

Understanding the Subprime Mortgage Crisis*

Yuliya Demyanyk

FRB St. Louis

Otto Van Hemert

NYU Stern

This Draft: October 9, 2007

First Draft: October 9, 2007

Abstract

We analyze the subprime mortgage crisis: an unusually large fraction of subprime mortgages originated in 2006 being delinquent or in foreclosure only months later. We utilize a loan-level database, covering about half of all US subprime mortgages, and identify two major causes. First, over the past five years, high loan-to-value borrowers increasingly became high-risk borrowers, in terms of elevated delinquency and foreclosure rates. Lenders were aware of this and adjusted mortgage rates accordingly over time. Second, the below-average house price appreciation in 2006-2007 further contributed to the crisis.

*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 Joost Driessen, William Emmons, Ralph Koijen, and Andrew Meyer. 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.

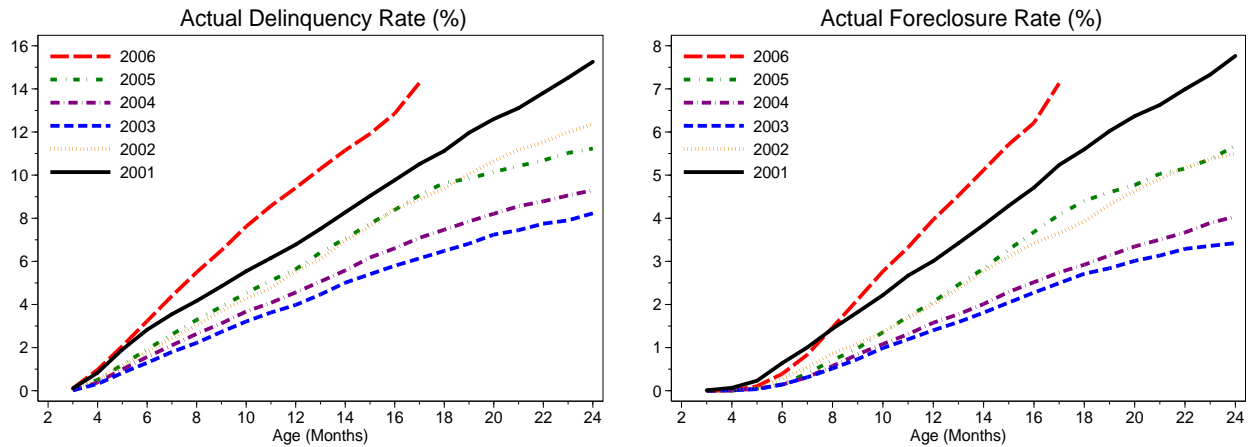
1 Introduction

The current subprime mortgage crisis, an unusually large fraction of subprime mortgages originated in 2006 being delinquent or in foreclosure only months later, has spurred massive media attention. Many different causes have been suggested. The goal of this paper is to “let the data speak” for revealing the causes of the crisis. To this end we utilize a loan-level database containing information about a one half of all US subprime mortgages originated between 2001 and 2006.

The relatively bad performance of vintage 2006 loans is illustrated in Figure 1. At every mortgage loan age, loans originated in 2006 show a higher delinquency rate (left panel) and a higher foreclosure rate (right panel) than loans originated in earlier years at the same ages. Noteworthy, 2001 was a fairly bad vintage year as well, ranking second, after 2006, both in terms of the delinquency and the foreclosure rate.

Figure 1: Actual Delinquency and Foreclosure Rate.

The figure shows the age pattern in the actual delinquency and foreclosure rate for the different vintages. Delinquency is defined as being 60+ days late with the monthly mortgage payment, in foreclosure, or real-estate owned. Foreclosure is defined as being in foreclosure or real-estate owned.



We document that the bad performance of the vintage 2006 loans was not confined to a particular segment of the subprime mortgage market. For example, fixed-rate, adjustable-rate, purchase-money, cash-out refinancing, low-documentation, and full-documentation loans all showed substantially higher delinquency and foreclosure rates for vintage 2006 loans compared to earlier vintages. This contrasts a widely-held believe that the subprime mortgage crisis is particular only to the adjustable-rate or low-documentation mortgages.¹

¹E.g., Bernanke (2007), Chairman of the Federal Reserve System, stated that “For subprime mortgages with fixed rather than variable rates, for example, serious delinquencies have been fairly stable at about 5-1/2 percent.” See the

We explore to what extent the subprime mortgage crisis can be attributed to different loan characteristics, borrower characteristics, and subsequent house price appreciation for vintage 2006 loans. For the empirical analysis, we run logit regressions with the probability of either delinquency or foreclosure being a function of the loan, borrower, and economic 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 compared to the earlier years, they cannot explain the unusually bad performance of vintage 2006 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%. However, because the average debt-to-income ratio was only 0.15 standard deviations higher in 2006 compared to previous years, it hardly contributes to explaining the bad performance of vintage 2006 loans. The only variable in the logit regression model that substantially contributed to the crisis is the low subsequent house price appreciation for vintage 2006 loans, which can explain about a 1% higher-than-average delinquency rate 12 months after origination.

