FEDERAL RESERVE BANK OF ST. LOUIS
SUPERVISORY POLICY ANALYSIS WORKING PAPER

Working Paper 2007-05

# Understanding the Subprime Mortgage Crisis

**Yuliya Demyanyk**
Economist
Federal Reserve Bank of St. Louis

**Otto Van Hemert**
Assistant Professor
New York University

This draft: August 2008

# Understanding the Subprime Mortgage Crisis

Yuliya Demyanyk, Otto Van Hemert*

This Draft: August 19, 2008
First Draft: October 9, 2007

**Abstract**

Using loan-level data, we analyze the quality of subprime mortgage loans by adjusting their performance for differences in borrower characteristics, loan characteristics, and house price appreciation since origination. We find that the quality of loans deteriorated for six consecutive years before the crisis and that securitizers were, to some extent, aware of it. We provide evidence that the rise and fall of the subprime mortgage market follows a classic lending boom-bust scenario, in which unsustainable growth leads to the collapse of the market. Problems could have been detected long before the crisis, but they were masked by high house price appreciation between 2003 and 2005.

*Demyanyk: Banking Supervision and Regulation, Federal Reserve Bank of St. Louis, P.O. Box 442, St. Louis, MO 63166, Yuliya.Demyanyk@stls.frb.org. Van Hemert: Department of Finance, Stern School of Business, New York University, 44 W. 4th Street, New York, NY 10012, ovanheme@stern.nyu.edu. The authors would like to thank Cliff Asness, Joost Driessen, William Emmons, Scott Frame, Xavier Gabaix, Dwight Jaffee, Ralph Koijen, Andreas Lehnert, Chris Mayer, Andrew Meyer, Toby Moskowitz, Lasse Pedersen, Robert Rasche, Matt Richardson, Stefano Risa, Bent Sorensen, Stijn Van Nieuwerburgh, James Vickery, Jeff Wurgler, and seminar participants at the Federal Reserve Bank of St. Louis, the Florida Atlantic University, the International Monetary Fund, the second New York Fed – Princeton liquidity conference, Lehman Brothers, the Baruch-Columbia-Stern real estate conference, NYU Stern Research Day, Capula Investment Management, AQR Capital Management, the Conference on the Subprime Crisis and Economic Outlook in 2008 at Lehman Brothers, Freddie Mac, Federal Deposit and Insurance Corporation (FDIC), US Securities and Exchange Comission (SEC), Office of Federal Housing Enterprise Oversight (OFHEO), Board of Governors of the Federal Reserve System, Carnegie Mellon University, Baruch, University of British Columbia, University of Amsterdam, the 44th Annual Conference on Bank Structure and Competition at the Federal Reserve Bank of Chicago, and the Federal Reserve Research and Policy Activities Pertaining to the Recent Turmoil in Financial Markets, Atlanta. The views expressed are those of the authors and do not necessarily reflect the official positions of the Federal Reserve Bank of St. Louis or the Federal Reserve System.

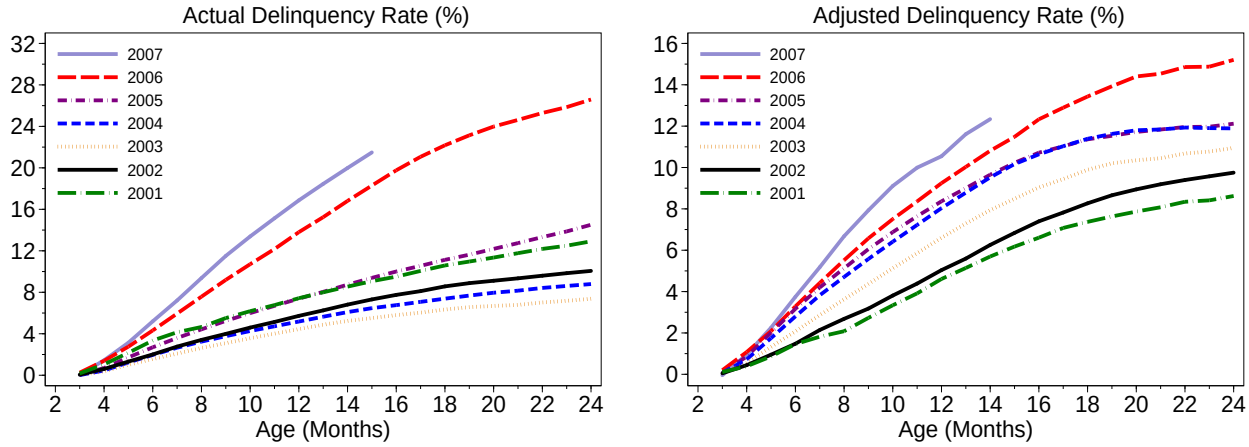
# 1 Introduction

The subprime mortgage crisis of 2007 was characterized by an unusually large fraction of subprime mortgages originated in 2006 and 2007 being delinquent or in foreclosure only months later. The crisis spurred massive media attention; many different explanations of the crisis have been suggested. The goal of this paper is to answer the question: "What do the data tell us about the possible causes of the crisis?" To this end we use a loan-level database containing information on about half of all U.S. subprime mortgages originated between 2001 and 2007.

The relatively poor performance of vintages 2006 and 2007 loans is illustrated in Figure 1 (left panel). At every mortgage loan age, loans originated in 2006 and 2007 show a much higher delinquency rate than loans originated in earlier years at the same ages.

Figure 1: Actual and Adjusted Delinquency Rate

The figure shows the age pattern in the actual (left panel) and adjusted (right panel) delinquency rate for the different vintage years. Delinquency is defined as being 60 or more days late with the monthly mortgage payment, in foreclosure, real-estate owned, or defaulted. The adjusted delinquency rate is obtained by adjusting the actual rate for year-by-year variation in FICO scores, loan-to-value ratios, debt-to-income ratios, missing debt-to-income ratio dummies, cash-out refinancing dummies, owner-occupation dummies, documentation levels, percentage of loans with prepayment penalties, mortgage rates, margins, house price appreciation since origination, composition of mortgage contract types, and origination amounts.




We document that the poor performance of the vintage 2006 and 2007 loans was not confined to a particular segment of the subprime mortgage market. For example, fixed-rate, hybrid, purchase-money, cash-out refinancing, low-documentation, and full-documentation loans originated in 2006 and 2007 all showed substantially higher delinquency rates than loans made the prior five years. This contradicts a

widely-held belief that the subprime mortgage crisis was mostly confined to hybrid or low-documentation mortgages.

We explore to what extent the subprime mortgage crisis can be attributed to different loan characteristics, borrower characteristics, and subsequent house price appreciation. The subsequent house price appreciation is measured as the MSA-level house price change between the moment of origination and the moment of loan performance evaluation. For the empirical analysis, we run logit regressions with the probability of delinquency being a function of these factors.

We find that loan and borrower characteristics are very important in terms of explaining the cross-section of loan performance. However, because these characteristics were not sufficiently different in 2006 and 2007 compared with the prior five years, they cannot explain the unusually weak performance of vintage 2006 and 2007 loans. For example, a one-standard-deviation increase in the debt-to-income ratio raises the probability of delinquency 12 months after origination by as much as 1.1 percentage points. However, because the average debt-to-income ratio was only 0.2 standard deviations higher in 2006 than its level in previous years, it contributes very little to explain the inferior performance of vintage 2006 loans. The only variable in the considered logit regression model that contributed substantially to the crisis is the low subsequent house price appreciation for vintage 2006 and 2007 loans, which can explain about a 2 to 4 percentage points higher-than-average delinquency rate 12 months after origination.[1] Due to geographical heterogeneity in house price changes, some areas have experienced larger-than-average house price declines and therefore have a larger explained increase in delinquency and foreclosure rates.[2]

We analyze the quality of loans based on their performance, adjusted for differences in observed loan characteristics, borrower characteristics, and subsequent house price appreciation. For the analysis, we compute the prediction error as the difference between the actual delinquency rate and the estimated probability of delinquency based on the logit regression model. In Figure 1 (right panel) we plot the adjusted delinquency rates, which are obtained by adding up the prediction errors and the weighted average actual rates. This ensures having the same weighted average for the actual (Figure 1, left panel) and adjusted (Figure 1, right panel) delinquency rates.

As shown in Figure 1 (right panel), the adjusted delinquency rates have been steadily rising for the past seven years. In other words, loan quality—adjusted for observed characteristics and subsequent house

[1] Other papers that research the relationship between house prices and mortgage financing include Genesove and Mayer (1997), Genesove and Mayer (2001), and Brunnermeier and Julliard (2007).

[2] Also, house price appreciation may differ in cities versus rural areas. See for example Glaeser and Gyourko (2005) and Gyourko and Sinai (2006).

price appreciation—deteriorated monotonically between 2001 and 2007. Interestingly, 2001 was among the worst vintage years in terms of actual delinquency and foreclosure rates, but is in fact the best vintage year in terms of the adjusted rates. High interest rates, low average FICO credit scores, and low house price appreciation created the "perfect storm" in 2001, resulting in a high actual delinquency rate; after adjusting for these unfavorable circumstances, however, the adjusted delinquency rates are low.

In addition to the monotonic deterioration of loan quality, we show that over time the average combined loan-to-value ratio increased, the fraction of low documentation loans increased, and the subprime-prime rate spread decreased. The rapid rise and subsequent fall of the subprime mortgage market is therefore reminiscent of a classic lending boom-bust scenario.[3] The origin of the subprime lending boom has often been attributed to the increased demand for so-called private-label mortgage-backed securities (MBSs) by both domestic and foreign investors. Our database does not allow us to directly test this hypothesis, but an increase in demand for subprime MBSs is consistent with our finding of lower spreads and higher volume. Mian and Sufi (2008) find evidence consistent with this view that increased demand for MBSs spurred the lending boom.

The logit regression specification used to compute the adjusted delinquency and foreclosure rates assumes that the regression coefficients on the different explanatory variables remain constant over time. We test the validity of this assumption for all variables and find that it is the most strongly rejected for the loan-to-value (LTV) ratio. High-LTV borrowers in 2006 and 2007 were riskier than those in 2001 in terms of the probability of delinquency or foreclosure, for given values of the other explanatory variables. Were securitizers aware of the increasing riskiness of high-LTV borrowers?[4] To answer this question, we analyze the relationship between the mortgage rate and LTV ratio (along with the other loan and borrower characteristics). We perform a cross-sectional ordinary least squares (OLS) regression, with the mortgage rate as the dependent variable, for each quarter from 2001Q1 to 2007Q2 for both fixed-rate mortgages and 2/28 hybrid mortgages. Figure 2 shows that the coefficient on the first-lien LTV variable, scaled by the standard deviation of the first-lien LTV ratio, has been increasing over time. We thus find evidence that securitizers were aware of the increasing riskiness of high-LTV borrowers, and adjusted mortgage

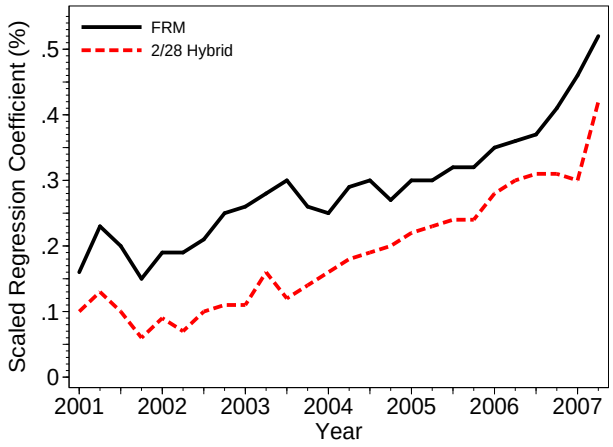
[3]Berger and Udell (2004) discuss the empirical stylized fact that during a monetary expansion lending volume typically increases and underwriting standards loosen. Loan performance is the worst for those loans underwritten toward the end of the cycle. Demirgüç-Kunt and Detragiache (2002) and Gourinchas, Valdes, and Landerretche (2001) find that lending booms raise the probability of a banking crisis. Dell'Ariccia and Marquez (2006) show in a theoretical model that a change in information asymmetry across banks might cause a lending boom that features lower standards and lower profits. Ruckes (2004) shows that low screening activity may lead to intense price competition and lower standards.

