Where's the Smoking Gun? A Study of Underwriting Standards for US Subprime Mortgages*

Geetesh Bhardwaj† Rajdeep Sengupta‡§

October 27, 2008

Abstract

The dominant explanation for the meltdown in the US subprime mortgage market is that lending standards dramatically weakened after 2004. Using loan-level data, we examine underwriting standards on the subprime mortgage originations from 1998 to 2007. Contrary to popular belief, we find no evidence of a dramatic weakening of lending standards within the subprime market. We show that while underwriting may have weakened along some dimensions, it certainly strengthened along others. Our results indicate that (average) observable risk characteristics on mortgages underwritten post-2004 would have resulted in a significantly lower ex post default if they were to be given a loan in 2001 or 2002. We show that while it is possible that underwriting standards in this market were poor to begin with, deterioration in underwriting post-2004 cannot be the explanation for collapse of subprime mortgage market.

JEL Codes: G21, D82, D86.

Keywords: mortgages, subprime, underwriting, crisis.

*Thanks to Mara Faccio, Gary Gorton, Geert Rouwenhorst and Dave Wheelock for their comments and suggestions on an earlier draft of this paper.

†Vice President, AIG Financial Products. The views expressed are those of the individual author and do not necessarily reflect the official positions of AIG Financial Products Corp.

‡Economist, Federal Reserve Bank of St. Louis. The views expressed are those of the individual author and do not necessarily reflect official positions of the Federal Reserve Bank of St. Louis, the Federal Reserve System, or the Board of Governors.

§Correspondence: Research Division, Federal Reserve Bank of St. Louis, P.O. Box 442, St. Louis, MO 63166-0442. Phone: (314) 444-8819, Fax: (314) 444-8731, Email: rajdeep.sengupta@stls.frb.org.

1 Introduction

> There is doubtless an unfortunate tendency among some, I hesitate to say most, bankers to lend aggressively at the peak of a cycle and that is when the vast majority of bad loans are made.[1]

Existing wisdom on financial crisis argues that the peak of the credit cycle is often associated with a weakening of lending standards. The hypothesis that "most bad loans are made in good times" has been viewed, by policymakers and academics alike, as one of the principal features of credit crises (Kindleberger and Aliber, 2005). While this is arguably true of most historical episodes of credit crises, the same reason has been put forward for having caused more recent events (Reinhart and Obtsfeld, 2008a). Academic research and policy initiatives on the current crisis in subprime mortgages in the US have argued that there was a significant decline in the underwriting standards adopted by subprime lenders. For example, the President's Working Group on Financial Markets (March, 2008) has concluded:

> The turmoil in financial markets was triggered *by a dramatic weakening of underwriting standards for U.S. subprime mortgages, beginning in late 2004*, and extending into early 2007.[2]

Much of the same sentiment was echoed in the popular press:

> Strange was becoming increasingly common: loans that required no documentation of a borrower's income. No proof of employment. No money down. "I was truly amazed that we were able to place these loans,"...[3]
>
> House prices levitated as mortgage underwriting standards collapsed. The credit markets went into speculative orbit, and an idea took hold. Risk, the bankers and brokers and professional investors decided, was yesteryear's problem.[4]

---

[1] Remarks by Chairman Alan Greenspan on Banking Supervision, before the Independent Community Bankers of America, March 7, 2001.
http://www.federalreserve.gov/boarddocs/speeches/2001/20010307/default.htm.

[2] Policy Statement on Financial Market Developments, March 2008 (Emphasis in the original).

[3] "The Bubble: How homeowners, speculators and Wall Street dealmakers rode a wave of easy money with crippling consequences", *The Washington Post,* June 15, 2008.

[4] "Why no Outrage?", *Wall Street Journal*, July 19, 2008.

Despite the analysis of these events from various perspectives, there has been very little economic analysis of the proposition of examining underwriting standards in the subprime mortgage market. At the cost of parsing the remark from the Policy Statement too literally, we examine two related questions. First, was there a dramatic weakening of underwriting standards? Second, did this weakening begin around late 2004? To examine these questions, we use loan level data on originations for subprime mortgages from 1998 to 2007. Our aim is to study the underlying distribution and evolution of borrower and mortgage (loan) characteristics in the subprime market with a view to identifying the "deterioration" in underwriting standards.

We argue that any study of underwriting standards in this environment needs to account for two important features of credit risk that has largely been ignored up to this point. The first takes into account the multidimensional nature of credit risk: Lenders often compensate for the increase in the ex ante risk of one borrower attribute by raising the requirement standards along another dimension. The second addresses the endogeneity problem that confronts the use of mortgage characteristics like loan-to-value (LTV) ratio and mortgage interest rate as explanatory variables in determining loan performance. While both borrower attributes and mortgage characteristics determine credit risk, the terms and conditions on the latter is largely determined by the former. We begin with a test for endogeneity bias adopting techniques in Chiappori and Salanie (2000). Following this, we study the determinants of mortgage characteristics (like LTV) and default in the subprime market by accounting for both features mentioned above. Finally, we devise a counterfactual technique to answer the question as to whether there was a decline in mortgage underwriting *within* the subprime market after 2004.

Our results indicate that there is no evidence of a dramatic change in underwriting standards in the subprime market, particularly for originations after 2004. Given the multidimensional nature of ex ante credit risk, it is difficult to emphasize weakening in terms of some attributes as a decline in overall underwriting standards. The results show that while underwriting may have weakened along some dimensions (e.g. lower documentation), it also strengthened in others (e.g. higher FICO scores).

The test of endogeneity bias presents evidence of a strong correlation, conditional on observable characteristics, between the individual's choice of loan-to-value ratio (coverage), and the ex-post occurrence of the event of delinquency (risk). When we account for this endogeneity problem, our estimation results fail to document evidence that can unambiguously suggest that underwriting standards within the subprime market deteriorated over this period. On the contrary, we find evidence of credible underwriting over this period which attempted to adjust riskier borrower attributes with lower LTV ratios and higher FICO scores.

Using counterfactual analysis, we argue that one can reject the null hypothesis that underwriting within the subprime mortgage market remained unaltered after 2004 in favour of the alternative that the underwriting standards did not decline. Our results seem to indicate that (average) observable risk characteristics on loans underwritten post-2004 would have registered a significantly lower ex post default in 2001 and 2002 than (average) observable risk characteristics on loans underwritten in their current years (2001 or 2002). Stated differently, if loans underwritten in 2005 (or 2006 or 2007) were originated in 2001 or 2002, then they would have performed significantly better *on average* than loans underwritten in 2001 or 2002. In light of this evidence, it is unclear how a deterioration in underwriting standards can be the dominant explanation of default and delinquency in the subprime market. Of course, our analysis does not rule out the hypothesis that underwriting standards in this market were probably poor to begin with. At the very least, unobservable risk characteristics and market conditions (like house price appreciation) had a greater role to play than was earlier believed.[5]

Despite noteworthy advances on the history and macro trends on financial crises, the current understanding on financial crises is at best "panoramic" (Reinhart and Rogoff, 2008b). This is largely due to the fact that there is very little microeconomic data, particularly at the level of individual loans, on earlier episodes of financial turmoil. This paper reverses the trend in its use of loan-level subprime mortgage data from First American Loan Performance (FALP)—the largest publicly available repository of data on individual subprime loans (see Mayer and Pence, 2007 for a study of databases on subprime loans). Along with some contemporaneous work reviewed below, this study stands among the few that have examined microeconomic trends before the onset of financial crises. We discuss these studies and our contribution to

[5] Of course, this begs the question as to why the market did not experience a meltdown earlier in its existence. We address this question in a companion paper, Bhardwaj and Sengupta (2008), which argues that the subprime market was sustained by prepayments during the boom in house prices.

this literature in the next section. In Section 3, we present a brief discussion of the underlying theoretical framework and formulate our tests for endogeneity bias. Summary data on underwriting with respect to borrower and mortgage characteristics are reported in Section 4. Section 5 provides the evidence on endogeneity bias and the estimation results on underwriting and loan performance in the subprime market. The counterfactual analysis is described in Section 6. Section 7 concludes.

2 Related Literature

The subprime mortgage crisis has generated a substantial literature, not least because of the ensuing turmoil in financial markets. Mian and Sufi (2008) show that disintermediation-driven loan originations led to increased risk-taking on the part of lenders from 2001 to 2005. They argue that the rapid expansion in the supply of credit is responsible for the house price boom in the early 2000s and the subsequent mortgage defaults in the last couple of years. This last finding is supported in other work on subprime mortgages (Gerardi et al., 2008). In a companion paper (Bhardwaj and Sengupta, 2008), we show that the boom in the subprime market was indeed sustained by prepayments during the period of appreciating house prices. Undoubtedly, when one considers the mortgage market as a whole, the emergence and growth of a non-prime segment obviously demonstrates enhanced risk tolerance on the part of mortgage lenders (Mian and Sufi, 2008; Gorton, 2008). However, the question of interest in this paper is whether there was a gradual shift over the years to a riskier consumer base *within* the subprime market.

Several studies have argued that the subprime mortgage market in the US witnessed a sharp decline in underwriting standards. These studies have based their arguments on the originate-to-distribute hypothesis, implying that underwriting standards declined as mortgage originators could successfully pass on credit risk through the process of securitization. This assumption appears exceptionally simplistic in the face of detailed evidence on the securitization process (Gorton, 2008).[6] Our concerns about these current findings are explained below.

[6] Gorton (2008) provides a telling critique on the originate-to-distribute hypothesis, arguing that the hypothesis fails to explain why such a crisis should occur only in the securitization of subprime mortgages but not for other assets that are also securitized. The idea that in a post-securitization scenario, originators retained no significant exposure to the risk of the underlying mortgages flies in the face of the bankruptcies of some of the top subprime

Keys et al. (2008) argue that industry thumb-rule practices meant that loans that were easier to securitize were often "let-off" with lax underwriting. They use FALP data on securitized loans to argue that the lender has a stronger incentive to screen the borrower more carefully at FICO score "620-" (than "620+"), where there is a higher likelihood that the borrower will end up on her balance sheet. Clearly, a credible test of their hypothesis requires data on loans that were not securitized but retained in the originators portfolio.[7] A higher proportion of loans at credit scores of "620+" could be a result of (1) a higher proportion of loans (both securitized and portfolio loans) approved around the FICO threshold of 620 and (2) a higher proportion of individuals above this threshold in the credit eligible population. To give us an idea as to why this may be important, we examine their hypotheses in greater detail. Figure 1 shows the distribution of FICO scores across different vintages for securitized first-lien subprime loans in the FALP data. Obviously, the FICO score of 620 is not the only important threshold. A similar pattern is observed for FICO thresholds of 550, 560, 580, 600, 620, 640 and 660. It is difficult to conceive that all of these thresholds were caused by an exogenous variation in the ease of loan securitization.[8] It is also likely that the emphasis on the 620 threshold maybe misplaced. After all, what is the rationale behind focusing on the threshold of 620 and ignoring the others?

Next, the claim that default rates are higher for borrower FICO scores above 620 also bears scrutiny. Keys et al.(2008) do not address the endogeneity problem that confronts the use of mortgage characteristics like loan-to-value ratio and mortgage interest rate as explanatory variables as determinants of loan performance.[9] This estimation procedure significantly undermines the findings in a large volume of the literature on the mortgage underwriting and default (Stein, 1995; Lamont and Stein, 1999, Brueckner, 2000 and Cutts and Van Order, 2005).[10] In Section 5.1 of this paper, we demonstrate that, at least in the case of subprime loan originations, interest

originators like Option One, Ameriquest and New Century. Gorton (2008) demonstrates how there is direct exposure to the originated risk and how there are implicit contracts that make such arrangements incentive compatible.

[7]The FALP data used in their paper (as well as ours) contains only securitized subprime loans.

[8]On the contrary, the pattern in Figure 1 seems to strongly suggest the existence of this peculiarity in the way FICO scores are distributed in the credit eligible borrower population.

[9]The estimation results in Demyanyk and van Hemert (2008), who find excessive loan-to-value ratios in subprime originations, also suffers from this endogeneity bias.

[10]The list is too numerous to mention here. See for example, Vandell et al. 1993, Archer et al. 2002, Ciochetti et al. 2002, 2003, Ambrose and Sanders 2003, Chen and Deng, 2003, Grovenstein et al. 2005.

rate and the loan-to-value ratio are endogenously determined by the underwriting process.

Finally, it is important to mention here that the proportion of loans in the FALP data with FICO score in the range 620-625 was less then 5% of the market for every single vintage from 1998 to 2007. Moreover, the majority of the subprime loans are below the threshold of 620. Loans with FICO score in the range less than 620 comprise more than 50% of all the subprime loans in the FALP data for every vintage from 1998 to 2007. Taken together, these facts make it difficult to verify the assertion that lax securitization around an ad hoc credit threshold of 620 can explain loan performance in the subprime mortgage market.

The analysis in this paper improves upon prior research in three ways. First, our study takes into account the multidimensional nature of credit risk arguing that any focus on a single borrower or mortgage characteristic is misleading. Second, we test for the endogeneity of mortgage characteristics in models linking underwriting standards to default and delinquency in subprime loans. Third, taking into account both the multidimensional nature of credit risk and the endogeneity bias ignored previous studies we evaluate the change in underwriting standards for various vintages of mortgage originations. We discuss the details of our methodology in the next section.

