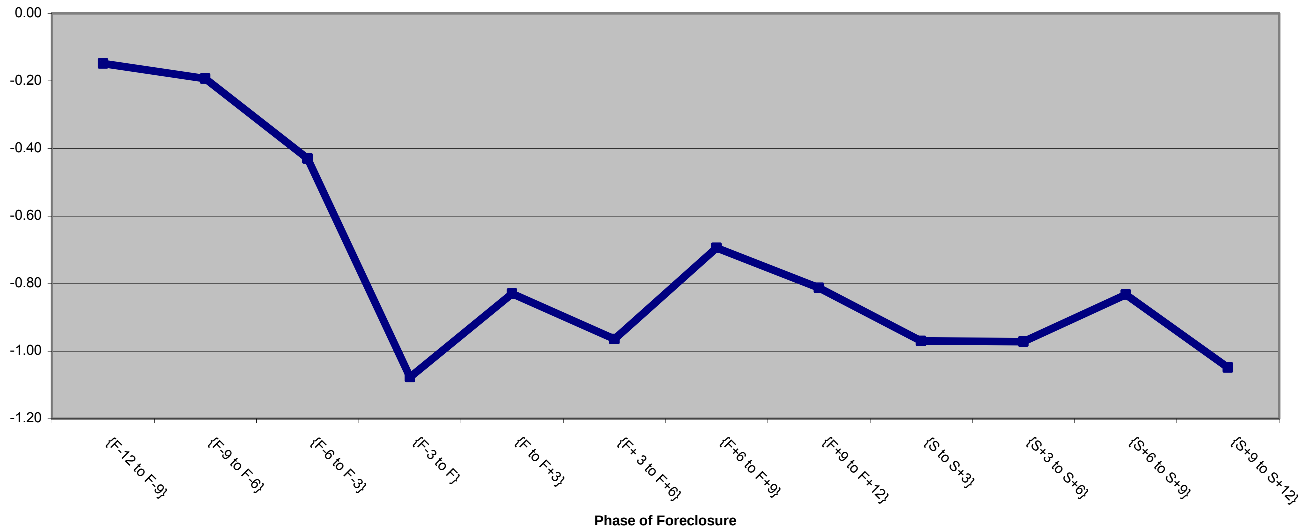


Figure 3 displays the estimated foreclosure discounts resulting from foreclosures within three hundred feet (Ring 1) of a non-distressed sale. The discount varies with the phase of the foreclosure, ranging from twelve months prior to the foreclosure sale until twelve months after the REO sale.