[4]For loans that are securitized (as are all loans in our database), the securitizer effectively dictates the mortgage rate charged by the originator.

rates accordingly.

Figure 2: Sensitivity of Mortgage Rate to First-Lien Loan-to-Value Ratio

The figure shows the effect of the first-lien loan-to-value ratio on the mortgage rate for first-lien fixed-rate and 2/28 hybrid mortgages. The effect is measured as the regression coefficient on the first-lien loan-to-value ratio (scaled by the standard deviation) in an ordinary least squares regression with the mortgage rate as the dependent variable and the FICO score, first-lien loan-to-value ratio, second-lien loan-to-value ratio, debt-to-income ratio, missing debt-to-income ratio dummy, cash-out refinancing dummy, owner-occupation dummy, prepayment penalty dummy, origination amount, term of the mortgage, prepayment term, and margin (only applicable to 2/28 hybrid) as independent variables. Each point corresponds to a separate regression, with a minimum of 18,784 observations.




We show that our main results are robust to analyzing mortgage contract types separately, focussing on foreclosures rather than delinquencies, and numerous different regression specifications like allowing for interaction effects between different loan and borrower characteristics. The latter includes taking into account risk-layering—the origination of loans that are risky in several dimensions, such as the combination of a high LTV ratio and a low FICO score.

As an extension, we estimate our regression model using data just through year-end 2005 and again obtain the continual deterioration of loan quality since 2001. This means that the seeds for the crisis were sown long before 2007, but detecting them was complicated by high house price appreciation between 2003 and 2005—appreciation that masked the true riskiness of subprime mortgages.

There is a large literature on the determinants of mortgage delinquencies and foreclosures, dating back to at least Von Furstenberg and Green (1974). Recent contributions include Cutts and Van Order (2005) and Pennington-Cross and Chomsisengphet (2007).[5] Other papers analyzing the subprime crisis

[5]Deng, Quigley, and Van Order (2000) discuss the simultaneity of the mortgage prepayment and default option. Campbell and Cocco (2003) and Van Hemert (2007) discuss mortgage choice over the life cycle.

include Gerardi, Shapiro, and Willen (2008), Mian and Sufi (2008), DellAriccia, Igan, and Laeven (2008), and Keys, Mukherjee, Seru, and Vig (2008). Our paper makes several novel contributions. First, we quantify how much different determinants have contributed to the observed high delinquency rates for vintage 2006 and 2007 loans, which led up to the 2007 subprime mortgage crisis. Our data enables us to show that the effect of different loan-level characteristics as well as low house price appreciation was quantitatively too small to explain the poor performance of 2006 and 2007 vintage loans. Second, we uncover a downward trend in loan quality, determined as loan performance adjusted for differences in loan and borrower characteristics as well as subsequent house price appreciation. We further show that there was a deterioration of lending standards and a decrease in the subprime-prime mortgage rate spread during the 2001–2007 period. Together these results provide evidence that the rise and fall of the subprime mortgage market follows a classic lending boom-bust scenario, in which unsustainable growth leads to the collapse of the market. Third, we show that continual deterioration of loan quality could have been detected long before the crisis by means of a simple statistical exercise. Fourth, securitizers were, to some extent, aware of this deterioration over time, as evidenced by changing determinants of mortgage rates.

The structure of this paper is as follows. In Section 2 we show the descriptive statistics for the subprime mortgages in our database. In Section 3 we present the econometric results and discuss explanatory factors for delinquency. In Section 4 we discuss the increasing riskiness of high-LTV borrowers, and the extent to which securitizers were aware of this risk. In Section 5 we analyze the subprime-prime rate spread and in Section 6 we conclude. We provide several additional robustness checks in the appendices.

## 2 Descriptive Analysis

In this paper we use the First American CoreLogic LoanPerformance (henceforth: LoanPerformance) database, which covers loan-level data on about 85 percent of all securitized subprime mortgages; more than half of the U.S. subprime mortgage market.[6] Since the first version of this paper in October 2007, LoanPerformance has responded to the request by trustees' clients to reclassify some of its subprime loans to Alt-A status. While it is not clear to us whether the pre- or post-classification subprime data is the most appropriate for research purposes, it is reassuring that our results proved to be robust to

[6] Mortgage Market Statistical Annual (2007) reports securitization shares of subprime mortgages each year from 2001 to 2006 equal to 54, 63, 61, 76, 76, and 75 percent respectively.

the reclassification. In this version we focus on the post-classification data. In Appendix A we provide more details on the reclassification of the LoanPerformance database and show the robustness of our main results to using pre-reclassification data.

There is no consensus on the exact definition of a subprime mortgage loan. The term subprime can be used to describe certain characteristics of the borrower (e.g., a FICO credit score less than 620),[7] lender (e.g., specialization in high-cost loans),[8] security of which the loan can become a part (e.g., high projected default rate for the pool of underlying loans), or mortgage contract type (e.g., no money down and no documentation provided, or 2/28 hybrid). The common element across definitions of a subprime loan is a high default risk. In this paper, subprime loans are those underlying subprime securities. We do not include less-risky Alt-A mortgage loans in our analysis. We focus on first-lien loans and consider the 2001 through 2008 sample period.

We first discuss the main characteristics of the loans in our database at origination. Second, we discuss the delinquency rates of these loans for various segments of the subprime mortgage market.

### 2.1 Loan Characteristics at Origination

Table 1 provides the descriptive statistics for the subprime mortgage loans in our database that were originated between 2001 and 2007. In the first block of Table 1 we see that the annual number of originated loans increased by a factor four between 2001 and 2006. The average loan size almost doubled over those five years. The total dollar amount originated in 2001 was $57 billion, while in 2006 it was $375 billion. In 2007, in the wake of the subprime mortgage crisis, the dollar amount originated fell sharply to $69 billion, primarily originated in the first half of 2007.

In the second block of Table 1, we split the pool of mortgages into four main mortgage contract types. Most numerous are the hybrid mortgages, accounting for more than half of all our subprime loans. A hybrid mortgage carries a fixed rate for an initial period (typically 2 or 3 years) and then the rate resets to a reference rate (often the 6-month LIBOR) plus a margin. The fixed-rate mortgage contract has become less popular in the subprime market over time and accounted for just 20 percent of the total number of

[7] The Board of Governors of the Federal Reserve System, The Office of the Controller of the Currency, the Federal Deposit Insurance Corporation, and the Office of Thrift Supervision use this definition. See e.g. http://www.fdic.gov/news/news/press/2001/pr0901a.html

[8] The U.S. Department of Housing and Urban Development uses HMDA data and interviews lenders to identify subprime lenders among them. There are, however, some subprime lenders making prime loans and some prime lenders originating subprime loans.

Table 1: Loan Characteristics at Origination for Different Vintages

Descriptive statistics for the first-lien subprime loans in the LoanPerformance database.

| | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 |
|---|---|---|---|---|---|---|---|
| | | | | *Size* | | | |
| Number of Loans (*1000) | 452 | 737 | 1,258 | 1,911 | 2,274 | 1,772 | 316 |
| Average Loan Size (*$1000) | 126 | 145 | 164 | 180 | 200 | 212 | 220 |
| | | | | *Mortgage Type* | | | |
| FRM (%) | 33.2 | 29.0 | 33.6 | 23.8 | 18.6 | 19.9 | 27.5 |
| ARM (%) | 0.4 | 0.4 | 0.3 | 0.3 | 0.4 | 0.4 | 0.2 |
| Hybrid (%) | 59.9 | 68.2 | 65.3 | 75.8 | 76.8 | 54.5 | 43.8 |
| Balloon (%) | 6.5 | 2.5 | 0.8 | 0.2 | 4.2 | 25.2 | 28.5 |
| | | | | *Loan Purpose* | | | |
| Purchase (%) | 29.7 | 29.3 | 30.1 | 35.8 | 41.3 | 42.4 | 29.6 |
| Refinancing (cash out) (%) | 58.4 | 57.4 | 57.7 | 56.5 | 52.4 | 51.4 | 59.0 |
| Refinancing (no cash out) (%) | 11.2 | 12.9 | 11.8 | 7.7 | 6.3 | 6.2 | 11.4 |
| | | | | *Variable Means* | | | |
| FICO Score | 601.2 | 608.9 | 618.1 | 618.3 | 620.9 | 618.1 | 613.2 |
| Combined Loan-to-Value Ratio (%) | 79.4 | 80.1 | 82.0 | 83.6 | 84.9 | 85.9 | 82.8 |
| Debt-to-Income Ratio (%) | 38.0 | 38.5 | 38.9 | 39.4 | 40.2 | 41.1 | 41.4 |
| Missing Debt-to-Income Ratio Dummy (%) | 34.7 | 37.5 | 29.3 | 26.5 | 31.2 | 19.7 | 30.9 |
| Investor Dummy (%) | 8.2 | 8.1 | 8.1 | 8.3 | 8.3 | 8.2 | 8.2 |
| Documentation Dummy (%) | 76.5 | 70.4 | 67.8 | 66.4 | 63.4 | 62.3 | 66.7 |
| Prepayment Penalty Dummy (%) | 75.9 | 75.3 | 74.0 | 73.1 | 72.5 | 71.0 | 70.2 |
| Mortgage Rate (%) | 9.7 | 8.7 | 7.7 | 7.3 | 7.5 | 8.4 | 8.6 |
| Margin for ARM and Hybrid Mortgage Loans (%) | 6.4 | 6.6 | 6.3 | 6.1 | 5.9 | 6.1 | 6.0 |

loans in 2006. In contrast, in the prime mortgage market, most mortgage loans are of the fixed-rate type.[9] In 2007, in the wake of the subprime mortgage crisis, it increased again to 28%. The proportion of balloon mortgage contracts jumped substantially in 2006, and accounted for 25 percent of the total number of mortgages originated that year. A balloon mortgage does not fully amortize over the term of the loan and therefore requires a large final (balloon) payment. Less than 1 percent of the mortgages originated over the sample period were adjustable-rate (non-hybrid) mortgages.

In the third block of Table 1, we report the purpose of the mortgage loans. In about 30 to 40 percent of cases, the purpose is to finance the purchase of a house. Approximately 55 percent of our subprime mortgage loans were originated to extract cash, by refinancing an existing mortgage loan into a larger new mortgage loan. The share of loans originated in order to refinance with no cash extraction is relatively small.