We compute the prediction error as the difference between the actual delinquency or foreclosure rate and the estimated probability based on the logit regression model. In Figure 2 we plot the corrected delinquency (left panel) and corrected foreclosure (right panel) 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) and corrected (Figure 2) delinquency and foreclosure rates. As shown in Figure 2, the corrected delinquency and foreclosure rates have been steadily rising for the past six years, with no particularly large innovation from 2005 to 2006. Interestingly, vintage 2001 is the second-worst vintage year in terms of the actual delinquency and foreclosure rates, but is in fact the best vintage year in terms of the corrected rates. The high interest rate and low average credit score of borrowers have contributed to the high actual, but the low corrected rate in 2001.

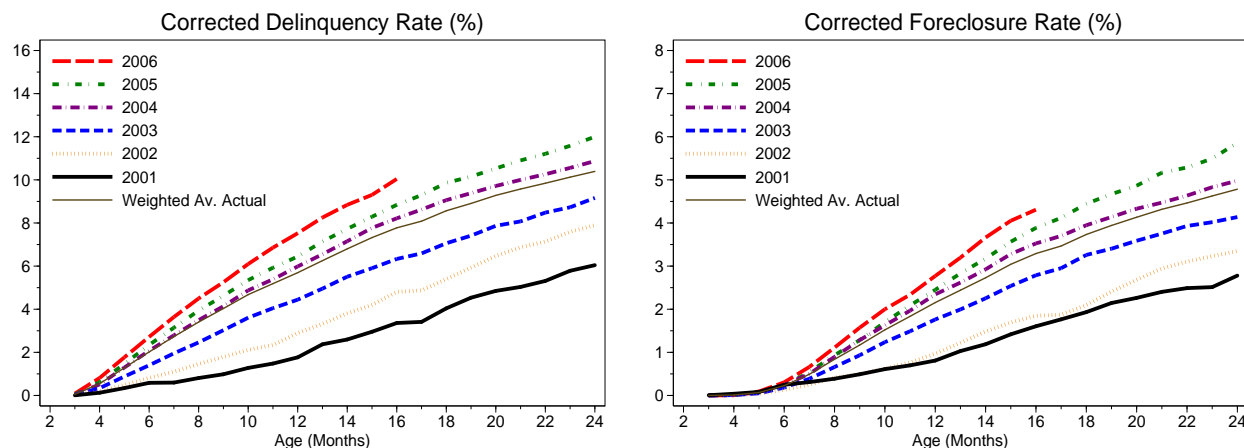
The logit regression specification used to compute corrected delinquency and foreclosure rates assumes that the regression coefficient on the different explanatory variables remains constant over time. We test the validity of this assumption and find that in particular for the loan-to-value (LTV) ratio this assumption is rejected by the data. High-LTV borrowers in 2006 are much riskier than were those in 2001 in terms of the probability of delinquency or foreclosure, for given values of the other explanatory variables. In fact, the rise in the corrected delinquency and foreclosure rate is almost exclusively caused by the worsening performance of loans with an LTV of 80% or greater.

Were lenders aware of the increasing riskiness of high-LTV borrowers? To answer this question, we run a cross-sectional OLS regression with the mortgage rate as a dependent variable, and loan

August 31, 2007 speech, available at <http://www.federalreserve.gov/newsevents/speech/Bernanke20070831a.htm>.

Figure 2: Corrected Delinquency and Foreclosure Rate.

The Figure shows the age pattern in the delinquency rate (left panel) and foreclosure rate (right panel) for the different vintages, after controlling for year-by-year variation in the average FICO score, loan-to-value ratio, debt-to-income ratio, documentation level, percentage loans with prepayment penalty, initial rate, margin, house price appreciation since origination, composition of mortgage contract type, and origination amount.



and borrower characteristics as the independent variables. We run one regression of this type for each quarter of 2001 through 2006 and three different contract types. Figure 3 shows the coefficient on the LTV variable scaled by the standard deviation of the LTV ratio. A striking pattern emerges. In 2001 lenders hardly charged a higher mortgage rate for the high-LTV borrowers. In 2006, a borrower with a one standard deviation above-average LTV ratio was charged a 0.3% higher interest rate. We thus find clear evidence that lenders were aware of the increasing riskiness of high-LTV borrowers, and adjusted mortgage rates accordingly.

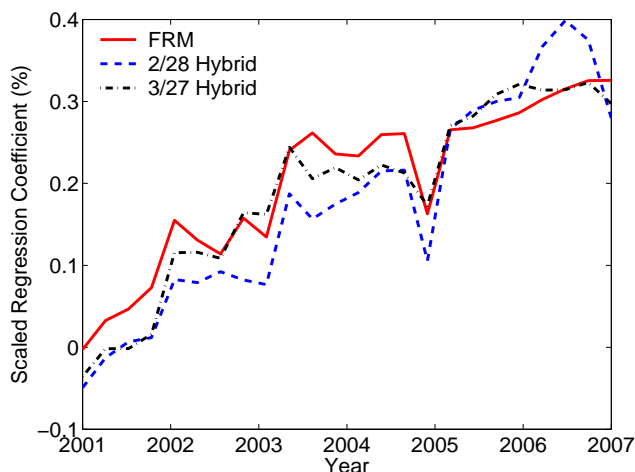
There is a large literature studying the determinants of mortgage delinquencies and foreclosures, dating back to at least von Furstenberg and Green (1974). Recent contributions include Cutts and van Order (2005) and Pennington-Cross and Chomsisengphet (2007).² To the best of our knowledge we are the first to quantify how much the different determinants have contributed to the observed high delinquency and foreclosure rates for vintage 2006 loans. In addition, we uncover a trend in the delinquency and foreclosure rate, after correcting for differences in loan characteristics, borrower characteristics, and the subsequent house price appreciation. Finally, we relate the rise in corrected delinquency and foreclosure rates to the increasing riskiness of high-LTV borrowers.

