Working Papers

Federal Reserve Bank of Boston

No. 07-15

Subprime Outcomes: Risky Mortgages, Homeownership Experiences, and Foreclosures

Kristopher Gerardi, Adam Hale Shapiro, and Paul S. Willen

Abstract:
This paper provides the first rigorous assessment of the homeownership experiences of subprime borrowers. We consider homeowners who used subprime mortgages to buy their homes, and estimate how often these borrowers end up in foreclosure. In order to evaluate these issues, we analyze homeownership experiences in Massachusetts over the 1989–2007 period using a competing risks, proportional hazard framework. We present two main findings. First, homeownerships that begin with a subprime purchase mortgage end up in foreclosure almost 20 percent of the time, or more than 6 times as often as experiences that begin with prime purchase mortgages. Second, house price appreciation plays a dominant role in generating foreclosures. In fact, we attribute most of the dramatic rise in Massachusetts foreclosures during 2006 and 2007 to the decline in house prices that began in the summer of 2005.

JEL Classifications: D11, D12, G21

Kristopher Gerardi is a research associate at the Federal Reserve Bank of Boston, and a Ph.D. candidate at Boston University. His e-mail address is kristopher.gerardi@bos.frb.org. Adam Hale Shapiro is a research associate at the Federal Reserve Bank of Boston, and a Ph.D. candidate at Boston University. His e-mail address is adam.shapiro@bos.frb.org. Paul S. Willen is a senior economist and policy advisor at the Federal Reserve Bank of Boston. His e-mail address is paul.willen@bos.frb.org.

This paper, which may be revised, is available on the web site of the Federal Reserve Bank of Boston at http://www.bos.frb.org/economic/wp/index.htm.

Chris Foote figured out how to identify subprime purchase mortgages in The Warren Group data using the HUD subprime lender list, which greatly simplified this project. We thank Jeff Fuhrer, Robert King, Lorenz Goette, Andreas Lehnert, Julio Rotemberg, and Geoff Tootell for helpful comments and suggestions. We also thank Tim Warren and Alan Pasnik of The Warrren Group and Dick Howe Jr., the Register of Deeds of North Middlesex County, Massachusetts, for providing us with data, advice, and insight, and thank Anthony Pennington-Cross for providing programs. Finally, we thank Elizabeth Murry for providing helpful comments and edits. Any remaining errors in the paper are our own.

This version: December 3, 2007

1 Introduction

In this paper, we analyze homeownership experiences in Massachusetts over the period 1989 to 2007. We pay particular attention to those ownerships that started with a subprime mortgage, exploring both the outcomes of these ownerships over the entire sample, and their role in the Massachusetts foreclosure crisis of 2007. We have two main findings. First, homeownerships that begin with a subprime purchase mortgage end up in foreclosure almost 20 percent of the time, or more than six times as often as experiences that begin with prime purchase mortgages. Second, house price appreciation plays a dominant role in generating foreclosures: homeowners who have suffered a 20 percent or greater fall in house prices are about fourteen times more likely to default on a mortgage compared to homeowners who have enjoyed a 20 percent increase. We attribute most of the dramatic rise in foreclosures in 2006 and 2007 in Massachusetts to the decline in house prices that began in the summer of 2005. Subprime lending played a role but that role was in creating a class of homeowners who were particularly sensitive to declining house price appreciation, rather than, as is commonly believed, by placing people in inherently problematic mortgages.

Our first point, about the outcome of subprime purchase mortgages addresses the many commentators who have questioned whether borrowers who need to use subprime loans to buy homes really should be buying homes at all. For example, financial historian Niall Ferguson writes:

> “Maybe, just maybe, not everyone is cut out to be a property owner. Maybe, just maybe, we should not be bribing and cajoling people at the margin into taking out mortgages and buying houses. And maybe, just maybe, a day of reckoning is approaching, when the costs of this policy will have to be borne not just by a minority of over-burdened households, but by everyone.” (July 15, 2007, Telegraph.co.uk)

Essentially, Ferguson and others argue that subprime borrowers end up in foreclosure too often. As far as what “too often” means, commentators make two different arguments, often simultaneously. One idea is that borrowers and lenders know the risks, but do not internalize the social costs of foreclosure that come from the foreclosure process itself, which leads to vacant properties, crime and homelessness, and from the inability of the government to commit not to bail out reckless borrowers (the “costs ... born by everyone”).[1] The

[1] Examples of studies that analyze negative social externalities of residential foreclosure include Immergluck and Smith (2006), and Apgar and Duda (2005).

second not completely inconsistent position is that borrowers don't understand the risks because lenders and realtors downplay them (or engage in "bribing and cajoling") and that a borrower who truly understood how likely he or she was to end up in foreclosure would never enter into the transaction in the first place.

Until this present analysis, one could not really talk sensibly about whether subprime purchasers defaulted "too often" because we simply did not know how often subprime borrowers defaulted. Previous analyses of subprime borrowers were performed using loan-level datasets, which allows one to measure the probability of foreclosure on an individual loan, but that paints a misleading picture of the incidence of foreclosure over the entire homeownership. For starters, most subprime loans are refinances of a previous mortgage of unknown type, so typically, we have no way of knowing whether a subprime loan played any role in the initial transition into homeownership – all we know is that the borrower refinanced into a subprime loan at some point. One can, alternatively, look at purchase mortgages alone, but doing so paints a deceivingly benign picture, since most subprime borrowers successfully refinance soon after purchase.[2]

The numbers have a "glass half full/glass half empty" quality to them. We estimate that about 18 perecent of people who finance home purchases with subprime mortgages will eventually experience foreclosure, which means that 82 percent will either remain in the home for at least twelve years or sell the property. By comparison, we estimate that borrowers who finance their home purchase with a prime mortgage is approximately 97 percent. The 18 percent figure comes from our estimated duration model, which provides predicted foreclosure hazards as a function of a list of explanatory variables. The model addresses both censoring issues and the biases introduced by the paucity of long samples.[3]

We find that homeownership outcomes are highly sensitive to the evolution of house prices and to the initial combined loan-to-value ratio at origination, while they are somewhat less sensitive to employment conditions. Since there is large variation in the values of these variables across different cohorts[4] of buyers, there is huge variation in the subprime outcomes. For some yearly Massachusetts cohorts, like those households who bought homes

[2]In Loan Performance (LP) data for Middlesex County, MA, we find that after 3 years, foreclosures have occurred on only 4 percent of subprime purchase mortgages and 88 percent of borrowers have prepaid their loan.

[3]For instance, of the 60,000 people who have used subprime mortgages to buy homes in Massachusetts in the last twenty years, about 4,800, or 8 percent, have lost their homes to foreclosure, meaning that 92 percent of these borrowers have either sold their house or are still in it. However, this number suffers from a serious right-censoring problem since the use of subprime mortgages for purchase is a relatively recent phenomenon; many current borrowers may default in the future.

[4]We use the terms "cohort" to refer to ownership experiences that begin in a specified year, or set of years (that is, the 2004 cohort refers to all ownership experiences that begin in 2004).

with subprime mortgages in 1998, we estimate that less than 6 percent will ever experience foreclosure, because they benefited from the state's historic run-up in house prices between 1993 and 2005, which to varying degrees was mirrored in many other housing markets across the U.S. For other subprime cohorts, such as those who took out loans in 2005, and those who might have taken out loans in the late 1980s and early 1990s had a subprime market existed, the percentage is much higher, because of subsequent declines in house prices that affected the value of the mortgage relative to the property's market value.

Our second point is that house price appreciation (HPA) is the main driver of foreclosures. The easiest way to see this is to look at aggregate data. Figure 1 shows that periods of exceptionally high HPA (2002-4) in Massachusetts are associated with exceptionally low numbers of Massachusetts foreclosures and vice versa (1989-1991 and 2005-2007). Cash flow problems at the household level, driven by job loss, for example, play a role, but only when HPA is low. For example, in 2001, a serious recession generated record numbers of delinquencies, a sign that many households had problems making payments, and also record numbers of foreclosures. However, the records were opposites; a high number of delinquencies, but a low number of foreclosures.

We argue that the relationship between foreclosures, prices, and cash flow problems is consistent with a simple model of the household decision to sell, default, or remain in the residence, which we develop in Section 2. We model mortgage default as an option, as is typical in the literature, but we depart from the literature by embedding it in a model of portfolio choice, in which households face a battery of realistic constraints: deposits earn less interest than a household pays on unsecured credit; and shorting the house is impossible. Unlike standard models in the literature, which focus only on house prices and interest rates as determinants of default, our model provides an important role for the individual household's unique financial situation.

The first key insight from the model is that negative equity is a necessary but not sufficient condition for default, because selling dominates defaulting if a borrower has positive equity. However, negative equity is not sufficient for default, because future house price appreciation may make it profitable to continue making mortgage payments. We argue that cash flow, or lack thereof, then determines whether it is optimal to default – if a borrower must make extreme sacrifices in term of current consumption to realize benefits in the distant future, it may not make sense to continue paying the mortgage, but if the sacrifice to current consumption is small, it may make sense to continue making payments.

The model predicts that if HPA is high, we will observe few foreclosures, even if people have cash flow problems, just as we see in the data. The theory suggests that a fall in HPA will not generate foreclosures unless we also have cash flow problems, but the bottom panel of Figure 1 shows that some fraction of borrowers (at least two percent) *always* have cash flow problems. That said, cash flow problems did appear to increase in 2006 and 2007, but, to prove our point, the scale of the increase in foreclosures is an order of magnitude larger than the increase in delinquencies.