In the final block of Table 1, we report the mean values for the variables that we will use in the regression analysis (see Table 2 for a definition of these variables). The average FICO credit score rose 20 points between 2001 and 2005. The combined loan-to-value (CLTV) ratio, which measures the value of all-lien loans divided by the value of the house, slightly increased over 2001–2006, primarily because of the increased popularity of second-lien and third-lien loans. The (back-end) debt-to-income ratio (if provided) and the fraction of loans with a prepayment penalty were fairly constant. For about a third of the loans in our database, no debt-to-income ratio was provided (the reported value in those cases is zero); this is captured by the missing debt-to-income ratio dummy variable. The share of loans with full documentation fell considerably over the sample period, from 77 percent in 2001 to 67 percent in 2007. The mean mortgage rate fell from 2001 to 2004 and rebounded after that, consistent with movements in both the 1-year and 10-year Treasury yields over the same period. Finally, the margin (over a reference rate) for adjustable-rate and hybrid mortgages stayed rather constant over time.

We do not report summary statistics on the loan source, such as whether a mortgage broker intermediated, as the broad classification used in the database rendered this variable less informative.

[9]For example Koijen, Van Hemert, and Van Nieuwerburgh (2007) show that the fraction of conventional, single-family, fully amortizing, purchase-money loans reported by the Federal Housing Financing Board in its Monthly Interest Rate Survey that are of the fixed-rate type fluctuated between 60 and 90 percent from 2001 to 2006. Vickery (2007) shows that empirical mortgage choice is affected by the eligibility of the mortgage loan to be purchased by Fannie Mae and Freddie Mac.

### 2.2 Performance of Loans by Market Segments

In Figure 1 (left panel) we showed that for the subprime mortgage market as a whole, vintage 2006 and 2007 loans stand out in terms of high delinquency rates (for variable definitions, see Table 2). In Figure 3, we again plot the age pattern in the delinquency rate for vintages 2001 through 2007 and split the subprime mortgage market into various segments. As the figure shows, the poor performance of the 2006 and 2007 vintages is not confined to a particular segment of the subprime market, but rather reflects a market-wide phenomenon.

In the six panels of Figure 3 we see that for hybrid, fixed-rate, purchase-money, cash-out refinancing, low-documentation, and full-documentation mortgage loans, the 2006 and 2007 vintages show the highest delinquency rate pattern. In general, vintages 2001 and 2005 come next in terms of delinquency rates, and vintage 2003 loans have the lowest delinquency rates. Notice that the scale of the vertical axis differs across the panels. The delinquency rates for the fixed-rate mortgages (FRMs) are lower than those for hybrid mortgages but exhibit a remarkably similar pattern across vintage years.

In Figure 4 we plot the delinquency rates of all *outstanding* mortgages. Notice that the fraction of FRMs that are delinquent remained fairly constant from 2005Q1 to 2007Q2. These rates are consistent with those used in an August 2007 speech by the Chairman of the Federal Reserve System (Bernanke (2007)), who said "For subprime mortgages with fixed rather than variable rates, for example, serious delinquencies have been fairly stable." It is important, though, to realize that this result is driven by an aging effect of the FRM pool, caused by a decrease in the popularity of FRMs over 2001-2006 (see Table 1). In other words, FRMs originated in 2006 in fact performed unusually poorly (Figure 3, upper-right Panel), but if one plots the delinquency rate of outstanding FRMs over time (Figure 4, left Panel), the weaker performance of vintage 2006 loans is masked by the aging of the overall FRM pool.

## 3 Empirical Analysis of Delinquency Determinants

In this section we investigate to what extent a logit regression model can explain the high levels of delinquencies for the vintage 2006 and 2007 mortgage loans in our database. All results in this section are based on a random sample of one million first-lien subprime mortgage loans, originated between 2001 and 2007.

Figure 3: Actual Delinquency Rate for Segments of the Subprime Mortgage Market

The figure shows the age pattern in delinquency rate for the different vintages. Each of the six panels focuses on a different segment of the subprime mortgage market.

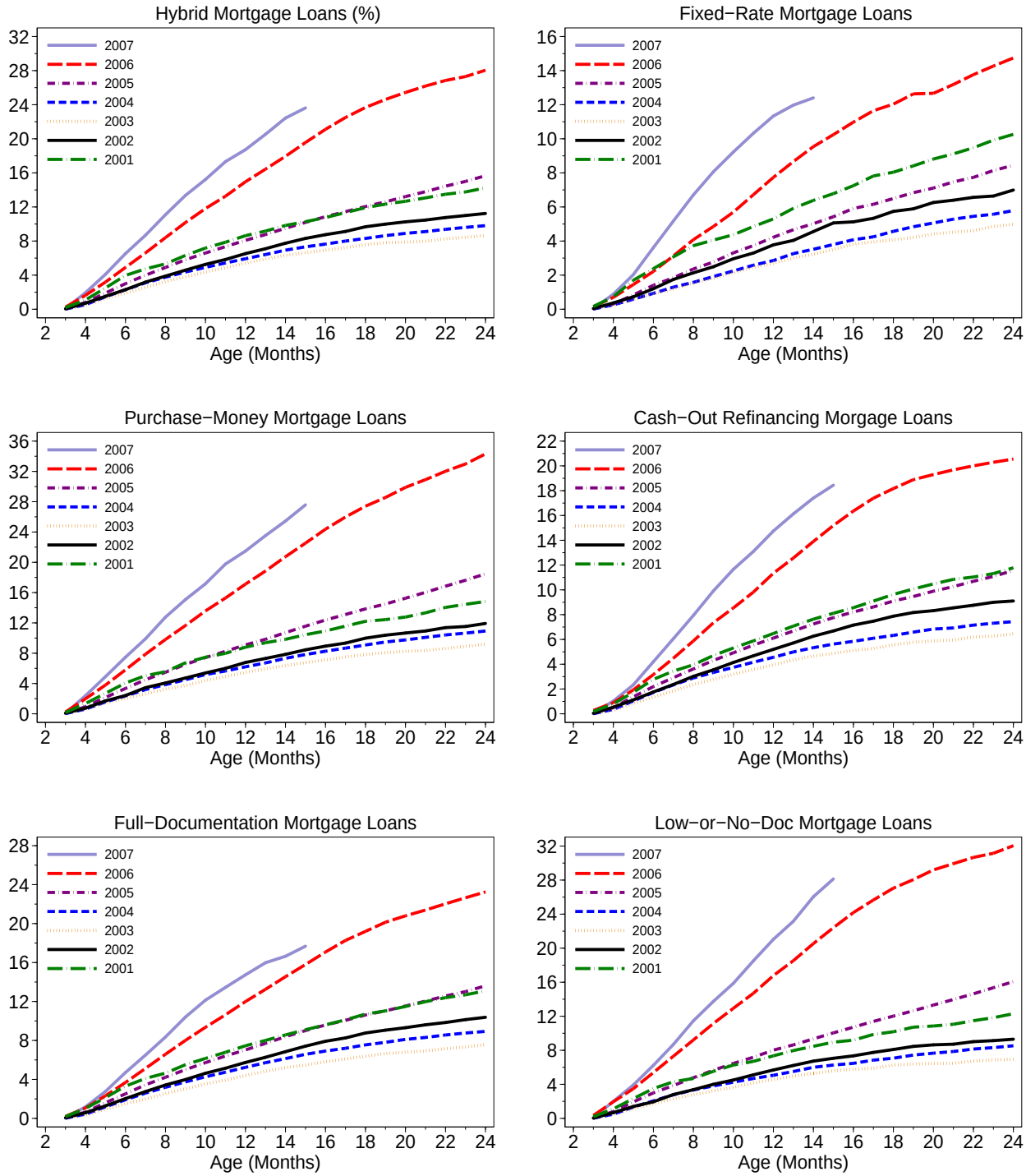
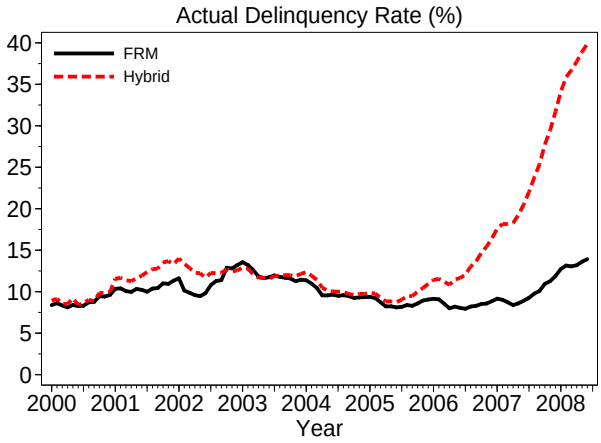


Figure 4: Actual Delinquency Rates of Outstanding Mortgages

The Figure shows the actual delinquency rates of all outstanding FRMs and hybrids from January 2000 through June 2008.




### 3.1 Empirical Model Specification

We run the following logit regression

$$\Pr(event) = \Phi(\beta' X), \tag{1}$$

where the *event* is delinquency of a subprime mortgage loan after a given number of months; $\Phi(x) = 1/(1+\exp(-x))$ is the logit function; $X$ is the vector of explanatory variables; and $\beta$ is the vector of regression coefficients. We will report the following statistics for each explanatory variable $i$:

$$\begin{aligned} marginal_i &= \Phi(\beta'\bar{X} + \beta_i\sigma_i) - \Phi(\beta'\bar{X}) && (2)\\ deviation01_i &= (\overline{X01}_i - \bar{X}_i)/\sigma_i && (3)\\ contribution01_i &= \Phi(\beta'\bar{X} + \beta_i(\overline{X01}_i - \bar{X}_i)) - \Phi(\beta'\bar{X}) && (4)\\ &\approx marginal_i \times deviation01_i && (5) \end{aligned}$$

where $\bar{X}$ is the vector with mean values, $\sigma_i$ is the standard deviation of the $i$-th variable, and $\overline{X01}_i$ is the mean value of the $i$-th variable for vintage 2001 loans. We define mean value, deviation, and contribution for vintage years other than 2001 in a similar fashion. Equation (5) emerges from a first-order Taylor approximation with the derivative of the logit function with respect to the $i$-th variable approximated by

$marginal_i$.[10] The *marginal* statistic measures the effect of a one-standard-deviation increase in a variable (from its mean) on the probability of an event. The *deviation* statistic measures the number of standard deviations that the mean value of a variable in a particular vintage year was different from the mean value measured over the entire sample. The *contribution* statistic measures the deviation of the (average) event probability in a particular vintage year from the (average) event probability over the entire sample that can be explained by a particular variable.

For any subgroup of loans, such as a particular vintage, we can determine the predicted probability of an event by computing:

$$predicted = \sum_{j=1}^{L} \Phi(\beta' X^j)/L, \tag{6}$$

where the superscript $j$ refers to the loan number and $L$ is the total number of loans in the subgroup.

### 3.2 Variable Definitions

Table 2 provides the definitions of the dependent and independent (explanatory) variables used in the empirical analysis. We use the delinquency dummy variable as the dependent variable for the main analysis and consider a foreclosure dummy variable in Appendix B. We define a loan to be delinquent if payments on the loan are 60 or more days late, or the loan is reported as in foreclosure, real estate owned, or default. We do not always observe for a terminated loan whether the loan was prepaid or there was a default. In those cases we classify a terminated loan as a default if in the prior month the loan was in foreclosure, and as a prepayment otherwise. In Appendix C we provide a robustness check by omitting all terminated loans.

The borrower and loan characteristics we use in the analysis are: the FICO credit score, the combined loan-to-value ratio, the value of the debt-to-income ratio (when provided), a dummy variable indicating whether the debt-to-income ratio was missing (reported as zero), a dummy variable indicating whether the loan was a cash-out refinancing, a dummy variable indicating whether the borrower was an investor (as opposed to an owner-occupier), a dummy variable indicating whether full documentation was provided, a dummy variable indicating whether there is a prepayment penalty on a loan, the (initial) mortgage rate, and the margin for adjustable-rate and hybrid loans.[11]

[10] Technically, we first change units by multiplying by $\sigma_i$ in Equation (2) and diving by $\sigma_i$ in Equation (3).