3 Theory and Estimation Methodology

3.1 Characterizing a weakening of underwriting standards

Mortgage underwriting refers to the process used by a mortgagee (lender) to assess the credit risk of the mortgagor (borrower). The process involves summarizing the ex ante risk of default from a profile of borrower attributes with the purpose of approving or denying the borrower's loan application. Therefore, underwriting is based on the borrower's *observable* characteristics *at the time of origination*.

Several characteristics of the borrower that need to be summarized to determine overall credit risk. Lenders are known to compensate for the increase in the ex ante risk of one borrower attribute by raising the requirement standards along another dimension. Stated differently, borrower credit risk is multidimensional. Accordingly, in order to define a decline in underwriting standards there is a need to "aggregate" each borrower characteristic to build a summary measure that fulfills a variety of desirable conditions. Needless to say, the solution to this aggregation problem has proved elusive. To the best of our knowledge, we are not aware of a single metric that adequately summarizes a variety of borrower characteristics. Therefore, we need to exercise caution when characterizing a decline in underwriting standards.

To begin with, approving loan applications of borrowers that would previously be considered uncreditworthy can be viewed as a weakening of underwriting standards.[11] Interestingly, the subprime market was primarily conceived to supply borrowers who would otherwise be denied loans in the prime market. Taken to its logical conclusion, one could view the emergence of subprime lending as a weakening of underwriting standards for the US housing market as a whole. Significantly, for loans older than 60 months in our sample, default probabilities on subprime mortgages have never been lower than 28 percent. These facts raise important questions about the viability of the subprime market as a whole. However, such questions are beyond the scope of this paper. For our purposes, it is important to keep in mind that our examination of a weakening in underwriting standards is relative to subprime mortgages of earlier vintages and not vis-à-vis mortgages in the prime market.

Another caveat, mentioned earlier, relates to the use of borrower and mortgage characteristics in determining the changes in underwriting standards. Some earlier studies used both borrower and mortgage characteristics as determinants of mortgage default.[12] However, it is important to note that mortgage characteristics are the outcome of the underwriting process. For example, Cutts and Van Order (2005) have shown that at least in the case of the subprime market, terms of the mortgage contract are determined by variations in borrower attributes. Consequently, treating mortgage terms and conditions as exogenous to the likelihood of mortgage default leads to endogeneity bias. For that reason, it is important to report how loan characteristics vary with changes in the borrower profile.

[11] Admittedly, this can arise with a decline in the cost of funds for lenders. In a model of entry and competition between asymmetrically informed lenders, Sengupta (2007) shows that it is optimal for an uninformed lender (entrant) to pool uncreditworthy borrowers in order to capture creditworthy borrowers away from the informed lender (incumbent). Interestingly, these pooling equilibria can only be sustained if the cost of funds is significantly low in the economy.

[12] See Quercia and Stegman (2000) for a survey of studies on mortgage default.

3.2 Theoretical Framework

The theoretical underpinnings of our estimation procedure are best illustrated in terms of anecdotal evidence presented in Table 4 in Cutts and Van Order (2005). This table was prepared from actual interest rates on offer for 30-year fixed rate mortgages in the subprime market by Option One Mortgage Corporation in 2002. In Table 1, we collect similar information from the Option One Mortgage Corporation website which shows the rates effective November 2007. These tables vividly summarize the origination process in the subprime market.[13] Note that for a given borrower type—characterized by the borrower's credit grade[14] and FICO score—the interest rates on offer vary with the downpayment on the loan. In other words, borrowers of "riskier type" have to put up more equity to qualify for the same interest rate.

Significantly however, some observers have argued that evidence of a decline in lending standards can be found in the terms and conditions on the mortgage contracts offered to borrowers. For example, an increase in mortgage characteristics like loan-to-value ratios is viewed as a decline in lending standards (Demyanyk and van Hemert, 2008). While it is true that mortgage characteristics determine the credit risk on the loan, it is important to realize that these mortgage characteristics are themselves the outcome of underwriting standards. Stated differently, borrower characteristics determine the terms and conditions on the loan contract. The conventional theoretical explanation behind this argument is based on the vast literature on asymmetric information, as pioneered by Akerlof (1970), Rotshchild and Stiglitz (1976) and many others. More recently, recent advances in empirical contract theory (Chiappori and Salanie, 2000; Chiappori et al. 2006) provide robust tests of asymmetric information, as predicted by the stylized theoretical models mentioned above.

For a more rigorous examination of endogeneity, we propose a test for asymmetric information using the predictions outlined in Chiappori and Salanie (2000). Chiappori et al.(2006) show that under both adverse selection and moral hazard, one should observe a positive corre-

[13] Not surprisingly, the differences in the two tables illustrate how mortgage originators have cut back on the loans on offer after the downturn in this market.

[14] The credit grade is assigned by the lender and typically depends on the lender's assessment of credit risk depending on borrower characteristics. Evidently, this grading process varies from lender to lender.

lation (conditional on observables) between risk and coverage.[15] If different mortgage contracts are actually sold to observationally identical borrowers, then the frequency of default among the subscribers to a contract should increase with the loan-to-value ratio on the mortgage. In a model of lender competition under adverse selection, where riskiness is an exogenous and unobservable characteristic of an agent, the correlation stems from the fact that high-risk agents are more likely to opt for the mortgage contract with the lower downpayment but a higher interest rate (Brueckner, 2000). Under moral hazard, the reverse causality would generate the same correlation: borrowers buying into mortgages with higher LTV for any unspecified or exogenous reasons are likely to exert less effort to repay the loan and therefore become riskier. Based on this outline, we can make the following inferences about the process of mortgage origination.

Firstly, conditional on observable risk, borrowers are offered menus of contracts varying in their interest rate and LTV requirements as given in Table 1. Borrower characteristics define borrower credit grade, which together with borrower credit score determines the menus of contracts available to the borrower. In terms of actual mortgage originations, this means that a borrower can choose among the contract terms given along a row in Table 1.

Secondly, within the menu of contracts on offer, contracts with a higher LTV typically come with a higher rate of interest.[16] This feature is critical to our understanding of the underwriting process. The borrower's downpayment on the mortgage determines the interest rate on the loan and vice-versa. Stated differently, we can use this feature to model the determinants of either of these terms.

3.3 Estimation Strategy and Test for Endogeneity Bias

Our test of asymmetric information is based on the conditional independence between the individual's choice of loan-to-value ratio and the ex-post occurrence of the event of delinquency, where the conditioning information is all borrower characteristics observable to the lender at

[15] See Chiappori et al. (2006) for details on the robustness of the positive correlation property under a variety of settings.

[16] Without a rigorous test, we are unable to assert whether this is also true for prime mortgages as well. Some prime borrowers are all too familiar with the notion that conditional on qualifying for a prime mortgage, mortgage rates are less responsive with changes in the loan to value ratio.

the time of origination of the loan. We begin with a brief description of the estimation strategy followed by a discussion on the test for endogeneity bias.

To derive testable predictions about the choice of loan-to-value ratio, we must deal with heterogeneity across borrowers. We will assume that the borrowers sharing the same values of observable characteristics are ex ante indistinguishable from the lenders point of view and therefore must have the same mortgage contracts on offer. To simplify our analysis, we assume that mortgage contracts in the subprime market can essentially be summarized by the following three features: product type (FRM or ARM), loan-to-value ratio and the interest rate on the loan. As discussed above, conditional on observable risk (as summarized from credit grade and scores), a borrower is free to choose among a menu of contracts which vary in these three attributes. A borrower's choice of LTV (together with his choice of product type) from among the menu of contracts on offer will naturally determine the interest rate on the loan. It follows from the theoretical framework outlined above that once we model the determinants of LTV and product type, modeling the choice of interest rate on the loan is redundant. Accordingly, we can focus our attention to simply the determinants of LTV and product type on the loan. In what follows, we assume that the LTV and product type are jointly determined by the borrower characteristics as given by the following equations:

$$Type^* = \mathbf{X}\boldsymbol{\delta}+\delta_{LTV}LTV + v \tag{1}$$

$$Type = \text{FRM}\,[Type^* > 0] \tag{2}$$

$$LTV = \mathbf{X}\boldsymbol{\gamma} + u \tag{3}$$

where the first and second equations are structural equations that determine product type, while the third equation is a reduced form equation for LTV.[17]

To derive testable predictions about the ex-post occurrence of default, we estimate the semi-parametric hazard rate regression for the 90-day delinquency event. The hazard function $h(t)$

[17]See Maddala (1983, Chapter 7) and Wooldridge (2002, Chapter 15) for a discussion of discrete response models with continuous endogenous explanatory variables.

is the instantaneous probability of delinquency at age t, and is given by

$$h(t) = \lim_{\Delta t \to 0} \frac{\Pr(t \le T < t + \Delta t | T \ge t)}{\Delta t} \tag{4}$$

Following Cox (1972), the semiparametric representation that we estimate takes the form

$$h(t) = h_0(t) \exp(\mathbf{X}\beta) \tag{5}$$

where $h_0(t)$ is baseline hazard function.

3.3.1 Testing endogeneity bias

For mortgages of every vintage, we set up a two-equation model, similar to the approach in Chiappori and Salanie (2000).

$$\begin{aligned} z_i &= X_i \gamma + u_i && (6) \\ h_i(t) &= h_0(t) \exp(X_i \beta) && (7) \end{aligned}$$

The first equation, identical to equation (3), is an ordinary least squares regression with LTV ratio as the dependent variable. The second equation, identical to equation (5), is a Cox proportional hazard rate regression model.[18]

The martingale residuals of the Cox model are calculated as

$$\hat{\eta}_i = \delta_i - \hat{H}_0(t) \exp(X_i \hat{\beta}) \tag{9}$$

[18] The object of interest in a Cox proportional hazard rate regression model is hazard ratio, that has the interpretation of a multiplicative change in the instantaneous probability of delinquency for a marginal change in a particular risk characteristic. Hazard ratio is analogous to the odds ratio in logistic regressions. Let $h(t|X)$ be the instantaneous probability of delinquency at age t conditional on other characteristics given by vector X. We can define the estimated hazard ratio (HR) for marginal change in risk characteristic x_i as

$$\begin{aligned} \widehat{HR}(t\,|x_i = x_i + \Delta x_i\,) &= \frac{h_0(t)\exp(x_1\widehat{\beta}_1 + x_2\widehat{\beta}_2 + \cdots + (x_i + \Delta x_i)\,\widehat{\beta}_i + \cdots)}{h_0(t)\exp(x_1\widehat{\beta}_1 + x_2\widehat{\beta}_2 + \cdots + x_i\widehat{\beta}_i + \cdots)} && (8) \\ &= \exp(\Delta x_i \widehat{\beta}_i). \end{aligned}$$

$$h(t|X, x_i = x_i + \Delta x_i) = h(t|X) * \widehat{HR}(t\,|x_i = x_i + \Delta x_i\,)$$

where $\hat{H}_0(t)$ is the estimated cumulative baseline hazard rate and δ_i is an indicator that takes the value 1 when a delinquency is recorded at loan age t for mortgage i and zero otherwise.

We estimate the two equations independently and compute the residuals $\widehat{u}_i$ and $\hat{\eta}_i$. Following Chiappori and Salanie (2000), the test statistic for the null of conditional independence $cov(\varepsilon_i, \eta_i) = 0$ is defined by:

$$W = \frac{(\sum_{i=1}^{n} \widehat{u}_i \hat{\eta}_i)^2}{\sum_{i=1}^{n} \widehat{u}_i^2 \hat{\eta}_i^2} \tag{10}$$

where W is distributed asymptotically as a $\chi^2(1)$.[19]

In addition to the test statistic described above, we construct a bootstrap confidence interval for testing the significance of correlation (conditional on observables) between risk and coverage.[20] The results of the test are reported in Section 5.1 of this paper. In what follows, we focus our attention on the subprime market for the ten year period 1998-2007. Given that the market evolved fairly rapidly over this period, it is interesting to record changes in underwriting standards by year. Therefore, throughout this paper, we report our results by year of origination (vintage).

4 Data and Summary Statistics

We analyze loan-level mortgage data from the Asset Backed Securities (ABS) Database of the FALP data repository. As of June 2008, this ABS Database, including both the Alt-A

[19] Chiappori and Salanie (2000) estimate a probit equation for the probability of accident in insurance markets and their test statistic is calculated by weighting each individual by days under insurance. In this case, we use the hazard rate regression for calculating the probability of default which explicitly takes the age of the mortgage into account. Furthermore, we estimate the probit model on the event of default and the test by weighting each mortgage by the age (in months) at the time of delinquency event. The results are qualitatively similar.

[20] The bootstrap methodology can be described as follows. Borrower characteristics on mortgage-i with LTV of z_i are denoted by X_i. Also, the age in months at which mortgage-i faces the 90-day delinquency event is denoted by y_i. Constructing the bootstrap confidence interval involves the following steps:

Step 1: We draw a bootstrap sample $(z^*, y^*, X^*) = \{(z_1^*, y_1^*, X_1^*), (z_2^*, y_2^*, X_2^*), \ldots, (z_n^*, y_n^*, X_n^*)\}$ with replacement from $(z_1, y_1, X_1), (z_2, y_2, X_2), \ldots, (z_n, y_n, X_n)$.