Subprime lenders created a group of borrowers that were much more likely to default for at least two reasons. First, while they did not invent zero-equity borrowing, they did allow a much larger fraction of borrowers to start homeownership with no cushion against negative HPA. Second, subprime lenders allowed borrowers with a history of cash flow problems and with monthly payments that exceeded fifty percent of current income to enter homeownership. Under the best of circumstances, subprime borrowers are at least five times as likely to become delinquent as prime borrowers.

We address the sources of foreclosure in a much more rigorous way in the paper, estimating a duration model of homeownership in which the two possible outcomes are foreclosure and sale. We use individual level data, following approximately 1.5 million homeownerships in all 351 Massachusetts towns over 18 years. For each homeownership, we measure HPA and we address negative equity by using the initial loan-to-value ratio on the property, which we can measure accurately because we have access to all loans on the property, including second liens. We measure HPA at the town-level by calculating a set of Case-Shiller, weighted-repeat-sales price indices for each town, and we address cash flow issues using town-level unemployment data from the BLS. We also control for town-level differences in median income and racial makeup.

An additional finding is that subprime mortgages did contribute significantly to the foreclosure crisis of 2006 and 2007, but in quantifying the impact of the subprime mortgage market on this problem, it is important to distinguish between subprime loans made for initial purchases, and subprime refinances of existing mortgages. Approximately 30 percent of the 2006 and 2007 foreclosures in Massachusetts were traced to homeowners who used a subprime mortgage to purchase their house. However, almost 44 percent of the foreclosures were of homeowners whose *last* mortgage was originated by a subprime lender. Of this 44 percent, approximately 60 percent initially financed their purchase with a mortgage from a prime lender. This result implies that a large factor in the crisis stemmed from borrowers who began their home ownership with a prime mortgage, but subsequently refinanced into a subprime mortgage. With respect to the public policy debate of whether or not the subprime market should be regulated, and in particular whether or not subprime borrowers should be allowed to purchase homes, it is important to distinguish between these two groups of borrowers. Since many Massachusetts foreclosures in 2006 and 2007 are traced to borrowers that initially financed their home purchase through a prime lender, and *then* transitioned into the subprime mortgage market, the current public policy debate improperly focuses on subprime borrowers as an entire group. Instead, the focus should be directed toward the group of borrowers who initially financed their homes with a subprime mortgage, which is exactly what we do in this paper. By examining these subprime lending experiences and outcomes in Massachusetts, we seek to better inform this national debate. While our specific analysis is necessarily confined to this state, the implications of our study go well beyond this state, and we believe are broadly applicable to the rest of the country.[5]

In Section 1.1 we discuss alternative definitions of a subprime mortgage, and the definition that we adopt in this paper. In Section 3, we introduce our empirical duration model of sale and foreclosure. In Section 4, we discuss our data. We divide our results into three sections: Section 5.1 focuses on the non-parametric Kaplan-Meier hazards, Section 5.2 on the output of our duration model, and Section 5.3 on specific results related to subprime mortgages. Section 6 concludes our analysis.

1.1 Subprime Mortgage Market

Unfortunately, there is no universally accepted definition of a subprime mortgage, so in order to analyze the subprime mortgage market, we must first choose an appropriate definition. The terms subprime borrower, subprime lender, and subprime mortgage, are often used interchangeably by analysts and researchers, but for the purposes of this paper, it is important to make a distinction.

The term subprime borrower has traditionally been applied to a borrower that is perceived to be more risky relative to the average borrower, usually because of a poor credit history. In the United States, a subprime borrower today typically refers to an individual with a FICO score below 620, who has become delinquent on some form of debt repayment in the previous 12 to 24 months, or who has even filed for bankruptcy in the last few years. There have always been small-scale venues for subprime borrowers; for instance, pawn shops and payday lenders have both existed for a long time. However, until recently, subprime borrowers were unable to systematically obtain large-scale loans such as mortgages. With reformed lending laws and increasingly sophisticated financial markets and instruments, a new mortgage lending channel emerged that serviced subprime borrowers in particular. It is unclear exactly when the subprime mortgage market truly began, but in 1993 the Depart-

[5] In Section 4.3, we compare the Massachusetts housing market to the national market.

ment of Housing and Urban Development began tracking the subprime mortgage market, and developed an annual list of subprime lenders, which were defined as mortgage lenders that specialized in lending to subprime borrowers.[6] This list is calculated each year by identifying mortgage lenders that originate a large percentage of "high-cost" loans, where HUD defines a "high-cost" loan to be a mortgage with an initial interest rate that is at least 300 basis points larger than the yield of a treasury bill with a comparable maturity period. The reasoning behind this method is that lenders charge higher interest rates to subprime borrowers to compensate for the elevated credit risk.

During its beginning stages, this subprime lending channel focussed almost exclusively on subprime borrowers, or borrowers with impaired credit. In this period, the term "subprime mortgage" referred to a loan made to a subprime borrower. However, with the rapid increase in mortgage securitization, as well as the persistent growth in house prices nationwide, the subprime lending channel soon expanded its credit to borrowers on other margins. These borrowers included households that did not want to produce a downpayment for a house purchase (so-called "zero-down" borrowers), households that did not wish to fully disclose their income and financial wealth (so-called "no doc" or "low doc" borrowers), households that wished to purchase a larger home than they otherwise could have purchased with financing from a prime lender, or households who wanted to do some combination of these actions (or even all them). Thus, a subprime mortgage evolved from a loan originated by a subprime lender to a borrower with a poor credit history, to a loan originated by a subprime lender to a borrower that was marginal and thus riskier relative to the average borrower on any one of a number of different dimensions. The point is that the subprime lending channel no longer focused on only credit impaired borrowers, but also on borrowers who might be considered prime based on their FICO score, but who were perceived to be elevated credit risks because of other characteristics.

Another common definition for a subprime mortgage comes from the secondary mortgage market. The secondary mortgage market consists of investors who purchase securities that are collateralized by residential mortgages. There are three broad types of securities, and they are referred to as prime, alt-a, and subprime. The three types are primarily distinguished by the credit risk of the underlying mortgages, with prime denoting mortgages with the least amount of risk, subprime denoting mortgages with the most amount of risk, and alt-a denoting mortgages with risk properties somewhere in between. A subprime mort-

[6]It is important to stress that most, but not all of a subprime mortgage lender's business involves servicing subprime borrowers, so that many subprime lenders also service prime borrowers, or borrowers with a sound financial credit history.

gage in this context refers to a loan placed in a pool of securitized mortgages that is labelled "subprime." The heterogeneity in the characteristics of borrowers obtaining mortgages from the subprime credit channel discussed above is also present in this definition of a subprime mortgage. To illustrate this heterogeneity, we obtained data on securitized mortgages from Loan Performance[7] for Massachusetts, Connecticut, and Rhode Island for the period 2000-2007. The average FICO score associated with subprime-labelled purchase mortgages over this period is 643, but the variance of the distribution is large. Approximately 25 percent of these mortgages had associated FICO scores over 680, 10 percent had FICO scores above 720, while another 25 percent had associated FICO scores below 600. Thus, mortgages designated as subprime by the secondary mortgage market are not necessarily loans made to borrowers with poor credit histories. If we look at a different dimension, such as initial loan-to-value (ltv) ratios, we find that the average initial ltv ratio for subprime mortgages with FICO scores above 680 is almost 92 percent (median is 95 percent), while the average ltv ratio for subprime mortgages with FICO scores below 620 is approximately 87 percent (median is 90 percent). This demonstrates how the subprime mortgage market no longer simply services borrowers with poor credit histories, but it also services borrowers who have more stable credit histories, but who are marginal, or perceived to be higher credit risks for different reasons, such as high initial ltv ratios.

The above discussion is meant to illustrate that a precise definition of a subprime mortgage, in terms of loan characteristics and borrower characteristics does not exist. Many analysts use a definition based on the secondary market's categorization of mortgages, some use HUD's definition of a "high-cost" mortgage, and finally others define a subprime mortgage to be a mortgage originated by a subprime lender, where a subprime lender is identified using HUD's annual list. For the purposes of this paper we use the latter definition, since one of the primary focuses of this paper is the subprime lending channel. That is, we are interested in quantifying the default risk of a borrower that achieved homeowner status through the subprime lending channel. While the types of borrowers who used this channel are a very heterogeneous group, the current foreclosure crisis has taken the hardest toll on subprime lenders in particular, and the current public policy debates are focused specifically on the subprime lending channel, and not necessarily on mortgages given to specific groups of borrowers. In future research, which we discuss in more detail in the concluding section, we will concentrate on different components of the subprime lending channel.

[7]Loan Performance data consists of non-agency, securitized mortgages (mortgages that are not securitized by the GSEs, but rather by privately companies).

2 A Two-Period Model of the Default Decision

Why would a household default on a home mortgage? According to what we will call the "frictionless option model" (FOM), a borrower should default if the value of his house H falls short of the value of the mortgage V on that house. This statement is often understood to mean that the borrower should default if the value of the house falls short of the unpaid principal balance U, that is, if the homeowner has negative equity. But, as Kau, Keenan and Kim (1994) explain,[8] this interpretation is wrong: V is the present value of future payments discounted using the market stochastic discount factor. Since borrowers have the option to refinance the mortgage and to default in the future, V is always less than U, so negative equity is not a sufficient condition for default.