The structure of this paper is the following. In section 2 we show the descriptive statistics for the subprime mortgage market. In section 3 we present the results from the logit regression

²Deng, 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.

Figure 3: Effect Loan-to-Value Ratio on Mortgage Rate.

The Figure shows the effect of the loan-to-value ratio on the mortgage rate. The effect is defined as the regression coefficient on the loan-to-value ratio (scaled by the standard deviation) in an ordinary least squares regression with the mortgage rate as dependent variable and as independent variables the FICO score, loan-to-value ratio, debt-to-income ratio, a dummy whether the debt-to-income ratio is missing, documentation dummy, prepayment penalty dummy, margin, origination amount, term of the mortgage, and prepayment term.



specification. In section 4 we discuss the increasing riskiness of high-LTV borrowers, and the extent to which lenders were aware of this. In section 5 we conclude.

2 Descriptive Statistics

In this paper we use a loan-level data set that covers about half of the US subprime mortgage market.³ We exclude Alt-A loans, which are generally considered to be less risky because only borrowers with a good credit history, usually reflected in a high credit score, can obtain such a loan. We focus on first-lien loans. We consider the 2001 to 2006 sample period.

We first discuss the main characteristics of the loans in our database at origination. Second, we discuss the delinquency rate on 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, originated between 2001 and 2006. The subprime mortgage market has grown dramatically since 2001. In the first block of Table 1 we see that the number of originated loans in each year more than

³Our dataset contains around 85% of all securitized subprime mortgages, which in turn make up about 60% of the total subprime mortgage market.

quadrupled, and the average loan size almost doubled over the sample period. The total dollar amount originated in 2001 was \$94 billion. In 2006 it was \$685 billion.

Table 1: Loan Characteristics at Origination for Different Vintages

Descriptive statistics for the first-lien subprime loans in the LoanPerformance database. We do not report other mortgage types, which amount to less than 0.1%.

	2001	2002	2003	2004	2005	2006
	<i>Size</i>					
Number of Loans (*1000)	624	974	1676	2743	3440	2646
Average Loan Size (*\$1000)	151	168	180	201	234	259
	<i>Mortgage Type</i>					
FRM (%)	41.4	39.9	43.3	28.2	25.1	26.1
ARM (%)	0.9	1.9	1.3	4.3	10.3	12.8
Hybrid (%)	52.2	55.9	54.7	67.3	62.0	46.2
Balloon (%)	5.5	2.2	0.8	0.2	2.6	14.9
	<i>Loan Purpose</i>					
Purchase (%)	35.1	33.9	32.9	42.0	45.7	45.4
Refinancing (cash out) (%)	52.1	51.2	51.6	47.9	45.7	44.8
Refinancing (no cash out) (%)	12.3	14.6	15.1	10.0	8.6	9.8
	<i>Variable Means</i>					
FICO Score	620.1	630.5	641.4	645.9	653.7	654.7
Loan-to-Value Ratio (%)	79.3	79.4	79.2	79.3	78.5	78.3
Debt-to-Income Ratio (%)	37.8	38.1	38.2	38.5	39.1	39.8
Missing Debt-to-Income Ratio Dummy (%)	41.6	44.1	38.3	35.1	39.2	31.7
Documentation Dummy (%)	68.5	63.4	59.8	57.2	51.8	44.7
Prepayment Penalty Dummy (%)	66.3	63.8	61.4	60.1	60.6	61.6
Initial Rate (%)	9.4	8.3	7.3	6.7	6.6	7.2
Margin for ARM and Hybrid Mortgage Loans (%)	6.2	6.3	5.9	5.3	5.0	4.9

In the second block of Table 1, we split the pool of mortgages into four main mortgage contract types. Most popular are hybrid mortgages, accounting for about half of all subprime loans. The fixed-rate mortgage contract has become less popular over time and accounts for 26% of the total number of loans in 2006. In contrast, in the prime mortgage market, the majority of mortgage loans is of the fixed-rate type.⁴ Balloon mortgage contracts have become more common recently,

⁴For 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 their Monthly

and accounted for 15% of the total number of mortgages originated in 2006. A balloon mortgage does not fully amortize over the term of the loan and thus involves a large final (balloon) payment. Around 13% of the mortgages originated in 2006 were adjustable-rate mortgages.