Step 2: From the bootstrap sample estimate equations (6) and (7), recover the OLS residuals on equation (6) and the martingale residuals in (9); and calculate the correlation between the two estimated residuals.

Step 3: Repeat the process B times to obtain the distribution of estimated correlation between risk and coverage.

& Nonprime market segments, provides data on more than 17 million individual mortgages.[21] According to FALP, its coverage of the overall subprime market, as of June 2008 (the data used in this paper) is in excess of 70 percent.[22] This database contains both subprime and Alt-A pools. For the purposes of this study, we restrict our analysis to subprime loans.[23] Loosely speaking, subprime pools include loans to borrowers with incomplete or impaired credit histories while Alt-A pools include loans to borrowers who generally have high credit scores but who are unable or unwilling to document a stable income history or are buying second homes or investment properties (Fabozzi, 2000).

FALP data include only those loans that were securitized in the ABS market, as opposed to loans that were retained by originators in their portfolios. Apart from various borrower and mortgage characteristics, it records all activity on the loan since securitization including repayment behavior. However, the data set is not without its limitations: First, there is little information on the households that had subprime mortgages. For example, there are no data on household debt, income, employment and demographics. Second, unlike other studies using mortgage data, the lack of identifiers in this database makes it difficult to match and combine these data with other databases to broaden the scope of analysis. Third, we do not have data on mortgage applications, and are therefore unable to compare approvals to loan applications that were denied. Finally, even for loans in the database, we are unable to track multiple liens or mortgages on the same property. Consequently, in what follows, our analysis will focus on first lien subprime loans in the ABS database.

Table 2 summarizes first lien subprime mortgages by product type for every year of origination from 1998 to 2007. The numbers give us the market shares for particular product types. We divide our sample according to fixed or adjustable rate mortgages (hereafter abbreviated as FRMs and ARMs).[24] For ARMs, lenders often employ an initial teaser rate that is lower than

[21] Our data are current up to June, 2008. LoanPerformance securities databases comprise the mortgage market's largest and most comprehensive mortgage securities data repository. For more details, see http://www.loanperformance.com/data-power/default.aspx

[22] For a more detailed description of the data, see Chomsisengphet and Pennington-Cross (2006).

[23] We classify a loan as a subprime loan if it belongs to a subprime pool in the ABS database. The industry classification of Subprime and Alt-A is at the pool level rather than on individual loan characteristics. Therefore, while Subprime and Alt-A loans each have distinct loan credit and documentation characteristics, it is possible for a Subprime pool to include a loan with characteristics more suitable for the Alt-A pool and vice versa.

[24] Fixed rate mortgages (FRMs) have an interest rate that is set (or locked) at the closing of the loan and

the fully indexed rate to attract borrowers to the product. For a hybrid-ARM, this teaser rate is often fixed for longer periods of time such as 2, 3 and even 5 years. To simplify classification over a very broad range of product types in the market, we define these products as ARM2, ARM3 and ARM5 respectively. As seen from Table 2, the subprime mortgage market comprises mainly three product types: FRM, ARM2 (which includes the hybrid 2/28 mortgage product) and ARM3 (which includes the hybrid 3/27).[25] All other product types make up a smaller fraction of the subprime mortgage market and their market share was on the decline for most of the sample period. As is evident from Table 2, there has been a clear shift over the years from FRMs to ARMs in the subprime market. The proportion of FRMs declined from more than half of the mortgages in the pools in 1998 to less than a fifth in 2006. At the same time there have been dramatic increases in ARM2 and ARM3.

In the previous section, we discussed the difficulties in determining whether there has been a decline in lending standards. However, even with these caveats in mind, we can make some simple assessments on changes in underwriting standards over this period. Our analysis proceeds in three steps. First, we document the trends in unconditional distributions for borrower characteristics like FICO. Next, we look at distributions of borrower FICO scores conditional on other borrower characteristics like documentation level.[26] For loan characteristics, we look at distributions of the loan-to-value ratio conditional on other mortgage terms like product type and borrower characteristics like FICO. Finally, we use regression techniques to determine how underwriting changed over the sample period for loans of different vintages. In doing so, we control for property and lender characteristics.

---

does not change over the life of the loan. However, rates are subject to change for an adjustable rate mortgages (ARM). ARMs typically reset annually and the periodic contractual rate is based on the index (an underlying reference rate like the LIBOR or COFI) and the margin (spread over the index).

[25] Not all ARM2 and ARM3 mortgages have a thirty-year maturity period. Therefore, while 2/28s and 3/27s make up the majority of loans in these two categories, they do not constitute all such loans.

[26] We also examine the conditional and unconditional distributions for other borrower characteristics, like occupancy type (owner-occupied, non-owner occupied (investor) or second home), loan purpose (purchase, refinance, cash-out refinance) etc. For the sake of brevity, these results are not reported here but are available on request.

4.1 Borrower Characteristics

Table 3 reports the unconditional distributions of borrower characteristics like FICO and documentation level of first-lien subprime loans from 1998 to 2007. The proportion of loans with no documentation is negligible throughout the sample, but that with low documentation steadily increased from 18.4 percent in 1999 to 37 percent in 2006. To the extent low-doc loans indicate a higher degree of uncertainty in borrower quality, the increasing proportion of low doc loans is suggestive of declining underwriting standards in the subprime mortgage market. However, the pattern is reversed when one considers the trend in borrower FICO score. The proportion of loans with a FICO score of less than 620 drops from close to 70 percent in the year 2000 to 50 percent in 2005. There is a corresponding increase in the proportion of loans for FICO-score in the range 620-659 and 660-719.

The unconditional distributions do not necessarily show a secular decline in lending standards. While the lending standards were lowered in terms of the documentation requirements needed to obtain a subprime mortgage, there was also an improvement in the average borrower quality as summarized by FICO scores. More important, these trends are discernible over the entire sample period and do not suggest anything particularly special about originations after 2004.

We supplement our univariate analysis by examining the distribution of FICO scores conditional on documentation level. A cross-sectional comparison for each year seems to indicate that borrowers with lower documentation have on average higher FICO scores (Table 4). Moreover, the proportion of borrowers in the lowest FICO-score category (< 620) has declined over the years. At the same time, there has been an increase in the proportion of borrowers in the 620-659 and the 660-719 range, especially for low-doc and no-doc loans.[27] When combined with the data from the unconditional distributions, these results suggest that although the proportion of low-doc loans was increasing over time, lenders sought to compensate the lack of documentation by seeking borrowers of higher quality, as determined by their FICO scores.

[27] We combined the low doc and no-doc categories in Table 2 to construct the conditional distributions in Table 3.

4.2 Mortgage Characteristics

Turning our attention to mortgage characteristics, we begin with the unconditional distribution of the loan-to-value ratio (LTV) at the time of loan origination.[28] As is evident from Table 5, there was a sharp rise in LTV values in the range (90,100] from a mere 3.2 percent in 1998 to 40.6 percent in 2006. Simultaneously, values in the range of less than or equal to 80% LTV have declined from a 68.2 percent in 1998 to a low of 35.7 percent in 2006. In short, there was a sharp increase in the lender's willingness to accommodate lower borrower equity in the subprime mortgage market. At the same time, lenders also transferred more of the interest rate risk onto borrowers. As has been shown previously in Table 2, the proportion of loans in FRMs declined whereas those for ARM2 and ARM3 recorded increases.[29]

Table 6 reports the distributions of borrower FICO score conditional on LTV. For a given vintage, mortgages with a smaller LTV have a lower FICO score on average. Just like in the case of loan documentation, there has been a shift of population from the lowest FICO group ($<$ 620) to the two intermediate FICO score groups (620-659 and 660-719) across the three LTV ranges. Finally, Table 7 reports the FICO distribution conditional on three product types: Fixed, ARM2 and ARM3. There is evidence of improvement in FICO scores over time and across all three product types.

The loan characteristics discussed above indicate that over the years, subprime borrowers had lower equity in their homes while moving towards ARMs. We argued earlier that lenders determine the terms and conditions of the mortgage contract based on their assessment of borrower credit risk as evaluated from a profile of borrower attributes. When one looks at the data in conjunction with borrower characteristics, it suggests that these adjustable rate mortgages and mortgages with higher LTV ratios were mostly likely underwritten to borrowers with higher FICO scores.

[28] Wherever available and reported in the FALP database, we use the cumulative LTV.

[29] Campbell and Cocco (2003) describe the FRM, without a prepayment option, as an "extremely risky contract" in terms of wealth risk, whereas the ARM is relatively safe because its real capital value is unaffected by inflation. On the other hand, risk of an ARM is the income risk of short-term variability in the real payments that are required each month. Therefore, it is difficult to assess credit risk purely in terms of product type.

4.3 FICO score and borrower risk profile

Based on the evidence presented in the summary Tables 2-7, it is difficult to argue, as some have claimed, that there was a secular decline in lending standards in terms of a borrower's observable risk characteristics. Despite exposing themselves to more credit risk on some borrower attributes (for example, by lowering documentation requirements), lenders seem to have attempted to offset this by increasing the average quality of borrowers (as measured by their credit scores) to whom such loans were made. To test this hypothesis in a multivariate framework, borrower FICO scores are regressed on other borrower attributes.[30]

The regression estimates presented in Table 8 summarize the equilibrium underwriting standards in terms of borrower characteristics on first lien subprime loans. These coefficients indicate the presence of underwriting efforts to control for overall credit risk by varying credit score requirements on loan approvals. For example, a large negative and significant coefficient on the full-doc dummy indicates that after controlling for other borrower attributes, a borrower with low or no documentation has a significantly higher FICO than a similar borrower providing full documentation on the loan. In the same way, owner-occupied and second home mortgages have significantly lower FICO scores when compared to non-owner (investor) occupied properties. Also, the FICO requirement for loan approval on owner occupied homes is lower than that for mortgages on second homes. Not surprisingly, mortgages on properties with greater value have progressively higher FICO scores. For loans of all vintages, property values in a lower quartile have on average a lower FICO than those property values in the immediately higher quartile. Evidently, refinances have a lower FICO on average than direct home purchases.

In summary, the regression results show that average borrower FICO is significantly higher for borrowers whose other attributes are arguably riskier. Moreover, as evidenced by the larger regression coefficients, the size of this adjustment appears to have increased over the years in our sample period. We conduct a statistical test of this hypothesis by estimating a fully interacted dummy variable model of the FICO equation estimated above. We regress borrower FICO

[30] We control for property type (dummies for single-family residence, condo, townhouse, co-operative, etc), property location (dummies for the state in which the property is located) and loan source (dummies for broker, realtor, wholesale, retail etc).

scores on other borrower attributes for all the vintages pooled together. The dummy variable takes the value one for all originations after a given calendar year, and zero otherwise. We report estimates of four specifications in Table 9, starting with interacted dummy for post-2002 originations and ending with the dummy for post-2005 originations. The estimated coefficient of 20.44 on the dummy variable for post-2004 vintage shows that the improvement in FICO score for originations between 2005 and 2007 was statistically as well as economically significant. In light of this evidence, it is again difficult to argue that there was a "dramatic weakening of underwriting standards" at least in terms of borrower attributes. Noticeably, there is little to suggest anything particularly remarkable about underwriting standards for originations after 2004, as suggested in the Policy Statement.

However, there are some caveats to our results. First, it needs to be reaffirmed that our results compare the underwriting standards for 2005-2007 vintages with those of previous originations, but only for the subprime market and not the mortgage market as a whole. Second, these results do not rule out the possibility that underwriting standards were poor to begin with and that they did not significantly improve on loans of later vintages. Third, based on the analysis thus far, we can only comment on the presence of credible underwriting (i.e., the appropriate sign on the coefficient). We cannot however comment on whether such underwriting was adequate in terms of the marginal rates of adjustment across different borrower attributes (i.e., the magnitude of the coefficient). Stated differently, we observe that the FICO scores on low documentation loans for all the vintages were on average higher than that on full documentation loans. However, we do not know if the difference in FICO of 19.14 points (as recorded on loans of 2006 vintage) as opposed to that of 9.24 points (as recorded on loans of 1998 vintage) was sufficient to cover for the increase in the borrower risk profile due to low documentation on loans. Finally, the preceding analysis seems to indicate a trend towards higher FICO scores alongside lower documentation and higher LTVs. This seems to suggest lenders' emphasis on FICO score as an adequate indicator of credit risk. We address this issue in greater detail in Section 5.2.