The FOM is useful in that it provides an explanation for why homeowners tolerate negative equity, and does well in explaining default behavior qualitatively, but not quantitatively. The problem is that the FOM assumes that the "economic environment facing a homeowner is described by two variables: the interest rate and the house value" (Kau, Keenan, and Kim, 1994). Essentially, the FOM implies that two different households who bought similar houses in the same market, at around the same time, on the same terms, will both default at the same time. Such a prediction simply isn't consistent with the data. Moreover, the deviations occur in a systematic way, as individual shocks like unemployment, divorce, and illness matter a lot – if one of our two borrowers lost a job, he or she would be much more likely to default.[9] To rectify this predictive deficiency, researchers have turned to two alternative explanations for variation in default rates among otherwise similar borrowers. The first incorporates "trigger events," – divorce, illness, and spells of unemployment are the typical examples – which make some borrowers more vulnerable to default. The second is the concept of "ruthless" default, which posits that to actually follow the advice of the FOM, a borrower must be cold-blooded and devoid of emotion. Neither of these explanations is satisfying from a theoretical standpoint, the former because it does not explain why the trigger events lead to default, and the latter because it appeals to a parameter termed, "ruthlessness," a concept which is difficult to define formally, let alone quantify.[10]

[8] The FOM model of Kau, Keenan and Kim applied the contingent claims framework of Black and Scholes (1973), Cox, Ingersoll, and Ross (1985), and Epperson et al. See Kau and Keenan (1995) for a survey of the mortgage pricing literature.

[9] See Vandell (1995) for a survey of the empirical default literature, and a discussion about the importance of individual variables in the default decision.

[10] The literature does stress the potential importance of transaction costs in modeling default. However, the concept of transaction costs is somewhat vague in the literature. Often transaction costs refer to the direct costs of defaulting, in terms of default penalties and mobility costs. In other cases such costs refer

We build a model that retains the basic structure of the FOM but yields the intuitive prediction that financially strapped borrowers are more likely to default, conditional on a given level of house prices and interest rates. To do this, we embed the financial contracts from the FOM into a portfolio choice model with constraints. As shown by He and Pearson (1991) and others, portfolio constraints imply that households use different stochastic discount factors to value assets. Put differently, individual household valuations of identical assets typically won't be identical. The upshot of this is that while the intuition that a borrower will default if the value of the mortgage exceeds the value of the house still remains true, both the value of the house and the value of the mortgage will differ across households depending on unique characteristics and the individual financial situation of the particular household. For example, less financial wealth reduces the value of the house relative to the value of the mortgage, and makes default more likely.

We now lay out our model and derive our basic result about default. At the end, we return to the FOM and show that our characterization of default, suitably adjusted, and the FOM characterization of default are the same.

We consider a two period model, in which the household enters period 0 living in a house with a market price H_0, and holds an interest-only mortgage with unpaid principal balance U, interest rate r_M, and with interest due in period zero. The household receives labor income y_0, and is assumed to have wealth w. For simplicity, we assume that there is only one type of house in the model, and that the house is also available for rent at price ρ. In period 0, the household may choose to 1) sell the house and rent, 2) default on the mortgage and rent, or 3) continue paying the mortgage and keep the house. In addition, the household simultaneously chooses an optimal level of consumption and saving, where it can borrow $\theta_B > 0$ dollars at unsecured interest rate R_B, or save $\theta_L \geq 0$ dollars at interest rate R_L. We assume that the household cannot refinance the mortgage. In period 1, the household receives income y_1 and the market value of the house is H_1. For simplicity, we assume that the household knows the value of the house in period 1 with certainty in period 0. This assumption does not change the main results derived from the model. If the price of the house is stochastic, then expectations of future house prices, instead of the actual present value, will matter in the household's default decision.[11] In period 1, the household can either sell the house or default on the mortgage.[12] After making this choice,

more broadly to frictions in the default decision.

[11] In the case of stochastic house prices, the household's default decision will also depend on the variance of the house price shock.

[12] For simplicity, we assume that there is no penalty associated with default. In reality, default may have a significant negative impact on the ability of a household to borrow in the future. If we impose a penalty

the household consumes its net savings and income, and the model ends. Formally, we consider three different budget constraints. If the household keeps the house, it faces this set of constraints:

$$c_0 = y_0 + w - r_m M_{-1} + \theta_B - \theta_L \quad (1)$$

$$c_1 = y_1 + (H_1 - U)^+ - \theta_U R_U + \theta_L R_L. \quad (2)$$

If the household elects to sell, it faces this set of constraints:

$$c_0 = y_0 + w + H_0 - U - \rho + \theta_B - \theta_L \quad (3)$$

$$c_1 = y_1 - \theta_U R_U + \theta_L R_L. \quad (4)$$

Finally, if it elects to default, the household's constraints are:

$$c_0 = y_0 + w - \rho + \theta_B - \theta_L \quad (5)$$

$$c_1 = y_1 - \theta_U R_U + \theta_L R_L. \quad (6)$$

For the case in which the household elects to keep the house in period 0, it will choose to sell in the final period if the house is worth more than the mortgage, and, if not, it will default. If instead, the household elects to default or sell the house in period 0, then we assume that it must rent a house of the same size (or value) at a price ρ (in order to obtain shelter services).

The first result here is that a borrower will default in the first period only if $H_0 < U$. In other words, negative equity is a necessary condition for default. However, by no means is negative equity a sufficient condition for default. The top panel of Figure 1, for example, shows that even in years in which house prices appreciated substantially, some borrowers did in fact default on home mortgages. High house price appreciation means that, in general, household equity goes up, but mortgage equity withdrawal, judgments, tax delinquency, and arrears on the mortgage can all lead to default. Overall, in the years with the highest house price appreciation, we see historically low levels of default.

We now focus on the interesting special case, where $H_0 < U$, $H_1 > U$, and $r_M U > \rho$. The first condition means that the household will never sell in the first period, since the house is worth less than the mortgage. The second two conditions imply that it is costly to keep the house, but there is some future benefit to doing so in the form of future price

in the model, the decision to default will be a function of the magnitude of the penalty.

appreciation. We define the return to keeping the house as:

$$R_K = \frac{(H_1 - U)^+}{r_M U - \rho}.$$

The interpretation of R_K is straightforward. The dividend on keeping the house is the payoff from selling the house in the next period, and the price is the cost of paying the mortgage less the rent the household would have to pay if it no longer owns the house. Under these conditions, we advance and prove the following proposition:

Proposition 1 *If*

$$R_B > R_K > R_L, \tag{7}$$

then there exists w^* *such that if* $w > w^*$*, the borrower keeps the house and if* $w < w^*$*, the borrower defaults. The implicit function* $\mathrm{w}^*(y_0, y_1, \beta, r_M, H, U)$ *has the following properties:* $w^*_{y_0} < 0$*,* $w^*_{y_1} > 0$*,* $w^*_\beta < 0$*,* $w^*_{y_0} < 0$*,* $w^*_H < 0$*,* $w^*_U > 0$*, and* $w^*_{r_M} < 0$*.*

Proof: See Appendix.

Proposition 1 shows how the incidence of foreclosure depends on household-level parameters. If we assume some distribution of households with respect to the exogenous parameters, then we can now say something about what determines the default rate. First, if we lower wealth, we get more defaults. Second, anything that reduces the relative value of future consumption (higher future income, lower current income, less patience) tends to increase the likelihood of a default decision that leads to a foreclosure. Third, as one would expect, increasing the mortgage interest rate r_M makes default and thus foreclosure more likely. Finally, reductions in rental prices make holding on to the house more expensive and increase the likelihood of default.

In the model, the effect of changes in house prices on foreclosures is somewhat more subtle. Holding H_1 constant, changes in H_0 actually have no effect on foreclosures. The logic for this is simple: a reduction in the value of the house makes no difference if the household is already out of the money (that is, if $H_0 < U$). What matters, in fact, is H_1; so the key here is beliefs about future prices rather than the current price. Essentially, holding onto the house makes sense in the model only if the household believes that somewhere down the line, the house will be worth more than the mortgage, because that is what justifies the premium paid over renting. In the end, H_0 matters only if one believes that there is a relationship between current and future housing prices. As most buyers pay close attention to recent trends in house prices, a reduction in H_0, conditional on $H_0 < U$, will likely increase the foreclosure rate.

Above, we argued that our model was closely connected to the FOM. According to the FOM, we need to compare the value of the mortgage with the value of the house. In the FOM, we can establish those values by looking at market prices, but in our model it is more complicated. Take the value of the house: the house is an asset that pays ρ this period and H_1 next period when the borrower sells the house. If the household can finance the house purchase out of savings, then the value is:

$$H = \rho + \frac{H_1}{R_L}.$$

Note that the value of the house to the borrower is *not* necessarily equal to the current market price of the house. Such an outcome would be an arbitrage opportunity in a frictionless world. If $H > H_0$, for example, one could borrow H_0 dollars at interest rate R_L, buy a house, and make a certain profit of $R_L(H - H_0)$. Another household that has to finance the house purchase by borrowing would value the house at:

$$H = \rho + \frac{H_1}{R_B}.$$

Using our definitions of the value of the house and the mortgage, we can show that condition (7) is equivalent to the FOM criterion that a sufficient condition for default is that $V > H$. For a household that can finance a house with accumulated savings, the condition $R_K > R_L$ and $V > H$ are equivalent. To see why, we can use the definitions of V and H to obtain:

$$r_M U + \frac{\min(U, H_1)}{R_L} < \rho + \frac{H_1}{R_L}.$$

Re-arranging and using the fact that $H_1 - \min(U, H_1) = (H_1 - U)^+$, yields:

$$R_L < \frac{(H_1 - U)^+}{r_M U - \rho} = R_K.$$

3 Empirical Model

3.1 Ownership Experiences

Our analysis is unique in that, to the best of our knowledge, this is the first study capable of tracking the same borrowers across different mortgage instruments for the same residential property. The empirical default literature has primarily used loan-level data to simultaneously model the decision either to default or to prepay a mortgage.[13] Instead of characterizing the prepayment and default probabilities of a single loan, in this paper we are able to characterize sale and default probabilities across the time horizons of entire "ownership experiences." In using this term, we are referring to the time that an individual household lives in a particular house. We believe this is a significant methodological contribution to modeling default behavior for a number of reasons.