In the third block of Table 1, we report the purpose of a mortgage loan. In about 30 to 45% of the cases the purpose is to finance the purchase of a house. In about 45 to 55% of the cases, the purpose is to extract cash by refinancing an existing mortgage loan into a larger new mortgage loan. Cash-out refinancing loans were particularly prevalent in 2001-2003. 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 means 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 has been improving over the sample period. The loan-to-value ratio, debt-to-income ratio (if provided), and fraction of loans with a prepayment penalty have been fairly constant. For about a third of all loans no debt-to-income ratio was provided, as captured by the missing debt-to-income ratio dummy variable. The share of loans with full documentation substantially decreased from 69% in 2001 to 45% in 2006.

The initial rate sharply decreased from 2001 to 2004, consistent with sharp declines in both the 1-year and 10-year Treasury yields over the same period. Between 2004 and 2006, both the 1-year and 10-year Treasury yield increased again, but the initial rate on subprime mortgages, remarkably, remained fairly constant. Also the margin (over a reference rate, often the 6-month LIBOR) for adjustable-rate and hybrid mortgages declined from 2001 to 2004 and remained fairly constant afterwards.

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

2.2 Empirical Delinquency Rates

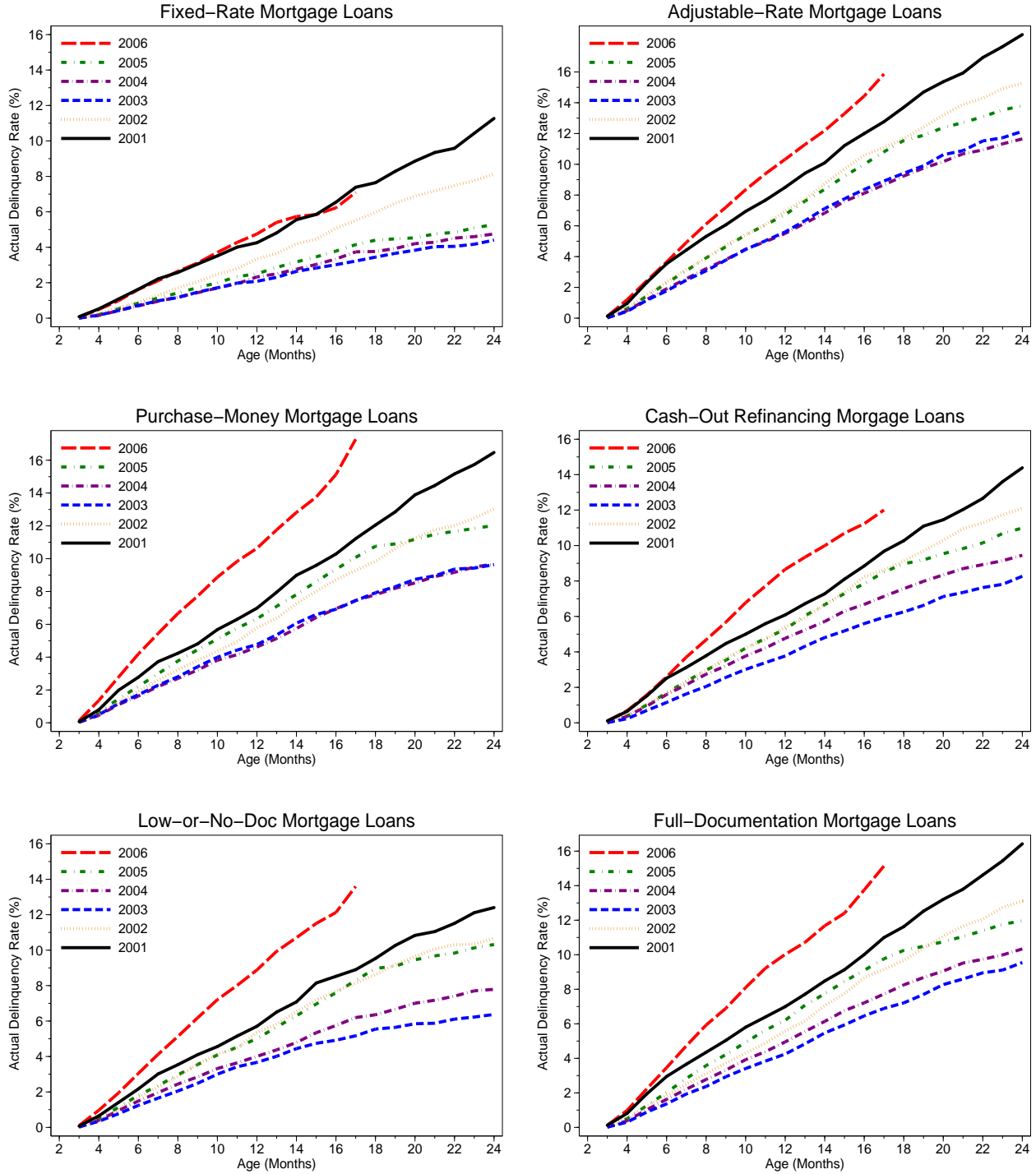
We define a loan to be delinquent if payments on the loan are more than 60 days late or the loan is in foreclosure (see also Table 2). In Figure 1 we show that for the subprime market as a whole the vintage 2006 stands out in terms of high delinquencies and foreclosures. In Figure 4, we again plot the age pattern in the delinquency rate for vintages 2001 through 2006, but split the subprime mortgage market into various segments. As the figure shows, the bad performance of the 2006 vintage is not confined to a particular segment of the subprime market, but rather reflects a market-wide phenomenon.