5 Results

5.1 The evidence on endogeneity bias

As discussed in Section 3, our test of endogeneity bias is based on the conditional independence between the individual's choice of loan-to-value ratio (coverage), and the ex-post occurrence of the event of delinquency (risk). Table 10 reports the conditional correlation between risk and coverage for all the vintages and the empirical bootstrap confidence interval (1^{st} and 99^{th} percentile). The results shows that the conditional correlation for all vintages is positive and significant. We arrive at the same conclusion if we look at the Chiappori and Salanie (2000) test statistic W in (10).[31]

Most articles on automobile insurance, like Chiappori and Salanie (2000), cannot reject the null of zero correlation between risk and coverage. It appears that for most conventional credit markets, there is little correlation between the coverage of a contract and the ex post riskiness of its subscribers (see references in Chiappori et al., 2006). Therefore, it is perhaps likely that the strong endogeneity bias in subprime markets is sufficiently weaker when it comes to other mortgage markets (like that for prime mortgages). However, for our purposes, the results confirm the endogeneity problem that confronts the use of mortgage characteristics like LTV ratio (and mortgage interest rate) as explanatory variables in determining of loan performance. Therefore, in what follows, we do not include these mortgage terms as explanatory variables.

5.2 Determinants of LTV

The OLS estimates of equation (3) for all loans are given in Table 11. Following our discussion in the previous section, we include all borrower attributes (including FICO score) as explanatory variables, but not mortgage characteristics like loan-to-value ratio and interest rates. In addition, we control for property type, property location and lender type.[32] The estimation results can be summarized as follows:

[31] These results are not reported here but are available on request.
[32] We perform the same exercise separately for ARMs and FRMs. The results are qualitatively similar.

1. We observe a "scale effect" in underwriting which required subprime borrowers in higher valued properties to have lower LTVs. This is reflected in the progressively lower coefficients for properties in higher valued quartiles showing that mortgages on higher property values have on average a lower loan-to-value ratio.

2. Owner-occupied homes have significantly higher LTVs than non-owner occupied homes. Here too, underwriting seems to have succeeded in getting non-owners (i.e., "investors") to make greater downpayment on loans of identical size.

3. Mortgages with full-documentation have significantly higher LTVs than low or no documentation loans. However, the magnitude of this coefficient declines over the sample period. Thus, underwriting attempts at tempering low-documentation loans with lower LTVs on average was getting weaker over the years.

4. Borrowers with lower FICO scores were also the ones with lower LTVs. But here the trend of adjustment of FICO scores with lower LTVs seems to have gotten stronger over the years. For latter vintages (2003-2007) lenders required 4 to 8 percent higher equity investment by the borrower to compensate a drop in FICO score by 100 points.[33]

5. No cash-out refinances have lower LTVs than purchases. This is hardly surprising given the property price inflation for most of our sample period. Also, LTVs are lowest in the case of cash-out refinances and highest for purchases. This result is explained below.

It is interesting to compare the signs of the coefficients in the LTV regression (Table 11) to those in the FICO regression (Table 7). The signs on the coefficients seem to indicate evidence of credible underwriting, given our a priori judgment of risk characteristics. For example, note that while full documentation is associated with a lower FICO score, borrowers providing full documentation on loans are allowed to make a lower downpayment. The important exception is the signs of coefficients on loan purpose. While refinances have lower FICO scores on average, borrowers refinancing loans also have lower LTVs. Typically, loans are refinanced with the original lender and because of a recorded payment history, mortgage refinances are considered to be less risky a priori. This could explain the lower FICO score on refinances. Explaining

[33] We normalize FICO scores by 100.

the LTV result requires a more nuanced view of subprime originations: Gorton (2008) shows that in the event of house price appreciation lenders can benefit even from a refinancing option, so long as the borrower does not extract to the full extent of the appreciated value.[34] This implies that lenders try to ensure that the borrowers retain sufficient equity in the property on a refinance, which could explain why refinances have lower LTVs on average than purchases.

In summary, we find evidence to suggest that the underwriting process attempted to adjust riskier borrower characteristics with lower LTVs. Again, there is little evidence to suggest any dramatic change in underwriting after 2004.

5.3 Default and Delinquency in Subprime Mortgages

Delinquency rates and the probability of surviving a delinquency are calculated by using the Kaplan and Meier (1958) product limit estimator. We begin this strictly empirical, non-parametric approach to survival and hazard function estimation by formalizing it in the current context of mortgage repayment behavior.

Following Kaplan and Meier (1958), the delinquency rate $D(t)$ at month t (the age of the mortgage in months) is defined as

$$D(t) = 1 - P(T > t) \tag{11}$$

where T is the age in months for the delinquency event (60 day, 90 day, or foreclosure) of a randomly selected mortgage and $S(t) \equiv P(T > t)$ is the survivor function or the probability of surviving the delinquency event beyond age t. Let $t_{(1)} < t_{(2)} < ... < t_{(k)}$ represent the ordered age in months at the time of delinquency event. For all these months, let n_i be the number of surviving mortgages just prior to month $t_{(i)}$. Surviving mortgages not only exclude the ones that have been delinquent, but also the ones that have been refinanced prior to age $t_{(i)}$. If d_i

[34] The lender now faces a less risky borrower who has built up equity in the house. Gorton (2008) argues that that subprime mortgages, the majority of which were hybrid-ARMs, were designed "to provide an implicit embedded option on house prices for the lender." Unwilling to speculate on house prices and borrower repayment behavior for long periods, lenders treated subprime mortgages as bridge-financing and sought the option to end the mortgage early. As a result, the fully-indexed rate is designed to be prohibitively high once it resets from the teaser rate, thereby essentially forcing a refinancing.

is the number of mortgages that go delinquent at age $t_{(i)}$, then the Kaplan-Meier estimator of surviving the event of delinquency is defined as

$$\hat{P}(T > t) = \prod_{i=1}^{k}(1 - \frac{d_i}{n_i}) \tag{12}$$

Following industry conventions, we define a mortgage to be in default if it records a 90-day delinquency event at any point in its repayment history.[35] On studying the probability of a 90-day delinquency event it is clear that defaults started to rise sharply in 2006 and 2007, primarily for originations between 2004 and 2007. To give an example, about 21 percent of mortgages originated in 1998 were delinquent by the fifth year (end of calendar year 2002) whereas the same proportion of defaults for 2004 originations occurred within three years (end of calendar year 2006). The numbers are even more striking when one considers that around 35 percent of mortgages originated in 2006 had defaulted by the end of 2007. Most of the commentary on subprime mortgages has sought to explain this significant increase in default probabilities by a weakening in lending standards for originations after 2004.

Table 12 reports the distribution of 90-day delinquency probabilities conditional on borrower FICO scores. The numbers are just as one would expect: delinquency probabilities of lower FICO-score groups for a given vintage are greater than that for higher FICO-score groups of that vintage.[36] Table 12 shows that this feature is true for loans of all vintages and across all four FICO-score groups. While there has been a significant increase in defaults over the years within a given FICO-score group, the trend is hardly monotonic. Almost always, mortgages of 2003 vintage show anomalous behavior that breaks away from this trend. What is perhaps remarkable about these numbers is the consistency that the numbers show across the different vintages and over the age of the mortgage.

At this point, it is important to recall several results from our analysis above. First, our analysis of summary data seems to indicate a trend towards higher FICO scores alongside lower documentation and higher LTVs. Next, we observed that at least in terms of borrower char-

[35] The results for 60-day delinquencies and foreclosures are qualitatively similar and are available on request.

[36] A similar trend is observed when we condition delinquencies on loan-to-value ratios. Again, for the sake of brevity, the results are not reported here but are available on request.

acteristics, average FICO score is significantly higher for borrowers whose other attributes are arguably riskier. Finally, we find evidence to suggest that the underwriting process attempted to adjust riskier borrower characteristics with lower LTVs. An important determinant of this adjustment is borrower FICO score and this adjustment strengthened over the years. Clearly, the evidence seems to indicate that over the years lenders became more willing to trust FICO scores as the best predictor of credit risk. Ex post, some industry experts have even faulted originators on this account:

> ... the crucial mistake many lenders made was relying on FICO credit scores to gauge default risk, regardless of the size of the down payment or the type of loan.[37]

However, if the summary evidence presented in Table 12 provides any indication of credit risk in the subprime market, it appears that the lenders were justified in their assessment.

For a more rigorous study of the determinants of default risk we estimate the hazard function in equation (7). We control for borrower attributes, lender characteristics, property type and property location. Table 13 reports the estimated hazard ratios for the Cox proportional hazard rate regressions conducted for all loans originated in a given calendar year. Our estimates show that a higher FICO score significantly lowers the probability of default.[38] Therefore, for originations in 2002, loans to borrowers with a 100 point higher FICO score reduces the probability of default by 59 percent. Requiring full documentation on the loan of 2002 vintage reduces the probability of default by 18 percent. In the same manner, the likelihood of default on the mortgage is reduced if the property is owner-occupied rather than for investment purposes and if the loan originated is a refinance as opposed to a direct purchase. The results are qualitatively similar for the different product types (FRMs and ARMs) and for originations of different vintages.[39] Clearly, a priori beliefs about the effect of individual borrower characteristics on credit risk turn out to be true: for example, full documentation and a higher credit score each reduce the probability of default.

[37] "The woman who called Wall Street's meltdown" - *Fortune Magazine*, Aug. 4, 2008.

[38] Just as in the previous section, the scores are normalized by dividing them by 100. Therefore, a change in the FICO hazard ratio corresponds to a 100 point increase in FICO score.

[39] The regression results for ARMs and FRMs are available on request.

Viewed independently, equation (7) tells us little about underwriting standards. On the other hand, when these regression results are examined in conjunction with previous regression results on borrower FICO and LTV of the mortgage, we are able to get a clearer picture of underwriting standards. Earlier, we showed evidence to suggest that the underwriting process attempted to adjust riskier borrower characteristics with higher FICO (Section 4.3) and lower LTVs (Section 5.2). Our earlier results also show that lenders adjusted higher LTVs with higher FICO scores and that the strength of adjustment increased over the years. Now, the hazard rate regressions show that ceteris paribus, FICO scores are an important determinant of ex post default. Taken together, there is strong evidence of credible mortgage underwriting: lenders tried to offset greater risk in terms of higher LTV and lower documentation by raising FICO scores at the time of loan origination because FICO scores are an important determinant of ex post default.

6 Counterfactual Analysis

Our analysis up to this point establishes that there has been no unequivocal decline in lending standards. However it does not establish that, when viewed from the standpoint of ex post default, aggregate lending standards did not decline. At the heart of this is the problem of aggregating a multidimensional profile of borrower attributes to a single metric that could summarize the overall credit risk of the borrower.

In this section, we attempt to get around this problem by using a counterfactual exercise. In so doing, we answer the following question: how would ex post default rates change if a mortgage originated to a "representative borrower" in 2005 were to be given a loan in 2001? To this end, we estimate the proportional hazard rate model for a particular vintage and then use the estimated relationship to evaluate the *estimated proportional hazard survivorship function* for a representative borrower from a different vintage (see Cameron and Trivedi, 2006 for further details).[40]

[40] Needless to say, the results of this counterfactual analysis are sensitive to the definition of the "representative borrower" of a particular vintage.

Let v be the index of vintage, $S_{v,0}(t)$ be the baselive survivor function, and $\mathbf{X}$ be the observable characteristic of the "representative borrower" of vintage v. The survivor function $S_v(t)$ for any vintage v and age of mortgage t, is the outcome of a mapping of observable borrower characteristics $\mathbf{X}$, and unobservable characteristics and market conditions captured by baseline survivor function $S_{v,0}(t)$.

$$S_v(t) = f\left(S_{v,0}(t), \mathbf{X}\right)$$

function f maps $\left(S_{v,0}(t), \mathbf{X}\right)$ into the range of $S_v(t)$.

For our purposes, the objective is to forecast the impact on the survivor function of vintage v_2 in the environment of vintage v_1.[41] In this specification, let $\mathbf{X}_1$ and $\mathbf{X}_2$ denote the "representative borrowers" of vintage v_1 and v_2 repsectively. If unobservable characteristics and market conditions captured by the baseline survivor function are applied on the different borrower characteristics, we can identify the effect of $\mathbf{X}_2$ on the survivor function in v_1 as follows:

$$S_{v_1}^{v_2}(t) = f\left(S_{v_1,0}(t), \mathbf{X}_2\right)$$

Such a counterfactual exercise helps us in testing the following hypothesis:

Null Hypothesis: Let $S_v(t)$ be the survivor function for vintage v and age of mortgage t, and $S_v^{\widetilde{v}}(t)$ be the counterfactual survivor function which is the result of the forecasting problem described above, then $S_v(t) \approx S_v^{\widetilde{v}}(t)$, for all t.

We proceed as follows. First, we estimate the Cox proportional hazard model in (7) for a given vintage v. Next, we calculate the estimated survivor function for the representative borrower of vintage v. Finally, we calculate the counterfactual survivor function for the representative borrower of a different vintage, say $\widetilde{v}$. Since our representative borrower is constructed to best reflect borrower characteristics of a particular vintage, we define characteristics of this representative borrower as follows. Any attribute of the representative borrower of vintage v is calculated as the average of the values of the attribute of all borrowers who originated loans in

[41] This problem is similar to **P-2** on program evaluation in Heckman and Vyltacil (2007).

year v. Therefore, if 28.6 percent of the sample had low or no documentation loans in 2002, the value of the "dummy" variable on documentation for 2002 vintage would be 0.286. Clearly, this is an oddity, but it is a simple way of summarizing the distribution of borrower characteristics.[42]

With these tools in place, we can now use our counterfactual analysis to test the null hypothesis that there was no dramatic weakening of underwriting standards beginning around late 2004. The null hypothesis is that mortgages approved after 2004 are equally likely to survive an event of default than those of earlier vintages, namely 2001, 2002 or 2003, in the environment of these vintages. The results of counterfactual analysis are summarized in Table 14 and Figures 2-4. Table 14 has three panels corresponding to the counterfactual exercises using survivor function estimates based on 2001, 2002, and 2003 data. The numbers in parentheses are the 95% confidence intervals for the estimated survivor function. The results show that if a representative borrower in 2006 (likewise for 2005 and 2007) had originated mortgages in 2001 and 2002, she would have performed significantly better than representative borrowers of vintages 2001 and 2002 respectively (Figure 2-3). Thus we can reject the null hypothesis in favour of the alternative that the underwriting standards actually improved in the latter vintages when compared to 2001 and 2002 vintages.