[11] We also studied specifications that included loan purpose, reported in Table 1, and housing outlook, defined as the house price accumulation in the year prior to the loan origination. These variables were not significant and did not materially change the regression coefficients on the other variables.

## Table 2: Variable Definitions

This table presents definitions of the variables used in the regression analysis. The first two variables are used as dependent variables. The other variables are used as independent variables. We report the expected sign for the independent variables in parentheses and sometimes provide a brief motivation.

| Variable (Expected Sign) | Explanation |
|---|---|
| Delinquency Dummy | Payments on the loan are 60 or more days late, or the loan is reported as in foreclosure, real estate owned, or defaulted. |
| Foreclosure Dummy | The loan is reported as in foreclosure, real estate owned, or defaulted. |
| FICO Score (-) | Fair, Isaac and Company (FICO) credit score at origination. |
| Combined Loan-to-Value Ratio (+) | Combined values of all liens divided by the value of the house at origination. A higher combined loan-to-value ratio makes default more attractive. |
| Debt-to-Income Ratio (+) | Back-end debt-to-income ratio, defined by the total monthly debt payments divided by the gross monthly income, at origination. A higher debt-to-income ratio makes it harder to make the monthly mortgage payment. |
| Missing Debt-to-Income Dummy (+) | Equals one if the back-end debt-to-income ratio is missing and zero if provided. We expect the lack of debt-to-income information to be a negative signal on borrower quality. |
| Cash-Out Dummy (-) | Equals one if the mortgage loan is a cash-out refinancing loan. Pennington-Cross and Chomsisengphet (2007) show that the most common reasons to initiate a cash-out refinancing are to consolidate debt and to improve property. |
| Investor Dummy (+) | Equals one if the borrower is an investor and does not owner-occupy the property. |
| Documentation Dummy (-) | Equals one if full documentation on the loan is provided and zero otherwise. We expect full documentation to be a positive signal on borrower quality. |
| Prepayment Penalty Dummy (+) | Equals one if there is a prepayment penalty and zero otherwise. We expect that a prepayment penalty makes refinancing less attractive. |
| Mortgage Rate (+) | Initial interest rate as of the first payment date. A higher interest rate makes it harder to make the monthly mortgage payment. |
| Margin (+) | Margin for an adjustable-rate or hybrid mortgage over an index interest rate, applicable after the first interest rate reset. A higher margin makes it harder to make the monthly mortgage payment. |
| House Price Appreciation (-) | MSA-level house price appreciation from the time of loan origination, reported by the Office of Federal Housing Enterprise Oversight (OFHEO). Higher housing equity leads to better opportunities to refinance the mortgage loan. |
| Product Type Dummies (+) | We consider four product types: FRMs, Hybrids, ARMs, and Balloons. We include a dummy variable for the latter three types, which therefore have the interpretation of the probability of delinquency relative to FRM. Because we expect the FRM to be chosen by more risk-averse and prudent borrowers, we expect positive signs for all three product type dummies. |
| Origination Amount (?) | Size of the mortgage loan. We have no clear prior on the effect of the origination amount on the probability of delinquency, holding constant the loan-to-value and debt-to-income ratio. |

In addition, we construct a variable that measures house price appreciation from the time of origination until the time we evaluate whether a loan is delinquent. To this end we use metropolitan statistical area (MSA) level house price indexes from the Office of Federal Housing Enterprise Oversight (OFHEO) and match loans with MSAs by using the zip code provided by LoanPerformance.[12]

We also considered the change in the unemployment rate from the period of origination until the period of loan performance evaluation, which we could only measure accurately at the state-level for the entire sample. It turned out that the unemployment variable mainly picked up the time trend in the delinquency rate. The relationship between the (trending) unemployment rate and the (trending) loan performance, however, is spurious. When vintage dummy variables are included in the regression, the unemployment rate becomes insignificant, both statistically and economically. We therefore decided to omit the change in the unemployment rate as an explanatory variable.

In Table 2 we report the expected sign for the regression coefficient on each of the explanatory variables in parentheses.

### 3.3 Determinants of Delinquency

Table 3 shows the results of the logit regression (Equation 1), where the *event* is delinquency 12 months after origination. The first column reports the explanatory variables. Column two reports the marginal effect of the explanatory variables (Equation 2) for the baseline case specification, in which we add a constant to the explanatory variables of column one. All marginal effects have the expected sign, as reported in Table 2. Except for the ARM dummy, all variables are significant at the 1% confidence level. The four explanatory variables with the largest (absolute) marginal effect and thus the most important for explaining cross-sectional differences in loan performance are the FICO score, the combined loan-to-value ratio, the mortgage rate, and the house price appreciation. According to the estimates, for example, a one standard deviation increase in the FICO score decreases the delinquency rate 12 months after origination by 2.33 percentage points. The product type has a relatively small effect on the performance of a loan, beyond what is explained by other characteristics. In Figure 3 we showed that FRMs experience a much lower delinquency rate than hybrid mortgages, which therefore must be driven by borrowers with better

[12] Estimating house price appreciation on the MSA-level, as opposed to the individual property level introduces a potential measurement error of this variable. To the best of our knowledge, there is no data available to estimate the size of this measurement error or to evaluate its impact on the results.

characteristics selecting into FRMs.[13] The pseudo R-squared statistic for the regression specification in column two is 10.2%.

In columns three and four we consider two alternative regression specifications: including both a constant and a trend, and including vintage year dummies. Comparing columns two to four, we see that the baseline case specification and these two alternative specifications lead to very similar marginal effects. The pseudo R-squared statistic is 10.7% for both the regression specifications in columns three and four. Hence adding a trend improves the fit compared to just including a constant. The (unreported) coefficient for the trend is positive and significant at the 1% confidence level. To gauge the economic significance we compute the predicted yearly percentage point increase in delinquency 12 months after origination using the regression coefficient of the trend, $\beta_{trend}$, as $\Phi(\beta'\bar{X} + \beta_{trend}) - \Phi(\beta'\bar{X}) = 0.79\%$. Adding vintage year dummies does not improve the fit further. The (unreported) values for the 7 vintage year dummies (2001, ..., 2007) are monotonically increasing over time. These results provide a first indication that loan quality deteriorated over time, after controlling for the effect of the explanatory variables in column one.

We explored numerous alternative regression specifications. First, we considered as explanatory variables those of the baseline case presented in Table 3, plus the 10 interaction and quadratic terms that can be constructed from the four most important explanatory variables: the FICO score, the CLTV ratio, the mortgage rate, and subsequent house price appreciation. Allowing for these additional terms, we take into account the effect of *risk-layering*—such as, for example, the effect of a combination of a borrower's low FICO score and a high CLTV ratio—on the probability of delinquency. It is in this case not a priori clear what the sign on the FICO-CLTV interaction variable is. A negative sign means that a low FICO and a high CLTV reinforce each other and give rise to a predicted delinquency probability that is higher than when the interaction is ignored. A positive sign could be explained by lenders who originate a low FICO and high CLTV loan only if they have positive private information on the loan or borrower quality. It turns out that the coefficient on the FICO-CLTV interaction term close to zero and insignificant. More certain is the sign we expect on the HPA-CLTV variable. Low house price appreciation is expected to especially give rise to a higher delinquency probability for a high CLTV ratio, because the borrower is closer to a situation with *negative equity* in the house (combined value of the mortgage loan larger than the market value of the house). Consistent with this intuition, we find a negative and significant (at the 1 percent level) coefficient on this interaction term for delinquency 12 months after origination. Allowing for the

[13] Consistent with this finding, LaCour-Little (2007) shows that individual credit characteristics are important for mortgage product choice.

## Table 3: Determinants of Delinquency 12 Months After Origination

The table shows the output of the logit regression (Equation 1), where the event is that a loan is delinquent 12 months after origination. The first column reports the explanatory variables. Columns two, three, and four report the marginal effect of the explanatory variables (Equation 2) for three different specifications: including a constant (baseline case), including a constant and a trend, including vintage year dummies. A "∗" indicates statistical significance at the 1% level. Columns five, six, and seven report the deviation of the average value of a variable in 2001, 2006, and 2007 from the average value over 2001–2007, expressed in number of standard deviations (Equation 3). Columns eight, nine, and ten report the contribution of a variable to explain a different probability of delinquency in 2001, 2006, 2007 (Equation 4), using the baseline case regression specification with a constant. We have the first-order approximation $contribution \approx marginal \times deviation$ (Equation 5).

| | **Marginal Effect, %** | | | **Deviations** | | | **Contribution, %** | | |
|---|---|---|---|---|---|---|---|---|---|
| Explanatory Variable | Constant | Trend | Dummies | 2001 | 2006 | 2007 | 2001 | 2006 | 2007 |
| FICO Score | −2.33∗ | −2.24∗ | −2.16∗ | −0.29 | 0.02 | −0.02 | 0.91 | −0.06 | 0.07 |
| Combined Loan-to-Value Ratio | 2.29∗ | 1.94∗ | 1.86∗ | −0.32 | 0.16 | 0.01 | −0.59 | 0.31 | 0.03 |
| Debt-to-Income Ratio | 1.16∗ | 1.04∗ | 1.05∗ | −0.21 | 0.23 | 0.02 | −0.22 | 0.24 | 0.03 |
| Missing Debt-to-Income Ratio | 1.01∗ | 0.99∗ | 1.01∗ | 0.16 | −0.19 | 0.03 | 0.15 | −0.17 | 0.03 |
| Cash-Out Dummy | −0.74∗ | −0.75∗ | −0.74∗ | 0.09 | −0.07 | 0.03 | −0.07 | 0.05 | −0.02 |
| Investor Dummy | 0.46∗ | 0.38∗ | 0.35∗ | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | −0.00 |
| Documentation Dummy | −0.94∗ | −0.84∗ | −0.80∗ | 0.25 | −0.08 | 0.01 | −0.25 | 0.08 | −0.01 |
| Prepayment Penalty Dummy | 0.26∗ | 0.33∗ | 0.35∗ | 0.10 | −0.05 | −0.07 | 0.02 | −0.01 | −0.02 |
| Mortgage Rate | 1.93∗ | 2.18∗ | 2.36∗ | 1.28 | 0.36 | 0.38 | 2.58 | 0.63 | 0.67 |
| Margin | 0.68∗ | 0.70∗ | 0.64∗ | −0.20 | 0.03 | −0.15 | −0.13 | 0.02 | −0.10 |
| House Price Appreciation | −2.04∗ | −1.71∗ | −1.79∗ | −0.31 | −0.76 | −1.21 | 0.82 | 2.25 | 3.91 |
| Hybrid Dummy | 0.22∗ | 0.19 | 0.23∗ | −0.14 | −0.28 | −0.49 | −0.03 | −0.06 | −0.10 |
| ARM Dummy | 0.04 | 0.03 | 0.04 | 0.02 | 0.01 | −0.01 | 0.00 | 0.00 | −0.00 |
| Balloon Dummy | 0.42∗ | 0.23∗ | 0.29∗ | −0.06 | 0.64 | 0.80 | −0.03 | 0.27 | 0.33 |
| Origination Amount | 1.06∗ | 0.89∗ | 0.91∗ | −0.46 | 0.22 | 0.26 | −0.44 | 0.22 | 0.26 |

interaction and quadratic terms did not substantially improve the overall fit, as measured by the pseudo R-squared statistic. Second, we considered as additional explanatory variable a dummy for the presence of the second-lien loan. This dummy variable had a positive significant effect on the predicted delinquency rate. However, it merely inherited some of the predictive power of the CLTV variable, while leaving the coefficients on the other variables as well as the overall fit virtually unaltered. Third, we considered as additional explanatory variable a dummy variable taking the value one whenever the CLTV equaled 80%, aimed to control for *silent seconds*, referring to a situation where an investor takes out a second-lien loan not reported in our database typically in combination with an 80% first-lien loan. This dummy variable was statistically significant but economically not very large and moreover hardly improved the overall fit. Fourth, we excluded the loans with not reported values of the debt-to-income ratios from the sample to make sure the measurement error associated with this variable does not lead to a significant bias of the results. The estimates based on the smaller subsample, in which debt-to-income variable has non-zero reported values, are statistically and economically similar to those based on the entire sample of loans.