First, it is unlikely that the probability of defaulting on a subsequent mortgage is independent of the risk associated with prior mortgages purchased by the household. For example, many borrowers choose to extract equity from their homes to smooth consumption by refinancing their mortgages; this is a practice called mortgage equity withdrawal. While undoubtedly, many borrowers doing so are simply consuming the returns to their housing investment as part of their optimal consumption plan, other households may have experienced a recent adverse transitory income shock, and are extracting their housing equity as a precautionary buffer. This latter scenario suggests that, in some cases, mortgage loans that end "successfully" in the form of prepayment may actually be a signal of financial distress. In such a case, the subsequent mortgage actually has a high probability of default. Thus, it is more informative to observe all of the mortgages issued to a given borrower on a given property to calculate unbiased probabilities of default.

Moreover, in the context of certain public policy questions, the probability of default associated with an ownership experience may be much more relevant than the probability of default of a single loan. This is especially true for the important public policy debate at the heart of this paper, which deals with the effects of the emergence of the subprime market on the home-ownership rate. As subprime borrowers prepay their mortgages at an extremely fast rate,[14] a dataset comprised of only individual loans will underestimate the cumulative number of defaults. By looking at an the entire duration of an ownership as opposed to an individual loan, we can calculate the cumulative probability of default even when a subprime mortgage is refinanced.

On the other hand, using loan-level data could also overestimate default probabilities for borrowers who purchase their homes with subprime mortgages, if the data do not distinguish between purchase and refinance loans. For instance, if many prime borrowers refinance into subprime mortgages as a result of financial duress, then subprime loans will overestimate default probabilities of borrowers who purchase homes with subprime mortgages.

[13]The term prepay is used in the literature to describe a situation when a borrower prematurely pays off a mortgage. Reasons for prepayment include the sale of the home, a refinance into a new loan, or early repayment of a loan.

[14]See Pennington-Cross and Ho (2006) for evidence.

Thus, for all of these reasons, we conduct our analysis over entire ownership experiences, rather than individual loans taken out at one point in the ownership cycle. In our data we find that the average number of mortgages over the life of completed ownerships is 2.7, which suggests the potential importance of this differentiation.

3.2 A Competing Risks Model of Home Ownership Termination

Most of the recent literature on the determinants of mortgage default has used loan-level data to simultaneously model default and prepayment decisions in a reduced-form framework. Deng, Quigley, and Van Order (2000), Deng and Quigley (2004), and Pennington-Cross and Ho (2006) are some recent examples of studies that have jointly analyzed the decisions to default and prepay jointly. These studies emphasize that the default option and prepayment option are linked by mutual exclusivity, since by exercising one choice, the borrower forfeits the opportunity to exercise the other one. In practice, the prepayment option is exercised for one of two reasons, either because the borrower is refinancing or because he or she is selling the property and moving. However, because of data limitations, the vast majority of studies do not distinguish between these two possibilities.

We have argued above that it is not possible to use loan-level data to answer the questions that we pose in this study. Thus, we use data on homeownership experiences instead, and estimate the joint probability of sale and default using a competing risks, proportional hazard, duration model. Sale and default are competing risks in the sense that these are the only possibilities (other than death) by which an ownership experience can end, and exercising one of these choices precludes the possibility of exercising the other. The model specification that we use is taken from Meyer (1995), and is very similar to the specification of Pennington-Cross and Ho (2006) and Yu (2006). It is a competing risks, proportional hazard, duration model that allows for time-varying covariates and unobserved heterogeneity, and is based on the studies of Han and Hausman (1990), Sueyoshi (1992), and McCall (1994). As in these papers, our data are observed in discrete intervals. In principle, we could estimate the model using monthly intervals, but in order to obtain precise house price indexes, it is necessary to use quarterly intervals. We utilize a specification that allows for a non-parametric baseline hazard, as well as a specification that parameterizes the baseline hazard as a third-order polynomial in the length of the ownership experience. In our present analysis we choose not to model unobserved heterogeneity.[15]

[15]Allowing for unobserved heterogeneity using a multi-dimensional, discrete distribution, such as the one used in Yu (2006) is an obvious straightforward extension. However, the computational time required to estimate the current specification without unobserved heterogeneity is already significant and would become

The hazard function is given by the probability that homeowner i terminates the ownership experience at time t, conditional on the fact that he or she has lived in the home until time t. In our model, there are two hazard functions corresponding to the two competing actions that the homeowner can take to end the ownership experience: default and sale. Formally, the hazard function for the ith action is given by

$$\lambda_i^r(t) = \lim_{\Delta t \to 0} \frac{P(t < T_i^r < t + \Delta t | T_i^r \geq t)}{\Delta t}, \tag{8}$$

where homeowner i can terminate the ownership by selling the home, $r = S$, or defaulting on the loan payment, $r = D$.

In our sample, the first realized termination time of owner experience i, is when the homeowner either sells the home or defaults on the loan, or when the end of the sample is reached. Letting T^D and T^S, denote discrete random variables representing the duration to default and sale, respectively, and T^C denote censoring due to the end of the sample, the first realized termination is $T_i^* = \min\{T_i^D, T_i^S, T_i^C\}$. The probability that homeowner i reaches T_i^* conditional on a vector of observed covariates, $X_i(t)$, is a function of both the default hazard, $\lambda_i^D(t|X_i(t))$, and the sale hazard, $\lambda_i^S(t|X_i(t))$, and is called the survival function:

$$V(T_i^*|X_i(t)) = \exp\left[- \int_{t=1}^{T_i^*} \left(\lambda_i^S(t|X_i(t)) + \lambda_i^D(t|X_i(t)) \right) dt \right]. \tag{9}$$

where

$$\lambda^r(t|X_i(t)) = \lambda_0^r(t) \exp(X_i(t)' * \beta^r), \quad r \in \{S, D\}. \tag{10}$$

This specification for the hazards, which is common in the literature, implicitly assumes that the default hazard $\lambda_i^D(t|X_i(t))$ and the sale hazard $\lambda_i^S(t|X_i(t))$ each take the same form as a function of the duration of the ownership period for all N ownership experiences in our sample. In other words, the analysis rests on the assumption that there is a hazard common to all homeownership experiences, $i \in N$ for default $\lambda_0^D(t)$, and for sale $\lambda_0^S(t)$. These hazards are referred to as the baseline hazards for default and sale, respectively. Our analysis estimates the effect of the covariates on the joint decision to sell and default as a proportion of the two respective baseline hazards. In our first specification of the model, we restrict the shape of the baseline hazard to a third-degree polynomial in the age of the ownership, such that $\lambda_0^r(t) = \exp(\alpha_0 + \alpha_{1,i}t + \alpha_{2,i}t^2 + \alpha_{3,i}t^3)$. We also estimate a specification

much greater with the addition of unobserved heterogeneity. Thus, in the current draft of this paper we choose to abstract from this extension, but we are pursuing it for a future draft.

of the model in which we use non-parametric baseline hazards, which entails estimating a set of dichotomous variables (one for each discrete interval, a quarter in our data).[16]

Since we do not observe the data continuously, but only in discrete intervals, we must account for this when forming the likelihood function. In what follows, we use the methods employed in Meyer (1995). Suppose we only observe T_i^* in the interval I_t, where $I_t = [t, t+1)$ for $t = 0, 1, ..., T^C - 1$ and $I_{T^C} = [T^C, \infty)$. If $T_i^* \epsilon I_t$, then we let $k_i = t$. We follow Sueyoshi (1992) and assume that any time-varying covariates are constant within the intervals I. This implies that the data available to the econometrician for the ith homeowner include $(k_i, \delta_i, X_i(t))$ where k_i is equal to the integer part of T_i^*, $X_i(t)$ denotes the path of the explanatory variables, and $\delta_i^r = 1$ when individual i is observed to fail due to the rth risk (default or sale), and $\delta_i^r = 0$ otherwise.

Under these assumptions, the probability of homeowner i surviving the rth risk in the interval $(t, t+1)$ is

$$\begin{aligned} P[T_i^{(r)} \geq t+1 | T_i^{(r)} \geq t] &= exp[-\int_t^{t+1} \lambda_i^r(s|X_i(s))ds] \\ &= exp\{-exp[\gamma^r(t) + X_i(t)'\beta^r]\}, \end{aligned} \tag{11}$$

where $\gamma^r(t) = \ln\{\int_t^{t+1} \lambda_0^r(s)ds\}$.