Conversely, the counterfactual using 2003 estimates shows that the loan performance of the representative borrower of 2006 vintage would have been worse than that of the representative borrower of the current (2003) vintage. However, there are no statistically significant differences in the loan performances between the representative borrowers of 2005 or 2007 vintages and that of the 2003 vintage (Figure 4). Therefore, in this case, we fail to reject the null hypothesis. The counterfactual analysis is strong evidence against the hypotheses of any weakening of underwriting standards.

[42] Needless to say, the results of this counterfactual analysis are sensitive to the definition of the "representative borrower" of a particular vintage. To test the robustness of our results, we adopt an alternative procedure. We adopt the first step as before. In the second step, we recover the estimated survivor function for all borrowers in year v. In the third step, we calculate the counterfactual survivor function for all borrowers who originated loans in year $\tilde{v}$. A final step involves averaging across all borrowers of a given vintage to obtain the actual and the counterfactual survivor functions for years v and $\tilde{v}$ respectively. The results are qualitatively similar.

7 Conclusion

We begin with pointing out some of the limitations in our study. First, it is extremely important to state that our conclusions are drawn from data available at the time of loan origination. Subsequent behavior of the borrower (e.g. originating a second lien on the property) is undeniably important in determining ex post delinquency and default. However, given our current limitations on data, we hope that this aspect would be covered in future lines of research.

Second, as with any empirical study, there is of course the possibility that there were borrower attributes observed by the lender, but that are not reported in the FALP data. Lack of data often makes it difficult to make a conclusive argument on some important characteristics, like for example, the debt to income ratio. Using HMDA data, Mian and Sufi (2008) report that aggregate mortgage debt to income ratios for entire zip codes have increased significantly in the borrower population. However, using the debt to income ratios in the FALP database on individual mortgages creates significant problems. First, there is almost no data on the front-end debt to income ratio. Second, even for the back-end ratio, the field is sparsely populated for earlier vintages in the FALP data. For the data that is available, we observe a trend of increasing (back-end) debt-to-income ratios. Again, our regression results show attempts to control for this increase by increasing other borrower attributes, namely the FICO score.

Third, some observers could raise the doubts about the veracity of the data. There is some anecdotal evidence that points to poor reporting and false documentation.[43] However, it is difficult to make this case over a repository of nine million loan observations.

Finally it needs to be mentioned that our examination of the underwriting standards is at level of the individual borrower and not at the level of the lending institution. We do not examine the hypothesis if, for example, originations of high-LTV mortgages were disproportionately high for a particular lending institution.

Nevertheless, this paper presents a contrarian perspective on underwriting standards in the subprime market. Our examination of the FALP data shows scant evidence of a decline

[43] Federal investigators are probing into allegations of fraud and misrepresentations by mortgage companies like Countrywide Financial Corp. See for example, "Loan Data Focus of Probe," Wall Street Journal, March 11 2008.

in underwriting standards. Moreover, our counterfactual analysis demonstrates that, at least on average, we can reject the hypothesis of no decline in underwriting standards in favour of improvement in underwriting standards. Of course, we cannot reject the premise that underwriting standards in the subprime market were poor to begin with. However, all this leads to the obvious questions as to what sustained the phenomenal growth in the subprime market for nearly a decade. And, of course, why did the subprime market collapse?

In a companion paper, Bhardwaj and Sengupta (2008), we attempt to answer these questions in sufficient detail. Following Gorton (2008), we argue that the subprime mortgage contracts were designed as "bridge-finance", providing the borrowers the incentive to graduate into a prime mortgage by building equity on their homes and improving their credit records. Bhardwaj and Sengupta (2008) find that, in the early years, a significant proportion of subprime mortgages were prepaid around the reset date. These prepayments were largely sustained by the boom in house prices in the United States from 1995 to 2006.

References

Ambrose, B. W., and Sanders, A.B. (2003). Commercial mortgage-backed securities: prepayment and default. Journal of Real Estate Finance and Economics, 26(2-3), 179-196.

Archer, W. R., Elmer, P., Harrison, D. M., and Ling, D. C. (2002). Determinants of multifamily mortgage default. Real Estate Economics, 30(3), 445-473.

Bhardwaj, G. and Sengupta, R. (2008). Did Prepayments Sustain the Subprime Mortgage market? Working Paper.

Brueckner, J. K. (2000). Mortgage default with asymmetric information. Journal of Real Estate Finance and Economics, Springer, 20(3), 251-74.

Campbell, J. Y. and Cocco, J. F. (2003). Household Risk Management And Optimal Mortgage Choice. The Quarterly Journal of Economics, 118(4), 1449-1494.

Chen, J., and Deng, Y. (2003). Commercial mortgage workout strategy and conditional default probability: evidence from special serviced CMBS loans. University of Southern California Lusk Center for Real Estate Working Paper No. 2003-1008.

Chiappori, P., and B. Jullien, B. Salanié and F. Salanié, (2006). "Asymmetric Information in Insurance: General Testable Implications", Rand Journal of Economics, 37(4), 2006.

Chiappori, P., and Salanie, B. (2000). Testing for asymmetric information in insurance markets. Journal of Political Economy, 108(1), 56-78.

Chomsisengphet, S. and Pennington-Cross, A., (2006). The Evolution of the Subprime Mortgage Market. Federal Reserve Bank of St. Louis Review, vol. 88(1), 31-56.

Ciochetti, B. A., Deng, Y., Gao, B., and Yao, R. (2002). The termination of commercial mortgage contracts through prepayment and default: a proportional hazards approach with competing risks. Real Estate Economics, 30(4), 595-633.

Ciochetti, B. A., Deng, Y., Lee, G., Shilling, J., and Yao, R. (2003). A proportional hazards model of commercial mortgage default with originator bias. Journal of Real Estate Finance and Economics, 27(1), 5-23.

Cox, D.R. (1972). Regression Models and Life-Tables (with Discussion). Journal of the Royal Statistical Society, Series B, 34, 187–220.

Cutts, A. and Van Order, R. (2005). On the economics of subprime lending. Journal of Real Estate Finance and Economics, 30(2), 167-196.

Demyanyk, Y. and Van Hemert, O. (2007). Understanding the subprime mortgage Crisis. Working Paper.

Fabozzi, F.J. ed. (2006). The Handbook of Mortgage-Backed Securities. New York: McGraw-Hill.

Gorton, G. (2008). The Panic of 2007. Manuscript Prepared for the Federal Reserve Bank of Kansas City, Jackson Hole Conference, August 2008 .

Grovenstein, R., Harding, J. P., Sirmans, C.F, Thebpanya, S., and Turnbull, G. K. (2005). Commercial mortgage underwriting: How well do lenders manage the risks? Journal of Housing Economics, 14(4), 355-383.

Heckman, J.J. and E.J. Vyltacil (2007), "Econometric evaluation of social programs, part I: causal models, structual models and econometric policy evaluation", Handbook of Econometrics, Vol 6B, Elsevier.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. Journal of the American Statistical Association, 53, 457-48.

Keys, B., Mukherjee, T., Seru, A. and Vig, V. (2008). Securitization and screening: evidence from subprime mortgage backed securities, mimeo, London Business School.

Kindleberger, C., and Aliber, R. (2005). Manias, panics, and crashes: a history of financial crises (Wiley Investment Classics). New York, NY: Wiley.

Lamont, O., and Stein, J. (1999). Leverage and house-price dynamics in U.S. cities. Rand Journal of Economics, 30(3), 498-514.

Maddala, G. S. (1983). Limited dependent and quantitative variables in econometrics. Cambridge: Cambridge University Press.

Mayer C. and Pence K. (2008) Subprime Mortgages: What, Where and to Whom?, NBER Working Paper No. W14083.

Mian, A., and Sufi, A. (2007). The consequences of mortgage credit expansion: evidence from the 2007 mortgage default crisis, unpublished manuscript, University of Chicago Graduate School of Business.

President's Working Group on Financial Markets (2008). Policy Statement on Financial Market Developments.

Quercia, R. G., and Stegman. S. (1992) Residential Mortgage Default: A Review of the Literature. Journal of Housing Research, Vol 3(2), 341–79.

Reinhart, C. M., and Rogoff S. K. (2008a), Is the 2007 U.S. Subprime Crisis So Different? An International Historical Comparison, American Economic Review, forthcoming.

Reinhart, C. M., and Rogoff S. K. (2008b), This Time is Different: A Panoramic View of Eight Centuries of Financial Crises, NBER Working Paper No. W13882.

Sengupta, R. (2007). Foreign Entry and Bank Competition, Journal of Financial Economics, 84(2), 502-28.

Stein, J. C. (1995). Prices and Trading Volume in the Housing Market: A Model with Down-Payment Effects. The Quarterly Journal of Economics, MIT Press, 110(2), 379-406.

Vandell, K. D., Barnese, W., Hartzell, D., Kraft, D., Wendt, W. (1993). Commercial Mortgage Defaults: Proportional Hazards Estimation Using Individual Loan Histories. Journal of the American Real Estate & Urban Economics Association, 21(4) 451-480.

Figure 1: Distribution of FICO scores for first-lien subprime loans

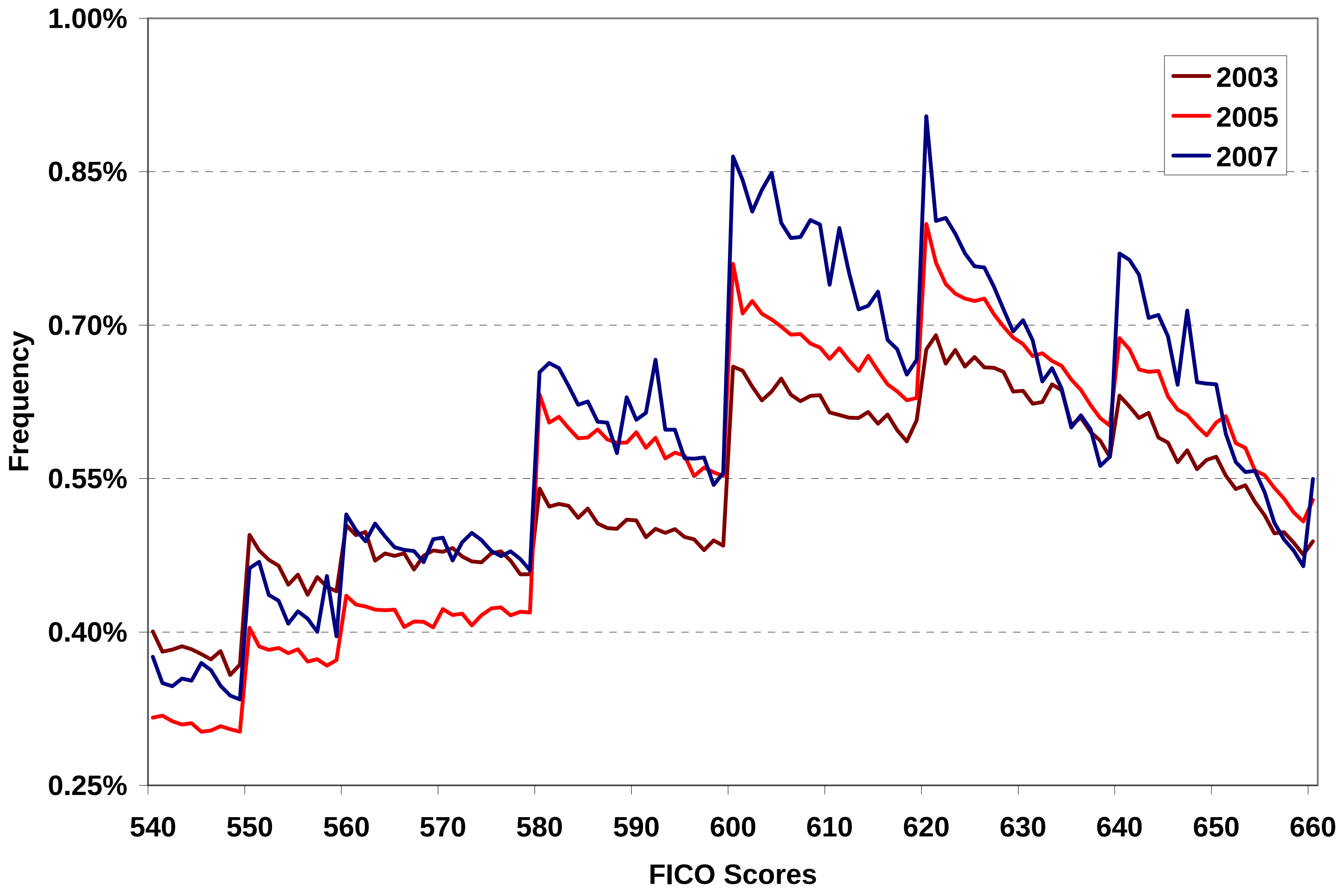

Table 1: Mortgage Pricing Sheet, Option one Mortgage Corporation

Rate sheet is for five year fixed mortgage with two year prepays charge. The worksheet assumes full documentation, one unit house, and loan amount in the range \$200,000 - \$417,000. In case of secondary financing (CLTV > LTV) and credit score less than 660 (or >=660) rate is adjusted upwards by 155 basis points (or 90 basis points).