In the upper-right panel, we see that for adjustable-rate mortgages, vintage 2006 clearly stands out with higher delinquency rates; 2001 being the second-worst vintage year. In the upper-left panel of Figure 4 we see that generally the delinquency rate on fixed-rate mortgages is lower than

Interest Rate Survey is around 70% in 2006.

Figure 4: Actual Delinquency Rate for Segments of the Subprime Mortgage Market.

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



on adjustable-rate mortgages. However, also for FRMs, the delinquency rate for vintage 2006 loans is substantially higher than for vintage 2002-2005. This result therefore contrasts recent remarks by the Chairman of the Federal Reserve System, Bernanke (2007), that “For subprime mortgages with fixed rather than variable rates, for example, serious delinquencies have been fairly stable at about 5-1/2 percent.”

In the other four panels of Figure 4 we see that also for purchase-money, cash-out refinancing, low-documentation, and full-documentation mortgage loans, vintage 2006 showed the highest delinquency age pattern. In general, vintage 2001 loans show the second-highest delinquency rates, and vintage 2003 the lowest delinquency rates.

3 Main Regression Analysis

In this section we investigate to what extent a logit regression model can explain the the high level of delinquencies and foreclosures for the vintage 2006 mortgage loans. The regression coefficients are assumed to be constant over time, in order not to trivially explain the time variation in the delinquency and foreclosure rates. All results in this section will be based on a random sample of one million first-lien subprime mortgage loans, originated between 2001 and 2006.

3.1 Regression Specification

We run the following regression

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

where the event is either delinquency or foreclosure of the 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, including a constant, and β is the vector of regression coefficients. We will report the following statistics for variable i :

$$marginal_i = \Phi(\beta' \bar{X} + \beta_i \sigma_i) - \Phi(\beta' \bar{X}) \tag{2}$$

$$deviation06_i = (\overline{X06}_i - \bar{X}_i) / \sigma_i \tag{3}$$

$$contribution06_i = \Phi(\beta' \bar{X} + \beta_i (\overline{X06}_i - \bar{X}_i)) - \Phi(\beta' \bar{X}) \tag{4}$$

$$\approx marginal_i \times deviation06_i \tag{5}$$

where \bar{X} is the vector with mean values, σ_i is the standard deviation of the i -th variable, and $\overline{X06}_i$ is the mean value of the i -th variable for vintage 2006 loans. Similarly we define $deviation01$ and $contribution01$ for vintage 2001 loans. The fourth line follows from a first-order Taylor approximation with the derivative of the logit function with respect to the i -th variable approximated by

*marginal*_{*i*}.⁵ The *marginal* statistic measures the effect of a one standard deviation increase in a variable on the probability of the event. The *deviation* statistic measures the number of standard deviations that the mean value of a variable in 2006/2001 was different from the mean value measured over the entire sample. The *contribution* statistic measures the deviation of the (average) event probability in 2006/2001 from the (average) event probability over the entire sample that can be explained by a particular variable.

For any subgroup of loans we can determine the predicted probability of the event by computing:

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

where the superscript j refers to the loan number, and L is the total number of loans.

3.2 Variable Definitions

Table 2 provides the definitions of the dependent and independent (explanatory) variables used in the empirical analysis. As dependent variables we use either a delinquency or a foreclosure dummy variable. We define a loan to be delinquent if payments on the loan are 60 or more days late, the loan is in foreclosure, or the associated house is the real estate owned. We define a loan to be in foreclosure if the reported status is in foreclosure or real estate owned. Once a foreclosure procedure for a loan is finalized and the loan balance becomes zero, we drop this loan from our analysis.

We use the following borrower and loan characteristics: the FICO credit score, the 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, a dummy variable indicating whether full documentation was provided, a dummy variable indicating whether there is a prepayment penalty on the loan, the initial mortgage rate, and the margin for adjustable-rate and hybrid loans.⁶

In addition, we construct a variable that measures the house price appreciation from the time of origination until the time we evaluate whether the loan is delinquent or in foreclosure. To this end we use metropolitan statistical area (MSA) level house price indexes from the Office of Federal Housing Enterprise Oversight (OFHEO). We match loans and MSAs by zip code.⁷

⁵Technically, we first change units by multiplying with σ_i in Equation (2) and diving by σ_i in Equation (3).

⁶We 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.

⁷We also considered the change in the unemployment rate from the period of origination, which we could only measure accurately at the state-level for the entire sample. The corresponding regression coefficient mainly picked up the time trend in the delinquency or foreclosure rate. This however we found to be a spurious relation, because when including vintage dummy variables the unemployment rate became insignificant.

In Table 2 we indicate the expected sign for the coefficient on the explanatory variables in parentheses.