### 3.4 Contribution to Explaining the Poor Performance of 2001, 2006, 2007

In the last three columns of Table 3 we report the contribution of the different explanatory variables to explaining the relatively high delinquency rates of loans originated in 2001, 2006, and 2007. Up to a first-order approximation, the contribution equals the marginal effect, presented in column two, times the average deviation from the sample mean of a variable in the respective years, presented in columns five to seven (see Equations 2–5 for formal definitions). First focussing on 2001, the mortgage rate was unusually high, the FICO score low, and the subsequent house price appreciation low. All three effects contributed to a high delinquency rate in 2001. In this sense one can say that loans originated in 2001 experienced the "perfect storm." For example the low average FICO score for 2001 can already explain a 0.91 percentage point increase in the delinquency rate 12 months after origination.

For vintages 2006 and 2007, low subsequent house price appreciation, in particular, contributed to their weak performance, and accounted for a 2 to 4 percentage point increase in delinquency rate 12 months after origination. The mean values in 2006 and 2007 for the other variables were not sufficiently different from the sample mean to contribute much to a different delinquency for loans originated in those years. It is worth noting that the high average CLTV ratio and the low fraction of loans with full documentation for vintage 2006 loans do not contribute much to the high observed delinquency rates for those loans.

We also computed the contributions of all explanatory factors for the other vintage years (not reported). For loans originated in 2003 and 2004, the high subsequent house price appreciation between 2003 and 2005 contributed to a lower actual delinquency rate. For example, the explained change in the delinquency rate 12 months after origination was –0.88 percentage points and –1.43 percentage points for 2003 and 2004, respectively. The house price appreciation variable had the largest (absolute) contribution among all variables considered for those years. Therefore, we can say that high house price appreciation between 2003 and 2005 masked the true riskiness of subprime mortgages.[14]

### 3.5 Adjusted Delinquency Rates

To examine to what extent the logit regression model is capable of explaining the large observed delinquency rates in 2006 and 2007, we plot the adjusted delinquency rates for different ages and different vintages in Figure 1 (right panel). The adjusted rate at a given age equals the prediction error (the actual rate minus the predicted rate) plus the weighted average rate over the 2001–2007 period, with weights equal to the number of loans originated in each year. The predicted delinquency rate is determined using Equation 6. We add up the weighted-average actual rates to facilitate the comparison with the actual rates plotted in Figure 1 (left panel). Interestingly, the adjusted delinquency rates have been increasing over the past seven years. In other words, loan quality deteriorated monotonically between 2001 and 2007. This picture is in sharp contrast with that obtained from actual rates, where 2003 was the year with the lowest delinquency rates, and 2001 was the year with the third-highest rates. In Subsection 3.3 we found a similar result: when adding a trend variable as explanatory variable, the associated regression coefficient implies a yearly increase of about 0.79 percentage points in the delinquency rate 12 months after origination. This amounts to a 4 to 5 percentage point increase over the 2001–2007 sample period that is due to the trend and thus not explained by the explanatory variables listed in column one of Table 3. The finding of a continual deterioration in loan performance also obtains when analyzing foreclosure rates (Appendix B), omitting terminated loans from the analysis (Appendix C), and analyzing hybrid mortgages and FRMs separately (Appendix D). Moreover, it obtains for the numerous alternative regression specifications discussed in Subsection 3.3 (not reported).

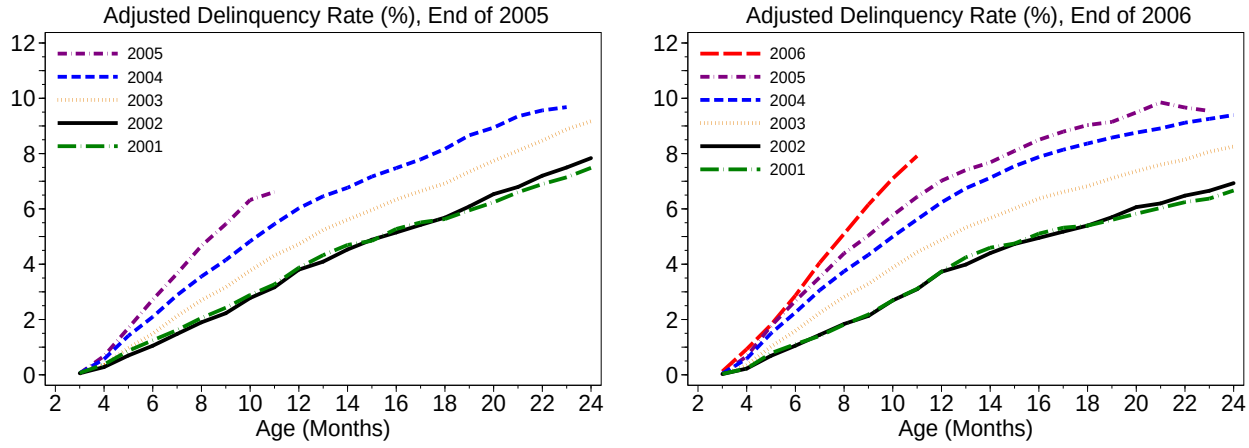
Next we study the following question: Based on information available at the end of 2005, was the dramatic deterioration of loan quality since 2001 already apparent? Notice that we cannot answer this

[14] Shiller (2007) argues that house prices were too high compared to fundamentals in this period and refers to the house price boom as a classic speculative bubble largely driven by an extravagant expectation for future house price appreciation.

question by simply inspecting vintages 2001 through 2005 in Figure 1 (right panel), because the computation of the adjusted delinquency rate for, say, vintage 2001 loans, makes use of a regression model estimated using data from 2001 through 2008. Hence, we re-estimate the logit regression model underlying Figure 1 (right panel) making use of only 2001–2005 data. The resulting age pattern in adjusted delinquency rates is plotted in Figure 5 (left panel). We again obtain the result that the adjusted delinquency rate rose monotonically from 2001. We therefore conclude that the dramatic deterioration of loan quality should have been apparent by the end of 2005. Figure 5 (right panel) depicts the situation when we use data available at the end of 2006. Again, the deterioration is clearly visible.[15]

Figure 5: Adjusted Delinquency Rate, Viewed at the End of 2005 and 2006. Delinquency is defined as being 60 or more days late with the monthly mortgage payment, in foreclosure, real-estate owned, or defaulted. The adjusted delinquency rate is obtained by adjusting the actual rate for year-by-year variation in FICO scores, loan-to-value ratios, debt-to-income ratios, missing debt-to-income ratio dummies, cash-out refinancing dummies, owner-occupation dummies, documentation levels, percentage of loans with prepayment penalties, mortgage rates, margins, house price appreciation since origination, composition of mortgage contract types, and origination amounts.

The figure shows the adjusted delinquency rate using data available at the end of 2005 (left panel) and 2006 (right panel).




[15]One reason why investors did not massively start to avoid or short subprime-related securities is that the timing of the subprime market downturn may have been hard to predict. Moreover, a short position is associated with a high cost of carry (Feldstein (2007)).

# 4 Non-Stationarity of the Loan-to-Value Effect

The logit regression specification used in Section 3 assumes that the regression coefficients are constant over time. That is, the effect of a unit change in an explanatory variable on the delinquency rate is the same in, for example, 2006 as it is in 2001, holding constant the values of the other explanatory variables. We test the validity of this assumption for all variables in our analysis by running cross-sectional OLS regressions for each calendar month from 2001 to 2006 and checking the stability of the regression coefficients. It turns out that the strongest rejection of a constant regression coefficient is for the CLTV ratio. In this section we first discuss this finding and then turn to the question of whether lenders were aware of the non-stationarity of the loan-to-value effect, by investigating the relationship between the loan-to-value ratio and mortgage rates over time.

## 4.1 Loan-to-Value Ratio and the Delinquency Rate

We consider three different CLTV value categories: CLTV<80%, CLTV=80%, and CLTV>80%, which account for about 28%, 15%, and 57% respectively of all loans originated in 2001–2007. Table 4 reports the actual minus the predicted delinquency rate for the different CLTV value categories and different vintage years, estimated using Equation 6. In other words, the table reports the average prediction error for the three CLTV subgroups discussed above and for each origination year of loans. A positive prediction error means that the actual delinquency rate was higher than the rate predicted by the logit regression model. Consistent with Figure 1 (right panel), the error increased over time. However, for the lowest CLTV group, the increase in the error was much smaller than that for the other groups and, in fact, had been fairly stable from 2004 onward. For a CLTV ratio of 80 percent, the increase in the error was 5.2 percentage points, and for the CLTV ratio above 80 percent, the increase was 8.8 percentage points. Therefore, high CLTV ratios were increasingly associated with higher delinquency rates, beyond what is captured by the logit regression model.

## 4.2 Loan-to-Value Ratio and the Mortgage Rate

The combined LTV ratio rather than the first-lien LTV ratio is believed to be the main determinant of delinquency, because it is the burden of all the debt together that may trigger financial problems for the borrower. In contrast, the first-lien LTV is the more important determinant of the mortgage rate on a

Table 4: Actual Minus Predicted Delinquency Rate

This table presents the actual minus the predicted delinquency rate 12 months after origination for different vintages and combined loan-to-value (CLTV) ratios.

| | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2007–2001 |
|---|---|---|---|---|---|---|---|---|
| $CLTV < 80\%$ | −2.2% | −2.1% | −0.7% | 0.4% | 0.2% | −0.5% | 0.5% | 2.8% |
| $CLTV = 80\%$ | −2.8% | −2.1% | −0.6% | 0.7% | 1.9% | 2.4% | 2.5% | 5.2% |
| $CLTV > 80\%$ | −4.6% | −3.4% | −1.4% | 0.1% | 0.1% | 1.6% | 4.2% | 8.8% |

first-lien mortgage, because it captures the dollar amount at stake for the first-lien lender.[16]

In this subsection we examine whether lenders were aware that high LTV ratios were increasingly associated with riskier borrowers. Specifically, we test whether the sensitivity of the lender's interest rate to the first-lien LTV ratio changed over time. We perform a cross-sectional OLS regression with the mortgage rate as the dependent variable and loan characteristics, including the first-lien LTV and second-lien LTV (CLTV minus first-lien LTV), as independent variables.[17] We perform one such regression for each calendar quarter in our sample period. We can only expect to get accurate results when using relatively homogeneous groups of loans, and therefore consider fully amortizing FRM and 2/28 hybrid loans separately. Together these two contract types account for about half of all mortgage loans in our database. Each cross-sectional regression is based on a minimum of 18,784 observations.