		LTV			
Grade	Credit Score	**65%**	**70%**	**75%**	**80%**
	700+	8.65	8.70	8.80	8.90
	660	8.75	8.80	8.90	9.00
AA+	620	9.00	9.05	9.15	9.25
	580	9.55	9.60	9.90	10.05
	540	10.45	10.70	10.90	11.15
	700+	9.35	9.40	9.50	9.60
	660	9.45	9.50	9.60	9.70
AA	620	9.70	9.75	9.85	9.95
	580	10.15	10.20	10.35	10.50
	540	10.70	10.95	11.00	11.25
	700+	9.45	9.50	9.60	9.70
	660	9.55	9.60	9.70	9.80
A	620	9.80	9.85	9.95	10.05
	580	10.25	10.30	10.45	10.60
	540	10.80	11.05	11.10	11.35
	700+	9.85	9.95	10.10	10.25
	660	10.05	10.15	10.35	10.45
B	620	10.40	10.55	10.75	10.80
	580	10.95	11.00	11.25	11.35
	540	11.55	11.7	11.95	

Option One Mortgage Corporation, west area rate sheet, effective 11/09/2007, downloaded on 07/03/2008, http://www.oomc.com/broker/broker_rateguide.asp

Table 2: Evolution of the Subprime Market (market share by product type).

Table summarizes first lien subprime mortgages by product type as fixed or adjustable rate mortgages (FRM and ARM) for every year of origination from 1998 to 2007. The numbers give us the market share for a particular product type. ARM2 and ARM3 denote hybrid-ARM products where the *teaser rate* is fixed for two and three years respectively. *Other* product types include ARM-other, Balloon, Two-Step, GPM, GEM and GPARM. The total number denotes the number of originations in each category.

Vintage	FRM	ARM2	ARM3	Other	Total Number
1998	51.34	26.55	4.52	17.59	252945
1999	38.87	29.35	19.21	12.57	369373
2000	32.58	43.29	14.78	9.35	399342
2001	31.69	48.69	12.44	7.17	498462
2002	28.36	54.85	12.63	4.16	755233
2003	33.57	52.60	11.38	2.45	1265536
2004	23.81	59.74	14.65	1.80	1921557
2005	18.69	65.43	13.24	2.64	2258155
2006	19.95	62.54	10.9	6.61	1766939
2007	27.55	50.64	10.32	11.49	315921
Total Number	2519608	5561509	1248407	473939	9803463

Table 3: Distribution of Document type and FICO by vintage

Borrower credit score at the time of loan origination is denoted by *FICO* (an industry standard developed by the Fair Isaac Corporation) with a number in the range 300-850. Loans coded by the source as with a non-blank documentation code are classified as *Full-doc* whereas those under a No doc program or prospectus are classified as *No doc*. Others are classified as *Low doc*.

	FICO				Documentation level		
Vintage	< 620	620-659	660-719	>= 720	Full doc	Low doc	No doc
1998	63.5%	19.6%	12.6%	4.3%	76.9%	22.8%	0.3%
1999	64.8%	19.1%	12.3%	3.9%	81.3%	18.4%	0.3%
2000	69.4%	17.8%	9.9%	2.8%	79.4%	19.9%	0.7%
2001	63.7%	20.3%	12.1%	3.9%	76.7%	22.9%	0.4%
2002	58.0%	22.2%	14.8%	5.0%	71.4%	28.0%	0.6%
2003	51.7%	23.9%	17.8%	6.6%	68.2%	31.2%	0.6%
2004	51.8%	24.3%	17.9%	6.1%	66.4%	33.3%	0.3%
2005	49.8%	25.7%	18.5%	6.0%	63.3%	36.5%	0.2%
2006	51.7%	27.0%	16.6%	4.8%	62.7%	37.0%	0.3%
2007	54.8%	26.4%	15.1%	3.7%	66.6%	33.1%	0.3%

Table 4: FICO distribution conditional on documentation level on loan by vintage

	Full doc loans				Low doc or No doc loans			
Vintage	< 620	620-659	660-719	>= 720	< 620	620-659	660-719	>= 720
1998	65.6%	18.9%	11.5%	4.0%	56.7%	21.7%	16.2%	5.4%
1999	67.4%	18.4%	10.9%	3.3%	53.3%	22.1%	18.2%	6.4%
2000	72.1%	16.9%	8.6%	2.4%	59.1%	21.3%	15.0%	4.6%
2001	67.8%	18.8%	10.0%	3.3%	50.2%	25.2%	18.7%	5.8%
2002	64.4%	20.2%	11.4%	4.0%	42.1%	27.2%	23.2%	7.5%
2003	58.4%	22.2%	13.9%	5.4%	37.3%	27.4%	26.2%	9.1%
2004	58.8%	22.5%	13.7%	5.0%	38.0%	27.8%	26.1%	8.1%
2005	58.7%	23.2%	13.6%	4.5%	34.4%	30.1%	26.9%	8.6%
2006	61.2%	23.8%	11.6%	3.4%	35.7%	32.3%	25.0%	7.1%
2007	60.9%	24.6%	11.6%	2.9%	42.5%	30.1%	22.0%	5.4%

Table 5: Distribution of Cumulative loan to Value Ratio (CLTV) by Vintage

Borrower credit score at the time of loan origination is denoted by FICO (an industry standard developed by the Fair Isaac Corporation) with a number in the range 300-850.

Vintage	CLTV $\leq$ 80	80 < CLTV $\leq$ 90	90 < CLTV $\leq$ 100	CLTV > 100
1998	68.2%	27.5%	3.2%	1.1%
1999	66.9%	28.7%	3.7%	0.7%
2000	63.4%	30.1%	6.3%	0.2%
2001	57.1%	33.4%	9.3%	0.2%
2002	55.1%	32.9%	12.0%	0.1%
2003	47.4%	30.3%	22.2%	0.1%
2004	42.8%	27.9%	29.2%	0.1%
2005	39.7%	25.3%	35.0%	0.1%
2006	35.7%	23.7%	40.6%	0.1%
2007	42.7%	29.4%	27.9%	0.0%

Table 6: Distribution of FICO scores conditional on CLTV by vintage

	CLTV ≤ 80				80 < CLTV ≤ 90				90 < CLTV ≤ 100			
Vintage	< 620	620-659	660-719	>= 720	< 620	620-659	660-719	>= 720	< 620	620-659	660-719	>= 720
1998	63.2%	18.4%	12.5%	5.9%	61.9%	21.1%	12.5%	4.4%	52.1%	22.2%	17.2%	8.4%
1999	65.1%	18.0%	12.1%	4.8%	63.9%	20.6%	11.8%	3.7%	44.2%	23.5%	23.0%	9.2%
2000	70.4%	16.5%	9.9%	3.2%	71.1%	18.1%	8.6%	2.2%	48.1%	29.3%	17.1%	5.6%
2001	66.0%	18.1%	11.7%	4.2%	65.8%	21.0%	10.6%	2.6%	43.9%	30.8%	18.9%	6.3%
2002	62.0%	19.3%	13.6%	5.2%	61.8%	21.9%	12.9%	3.4%	30.1%	36.1%	25.2%	8.5%
2003	59.2%	19.4%	15.1%	6.3%	55.8%	23.7%	15.7%	4.7%	30.2%	33.6%	26.5%	9.7%
2004	61.9%	19.2%	13.7%	5.2%	57.5%	23.2%	15.0%	4.3%	31.0%	32.9%	27.0%	9.0%
2005	60.6%	20.6%	13.8%	5.0%	55.9%	23.6%	15.8%	4.7%	32.8%	33.2%	25.9%	8.2%
2006	65.0%	19.7%	11.4%	3.9%	60.3%	23.3%	12.9%	3.4%	34.9%	35.4%	23.3%	6.4%
2007	68.2%	19.2%	9.8%	2.7%	57.6%	26.4%	13.2%	2.7%	32.1%	37.2%	24.6%	6.1%

Table 7: FICO distribution conditional on Product type

	Fixed				ARM 2				ARM 3			
Vintage	< 620	620-659	660-719	>= 720	< 620	620-659	660-719	>= 720	< 620	620-659	660-719	>= 720
1998	59.5%	20.7%	14.3%	5.5%	68.6%	18.2%	10.4%	2.8%	68.1%	18.7%	10.4%	2.8%
1999	59.7%	20.7%	14.6%	5.1%	71.2%	17.1%	9.4%	2.3%	68.2%	18.5%	10.6%	2.7%
2000	63.8%	20.2%	12.0%	4.0%	72.8%	16.5%	8.7%	2.0%	71.1%	17.3%	9.3%	2.4%
2001	54.7%	22.7%	16.2%	6.4%	69.6%	18.7%	9.4%	2.2%	65.3%	20.7%	11.1%	2.9%
2002	43.6%	25.5%	22.1%	8.8%	65.6%	20.0%	11.3%	3.1%	60.5%	24.1%	12.1%	3.2%
2003	38.9%	25.7%	24.5%	10.9%	59.4%	22.7%	13.9%	4.0%	57.3%	24.5%	14.3%	3.9%
2004	41.7%	25.8%	22.5%	10.0%	56.0%	23.6%	15.9%	4.5%	54.7%	24.4%	16.5%	4.4%
2005	45.6%	26.3%	20.3%	7.8%	51.9%	25.7%	17.4%	5.0%	49.8%	25.4%	18.9%	5.9%
2006	50.4%	26.3%	17.3%	6.0%	53.9%	27.0%	15.3%	3.8%	48.9%	27.7%	18.1%	5.2%
2007	56.3%	25.1%	14.7%	3.9%	56.1%	27.0%	14.0%	2.9%	50.8%	27.6%	17.1%	4.5%

Table 8: Regression of Credit Score (FICO) on Other Borrower Characteristics

Table reports OLS estimates with borrower FICO score as the left-hand side variable and other borrower characteristics as regressors. We control for property type (dummies for single-family residence, condo, townhouse, co-operative, etc), property location (dummies for the state in which the property is located) and loan source (dummies for broker, realtor, wholesale, retail etc.). *Home Value n*th *Quartile* is a dummy that equals one if the value of the property lies in the *n*-th quartile of all property values in the data and zero otherwise.

	Vintage									
Variable	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**
Intercept	664.59***	658.6***	644.64***	667.04***	698.15***	717.3***	679.74***	702.39***	699.13***	704.13***
Number of Units	-0.03	6***	4.1***	2.45***	4.51***	-2.49***	-0.91***	-0.94***	2.22***	-5.09***
Full- Doc	-9.24***	-16.81***	-15.15***	-18.48***	-22.04***	-19.41***	-17.69***	-18.83***	-19.14***	-16.83***
Owner Occupied	-24.74***	-25.09***	-26.81***	-24.32***	-27.59***	-32.44***	-33.6***	-32.15***	-31.46***	-32.42***
Second Home	-12.43***	-8.96***	-3.73***	-3.12***	-8.49***	-13.02***	-14.39***	-7.72***	-8.62***	-15.71***
Refinance (Cash Out)	-6.71***	-9.79***	-16.95***	-16.75***	-27.99***	-34.35***	-37.01***	-34.29***	-32.98***	-31.67***
Refinance (No Cash Out)	-5.64***	-12.5***	-19.13***	-17.8***	-20.21***	-22.09***	-22.24***	-19.46***	-18.43***	-23.68***
Home Value First Quartile	-9.14***	-7.43***	-7.26***	-13.29***	-11.24***	-13.53***	-13.09***	-14.14***	-13.88***	-12.74***
Home Value Second Quartile	-5.51***	-5.02***	-5.36***	-9.17***	-7.35***	-8.87***	-8.21***	-8.27***	-8.76***	-8.28***
Home Value Third Quartile	-3.63***	-3.16***	-3.59***	-5.47***	-5.78***	-7.24***	-6.69***	-6.3***	-6.59***	-5.34***
Property type Dummies	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included
Lender type Dummies	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included
Property State dummies	Included	Included	Included	Included	Included	Included	Included	Included	Included	Included
Adjusted R-Square	0.036	0.058	0.077	0.088	0.134	0.153	0.168	0.169	0.175	0.148

The symbols ***, ** and * denote statistical significance at 1-percent, 5-percent and 10-percent levels respectively.