3.3 Results

Table 3 reports the results from the logit regression described in Subsection 3.1 with the probability of delinquency (panel A) and foreclosure (panel B) 12 months after origination as event. The explanatory variables are listed in the first column. First focussing on the marginal effect of the regression coefficients, defined in Equation 2, we see that the initial rate is the most important variable in terms of explaining cross-sectional differences. The positive sign means that a higher initial rate increases the probability of delinquency (panel A) and for foreclosure (panel B), as one would expect. Based on the point estimate in panel A: a one standard deviation increase in the initial rate, increases the probability of delinquency (12 months after origination) by 2.16%.

The second most important marginal effect is associated with the FICO credit score, which works in the opposite direction. A higher FICO score decreases the probability of delinquency (panel A) and foreclosure (panel B), as one would expect. In general all the variables have the predicted sign (as discussed in Table 2) and all variables are significant at the 1% significance level. We also experimented with several interaction and quadratic terms, which yielded very similar results, both qualitatively and quantitatively.

In panel C of Table 3 we report the “Deviation” of the 2001/2006 average value of a variable compared to the the average value over the whole sample, 2001-2006, expressed in standard deviations (see Equation 3). In particular the initial rate was high in 2001, and the house price appreciation was low for 2006.

The “Contribution” of a variable to explain why the delinquency/foreclosure rate in 2001/2006 differed from the average over 2001-2006 is reported in panels A and B. Up to a first-order approximation, the contribution equals the marginal effect times the deviation. We see that the high initial rate for vintage 2001 loans was the major reason for high observed delinquency and foreclosure rates for that year.

For 2006, the largest contribution is due to the low house price appreciation, accounting for a 1.1% higher delinquency rate and a 0.6% higher foreclosure rate, 12 months after origination. None of the other variables substantially deviated in 2006, and therefore none of the other variables contributed much to a different delinquency or foreclosure rate in 2006.

To examine to what extent the logit regression model is capable of explaining the large observed delinquency and foreclosure rates in 2006, we compare the corrected delinquency and foreclosure rates for different ages and different vintages in Figure 2. The corrected rate equals the actual minus the predicted rate plus the weighted average rate over 2001-2006, with weights equal to the number of loans originated in each year. The predicted delinquency and foreclosure rate is

Table 2: Variable Definitions

This Table presents definitions of the variables used in the regression analyses. The first two variables are used as dependent variables. The other variables are used as independent variables, with the expected sign of the effect on the dependent variable in parentheses.

Variable (Expected Sign)	Explanation
Delinquency Dummy	Payments on the loan are 60+ days late, the loan is in foreclosure, or the loan is real estate owned.
Foreclosure Dummy	The loan is in foreclosure or is real estate owned.
FICO Score (-)	Fair, Isaac and Company (FICO) credit score at origination. We expect a high FICO score to be associated with a lower probability of delinquency and foreclosure.
Loan-to-value ratio (+)	Value of the loan divided by the value of the house at origination. A higher loan-to-value ratio makes default more attractive which we expect to be associated with a higher probability of delinquency and foreclosure.
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, which we expect to be associated with a higher probability of delinquency and foreclosure.
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, which is associated with a higher probability of delinquency and foreclosure.
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, which is associated with a lower probability of delinquency and foreclosure.
Prepayment Penalty Dummy (+)	Equals one if there is a prepayment penalty and zero otherwise. We expect that a prepayment penalty makes refinancing less attractive leading to a higher probability of delinquency and foreclosure.
Initial rate (+)	Initial interest rate as of the loans first payment date. A higher interest rate makes it harder to make the monthly mortgage payment, which we expect to be associated with a higher probability of delinquency and foreclosure.
Margin (+)	Margin for 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, which we expect to be associated with a higher probability of delinquency and foreclosure.
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 gives better opportunities to refinance the mortgage loan, which we expect to be associated with a lower probability of delinquency and foreclosure.
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 or foreclosure relative to an 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 foreclosure and delinquency, holding constant the loan-to-value and debt-to-income ratio.

Table 3: Regression Results

The table reports the output of the the logit regression defined in Equation 1, where the event is that a loan is delinquent (panel A) or in foreclosure (panel B), 12 months after origination. The first column reports the explanatory variables (constant not reported). In panels A and B we report the marginal effect of a variable, defined in Equation 2, and the contribution of a variable to explain a different probability of delinquency or foreclosure in 2001/2006, defined in Equation 4. Panel C reports the deviation of the 2001/2006 value of a variable from the average over 2001-2006, defined in Equation 3. We have the first-order approximation $contribution \approx marginal \times deviation$. All coefficients are statistically significant at the 1% level.