This implies that the likelihood of the data is given by

$$L(\gamma, \beta) = \prod_{i=1}^{N}([1-exp\{-exp[\gamma^D(k_i)+X(k_i)'\beta^D]\}]^{\delta_i^D} * [1-exp\{-exp[\gamma^S(k_i)+X(k_i)'\beta^S]\}]^{\delta_i^S} * H(k_i|X_i) \tag{12}$$

where

$$H(k_i|X_i) = \prod_{t=0}^{k_i-1} exp\{-(exp[\gamma^D(t) + X_i(t)'\beta^D] + exp[\gamma^S(t) + X_i(t)'\beta^S])\}. \tag{13}$$

4 Dataset

Our data come from the The Warren Group, and are comprised of historical registry of deeds records from January 1987 through August 2007 for the entire state of Massachusetts,

[16]Estimating non-parametric baseline hazards using maximum likelihood methods is a much more computationally intensive exercise. For example, if ownership experiences last a maximum of 40 quarters, then the non-parametric approach involves estimating almost 80 additional parameters, while the polynomial approach involves only 6 extra parameters (in the third-order case).

as well as 2006 and 2007 Massachusetts assessor data. The registry of deeds records contain information on all residential home sales and mortgage originations, allowing us to track every mortgage issued on a given individual property over our sample period,[17] while the assessor data contain information regarding characteristics of the property. In the data, we see transaction amounts and dates for mortgages and property sales, but we do not have information on mortgage characteristics such as the type of mortgage or the contracted interest rate. The data do contain information about the identity of the mortgage lender, which we use below to construct indicators for subprime mortgages. Information regarding the type of sale is also found in the data. Thus, we can distinguish between normal sale transactions, nominal sale transactions such as transfers among family members, and foreclosure sales.

In this paper we use foreclosure sales as a proxy for default. Foreclosure sales in the data are final transactions, in the sense that they signify the eviction of the household or homeowner from the property. The data also cover foreclosure petitions going back to 2004. Foreclosure petitions are public notices declaring the initiation of foreclosure proceedings that by law mortgage lenders are required to make in Massachusetts. Petitions are usually filed once the borrower has become delinquent on three monthly mortgage payments. While a foreclosure petition is certainly a sign of serious delinquency, it is not a good indication of default, as there are many instances in the data in which a borrower receives a foreclosure petition, but does not end up in foreclosure proceedings. The typical time between the filing of a petition and a foreclosure deed (if it occurs) is about 2 months. Thus, a borrower who receives a petition has some time to decide whether or not to catch up on the late mortgage payments and forestall default. For this reason, we believe that the appropriate proxies for default in our data are foreclosure sales.

The fact that we see every sale transaction in the data allows us to construct the ownership experiences defined and discussed above, by identifying sales of the same property in the data. Figure 2 contains an example of how we construct ownership experiences.

Between January 1987 and August 2007 we observe more than 6 million mortgages and more than 3 million ownership experiences in Massachusetts. Table 1 lists the number of sales and foreclosures by year as well as by cohort, beginning in 1989.[18] Our data encompasses the housing bust of the early 1990s in the Northeast, and this event is apparent from increased foreclosure numbers. There is a large increase in foreclosures beginning in

[17] Residential properties include condominiums, single-family homes, and multi-family homes.

[18] The data on sales are fairly reliable before 1989; however, the number of foreclosures in the data before 1989 seems implausibly low, so we restrict our analysis of foreclosures to post-1989, inclusive.

1991 and peaking in 1992, with over 9,000 foreclosures statewide. Data from the housing boom that took place in the early 2000s display the opposite extreme. In this period foreclosures dropped significantly, reaching a low point of fewer than 600 in 2003. We see evidence of the current foreclosure crisis at the very end of our sample. Foreclosures in the first three quarters of 2007 approach the levels witnesses in the early 1990s.

The top panel of Figure 1 displays a time series plot of the foreclosure rate. Here we have simply divided the foreclosure numbers from Table 1 by the number of active ownership experiences in each year. The top panel of Figure 1 also contains a graph of annualized house price growth over our sample period.[19] This figure clearly shows the two housing market cycles that the New England economy has experienced over the past 20 years. The two series are negatively correlated, with the foreclosure rates moving inversely, at a slight lag with house price growth. In 1991, house price growth reached a low of almost -9 percent, and the foreclosure rate peaked one year later in 1992 at more than 0.6 percent of active ownership experiences. The latest figures for 2007 show the state's house price growth and the foreclosure rate approaching the levels of the early 1990s, as growth is approaching -5 percent, while the foreclosure rate is approaching 0.45 percent.

4.1 Explanatory Variables

While the theoretical model discussed above is too simple to yield formal, testable predictions, it is useful in identifying the types of variables that should be expected to impact a borrower's decision to default. It tells us, for example, that a borrower's current wealth, income, and housing equity should all be important determinants of default risk. The wealth threshold derived from the model is a function of both current and future income. This threshold also suggests that income variability, although formally left out of our model, will also influence the decision to default.[20] Unfortunately our current dataset does not contain borrower-specific demographic or financial information.[21] Thus, we attempt to proxy for these borrower-specific variables with more aggregated information at the town level or the

[19] Our detailed data of repeat sales allow us to use the Case-Shiller weighted-repeat-sales (WRS) methodology to calculate house price indexes at various levels of aggregation. Figure 1 displays house prices at the state level, but for the purposes of estimation, we are able to calculate house price indexes at the city level for approximately two-thirds of the cities in Massachusetts. A detailed discussion of the WRS methodology and our aggregation assumptions can be found in Appendix A.

[20] Income variability has been identified in the literature as having an important impact on the probability of default. Herzog and Early (1970) found that borrowers in occupations with greater income volatility were more likely to be delinquent than other borrowers.

[21] We are currently attempting to merge with our dataset individual and loan-level information from HMDA and from Loan Performance. In the next draft of this paper, we hope to have this accomplished.

zip code level.

4.1.1 Initial Owner Equity and Cumulative House Price Appreciation

Previous studies in the literature, such as Deng, Quigley, and Van Order (2000), use house price indexes at the MSA-level and the unpaid mortgage balance computed from the contract terms of each loan to estimate the probability of negative homeowner equity, which proxies for whether or not the default option is "in the money." While our dataset does not include the necessary contract information to calculate the remaining mortgage balance each period, we argue that even if it did, including such information in our estimation would introduce endogeneity issues. Since we perform our estimation over entire ownership experiences, to compute net homeowner equity in each period would require including mortgage balance information for every loan obtained over the duration of each ownership. As we discussed above, the choice to refinance, as well as the choice of the refinanced mortgage terms, are endogenous decisions. Many borrowers, for example, who suffer negative income shocks or other adverse life events may choose to extract equity and obtain larger subsequent mortgages. In these cases the mortgage balances would simply be conveying information about negative income shocks or other adverse shocks affecting the individual borrower. Thus, including the time profile of net equity in our estimation would result in biased estimates. Instead, we choose to include initial homeowner equity, and the true exogenous component of net equity over time, house price appreciation.

Our dataset has enough information to construct reasonable proxies for initial net housing equity and cumulative house price appreciation. Since we see dates and amounts in our data, we are able to calculate initial loan-to-value ratios (LTV) for each ownership experience in which we observe a purchase transaction (meaning the ownership experiences that are not left-censored). These initial ltvs are cumulative in the sense that we observe all of the originating mortgages issued at the time of purchase, so if there is a second or even a third mortgage that accompanies the first mortgage, we will add these amounts to our LTV calculation.[22] In addition to initial ltvs, we use repeat-sale price indexes to calculate from the date of purchase the average cumulative price appreciation at each quarter in the town containing each property in our data.

Table 2 displays summary statistics for initial ltvs by year. The table contains means and medians for all ownership experiences initiated in a given year, as well as for ownership experiences that ended in a foreclosure. The differences are substantial. For example,

[22]Many studies have found evidence that initial LTV ratios alone are significant predictors of default. Early examples in the literature include von Furstenberg (1969) and Campbell and Dietrich (1983).

for ownership experiences beginning in 2003, the average initial LTV ratio was 0.81, and the median was 0.85. In contrast, for the ownership experiences that began in 2003 and subsequently defaulted, the average initial LTV ratio was 0.93, and the median was 0.95.

4.1.2 Investors

In addition to the financial characteristics associated with an individual borrower, the reason for purchasing the residential property should also impact upon default. Borrowers who are not owner-occupants, but who purchased the property strictly for investment purposes, are likely greater default risks, *ceteris paribus*. Since non-owner occupants do not face mobility costs and do not have an emotional stake in the property, their cost of default is likely lower relative to the cost to owner occupants. This is also apparent from the model, as an investor who defaults would not have to pay ρ to rent a home. We do not have direct information in the data to distinguish properties bought to serve as primary residences or investments; however, we can proxy for these different purposes using property characteristics. The Warren Group data contains a limited amount of information on the characteristics of each property, allowing us to distinguish between single-family homes, multi-family homes, and condominiums. We hypothesize that owners of multi-family properties are more likely to be investors, as multi-family units provide a stream of rental income. We also believe that condominiums may proxy in part for investors and real estate speculators, since the condominium market is often hit hardest during a housing bust.

4.1.3 Labor Market Conditions

Besides including initial ltvs, cumulative price appreciation, and property characteristics, we obtain information regarding certain characteristics of the town or census tract where each property is located. Since we know the exact location of each property, we can group properties by town/city as well as by zip code, allowing us to merge our data with data from the Census Bureau and the Bureau of Labor Statistics (BLS). We obtain town-level data on monthly unemployment rates going back to 1990 from the BLS. Unemployment rates have been used previously in the literature as a proxy for income volatility.[23] In addition, for some households, periods of unemployment turn out to be permanent income shocks, which our portfolio-choice model tells us will affect the default decision.

[23] Williams, Beranek, and Kenkel (1974) found that unemployment rates in Pittsburgh had a positive effect on default. Campbell and Dietrich (1983) and Deng, Quigley, and van Order (2000) also found evidence of a positive effect of the unemployment rate on default.