Table 9: Fully Interacted dummy variable Regression of Credit Score (FICO) on Other Borrower Characteristics

Table reports OLS estimates of a filly interacted dummy variable regression of borrower FICO scores on other borrower attributes, for all the vintages pooled together; the dummy variable is turned on for latter vintages. We report four versions of this equation where dummy variable is turned on for post-2002 to post-2005 vintages.

	Dummy = 1 if vintage			
Variable	>=2003	>=2004	>=2005	>=2006
Intercept	671.27***	680.09***	680.64***	685.51***
Dummy	20.45***	8.49***	20.44***	14.94***
Number of Units	4.53***	3.05***	0.57***	0.24***
Number of Units x Dummy	-5.41***	-3.8***	-1.01***	0.89***
Full- Doc	-19.52***	-20.76***	-20.02***	-20***
Full- Doc x Dummy	0.94***	2.38***	1.22***	1.19***
Owner Occupied	-26.39***	-28.45***	-30.51***	-30.84***
Owner Occupied x Dummy	-6.16***	-4.09***	-1.36***	-0.71***
Second Home	-8.86***	-10.98***	-12.51***	-10.3***
Second Home x Dummy	-1.93***	0.560	3.87***	0.610
Refinance (Cash Out)	-18.32***	-24.31***	-29.66***	-31.3***
Refinance (Cash Out) x Dummy	-16.4***	-10.45***	-4***	-1.56***
Refinance (No Cash Out)	-16.58***	-19.45***	-22.37***	-22.77***
Refinance (No Cash Out) x Dummy	-4.1***	-1.05***	2.7***	2.96***
Home Value First Quartile	-8.18***	-7.45***	-8.17***	-9.34***
Home Value First Quartile x Dummy	-5.33***	-6.11***	-5.7***	-4.29***
Home Value Second Quartile	-5.08***	-4.46***	-4.87***	-5.47***
Home Value Second Quartile x Dummy	-3.22***	-3.8***	-3.5***	-3.18***
Home Value Third Quartile	-6.64***	-6.18***	-5.46***	-4.87***
Home Value Third Quartile x Dummy	-3.82***	-4.29***	-4.85***	-5.12***
Adj R-Sq	0.1545	0.1478	0.1435	0.1416

The symbols ***, ** and * denote statistical significance at 1-percent, 5-percent and 10-percent levels respectively.

Table 10: Test of endogeneity bias and bootstrap confidence interval

Tabulated entries are estimated correlation (conditional on observables) between risk and coverage, and 1%, 99% level bootstrap critical values. Results are based on 100 bootstrap replications.

Vintage	1% CV	Correlation Coefficient	99% CV
1998	0.033	0.038	0.043
1999	0.038	0.041	0.045
2000	0.039	0.042	0.046
2001	0.056	0.060	0.063
2002	0.057	0.059	0.061
2003	0.067	0.069	0.071
2004	0.089	0.090	0.091
2005	0.127	0.134	0.138
2006	0.167	0.168	0.171
2007	0.150	0.153	0.157

Table 11: Determinants of Loan to Value Ratio

The dependent variable here is the loan to value ratio at the time of origination. We control for property type (dummies for single-family residence, condo, townhouse, co-operative, etc), property location (dummies for the state in which the property is located) and loan source (dummies for broker, realtor, wholesale, retail etc.). *Home Value n*th *Quartile* is a dummy that equals one if the value of the property lies in the *n*-th quartile of all property values in the data and zero otherwise.

Variable	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**
FICO	0.71***	1.16***	1.73***	2.01***	3.02***	3.5***	4.41***	4.86***	5.23***	6.22***
Number of Units	0.02	0.06	0.45***	0.71***	0.98***	0.62***	0.04***	0.14***	0.16***	-0.8***
Full- Doc	4.75***	4.27***	5.59***	4.49***	3.34***	2.84***	1.74***	1.34***	0.95***	1.61***
Owner Occupied	3.72***	4.08***	4.15***	4.53***	4.75***	5.67***	5.47***	5.08***	5.5***	6.01***
Second Home	-2.81***	-1.51***	-0.48*	-1.49***	-0.29	-0.8***	-0.77***	0.06	0.17*	0.5*
Refinance (Cash Out)	-7.17***	-7.18***	-8.06***	-8.27***	-7.55***	-10.35***	-11.25***	-12.26***	-13.88***	-13.57***
Refinance (No Cash Out)	-4.53***	-4.9***	-5.99***	-6.03***	-5.1***	-8.25***	-9.4***	-9.09***	-9.86***	-10.69***
Home Value First Quartile	1.01***	0.15*	0.05	2.36***	3.4***	4.62***	4.19***	3.6***	2.94***	4.07***
Home Value Second Quartile	1.25***	0.69***	0.76***	2.5***	3.25***	4.07***	3.62***	2.93***	2.35***	2.89***
Home Value Third Quartile	0.95***	0.51***	0.69***	2.27***	2.91***	3.05***	2.46***	1.48***	1.27***	1.86***
Adjusted R-Squared	0.10	0.10	0.14	0.15	0.16	0.24	0.29	0.31	0.34	0.31

The symbols ***, ** and * denote statistical significance at 1-percent, 5-percent and 10-percent levels respectively.

Table 12: Probability of a 90 day delinquency conditional on FICO

Table reports the delinquency rate for all the vintages, for loans grouped by their FICO score. Delinquency rate is defined in section 5.2 as one minus Kaplan and Meier (1958) survivor function. Delinquency rate is thus one minus the probability of surviving the delinquency event beyond the given age in months.

	FICO: < 620											FICO: 620-659										
	Calendar Year Ending											Calendar Year Ending										
Vintage	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	June-2008	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	June-2008
1998	3.7%	10.6%	16.5%	22.0%	26.2%	30.3%	34.8%	39.4%	44.7%	49.3%	51.4%	1.5%	5.3%	9.2%	13.0%	16.2%	19.3%	23.0%	26.5%	29.7%	32.9%	35.6%
1999		4.0%	11.4%	18.6%	24.6%	30.1%	35.4%	40.9%	45.3%	50.0%	52.1%		1.4%	5.4%	10.0%	14.3%	18.5%	22.6%	27.2%	30.4%	34.0%	36.1%
2000			5.8%	15.3%	23.2%	31.2%	37.9%	44.1%	49.5%	54.8%	57.2%			1.8%	6.7%	12.5%	18.6%	24.2%	29.6%	34.1%	38.2%	41.3%
2001				4.9%	14.3%	24.3%	33.5%	41.8%	48.2%	53.8%	56.4%				2.0%	7.0%	14.3%	21.7%	28.5%	33.9%	38.5%	41.9%
2002					4.2%	14.2%	25.1%	36.1%	43.9%	50.7%	53.3%					1.9%	7.8%	15.3%	23.6%	30.3%	35.8%	38.4%
2003						3.7%	12.6%	23.7%	33.3%	41.0%	43.6%						1.7%	6.7%	14.0%	20.7%	26.9%	29.0%
2004							4.9%	15.9%	28.7%	42.0%	46.1%							2.3%	8.5%	18.5%	30.4%	34.2%
2005								7.0%	22.4%	44.2%	51.8%								4.0%	15.8%	38.9%	46.5%
2006									12.0%	38.9%	49.8%									9.7%	35.0%	48.0%
2007										15.6%	29.7%										11.8%	25.4%

	FICO: 660-719											FICO: >= 720										
	Calendar Year Ending											Calendar Year Ending										
Vintage	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	June-2008	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	June-2008
1998	0.9%	3.5%	6.0%	8.4%	10.8%	13.0%	15.3%	17.6%	20.0%	22.5%	24.2%	0.5%	1.6%	2.6%	3.4%	4.2%	5.3%	6.4%	7.9%	9.4%	11.9%	-
1999		0.8%	2.9%	5.9%	9.1%	12.4%	15.6%	19.0%	21.4%	24.7%	26.0%		0.7%	1.7%	3.1%	4.3%	5.8%	7.9%	10.1%	12.0%	13.1%	-
2000			1.2%	4.5%	8.2%	12.1%	16.4%	21.2%	24.7%	28.5%	30.8%			0.9%	3.0%	5.5%	7.8%	10.2%	12.7%	15.1%	17.8%	18.7%
2001				1.2%	4.3%	9.0%	13.6%	18.5%	22.5%	26.1%	27.8%				1.1%	2.6%	4.9%	7.2%	9.0%	10.4%	12.1%	12.3%
2002					1.2%	4.9%	9.6%	14.8%	18.7%	22.8%	23.9%					1.0%	2.9%	5.1%	7.1%	9.1%	11.4%	11.6%
2003						1.1%	4.0%	8.1%	11.8%	15.6%	17.0%						0.8%	2.2%	4.2%	5.8%	7.3%	7.9%
2004							1.4%	5.4%	12.3%	21.1%	23.5%							1.2%	3.5%	7.1%	11.1%	12.0%
2005								2.6%	10.8%	32.5%	39.9%								2.2%	7.6%	23.1%	26.9%
2006									7.4%	29.8%	44.4%									6.7%	23.1%	33.9%
2007										10.4%	22.2%										9.2%	17.3%

Table 13: Estimated Cox proportional hazard rate regression: Hazard Ratio for 90 day delinquency event

This table reports the estimated hazard ratios for the Cox proportional hazard rate regressions conducted for all loans originated in a given calendar year. We control for property type (dummies for single-family residence, condo, townhouse, co-operative, etc), property location (dummies for the state in which the property is located) and loan source (dummies for broker, realtor, wholesale, retail etc.). *Home Value n*th *Quartile* is a dummy that equals one if the value of the property lies in the *n*-th quartile of all property values in the data and zero otherwise.

Variable	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**
FICO	0.5327***	0.4824***	0.4484***	0.4386***	0.4115***	0.3645***	0.3919***	0.4682***	0.5131***	0.5165***
Number of Units	1.0635***	0.9637***	1.0163***	1.0269***	1.0523***	1.126***	1.0044***	1.0165***	1.0363***	1.0576***
Full- Doc	0.8332***	0.8945***	0.8741***	0.8626***	0.82***	0.7458***	0.7487***	0.6799***	0.6269***	0.6297***
Owner Occupied	0.8478***	0.8045***	0.8097***	0.8051***	0.8181***	0.7853***	0.7482***	0.7614***	0.7475***	0.7204***
Second Home	0.5677***	0.5722***	0.6278***	0.551***	0.5752***	0.5885***	0.5923***	0.6893***	0.6656***	0.6347***
Refinance (Cash Out)	0.8261***	0.8112***	0.7607***	0.6614***	0.6419***	0.5404***	0.5162***	0.4954***	0.5274***	0.5094***
Refinance (No Cash Out)	0.8756***	0.9189***	0.9168***	0.7928***	0.7501***	0.5839***	0.5366***	0.5421***	0.587***	0.5361***
Home Value First Quartile	1.1462***	1.0126	0.9051***	0.9278***	0.9496***	1.0195*	0.8739***	0.7041***	0.619***	0.6035***
Home Value Second Quartile	1.0514***	0.9632**	0.9378***	0.9025***	0.913***	0.9628***	0.8398***	0.6908***	0.6408***	0.6345***
Home Value Third Quartile	0.9763	0.9545***	0.9405***	0.9086***	0.8948***	0.9312***	0.8621***	0.8351***	0.8076***	0.8054***
LR test H_0: $\beta = 0$	13165	20912	22027	27222	43346	80949	119397	141258	125578	15385
(p-value)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

The symbols ***, ** and * denote statistical significance at 1-percent, 5-percent and 10-percent levels respectively.