Explanatory Variable	Panel A: Delinquency Rate		Panel B: Foreclosure Rate		Panel C: Deviations	
	Marginal Effect (%)	Contribution 2001 (%)	Marginal Effect (%)	Contribution 2001 (%)	Deviation 2001	Deviation 2006
Fico Score	-1.51	0.85	-0.36	0.18	-0.39	0.12
Loan-to-Value Ratio	0.77	0.02	0.31	0.01	0.02	-0.05
Debt-to-Income Ratio	1.12	-0.07	0.33	-0.02	-0.08	0.15
Missing Debt-to-Income Ratio	1.00	0.05	0.28	0.01	0.05	-0.12
Documentation Dummy	-0.68	-0.21	-0.30	-0.09	0.28	-0.21
Prepayment Penalty Dummy	0.23	0.03	0.08	0.01	0.11	0.01
Initial Rate	2.16	2.65	1.01	1.26	1.18	0.03
Margin	0.45	-0.04	0.13	-0.01	-0.09	0.01
House Price Appreciation	-1.03	0.59	-0.49	0.32	-0.45	-0.78
Hybrid Dummy	0.65	-0.07	0.34	-0.04	-0.13	-0.19
ARM Dummy	0.28	-0.06	0.17	-0.04	-0.23	0.25
Balloon Dummy	0.33	0.04	0.16	0.02	0.13	0.55
Origination Amount	0.65	-0.21	0.25	-0.08	-0.37	0.28

determined using Equation 6. We add up the weighted average rate to facilitate comparison with Figure 1 which plots the actual rates. Interestingly, both the corrected delinquency and foreclosure rates have been increasing over the last five years. This is in sharp contrast with the actual rates depicted in Figure 1, where 2003 was the year with the lowest rate, and 2001 was the year with the second-highest rates.

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

The logit regression specification we used up to now assumes that the regression coefficients are constant over time. That is, the effect of a unit change in an explanatory variable on the delinquency or foreclosure rate is the same in, for example, 2006 and 2001, holding constant the values of the other explanatory variables. We can test the validity of this assumption for a particular variable by interacting this variable with vintage (origination year) dummy variables and checking whether the logit regression coefficients for these interaction terms are approximately equal. It turns out that the strongest rejection of a constant regression coefficient is for the LTV ratio. In this section we first discuss this finding and show that it was an important determinant of the subprime mortgage crisis. Next we turn to the question whether lenders were aware of the non-stationarity of the loan-to-value effect by investigating the relation between the loan-to-value ratio and mortgage rates over time. For this analysis we focus on the delinquency rate.

4.1 Loan-to-Value Ratio and the Delinquency Rate

We first consider a regression specification similar to the one in table 3, panel A (delinquency rate), but with the LTV variable replaced by interaction variables between the LTV and the vintage dummies for each origination year, from 2001 through 2006. For 2001 the marginal effect of the corresponding interaction variable is only 0.31%. The marginal effect is monotonically increasing in time, and amounts to a substantial 1.95% in 2006.⁸ Hence, in 2001 a high loan-to-value ratio hardly gave rise to an increased probability of delinquency, holding constant the values of the other variables like FICO score and house price appreciation. In 2006, a high loan-to-value ratio did give rise to a substantially higher probability of delinquency.

To further investigate the importance of the loan-to-value ratio on the probability of delinquency, we first illustrate in Figure 5 the distribution of the loan-to-value ratio for all first-lien loans. About one-third has an LTV smaller than 80%, about one-third has an LTV of exactly 80%, and about one-third has an LTV greater than 80%. This distribution has been slightly changing over time. In 2001 the percentage of loans in these three LTV categories was 43%, 20%, and 38%

⁸The marginal effect measures the effect of a one standard deviation increase in a variable, starting from a situation where all explanatory variables are at their mean values.

respectively. In 2006 it was 32%, 43%, and 25% respectively, and hence twice as many loans with an LTV of exactly 80%. The average loan-to-value ratio has hardly changed over time, as can also be seen from Table 1.

Figure 5: Cumulative Distribution of Loan-to-Value Ratios.

The Figure shows the cumulative distribution of the loan-to-value ratios for all first-lien subprime mortgage loans originated between 2001 and 2006. Loans with a loan-to-value ratio greater than 100 account for less than 0.5% of all loans and are omitted from the Figure.

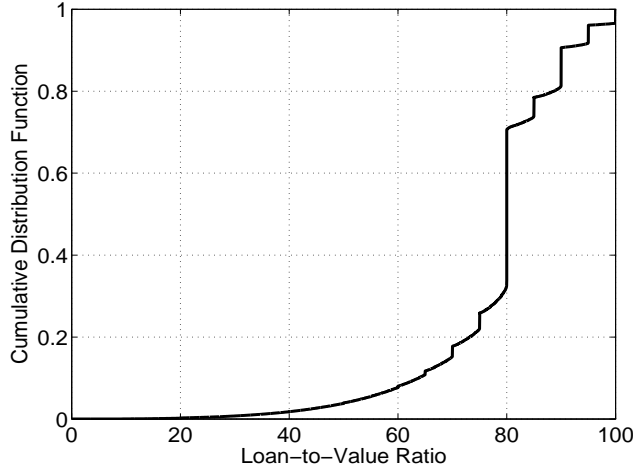


Table 4 reports the actual minus predicted delinquency rate for different loan-to-value categories and different vintage years. In other words, the table reports the average prediction error for the three LTV subgroups discussed above. A positive prediction error means that the actual delinquency rate was higher than the rate predicted by the logit regression model. Consistent with Figure 2, the error increases over time. However, for the low LTV group the increase in the error has been much smaller and in fact, it has been slightly decreasing from 2004 onwards. For the LTV ratios of 80% and higher, the increase in the error has been around 5%. This again illustrates that the high LTV ratios have been increasingly associated with the higher delinquency rates, beyond what is captured by the logit regression model. Finally notice that in Table 4 that the loans with an LTV ratio of exactly 80% have performed particularly worse than predicted by the logit regression model from 2003 onwards.⁹ This can be explained by the increasing use of piggy-back loans, which are taken out in addition to a first-lien loan. Borrowers with an LTV larger than 80 have to pay private mortgage insurance and therefore an LTV of exactly 80 is an important threshold.