Figure 2 shows the regression coefficient on the first-lien LTV ratio for each quarter from 2001Q1 through 2007Q2.[18] We scaled the coefficients by the standard deviation of the first-lien LTV ratio, and they can therefore be interpreted as the changes in the mortgage rates when the first-lien LTV ratios are increased by one standard deviation. In the fourth quarter of 2006, a one-standard-deviation increase in the first-lien LTV ratio corresponded to about a 30-basis-point increase in the mortgage rate for 2/28 hybrids and about a 40-basis-point increase for FRMs, keeping constant other loan characteristics. In contrast, in the first quarter of 2001, the corresponding rate increase was 10 and 16 basis points respectively. This provides evidence that lenders were to some extent aware of high LTV ratios being increasingly associated

[16] This is confirmed by our empirical results. To conserve space the results are not reported.

[17] Specifically, we use the FICO score, first-lien loan-to-value ratio, second-lien loan-to-value ratio, debt-to-income ratio, a dummy for a missing debt-to-income ratio, a cash-out refinancing dummy, a dummy for owner occupation, documentation dummy, prepayment penalty dummy, margin, origination amount, term of the mortgage, and prepayment term as the right-hand-side variables.

[18] Our data extends to 2007Q3, but due to a near shutdown of the securitized subprime mortgage market we lack statistical power in this quarter.

with risky borrowers.[19] In Appendix E we show that this result is robustness to allowing for a non-linear relation between the mortgage rate and the first-lien LTV ratio. Finally, notice that the effect of a one-standard-deviation increase in the first-lien LTV ratio on the 2/28 mortgage rate increased substantially in the wake of the subprime mortgage crisis: from 30 basis points in 2007Q1 to 42 basis points in 2007Q2.

# 5 Subprime-Prime Rate Spread

In general, interest rates on subprime mortgages are higher than on prime mortgages to compensate the lender for the (additional) default risk associated with subprime loans. In this section we analyze the time series of the subprime-prime rate spread, both with and without adjustment for changes in loan and borrower characteristics. We focus on fixed-rate mortgages for this exercise. For hybrid mortgages the subprime-prime comparison is more complicated because (i) both the initial (teaser) rate and the margin should be factored in, and (ii) we don't have good data on the prime initial rates and margins.

In Figure 6 we show the actual subprime-prime rate spread, defined in the Equation (8) below. The subprime rate is from the LoanPerformance database, calculated for this exercise as the average across individual loans mortgage rate at the time of origination for each calendar month; the prime rate is the contract rate on fixed-rate mortgages reported by the Federal Housing Finance Board (FHFB) in its Monthly Interest Rate Survey.[20] The subprime-prime spread—the difference between the average subprime and the prime rates—decreased substantially over time, with the largest decline between 2001 and 2004, which coincides with the most rapid growth in the number of loans originated (see Table 1). In Figure 6 we also plot the yield spread between 10-year BBB and AAA corporate bonds, which we obtained from Standard and Poor's Global Fixed Income Research. Compared to the corporate BBB-AAA yield spread, the actual subprime-prime rate spread declined much more and more steadily, hence the decline cannot just be attributed to a change in the overall level of risk aversion.

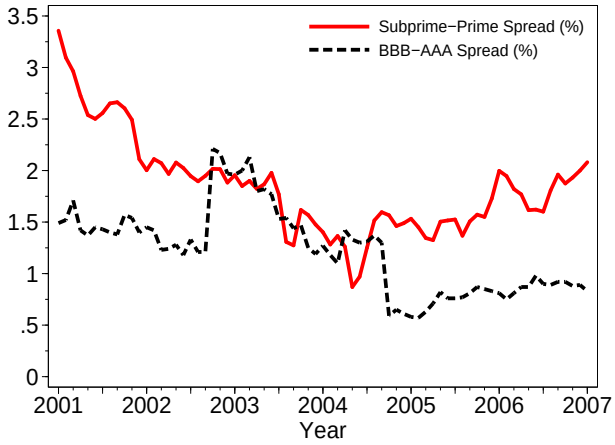
We perform a cross-sectional OLS regression with the loan-level spread as the dependent variable and the prime rate and various subprime loan and borrower characteristics as the explanatory variables, using

[19] The effects of other loan characteristics on mortgage rates have been much more stable over time, as unreported results suggest.

[20] Available at http://www.fhfb.gov/GetFile.aspx?FileID=6416.

Figure 6: FRM Rate Spread and Corporate Bond Yield Spread

The figure shows the FRM subprime-prime rate spread and the yield spread between 10-year BBB and AAA corporate bonds.




data from 2001 through 2006.[21]

$$spread_{it} = \beta_0 + \beta_1 prime_t + \beta_2' characteristics_{it} + error_{it}, \tag{7}$$

$$spread_{it} = subprime_{it} - prime_t \tag{8}$$

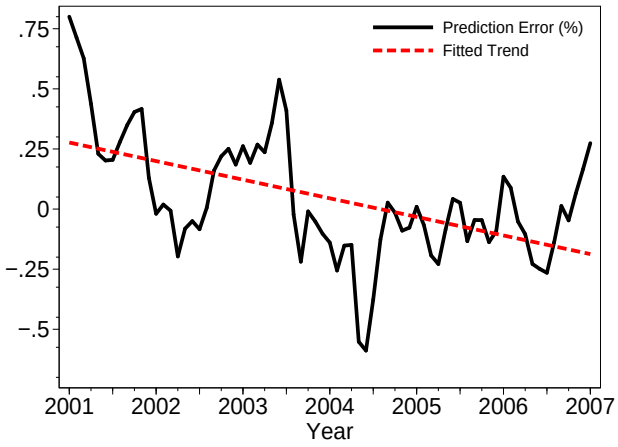
Notice that the $\beta_1 prime_t$ term corrects for the fact that the spread is affected by the prime rate itself, and thus changes over the business cycle, because a higher prime rate increases the default probability on subprime loans for a given spread. In Figure 7 we plot the prediction error, averaged per origination month $t$, along with a fitted linear trend.

The downward trend in Figure 7 indicates that the subprime-prime spread, after adjusting for differences in observed loan and borrower characteristics, declined. In Figure 1 (right panel) we showed that loan quality, obtained by adjusting loan performance for differences in loan and borrower characteristics and subsequent house price appreciation, deteriorated over the period, and thus the (adjusted) riskiness of loans rose. Therefore, on a per-unit-of-risk basis, the subprime-prime mortgage spread decreased even more than the level of the spread.

[21] The explanatory factors in the regression are the FICO credit score, a dummy variable that equals one if full documentation was provided, a dummy variable that equals one if prepayment penalty is present, origination amount, value of debt-to-income ratio, a dummy variable that equals one if debt-to-income was not provided, a dummy variable that equals one if loan is a refinancing, a dummy variable that equals one if a borrower is an investor, loan-to-value ratio based on a first-lien, and loan-to-value ratios based on a second, third, etc. liens if applicable.

Figure 7: Prediction Error in the Subprime-Prime Rate Spread

The figure shows the prediction error in the subprime-prime rate spread, determined in a regression of the spread on the prime rate and the following loan and borrower characteristics: FICO credit score, a dummy variable that equals one if full documentation was provided, a dummy variable that equals one if a prepayment penalty is present, origination amount, value of debt-to-income ratio, a dummy variable that equals one if debt-to-income was not provided, a dummy variable that equals one if the loan is a refinancing, a dummy variable that equals one if a borrower is an investor, loan-to-value ratio based on a first lien, and loan-to-value ratio based on a second, third, etc. liens if applicable.




# 6 Concluding Remarks

The subprime mortgage market experienced explosive growth between 2001 and 2006. Angell and Rowley (2006) and Kiff and Mills (2007), among others, argue that this was facilitated by the development of private-label mortgage backed securities, which do not carry any kind of credit risk protection by the Government Sponsored Enterprises. Investors in search of higher yields kept increasing their demand for private-label mortgage-backed securities, which also led to sharp increases in the subprime share of the mortgage market (from around 8 percent in 2001 to 20 percent in 2006) and in the securitized share of the subprime mortgage market (from 54 percent in 2001 to 75 percent in 2006).

In this paper we show that during the dramatic growth of the subprime (securitized) mortgage market, the quality of the market deteriorated dramatically. We analyze loan quality as the performance of loans, adjusted for differences in borrower characteristics (such as credit score, level of indebtedness, ability to provide documentation), loan characteristics (such as product type, amortization term, loan amount, interest rate), and subsequent house price appreciation.

The decline in loan quality has been monotonic, but not equally spread among different types of bor-

rowers. Over time, high-LTV borrowers became increasingly risky (their adjusted performance worsened more) compared to low-LTV borrowers. Securitizers seem to have been aware of this particular pattern in the relative riskiness of borrowers: We show that over time mortgage rates became more sensitive to the LTV ratio of borrowers. In 2001, for example, a borrower with a one standard deviation above-average LTV ratio paid a 10 basis point premium compared to an average LTV borrower. In contrast, in 2006 the premium paid by the high LTV borrower was around 30 basis point.

In principal, the subprime-prime mortgage rate spread (subprime mark-up) should account for the default risk of subprime loans. For the rapid growth of the subprime mortgage market to have been sustainable, the increase in the overall riskiness of subprime loans should have been accompanied by an increase in the subprime mark-up. In this paper we show that this was not the case: Subprime mark-up—adjusted and not adjusted for changes in differences in borrower and loan characteristics—declined over time. With the benefit of hindsight we now know that indeed this situation was not sustainable, and the subprime mortgage market experienced a severe crisis in 2007. In many respects, the subprime market experienced a classic lending boom-bust scenario with rapid market growth, loosening underwriting standards, deteriorating loan performance, and decreasing risk premiums.[22] Argentina in 1980, Chile in 1982, Sweden, Norway, and Finland in 1992, Mexico in 1994, Thailand, Indonesia, and Korea in 1997 all experienced the culmination of a boom-bust scenario, albeit in different economic settings.

Were problems in the subprime mortgage market apparent before the actual crisis showed signs in 2007? Our answer is yes, at least by the end of 2005. Using the data available only at the end of 2005, we show that the monotonic degradation of the subprime market was already apparent. Loan quality had been worsening for five years in a row at that point. Rapid appreciation in housing prices masked the deterioration in the subprime mortgage market and thus the true riskiness of subprime mortgage loans. When housing prices stopped climbing, the risk in the market became apparent.

[22] A more detailed discussion, theory, and empirical evidence on such episodes is available in Dell'Ariccia and Marquez (2006), Demirgüç-Kunt and Detragiache (2002), Gourinchas, Valdes, and Landerretche (2001), and Kamisky and Reinhart (1999), among many others.

# References


Angell, Cynthia, and Clare D. Rowley, 2006, Breaking new ground in U.S. mortgage lending, FDIC: Outlook Summer 2006 Federal Deposit Insurance Corporation.

Berger, Allen, and Gregory Udell, 2004, The institutional memory hypothesis and the procyclicality of bank lending behavior, *Journal of Financial Intermediation* 12, 458–495.

Bernanke, Ben, 2007, Housing, housing finance, and monetary policy, Speech Federal Reserve Bank of Kansas City's Economic Symposium.

Brunnermeier, Markus, and Christian Julliard, 2007, Money illusion and housing frenzies, *Review of Financial Studies* forthcoming.

Campbell, John Y., and João F. Cocco, 2003, Household risk management and optimal mortgage choice, *Quarterly Journal of Economics* 118, 1449–1494.

Cutts, Amy Crews, and Robert A. Van Order, 2005, On the economics of subprime lending, *The Journal of Real Estate Finance and Economics* 30, 167–196.

DellAriccia, Giovanni, Deniz Igan, and Luc Laeven, 2008, Credit booms and lending standards: Evidence from the subprime mortgage market, Working paper.