Figure 2: Counterfactual analysis for 2001 vintage

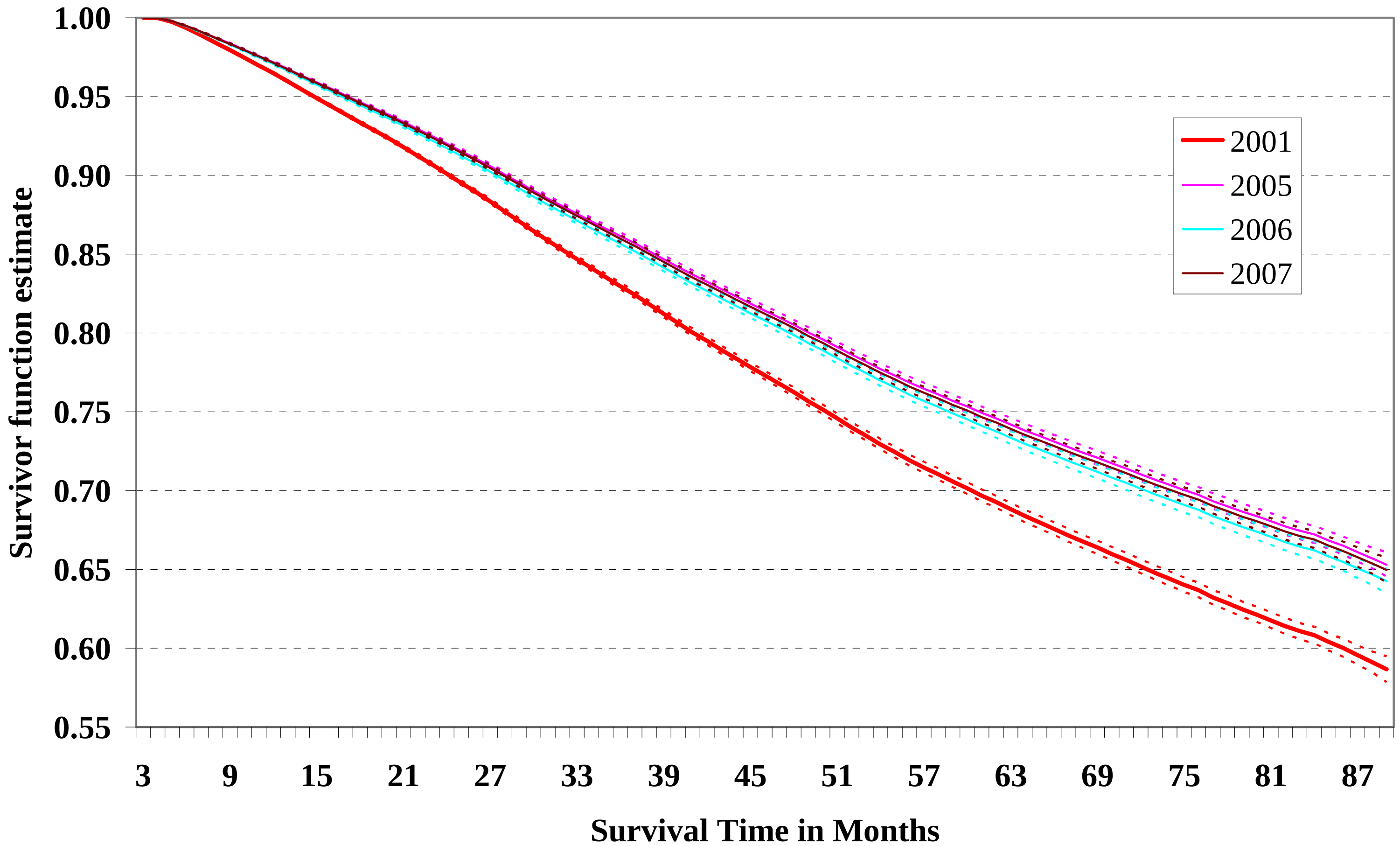

Figure 3: Counterfactual analysis for 2002 vintage

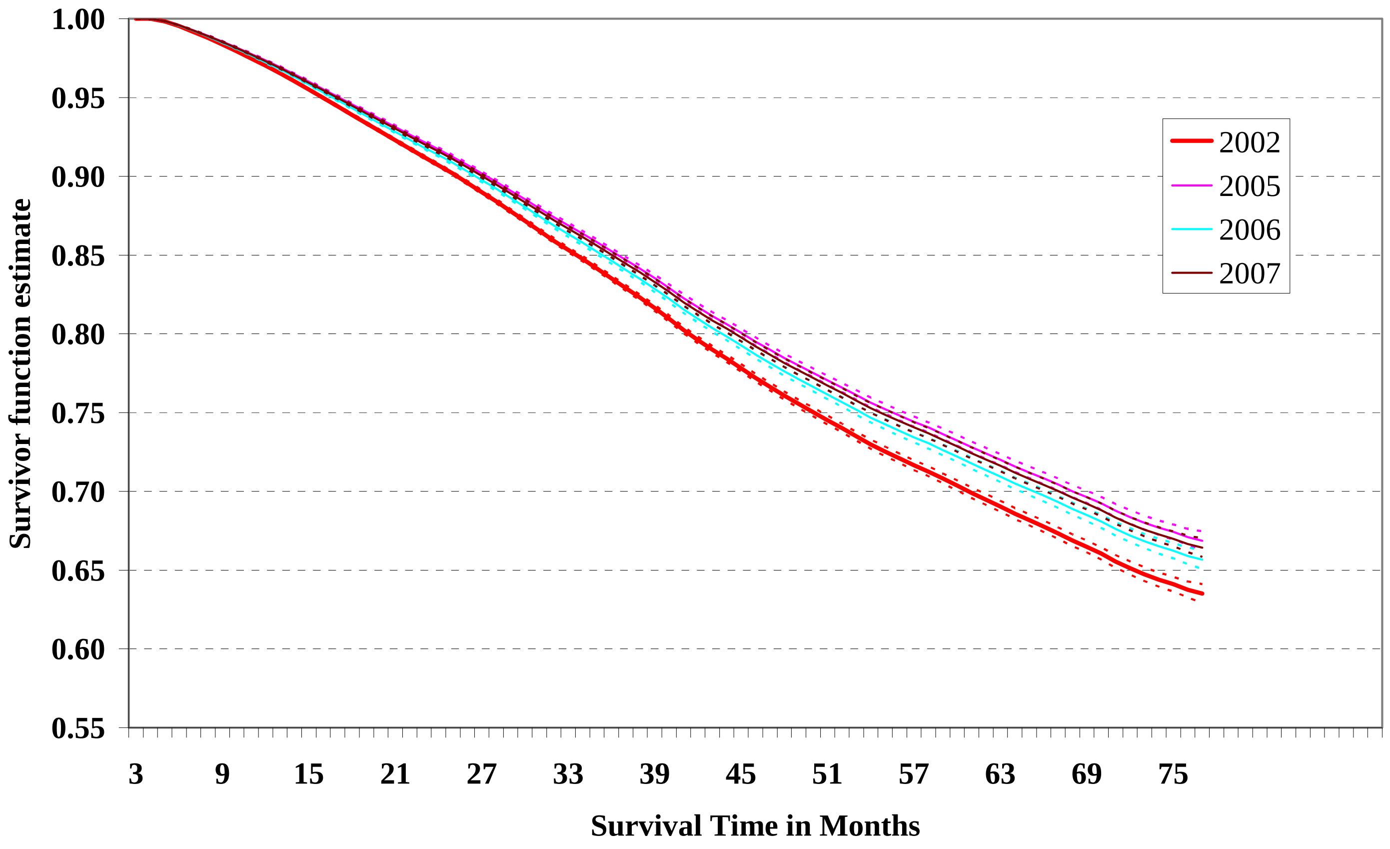

Figure 4: Counterfactual analysis for 2003 vintage

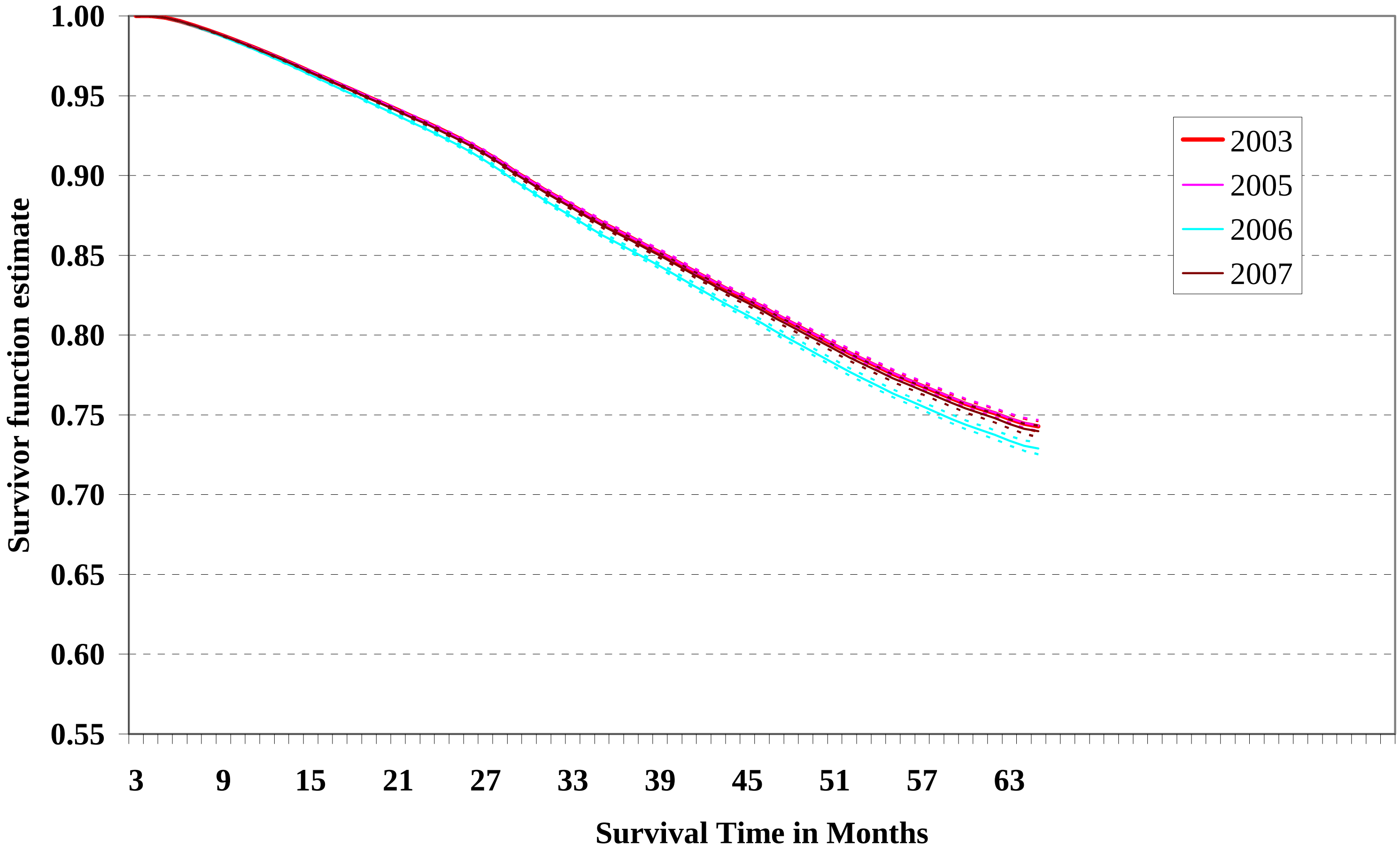

Table 14: Counterfactual Survival analysis

Three panels report numbers corresponding to counterfactual exercise using survivor function estimates based on 2001, 2002, and 2003 data. The numbers in the brackets are lower and upper confidence limits at 95 % confidence for the estimated survivor function.

		Panel 1: Counterfactual Analysis 2001		
Age of Loan (Months)	Survivor Function 2001	Counterfactual Survivor Function 2005	Counterfactual Survivor Function 2006	Counterfactual Survivor Function 2007
12	0.965	0.972	0.971	0.971
	(0.964, 0.966)	(0.971, 0.972)	(0.97, 0.971)	(0.971, 0.972)
24	0.901	0.920	0.917	0.919
	(0.9, 0.902)	(0.919, 0.921)	(0.916, 0.919)	(0.918, 0.921)
36	0.830	0.861	0.857	0.860
	(0.828, 0.832)	(0.859, 0.864)	(0.854, 0.859)	(0.858, 0.862)
48	0.763	0.805	0.799	0.803
	(0.76, 0.765)	(0.802, 0.808)	(0.796, 0.802)	(0.8, 0.806)
60	0.702	0.753	0.745	0.751
	(0.698, 0.705)	(0.75, 0.757)	(0.741, 0.749)	(0.747, 0.755)
		Panel 2: Counterfactual Analysis 2002		
Age of Loan (Months)	Survivor Function 2002	Counterfactual Survivor Function 2005	Counterfactual Survivor Function 2006	Counterfactual Survivor Function 2007
24	0.970	0.974	0.972	0.973
	(0.97, 0.971)	(0.973, 0.974)	(0.972, 0.973)	(0.973, 0.974)
36	0.907	0.917	0.913	0.916
	(0.906, 0.908)	(0.916, 0.918)	(0.912, 0.915)	(0.915, 0.917)
48	0.835	0.853	0.846	0.850
	(0.834, 0.837)	(0.851, 0.854)	(0.844, 0.848)	(0.848, 0.852)
60	0.761	0.785	0.776	0.782
	(0.758, 0.763)	(0.782, 0.787)	(0.773, 0.779)	(0.779, 0.784)
72	0.704	0.732	0.722	0.729
	(0.7, 0.707)	(0.729, 0.736)	(0.719, 0.726)	(0.725, 0.732)
		Panel 3: Counterfactual Analysis 2003		
Age of Loan (Months)	Survivor Function 2003	Counterfactual Survivor Function 2005	Counterfactual Survivor Function 2006	Counterfactual Survivor Function 2007
36	0.977	0.977	0.975	0.977
	(0.977, 0.977)	(0.977, 0.977)	(0.975, 0.976)	(0.976, 0.977)
48	0.929	0.929	0.924	0.928
	(0.928, 0.929)	(0.928, 0.93)	(0.923, 0.925)	(0.927, 0.928)
60	0.866	0.866	0.858	0.864
	(0.864, 0.867)	(0.865, 0.867)	(0.856, 0.859)	(0.863, 0.865)
72	0.808	0.808	0.797	0.805
	(0.806, 0.81)	(0.806, 0.81)	(0.795, 0.799)	(0.803, 0.807)
84	0.757	0.757	0.744	0.754
	(0.754, 0.759)	(0.755, 0.76)	(0.741, 0.746)	(0.751, 0.757)