4.1.4 Demographics

We obtain demographic information at the census tract level from the 2000 U.S. Census, including median household income and the percentage of minority households. Previous studies in the literature have found evidence of racial discrimination in the mortgage lending business. These studies show that minority households are less likely than others to obtain approval for a mortgage, and those that are successful in obtaining a mortgage are often given a loan with inferior terms relative to an equivalent non-minority household.[24] Thus, using the intuition from our simple model, we would expect to observe higher default rates in tracts with larger percentages of minority households, *ceteris paribus*.

4.1.5 Interest Rates

The mortgage default literature has also identified the difference between the par value and the market value of a loan as a significant determinant of default risk. When the value of the mortgage rises above the outstanding mortgage balance, the probability of default increases.[25] The difference between par and market value can be negative only when the prevailing market interest rate that the borrower is qualified to receive, falls below the original contract rate of the mortgage. Since we lack information about contracted interest rates, we proxy for this effect using the prevailing aggregate interest rates. In the estimation, we include the 6-month LIBOR rate, which is a short-term interest rate that has become a very popular index for adjustable-rate mortgages, especially in the subprime mortgage market. In addition, we also include the 10-year T-bill rate to capture changes in the market values of fixed-rate mortgages.

4.1.6 Subprime Mortgages

Finally, our dataset has very accurate information about the identity of each mortgage lender. We are able to use the list of subprime mortgage lenders (by year) constructed by the Department of Housing and Urban Development (HUD) to identify subprime mortgages. The HUD subprime lender list began in 1993 and is updated every year. While this list is not a perfect representation of the subprime mortgage market, it does provide a very good representation of subprime lenders.[26] Thus, as we discussed in Section 1.1, the results in

[24] See Munnell et al. (1996).

[25] See Quercia and Stegman (1992) for a discussion of this effect.

[26] It is the same method used by many researchers in the industry, including, for example, the Mortgage Banker's Association.

this paper about the subprime mortgage market should be interpreted as applying to the subprime lending channel. In terms of the public policy debate, it is the subprime lending channel that has been called into question, and thus basing our definition of a subprime mortgage on the identity of the lender, is a reasonable approach to illuminating this issue. Table 3 displays a list of the top ten subprime lenders in our data, in terms of the number of subprime purchase mortgages originated. These ten lenders accounted for almost 70% of the total number of subprime purchase mortgages that were originated in Massachusetts. As of November 2007, eight of the ten lenders have shutdown their operations.

The lenders on the HUD list are considered to be subprime because the majority of their business consists of the origination of high cost mortgages. However, some of these lenders also originate low cost loans. In Appendix B we take a closer look at this issue and perform a robustness check, where we use the limited interest rate information that is available in our data for adjustable-rate mortgages, to construct a definition of a subprime mortgage based on interest rate spreads, similar to the methods used by HMDA analysts.

4.2 Subprime Statistics

Beginning in 1993, Table 4 displays the size of the subprime market in terms of the percentage of purchases financed with subprime mortgages each year (column 1), as well as the subprime market's contribution to the total number of foreclosures in Massachusetts. In terms of purchase mortgages, the peak of subprime lending occurred in the 2004–2006 period, when between 10 and 15 percent of purchase mortgage originations were made by subprime lenders. The last three columns in Table 4 show the disproportionate impact that the subprime mortgage market has had on the recent rise in foreclosures. In 2006 and 2007, approximately 30 percent of Massachusetts foreclosures were on ownerships initially financed with subprime mortgages, up from only 10 percent in 2004 and 2005. While these percentages are high, we see an increase when looking at the percentage of recent foreclosures on borrowers whose last mortgage was issued by a subprime lender. In 2006 and 2007, this percentage was approximately 45 percent. If we broaden the criteria even further and consider the contribution from borrowers who obtained a subprime mortgage at any point in their ownership experience, the foreclosure percentage increases to approximately 59 percent.

This distinction emphasizes the importance of differentiating between different segments of the subprime market, and specifically between purchase loans and refinance loans. For ownership experiences that begin with mortgages obtained from a prime lender, subprime refinances are often a signal of financial distress, especially for borrowers that extract equity with a subprime refinance. It is likely that in the absence of a subprime market, many of those borrowers that ended up defaulting would have defaulted on their previous prime mortgages. This point is important in the context of the current public policy debate regarding whether or not subprime borrowers should be allowed to obtain mortgages and purchase homes, since borrowers that refinanced from a prime to a subprime lender should not be included in the discussion. Thus, we restrict our subprime analysis to ownerships that initially financed their home purchase with a subprime mortgage.

4.3 Comparison to U.S. Subprime Market

While we only use Massachusetts data in this study, we believe that many of the results can be generalized to the national subprime mortgage market. In this section we will present some basic facts regarding the relative size of the subprime mortgage market, the time-series of subprime foreclosure rates, and finally the time-series of house prices in Massachusetts, in comparison to the nation as a whole, as well as to California, which is the state with the largest subprime mortgage market. The purpose of this presentation is to give a sense of just how representative the Massachusetts subprime mortgage market is to the national subprime mortgage market.

Since 2001, Massachusetts has consistently been among the top 15 states in terms of subprime market shares. Table 5 displays the share of subprime loans originated between 2001 and 2005 in Massachusetts. The share is consistently between 2 and 3 percent each year. In comparison, we also show the market share of California, the state with the largest subprime mortgage market over the past decade. California accounts for approximately one-quarter of subprime mortgage originations nationwide. Despite the difference in the size of the subprime mortgage markets, the time series of foreclosure rates associated with subprime mortgages are actually very similar in the two states. Figure 3 displays a measure of the foreclosure rate for California, Massachusetts, and the average for the nation as a whole, over the past decade. Foreclosure rates for both California and Massachusetts display very similar patterns, as rates in both states were well below the national average until 2005 and 2006, at which point they increased dramatically to levels well above the national average.

Finally, Figure 4 displays house price indexes for Massachusetts, California, and the United States as a whole, over the period of our sample. While there is a discrepancy in house price levels between Massachusetts and California, the figure shows very similar dynamics in the two house price series over the past two decades. In comparison to national house prices, prices in both states have traditionally been both higher and more volatile.

4.4 The Decision to Sell

While our main focus in this paper is foreclosure, and specifically modeling default, the competing risks nature of our estimation means that we must also address a household's decision to sell. The mobility literature emphasizes the importance of life-cycle factors and factors that affect housing demand in the decision to sell and change residence.[27]

Variables such as household size, age, marital status, income, wealth, and education have been found to impact a household's decision to move. Unfortunately, our data do not contain such information at the household level. Thus, we choose to include the same list of explanatory variables to model the decision to sell as we include to model the decision to default. It is unclear what effect initial ltvs and cumulative house price appreciation should have on the decision to sell. A study of U.K. homeowners (Henly, 1998), finds that households with negative equity have less mobility than those with positive equity. This result suggests that house price appreciation may have a positive impact on the decision to sell. Henley also finds weak evidence of a negative impact on mobility from regional unemployment rates.

5 Results

5.1 Non-parametric hazards

To gain insight about default and sale probabilities from our data, we calculate the non-parametric Kaplan-Meier estimates by length of the ownership experience.[28] We use

[27]See, for example, Henderson and Ioannides (1989), Henly (1998), and Chan (1996).

[28]The Kaplan-Meier estimates are calculated as follows: We assume that hazards occur at discrete times t_j where $t_j = t_0 + j$, $j = 1, 2, ..., J$. If we define the number of loans that have reached time t_j without being terminated or censored as n_j, and the number of terminations due to risk k at t_j as d_{kj}, then the Kaplan-Meier estimates of the hazard and survival function is

$$\lambda_k(t_j) = \frac{d_{kj}}{n_j}, S(t_j) = \prod_{t_j \leq t} (\frac{d_j}{n_j}).$$

The cumulative incidence function for cause k is

$$I_k(t_j) = \sum_{i=1}^{j} \lambda_k(t_i) S(t_i).$$

a quarterly interval length in our subsequent analysis, as this is the finest partition with which we are able to estimate the house price indexes.

Figures 5 and 6 contain the Kaplan-Meier hazard functions for default and sale, respectively, using our entire sample of data.[29] We truncate the graphs at 60 quarters (15 years), since there are not enough ownership experiences longer than 60 quarters to obtain precise estimates. The default hazard exhibits positive duration dependence in the first 5 years of the ownership experience, peaking at a quarterly rate of almost 0.17 percent, and then negative duration dependence for the remainder of the ownership period. This hump-shaped pattern is consistent with findings from the mortgage default literature.[30] The sale hazard is shaped much differently, although it also exhibits positive duration dependence early in the ownership experience (approximately two years), and negative duration dependence later.

Figures 7–10 display Kaplan-Meier conditional hazard rates for default, broken down categorically for some of the explanatory variables discussed in Section 4.1. This exercise should provide some initial insight as to whether these variables have any potential to explain variations in foreclosure rates. In Figure 7 we see that hazard rates are much higher for ownership experiences that begin with extremely high LTV ratios (1 or greater). This effect seems particularly strong early on in the ownership period. For example, conditional on surviving for two years, an ownership experience that begins with an initial LTV ratio of one or greater is four times more likely to default relative to an experience that begins with an LTV between 0.9 and 1. It does not appear that initial ltvs between 0.9 and 1.0, and between 0.8 and 0.9 have systematically different hazard rates, but initial ltvs of 0.8 or lower do appear to have slightly lower hazard rates.