Figure 4

Contagion Effect by Phase of Foreclosure

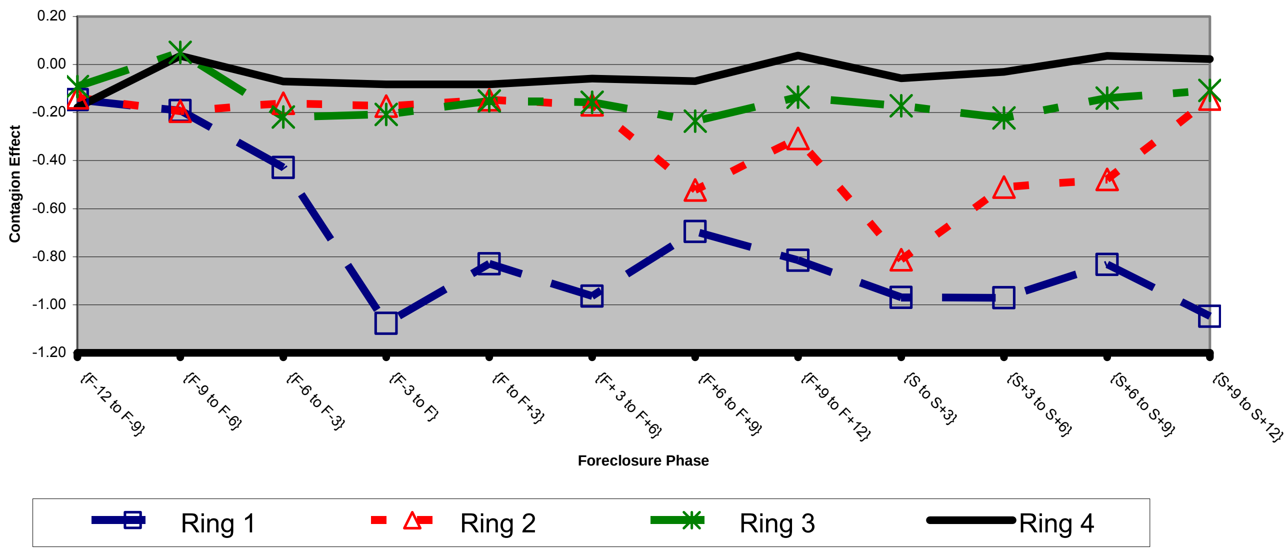
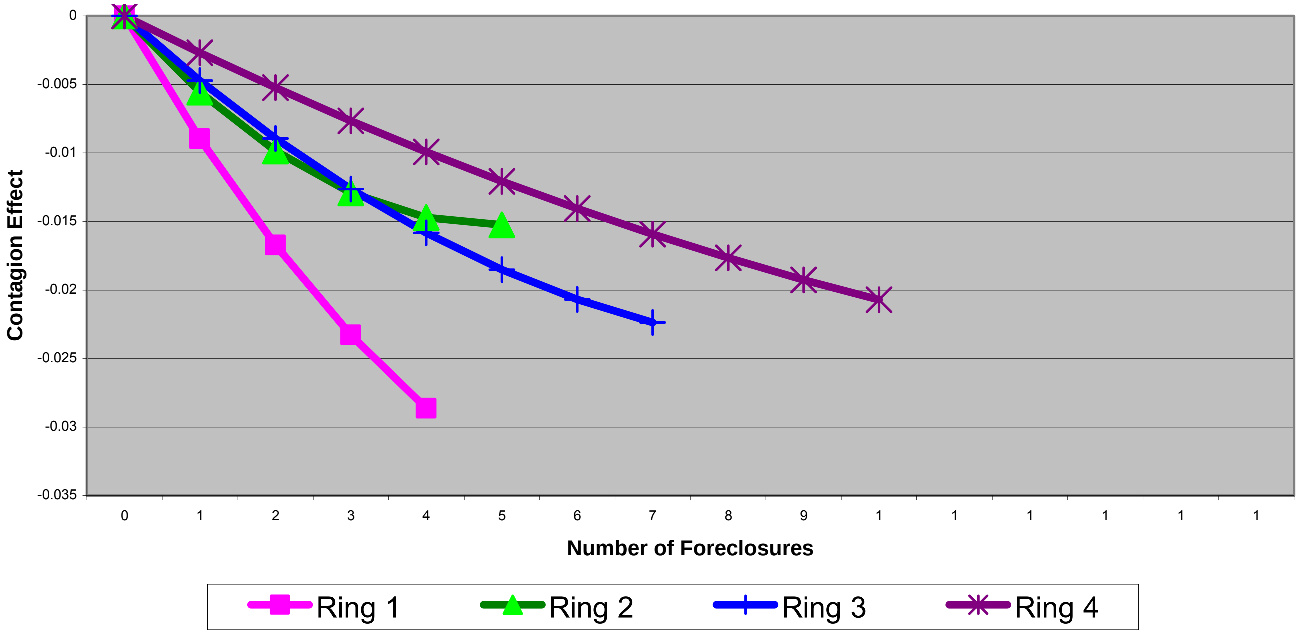