Dell'Ariccia, Giovanni, and Robert Marquez, 2006, Lending booms and lending standards, *Journal of Finance* 61, 2511–2546.

Demirgüç-Kunt, Asli, and Enrica Detragiache, 2002, Does deposit insurance increase banking system stability? An empirical investigation, *Journal of Monetary Economics* 49, 1373–1406.

Deng, Yongheng, John M. Quigley, and Robert A. Van Order, 2000, Mortgage terminations, heterogeneity and the exercise of mortgage options, *Econometrica* 68, 275–308.

Feldstein, Martin S., 2007, Housing, credit markets and the business cycle, NBER working paper 13471.

Gabaix, Xavier, Arvind Krishnamurthy, and Olivier Vigneron, 2007, Limits of arbitrage: Theory and evidence from the mortgage-backed securities market, *Journal of Finance* 62, 557–596.

Genesove, David, and Christopher Mayer, 1997, Equity and time to sale in the real estate market, *American Economic Review* 87, 255–269.

Genesove, David, and Christopher Mayer, 2001, Loss aversion and seller behavior: Evidence from the housing market, *Quarterly Journal of Economics* 116, 1233–1260.

Gerardi, Kristopher, Adam Shapiro, and Paul Willen, 2008, Subprime outcomes: Risky mortgages, homeownership experiences, and foreclosures, Working paper.

Glaeser, Edward, and Joseph Gyourko, 2005, Urban decline and durable housing, *Journal of Political Economy* 113, 345–375.

Gourinchas, Pierre-Oliver, Rodrigo Valdes, and Oscar Landerretche, 2001, Lending booms: Latin America and the world, *Economia* 1, 47–99.

Gyourko, Joseph, Mayer Christopher, and Todd Sinai, 2006, Superstar cities, NBER working paper 12355.

Kamisky, Graciela L., and Carmen M. Reinhart, 1999, The twin crises: The causes of banking and balance-of-payment problems, *American Economic Review* 89, 473–500.

Keys, Benjamin J., Tanmoy Mukherjee, Amit Seru, and Vikrant Vig, 2008, Did securitization lead to lax screening? evidence from subprime loans, Working paper.

Kiff, John, and Paul Mills, 2007, Money for nothing and checks for free: Recent developments in U.S. subprime mortgage markets, IMF working paper.

Koijen, Ralph, Otto Van Hemert, and Stijn Van Nieuwerburgh, 2007, Mortgage timing, NBER working paper 13361.

LaCour-Little, Michael, 2007, The home purchase mortgage preferences of low-and-moderate income households, *Real Estate Economics* forthcoming.

Mian, Atif, and Amir Sufi, 2008, The consequences of mortgage credit expansion: Evidence from the 2007 mortgage default crisis, Working paper.

Pennington-Cross, Anthony, and Souphala Chomsisengphet, 2007, Subprime refinancing: Equity extraction and mortgage termination, *Real Estate Economics* 35, 233–263.

Ruckes, Martin, 2004, Bank competition and credit standards, review of financial studies, *Review of Financial Studies* 17, 1073–1102.

Shiller, Robert J., 2007, Understanding recent trends in house prices and home ownership, NBER working paper 13553.

Van Hemert, Otto, 2007, Household interest rate risk management, Working paper.

Vickery, James, 2007, Interest rates and consumer choice in the residential mortgage market, Working paper.

Von Furstenberg, George M., and R. Jeffery Green, 1974, Home mortgage delinquencies: A cohort analysis, *Journal of Finance* 29, 1545–1548.

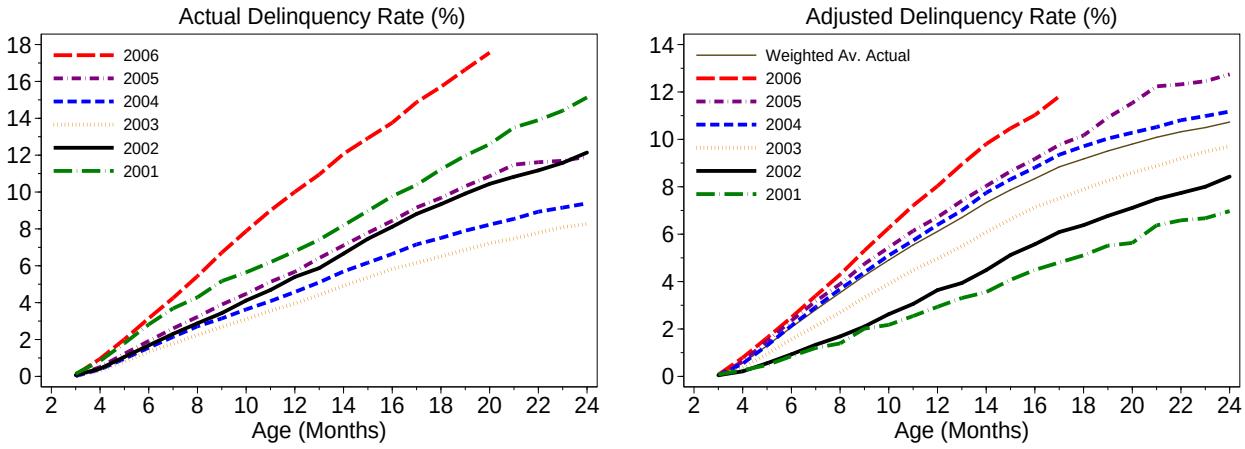
# A Reclassification of Loan-Level Data

The main database we utilize in this paper is the loan-level data on US securitized subprime mortgage loans, provided by First American LoanPerformance. Since the first version of this paper, October 9th 2007, LoanPerformance retroactively reclassified a large number of negative amortization loans from subprime to Alt-A. This reclassification was requested by the clients of the trustees of the mortgage pools. Most of the reclassification occurred over the first months of 2008. Analyzing the data before and after the reclassification we find that the reclassified loans had on average a larger balance and performed better in terms of lower delinquency rates. We checked extensively with LoanPerformance that the set of loans used in the current version of the paper corresponds to the intended post-reclassification data, as of May 2008.

It is not clear whether the data before or after reclassification is most suitable for the purpose of researching the subprime crisis that occurred before the reclassification. It is however reassuring that our main results are robust to the reclassification. To illustrate, Figure 8 shows the actual (left panel) and adjusted (right panel) delinquency rates as reported in earlier versions of this paper. It reflects the situation before the reclassification. Notice in the right panel that again the adjusted delinquency rate has been increasing monotonically over the past six years.

Figure 8: Pre-Reclassification Actual and Adjusted Delinquency Rate

The figure shows the age pattern in the actual (left panel) and adjusted (right panel) delinquency rate for the different vintage years prior to the reclassification of loans by LoanPerformance. Delinquency for this figure is defined as being 60 days or more late with the monthly mortgage payment, in foreclosure, or real-estate owned. The adjusted delinquency rate is obtained by adjusting the actual rate for year-by-year variation in FICO scores, loan-to-value ratios, debt-to-income ratios, missing debt-to-income ratio dummies, cash-out refinancing dummies, owner-occupation dummies, documentation levels, percentage of loans with prepayment penalties, mortgage rates, margins, house price appreciation since origination, composition of mortgage contract types, and origination amounts.

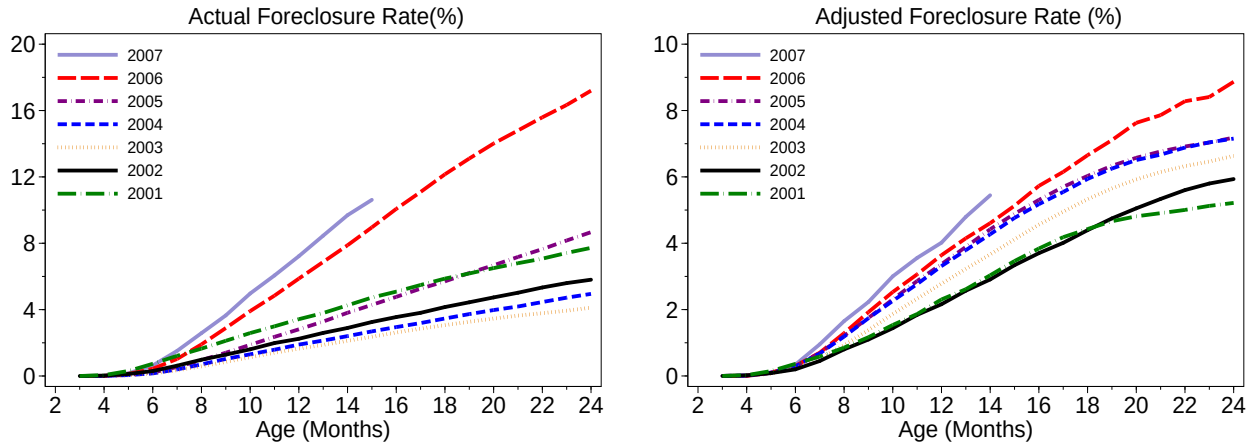


# B Foreclosure Rates

In this Appendix we show that the continual deterioration of adjusted loan performance over the 2001–2007 period is robust to using foreclosure, instead of delinquency, as measure for loan performance. Foreclosure is defined as a loan being in foreclosure, real-estate owned, or defaulted. In Figure 9 we present actual (left panel) and adjusted (right panel) foreclosure rates. The actual foreclosure rate for age six months and younger is close to zero, in contrast to the actual delinquency rate at this age, presented in Figure 1 (left panel). For older ages the actual foreclosure rate is roughly speaking twice as low as the actual delinquency rate. Notice that besides these two differences between actual delinquency and foreclosure rates, the age patterns for the different vintages is remarkably similar for delinquency and foreclosure. In particular, the ranking of the vintage years in terms of loan performance is the same. The adjusted foreclosure rates presented in Figure 9 (right panel) are again increasing in vintage year, but different vintages are relatively closer in terms of adjusted performance than for our analysis of adjusted delinquency rates, Figure 1 (right panel).

Figure 9: Actual and Adjusted Foreclosure Rates

The figure shows the age pattern in the actual (left panel) and adjusted (right panel) foreclosure rate for the different vintage years. Foreclosure is defined as a loan being in foreclosure, real-estate owned, or defaulted. The adjusted foreclosure rate is obtained by adjusting the actual rate for year-by-year variation in FICO scores, loan-to-value ratios, debt-to-income ratios, missing debt-to-income ratio dummies, cash-out refinancing dummies, owner-occupation dummies, documentation levels, percentage of loans with prepayment penalties, mortgage rates, margins, house price appreciation since origination, composition of mortgage contract types, and origination amounts.




# C Delinquency Rate without Terminated Loans

For the regressions in Section 3 we used the probability that a loan of a particular age is delinquent as dependent variable. For this exercise we included terminated loans and subsumed loans terminated due to default in our definition of delinquency. In certain cases the inference whether a terminated loan is a defaulted or prepaid loan was based on whether the loan was in

foreclosure or not the month prior to termination. In this appendix we confirm that our main results are robust to excluding all terminated loans and thus examine delinquency conditional on the loan not being terminated before the age of interest.

Since the focus in this paper is on young loans (age below 24 months), we do not expect terminations to play an important role. The fraction of loans that terminate in the first 12 months is 10%, 11%, 15%, 18%, 16%, 12%, 7% for 2001–2007 vintage loans respectively. This fraction peaks in 2004, during the heyday of the house price boom, which created favorable refinancing opportunities. Remarkably, it is relatively low for vintage 2001 loans, which were originated prior to interest rate declines. This result might be particular for loans of age below 24 months. For hybrid mortgage loans many refinances may occur at or after the moment of the rate reset, which is typically at a loan age of 24 or 36 months.[23]

We re-estimate the results presented in Table 3—analyzing the effects of loan characteristics, borrower characteristics, and economic circumstances on the probability of delinquency—but this time exclude loans that are terminated prior to the moment of loan performance evaluation.