Figure 8 performs the same exercise with the cumulative appreciation of housing prices since the start of the ownership experience. Cumulative appreciation rates are broken down into four categories: greater than 20 percent, between 0 and 20 percent, between -20 percent and 0 percent, and, lower than -20 percent. Hazard rates between these groupings differ dramatically. Conditional on lasting for five years, ownerships that experience -20 percent cumulative price appreciation or worse have a 0.70 percent probability of defaulting; ownerships that experience between -20 and 0 percent appreciation have a 0.30 percent probability of defaulting; ownerships that experience between 0 and 20 percent cumulative appreciation have a 0.10 percent probability of defaulting; and ownerships that experience

[29] In estimating the default (sale) hazard, we treat sales (defaults) as "censored" data, based on the implicit assumption that they are independent of each other. However, in our multi-variate analysis below, we relax this assumption.

[30] For example, von Furstenberg (1969) found that mortgage risk increased with the age of the mortgage, up to 3 or 4 years after origination, after which it declined.

more than 20 percent appreciation have a 0.05 percent probability of defaulting. Figure 9 displays differences in hazard rates for various unemployment groups. From the plot, it appears that borrowers living in areas with very high unemployment rates (greater than 7 percent) are much more likely to default.

Figure 10 displays the differences in hazard rates for borrowers that financed their house purchase using a mortgage from a subprime lender versus those that obtained a mortgage from a prime lender. The difference is substantial. The hazard rate for borrowers using prime mortgages never reaches 0.08 percent, while rates peak at approximately 0.7 percent after two years for borrowers using subprime mortgages. These non-parametric hazard rates imply that the cumulative survival (success) rate after 12 years for borrowers who finance their home purchase with a mortgage from a subprime lender is approximately 87 percent, while the rate for those using a prime lender is more than 98 percent. Figure 11 displays the Kaplan-Meier cumulative incidence rate for the subprime and prime categories, which is just the inverse of the cumulative survival rate.

5.2 Competing Risks Model

In order to ease the computational burden of the maximum likelihood estimation, we were forced to make a few assumptions. Our full data set includes all Massachusetts ownership experiences financed by a mortgage over the past 20 years. This accounts for approximately 1.6 million ownership experiences. After removing observations for which any of our explanatory variables were missing, this translates into a data set of approximately 1 million ownership experiences.[31] We chose to truncate ownership experiences at 48 quarters (12 years), since we do not see many ownerships endure longer than 12 years in our data, and those we do see all come from the early period of our dataset. We were afraid that a small number of these long ownerships could have a disproportionately large impact on our estimates of the tails of the polynomial baseline hazards. Also, this assumption proved to significantly speed up convergence in our maximum likelihood routine. To further facilitate estimation, we chose to use a 20 percent random sample of our data.[32]

[31] House prices were the most important constraint, as we were able to estimate the weighted-repeat-sale house price indexes for only about one-quarter of the 351 towns in Massachusetts. However, these towns did account for about two-thirds of the Massachusetts ownerships in our data. In future work, we will attempt to use reasonable assumptions and group towns together, in order to maximize our sample.

[32] In order to estimate the model, we must put the data into "long-form," which entails expanding the data set to include observations for each quarter that each ownership experience is active. For example, if we have an ownership experience that began in the first quarter of 2000 and ended in the first quarter of 2007, we would need to create 29 quarterly observations. Thus, our data set of 1 million ownerships becomes almost 20 million observations, and estimation becomes computationally burdensome.

Table 7 contains the estimation results for our competing risks model of default and sale. The first column displays estimates corresponding to the default and sale parameters when we parameterize the baseline hazard as a third-order polynomial. The direction of the signs of the coefficient estimates for default are consistent with our expectations, and with evidence from prior studies. The estimated effects of the initial LTV, short-term interest rates (LIBOR), and the unemployment rate on the default decision are positive. Borrowers who finance their house purchase with a mortgage from a subprime lender are more likely to default than those who used a prime lender. Borrowers who purchase a condominium or a multi-family property are more likely to default than borrowers who purchased a single-family home. Finally, an increase in cumulative house price appreciation is estimated to have a negative effect on the probability of default.

While the signs of the coefficient estimates seem very reasonable, we must also verify that the magnitudes are sensible. Since we assume a proportional hazard, the magnitude of each estimate in Table 7 is interpreted as a semi-elasticity, or the proportional increase in the overall hazard rate due to a unit increase in each covariate. For the polynomial baseline specification, the variables that have the highest impact on defaults are the initial LTV, cumulative house price accumulation, subprime lender indicator, and the property type variables. A 10 percentage point increase in the initial LTV is estimated to increase the default hazard by almost 33 percent, while a 10 percentage point increase in cumulative house price appreciation is estimated to decrease the default hazard by almost 27 percent. A $10,000 dollar increase in median family income at the Massachusetts zip code-level is estimated to decrease the default hazard by more than 16 percent. A 10 percentage point increase in the number of minority households increases the probability of default by about 9 percent. Owners of condominiums and multi-family homes are estimated, respectively, to have 67 percent and 86 percent higher conditional default probabilities than owners of single-family homes. Finally, financing a house purchase with a mortgage from a subprime lender is estimated to increase the default hazard by approximately 715 percent. This is a large effect, but not an entirely unexpected one, especially in light of the Kaplan-Meier non-parametric default hazards in Figure 10. To the extent that the subprime lender indicator is a reasonable proxy for subprime borrowers, then our theoretical model, as well as previous studies in the literature, tell us that we should expect much higher default rates for marginal borrowers with poor credit histories.

The third column in Table 7 reports the estimates for the sale hazard when we parameterize the baseline hazard as a third-order polynomial in the length of the ownership. Higher initial loan-to-value ratios are estimated to increase the probability of sale. Subprime ownerships are approximately 38 percent more likely to sell at any point in the ownership than prime ownerships. Condominiums are also estimated to be more likely to be sold compared with single- and multi-family homes. This result is not surprising to the extent that condominiums tend to be occupied by households that are in the early or late stages of their life-cycle. The other notable result for sales is the significant, negative effect from an increase in unemployment. A 1 percentage point increase in the unemployment rate is estimated to decrease the probability of sale by approximately 7 percent.

5.2.1 Measuring the Fit of the Model

To get a rough idea of how well the model fits the default data, we have used the predicted probabilities from the model to calculate the expected number of defaults at each quarter of the ownership experience. The predicted probability of default for the ith ownership at quarter j can be calculated using the estimates from the competing risk model with the expression for the default hazard:

$$\widehat{\lambda}_{ij} = \exp(\widehat{\lambda}_{i0} + x'_{ij}\widehat{\beta}). \tag{14}$$

Given the predicted probabilities for each ownership, we can calculate a predicted hazard rate at each quarter of the ownership:

$$\widehat{\lambda}_j = \frac{\sum_{i=1}^{N} \widehat{\lambda}_{ij}}{n_j} = \frac{\widehat{d}_j}{n_j}, \tag{15}$$

where $\widehat{d}_j$ is the expected number of defaults at age j, and n_j is the number of borrowers "at-risk." Figure 12 plots the predicted hazard rate $\widehat{\lambda}_j$ against the non-parametric, Kaplan-Meier hazard rate. The model seems to fit the data fairly well. The largest difference between the model and the data seems to be within the first three or four years of the ownership experience. The model over-predicts default probabilities in the first year of the ownership, but then under-predicts default probabilities in the second and third years of the ownership.

Figure 13 displays this same exercise for some of the specific cohorts of ownership experiences in our data that ranges from 1989 to 2007. While the model does not fit the data perfectly, it appears to do a very good job at getting the magnitudes and general shape of the default hazards correct. Borrowers who began their ownership in 1989 experienced the highest default hazard. This risk peaked at almost 0.6 percent after four years, which corresponds to the lowest point reached by Massachusetts house prices (levels) in 1993, and was more than double the default hazard peak for the 2004 cohort. In terms of the root-mean-squared error, the model performs the worst for ownerships initiated in 1989.[33]

Figure 14 displays the model's in-sample goodness of fit with respect to subprime ownership defaults. The figure displays the actual number of subprime foreclosures in each year in our sample of subprime borrowers versus the expected number of foreclosures $\widehat{d_j}$ predicted by the model of this group of borrowers. The figure shows that the model is not perfect, as it seems to slightly over-predict foreclosures from 2002–2005, and then under-predicts defaults in 2006 and 2007. However, it does a good job of replicating the general pattern of defaults between 1993 and 2007, and it does predict the dramatic rise in foreclosures beginning in 2005 that we see in the data.

5.2.2 Causality Issues

The FOM, as well as our theoretical model, concludes that negative equity is a necessary (but not a sufficient) condition for default. This finding, combined with the assumption that initial mortgage balances are not larger than house values, means that negative house price appreciation is a necessary condition for default. The estimation results from our empirical model confirm that house price appreciation has a strong, negative effect on the probability of foreclosure. That is, a decrease in cumulative house price appreciation from the time of purchase significantly increases the probability of foreclosure for a given ownership. One potential concern with this empirical finding is the direction of causality. The empirical duration model assumes that the causality runs from house prices to foreclosures, and thus, that house prices are exogenous in the model. However, if the direction of causality is the opposite, then the estimation may suffer from an identification problem, and our interpretation of the estimated parameters would be incorrect. For example, a large concentration of foreclosures in a given neighborhood could adversely affect the values of surrounding homes in that neighborhood. It is not unreasonable to argue that this may be an issue, since negative, social externalities of residential foreclosure have been shown to exist in the literature. However, it must be stressed that the explanatory variable in the duration model is cumulative house price appreciation since the home purchase for a given borrower, and is *not* simply quarterly house price growth at the town-level. Thus, this variable will be different for ownerships in the same town who purchased their homes in different time periods. In other words, there is significant cross-sectional variation in this variable across ownerships in the same town, as a result of different periods of ownership. This should

[33] The root-mean-squared error is calculated by taking the difference between actual defaults each period and defaults predicted by the model.

substantially mitigate any bias induced by a negative effect of foreclosures on house price growth resulting from possible negative externalities at the town-level of a foreclosure boom on property values. To address any remaining concern of reverse causality, we perform a formal Granger causality test on aggregate house price growth and foreclosure rates, which we discuss further below, and in addition, discuss some empirical observations that provide support for our assumption of the direction of causality.