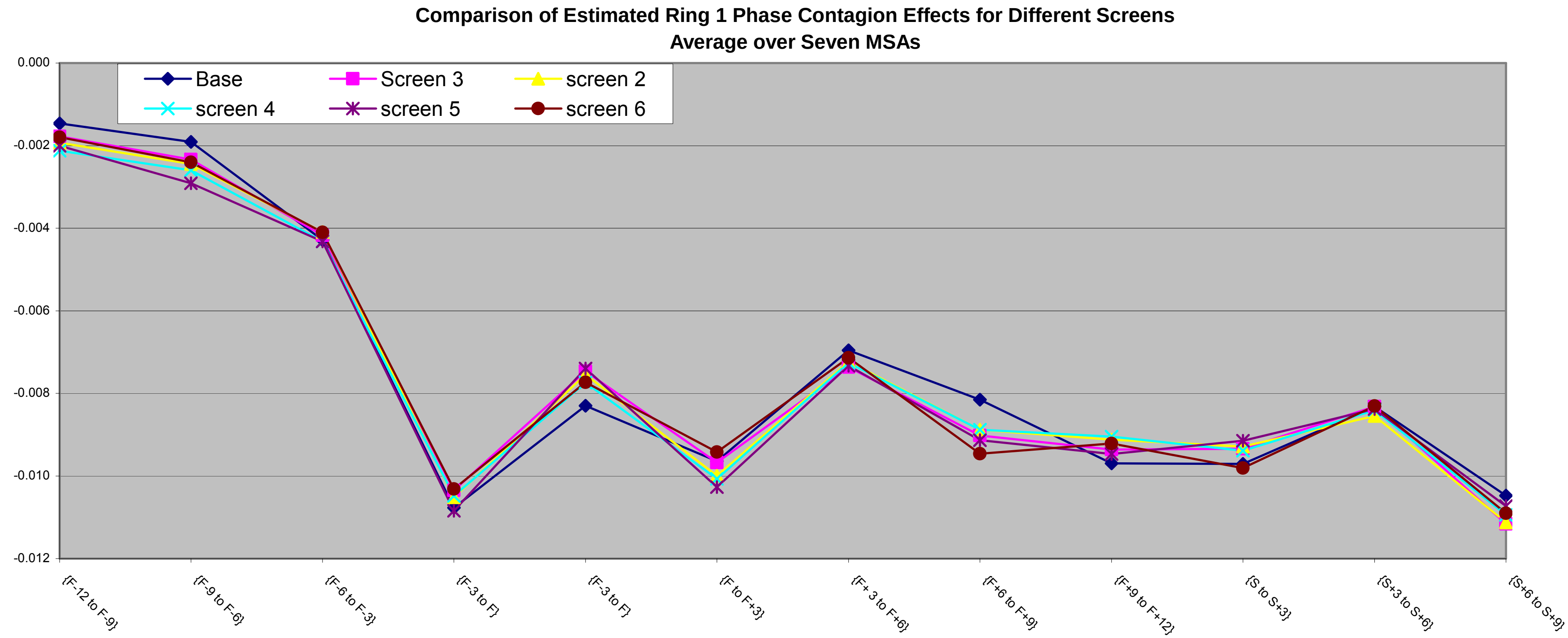
Figure 4 compares the estimated contagion effect by phase of foreclosure for all four Rings. Ring 1 contains all properties within three hundred feet of the non-distressed sale. Ring 2 contains all properties greater than three hundred feet and less than five hundred feet from the non-distressed sale. Ring 3 includes properties between five hundred feet and one thousand feet and Ring 4 contains properties from one thousand feet to two thousand feet. The plotted phase effects represent the average estimated effect over the seven different MSAs. The individual MSA effects vary and are reported in Tables 4 and 5 for Rings 1 and 2.

Figure 5
Effect of Multiple Foreclosures by Ring
Model Specification Estimates a Quadratic Effect



Note: Each line in Figure 5 shows how the estimated contagion discount varies with the number of nearby foreclosed properties within each of the specified rings around the subject property. The lines for all r are truncated to avoid extrapolating the effect beyond the range of the data used to estimate the models. Ring 1 includes all foreclosures less than 300 feet from the subject property. Ring 2 includes foreclosures that are 300 to 500 feet from the subject. Ring 3 includes foreclosures that are between 500 and 1000 feet from the subject, while ring 4 includes foreclosures that are between 1000 and 200 feet from the subject property.

Figure 6



Notes:
Figure 6 plots the average estimated Ring 1 phase effect from twelve months before the foreclosure sale to twelve months after the REO sale date. Ring 1 includes all properties within three hundred feet of the non-distressed property sale. The estimated effects attributable to properties that take longer than twelve months after the foreclosure sale to be sold to a third party are excluded because the sample sizes are small for that phase window and the estimates are not significant. The Base screen refers to the standard screen applied to repeat sales pairs used throughout most of the paper. The various screens are described in Table 7. Each screen is more restrictive (i.e., excludes more observations) than the Base screen. Each line in the figure represents the average estimated effect of all seven MSAs, where each model was estimated on a sample of repeat sales pairs that passed the indicated screen.