⁹We also confirmed this result by adding a dummy variable taking the value one if the LTV ratio is exactly 80% to the logit regression specification. The coefficient on this variable is statistically significant. However, adding this dummy does not change the result that the actual delinquency rate in 2006 is generally underpredicted.

Table 4: Actual Minus Predicted Delinquency Rate

This table presents the actual minus predicted delinquency rate 12 months after origination for different vintages and loan-to-value (LTV) ratios. The same logit regression specification as in Table 3 panel A is used.

	2001	2002	2003	2004	2005	2006
<i>LTV</i> < 80%	-1.3%	-1.9%	-1.2%	0.1%	-0.2%	-0.3%
<i>LTV</i> = 80%	-1.7%	-1.9%	-0.6%	0.8%	1.5%	3.9%
<i>LTV</i> > 80%	-3.0%	-3.0%	-1.6%	0.2%	0.5%	1.6%

4.2 Loan-to-Value Ratio and the Mortgage Rate

In this subsection we seek to answer the question whether lenders were aware of high LTV ratios being increasingly associated with riskier borrowers. We perform a cross-sectional ordinary least squares regression with the mortgage rate as the dependent variable and loan characteristics as independent variables.¹⁰ We perform one such regression for each calendar quarter in our sample period. We can only expect to get accurate result when using relatively homogeneous groups of loans, and therefore consider fully-amortizing FRM, 2/28 hybrid, and 3/37 hybrid loans separately. Together these three contract types account for well over half of all mortgage loans in our database. Each cross-sectional regression is based on over 10,000 observations.

Figure 3 shows the regression coefficient on the LTV ratio for each quarter from 2001 to 2006. We scaled the coefficients by the standard deviation of the LTV ratio, and it can therefore be interpreted as the change in the mortgage rate when the LTV is increased by one standard deviation. In the fourth quarter of 2006, a one standard deviation increase in the LTV ratio corresponded to a 30 to 40 basis point increase in the mortgage rate, keeping constant other loan characteristics. In contrast, in the first quarter of 2001, the FRM rate was hardly dependent on the LTV ratio, while the interest rate on a hybrid mortgage even slightly negatively depended on the LTV ratio. This provides evidence that lenders were well aware of high LTV ratios being increasingly associated with risky borrowers.¹¹

¹⁰Specifically, we use the FICO score, loan-to-value ratio, debt-to-income ratio, a dummy whether the debt-to-income ratio is missing, documentation dummy, prepayment penalty dummy, margin, origination amount, term of the mortgage, and prepayment term as the right-hand-side variables.

¹¹Unreported results reveal that the effect of other loan characteristics on the mortgage rate has been much more stable over time.

5 Conclusion

In this paper we find evidence for two major causes of the subprime mortgage crisis. First, high loan-to-value borrowers increasingly became high-risk borrowers, in terms of elevated delinquency and foreclosure rates. Lenders were aware of this and adjusted their mortgage rates accordingly over time. Second, the low house price appreciation in 2006-2007 also negatively impacted the performance of vintage 2006 loans.

References

- BERNANKE, B. (2007): “Housing, Housing Finance, and Monetary Policy,” Speech, Federal Reserve Bank of Kansas City’s Economic Symposium.
- CAMPBELL, J. Y., AND J. F. COCCO (2003): “Household Risk Management and Optimal Mortgage Choice,” *Quarterly Journal of Economics*, 118(4), 1449–1494.
- CUTTS, A. C., AND R. A. VAN ORDER (2005): “On the Economics of Subprime Lending,” *The Journal of Real Estate Finance and Economics*, 30(2), 167–196.
- DENG, Y., J. M. QUIGLEY, AND R. A. VAN ORDER (2000): “Mortgage Terminations, Heterogeneity and the Exercise of Mortgage Options,” *Econometrica*, 68(2), 275–308.
- KOLJEN, R., O. VAN HEMERT, AND S. VAN NIEUWERBURGH (2007): “Mortgage Timing,” Working paper, New York University.
- PENNINGTON-CROSS, A., AND S. CHOMSISENGPHET (2007): “Subprime Refinancing: Equity Extraction and Mortgage Termination,” *Real Estate Economics*, 35(2), 233–263.
- VAN HEMERT, O. (2007): “Household Interest Rate Risk Management,” Working paper, New York University.
- VON FURSTENBERG, AND R. J. GREEN (1974): “Home Mortgage Delinquencies: A Cohort Analysis,” *Journal of Finance*, 29(5), 1545–1548.