Using this alternative definition for delinquency, all explanatory variables we use in the regression (see Table 3) have the same sign and remain statistically significant, except for the ARM dummy. The mortgage rate, FICO score, the CLTV ratio, and the subsequent house price appreciation remain the most important explanatory variables, measured by the absolute size of the marginal effect.

In Figure 10 we plot the actual delinquency rate (left panel) and adjusted delinquency rate (right panel) based on the definition above that excludes terminated loans. In general the delinquency rate is somewhat higher than based on the baseline case definition of delinquency, in particular at older loan ages, see Figure 1 (both panels). Excluding terminations due to a prepayment will increase the delinquency rate compared to the baseline case. Excluding terminations due to a default have the opposite effect. Hence we find that the prepayment effect dominates the default effect. Importantly, both for the actual and the adjusted delinquency rate the order of the lines for the different vintage years remain unchanged. We again have the result that the adjusted delinquency rate has been steadily rising over the past seven years.

# D Adjusted Delinquency Rate for Hybrids and FRMs Separately

In this Appendix we show that the continual deterioration of adjusted loan performance over the 2001–2007 period also obtains when performing a separate regression analysis for the main contract types, as opposed to the baseline case in the main text where we perform a regression for all loans, but include contract type dummies in the regression specification. Figure 11 shows the adjusted delinquency rate for the two main contract types: 2/28 hybrids and FRMs. For both contract types, the adjusted delinquency rates have been monotonically increasing over time. Except for a level difference, the age pattern for the different vintage years looks very much the same for the two contract types.

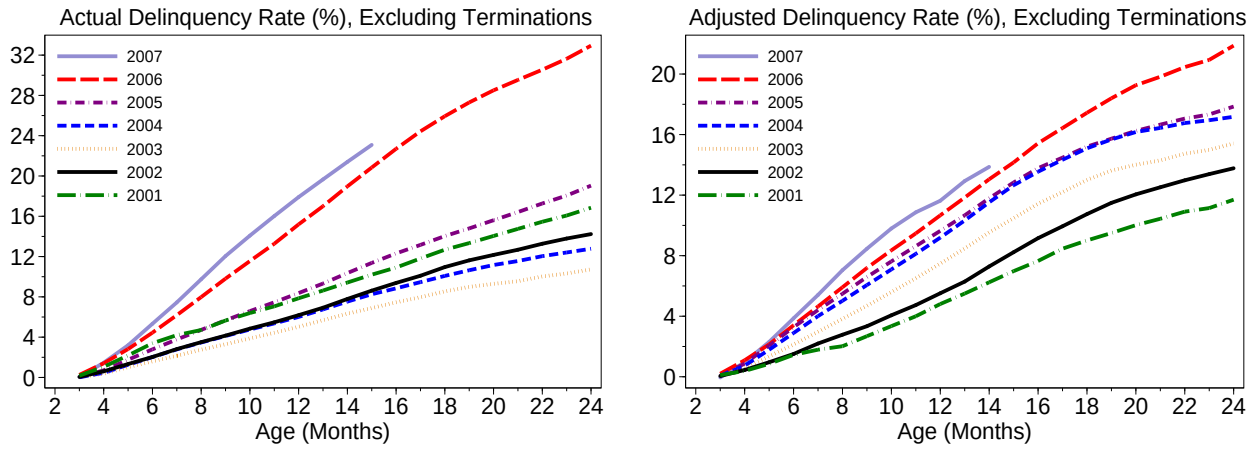
# E Non-Linearity in the Sensitivity of the Mortgage Rate to the LTV

In Figure 2 we plotted the sensitivity of the fixed-rate and 2/28 hybrid mortgage rates to the first-lien LTV ratio. The sensitivity is defined as the regression coefficient on the first-lien LTV (scaled by the standard deviation) in a regression

[23] Gabaix, Krishnamurthy, and Vigneron (2007) argue that the risk associated with mortgage prepayments is priced in the mortgage-backed securities market.

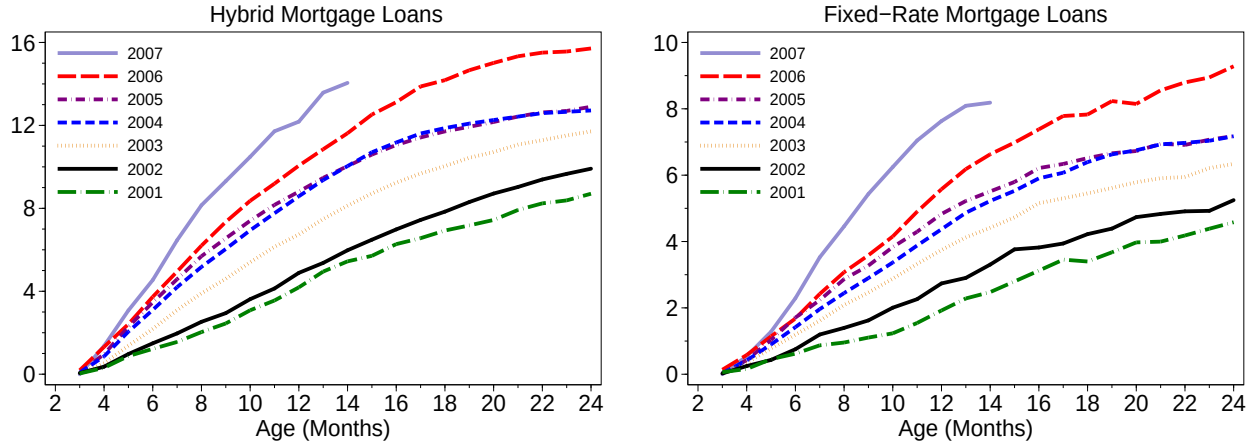
## Figure 10: Actual and Adjusted Delinquency Rates, Excluding Terminated Loans

The figure shows the actual delinquency rates (left panel) and adjusted delinquency rates (right panel) excluding terminated (prepaid and defaulted) loans for the different vintage years. Delinquency is defined as being 60 days or more late with the monthly mortgage payment, in foreclosure, real-estate owned or defaulted. The adjusted delinquency rate is obtained by adjusting the actual rate for year-by-year variation in FICO scores, loan-to-value ratios, debt-to-income ratios, missing debt-to-income ratio dummies, cash-out refinancing dummies, owner-occupation dummies, documentation levels, percentage of loans with prepayment penalties, mortgage rates, margins, house price appreciation since origination, composition of mortgage contract types, and origination amounts.




## Figure 11: Adjusted Delinquency Rates for Hybrids and FRMs Separately

The figure shows the adjusted delinquency rates based on hybrid mortgages (left panel) and FRMs (right panel) separately. Delinquency is defined as being 60 days or more late with the monthly mortgage payment, in foreclosure, real-estate owned or defaulted. The adjusted delinquency rate is obtained by adjusting the actual rate for year-by-year variation in FICO scores, loan-to-value ratios, debt-to-income ratios, missing debt-to-income ratio dummies, cash-out refinancing dummies, owner-occupation dummies, documentation levels, percentage of loans with prepayment penalties, mortgage rates, margins, house price appreciation since origination, and origination amounts.

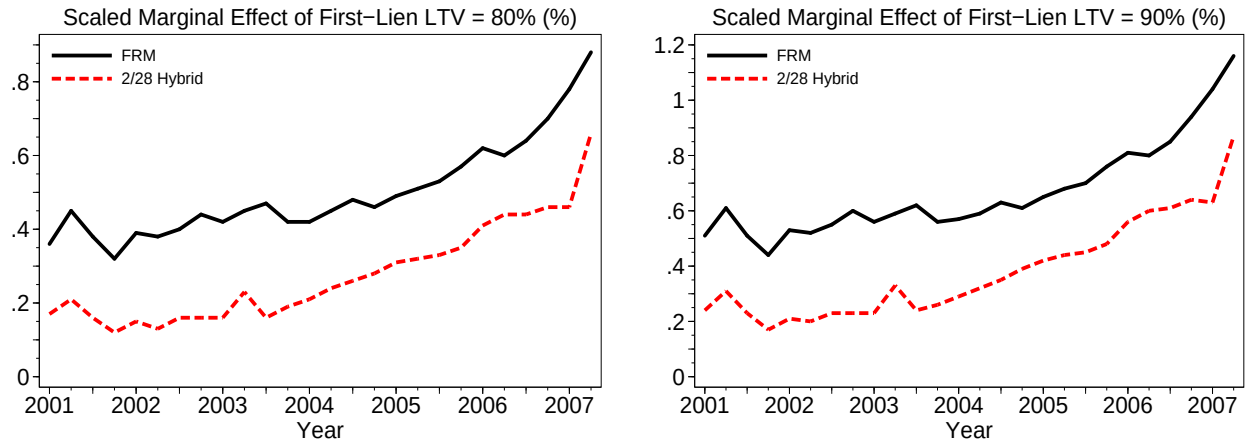


with the mortgage rate as dependent variable and the first-lien LTV, the second-lien LTV, and the other loan and borrower characteristics listed in Subsection 4.2, as independent variables.

In this appendix we study the robustness of this result to adding the square of the first-lien LTV and the square of the second-lien LTV as independent variables, therefore allowing for a non-linear functional form. In Figure 12 we report the resulting scaled marginal effect of the first-lien LTV for fixed-rate and 2/28 hybrid mortgages evaluated at a first-lien LTV of 80 percent (left panel) and 90 percent (right panel). Without non-linear terms the marginal effect is simply given by the regression coefficient. This is what we plotted in Figure 2. With the quadratic terms, the marginal effect is given by $\beta_{LTV} + 2\beta_{LTV^2}X$, where the $\beta$s are the regression coefficients and $X$ is the first-lien LTV ratio at which the marginal effect is evaluated.

Figure 12: Sensitivity of Mortgage Rate to First-Lien LTV Ratio Allowing for Non-Linearity

The figure shows the scaled marginal effect of the first-lien loan-to-value (LTV) ratio on the mortgage rate for first-lien fixed-rate and 2/28 hybrid mortgages, evaluated at a first-lien LTV of 80% (left panel) and 90% (right panel). The effect is determined using an OLS regression with the interest rate as dependent variable and the FICO score, first-lien LTV (and the square), second-lien LTV (and the square), debt-to-income ratio, missing debt-to-income ratio dummy, cash-out refinancing dummy, owner-occupation dummy, prepayment penalty dummy, origination amount, term of the mortgage, prepayment term, and margin as independent variables.




As shown in Figure 12, the marginal effect is rising over time, consistent with the baseline case results presented in Figure 2. Moreover, we find that there is a statistically and economically significant non-linear effect of the first-lien LTV on the mortgage rate. Comparing the left and right panels in Figure 12, the higher the first-lien LTV ratio, the more sensitive is the mortgage rate to changes in the first-lien LTV. The largest difference between the results based on specifications with and without non-linearity is observed for 2/28 hybrid mortgages in 2007 at a first-lien LTV of 90 percent (right panel). The scaled marginal effect increases by 27 basis points over the course of 3 months in 2007 when a model allows for non-linearity. In contrast, ignoring the non-linearity, as in Figure 2, the increase in the scaled marginal effect is only 13 basis points.