Our econometric test for Granger causality follow's directly from the autoregressive specification in Hamilton (1994).[34] We perform the estimation using quarterly data over the period 1975-2007, and include four lags of aggregate house price appreciation, the aggregate foreclosure rate, the unemployment rate, price inflation (CPI), wage-inflation, the Federal Funds Rate, and the output gap. We find a p-value of 0.60 from the F-test associated with the null hypothesis that foreclosures do not Granger-cause house price appreciation, and a p-value of 0.05 from the F-test associated with the null hypothesis that house price appreciation does not Granger-cause foreclosure rates. These results support our assumption that the direction of causality runs from house price appreciation to foreclosures. We view the inclusion of the unemployment rate in the Granger-causality test as especially important for validating the exogeneity assumption of cumulative house price appreciation in the empirical duration model. Our concern is that if house price movements reflect some other macroeconomic influence that is the true cause of foreclosures, but which is left out of the Granger-causality test, then our causal interpretation would not be valid. Perhaps the variable that best fits this description is unemployment. The intuition is that a negative unemployment shock in a given location could directly lead to cash-flow problems for households, thus causing a rise in foreclosures. At the same time, high levels of unemployment could result in migration out of the area, which could in turn result in a fall in house prices. However, the inclusion of the unemployment rate in the Granger-causality test ensures that movements in house prices are not merely reflections of variation in unemployment.[35]

Further support for our causality assumption is found in Figure 1. The figure shows that delinquencies, defined as a missed mortgage payment, are highly correlated with the Massachusetts business cycle, but that foreclosure rates are not as correlated with the business cycle. In recessionary periods, such as the early 1990s, and the early 2000s, we see large increases in delinquency rates. However, the key point is that we only see increases in foreclosure rates in periods of house price depreciation. For example, in the recessionary period

[34] Hamilton (1994), section 11.2

[35] In addition, unemployment rates at the town-level are included as a control variable in the duration model.

of 2001, delinquency rates increased significantly, but foreclosure rates actually decreased. In contrast, during the recessionary period of 1991, both delinquency and foreclosure rates increased significantly. One of the big differences between these periods is the behavior of house price appreciation. In the 2001 recessionary period, house price growth was positive and increasing, while in the 1991 recessionary period, house price growth was negative and decreasing. This makes sense, since net equity is likely positive in periods of positive house price appreciation, and thus a household experiencing cash flow problems can either sell their house or extract equity from their home by refinancing. In periods of negative house price growth, households suffering cash flow problems with negative equity in the home will be unable to sell or refinance, and will be forced to default. Further evidence supporting this claim can be seen in the behavior of delinquency rates and foreclosure rates during the mid 1990s. In this period we see delinquency rates rise, but again foreclosure rates are falling as house prices are rising. These empirical observations are completely consistent with the theoretical prediction that negative equity is a necessary condition for default. Thus, based upon the implications from our theoretical model and the evidence presented above, we believe that the assumption in the empirical model regarding the exogeneity of house prices is reasonable.

5.3 Subprime Analysis

In Figure 10, the Kaplan-Meier non-parametric estimates of the default hazard for home purchases that were initially financed with a subprime mortgage versus the estimates for those that were financed with a prime mortgage are substantially different. The subprime default hazard is approximately 10 times larger than the prime default hazard in the figure, which translates into a significant difference between the two categories in terms of the cumulative incidence of defaults. Figure 11 displays the implied cumulative incidence for the subprime and prime lending categories, and shows that after 12 years, approximately 13 percent of ownerships from our subprime sample have defaulted, while fewer than 2 percent of the prime sample have defaulted.

The problem with the non-parametric estimates is that these do not control for the other covariates (except for the subprime dummy). However, we can use the estimated model to perform a similar calculation and control for the effects of all of the explanatory variables. Specifically, the model allows us to obtain predicted default probabilities for a representative ownership experience from a specific cohort or group of cohorts c. To do this, we input the average covariate values of cohort c at each quarter in the life of the ownership, and obtain a predicted default hazard at each quarter of the ownership j from the estimated model. The predicted representative default hazards at each quarter j are given by

$$\widehat{\lambda}_j^c = \exp(\widehat{\lambda}_{j0} + \overline{x_j^c}'\widehat{\beta}), \tag{16}$$

where $\overline{x_j^c}$ is the vector of covariate averages at ownership age j of cohort c. We then calculate the cumulative probability of default for cohort c:

$$\widehat{I}_j^c = \widehat{\lambda}_j^c \sum_{k=1}^{j} (1 - \widehat{\lambda}_k^c). \tag{17}$$

In the upper left panel of Figure 15 we calculate the cumulative probability of default of subprime financed ownerships, using only data from ownerships financed by subprime loans, which entails excluding the subprime dummy from $\widehat{\beta}$. That is, we input $\overline{x_j^{sp}}$ into equation 16, where $c = sp$ indicates that we are averaging over subprime (sp) borrowers.[36] The plot shows that after 12 years, the cumulative probability of default of a subprime-financed ownership is approximately 13 percent, which corresponds exactly to the non-parametric estimate in Figure 11. This implies that the probability of a "positive" outcome of a subprime financed ownership is 87 percent, where positive is defined as either electing to sell or to remain in the home.

One problem with this exercise is that it may not be representative to use only data from subprime ownerships. In fact, this statistic may significantly underestimate the actual default risk associated with a subprime financed ownership, because the subprime mortgage market has been in existence only since the mid-1990s. The first column of Table 4 displays the percentage of ownership experiences each year that were financed with a mortgage from a subprime lender. This percentage was effectively zero before 1994, and less than 3 percent for every year up until 2003. In 2004 it reached 10 percent, and it peaked at almost 15 percent in 2005. During this time period, the Massachusetts housing market experienced large and persistent price appreciation, as depicted in Figure 1. From our simple theoretical model, and from the empirical estimates in Table 7, we know that cumulative house price appreciation is an important determinant of default risk. Furthermore, the historic incidence of foreclosure has been extremely low in times of significant price appreciation, which has certainly been the case in Massachusetts between the mid–1990s and the mid–2000s. Thus, with the exception of the past year or so, the subprime mortgage market has experienced good luck in the form of persistent, positive house price growth.

[36] This is the exact model counterpart to the Kaplan-Meier estimates in Figure 11.

A more informed calculation would be to compute the covariate averages using all ownership experiences over the 18-year period that our data span—which includes intervals of both positive and negative house price appreciation in Massachusetts—and also include the estimate on the subprime dummy variable. That is, we input $\overline{x_j^{all}}$ into equation 16, where $c = all$ indicates that we are averaging over the entire sample of borrowers. However, here we separate the hazards between prime and subprime financed ownerships for each quarter of ownership, by setting the subprime indicator variable to 1 and 0, respectively. The upper right panel of Figure 15 displays the results of this exercise. Taking averages of the explanatory variables across all ownership experiences increases the cumulative probability of default after 12 years, from 13 percent to approximately 18 percent for ownerships financed by subprime lenders. In comparison, the cumulative probability of default after 12 years for ownerships financed by prime lenders is only 3 percent. The reason for the increase in cumulative default probabilities is that we are now including in our calculation periods of large, persistent house price depreciation (early 1990s), as well as periods of very high unemployment (late 1980s and early 1990s).

6 Conclusion

In this paper, we use data encompassing two housing cycles in Massachusetts to document the foreclosure incidence of ownerships financed with a subprime mortgage versus ownerships financed with a prime mortgage. Rather than following the traditional methodology in the literature of estimating the determinants of default for single mortgages issued at purchase, we estimate the determinants of default for the entire duration of ownership. We also depart from the literature in our perspective and treatment of the default decision. We build a model that retains the basic structure of the frictionless option model, but yields the intuitive prediction that financially strapped borrowers are more likely to default, conditional on a given level of house prices and interest rates. To do this, we embed the default decision in a simple, two-period, consumption-portfolio choice model with realistic frictions. The presence of financial frictions in housing markets such as different borrowing and lending rates, and short-sale constraints imply that the standard option value approach taken to determine default probabilities is deficient. The basic difference between our model and the FOM, is that the value of the house to the borrower is *not* necessarily equal to the current market price of the house. As a result, household-specific variables such as income, wealth, and preferences, which determine a household's valuation of the property, play important roles in the decision to default.

Using a competing risks, proportional hazard, duration model of default and sale, and using a set of proxies to control for variables that our theoretical model emphasizes play an important role in the default decision, we estimate that homeownerships financed by subprime mortgages are five to six times more likely to default than are homeownerships financed by prime mortgages, on average, at any point in the life of the ownership. Furthermore, using data from our entire 18-year sample, we estimate that within 12 years of purchasing a home, the cumulative probability of default for a subprime borrower is approximately 18 percent, compared with 3 percent for prime borrowers. Put differently, we estimate the probability of success over 12 years for a residential housing purchase initially financed with a subprime mortgage is 82 percent, where we define "success" as either the ability to remain in the house and continue servicing the monthly mortgage payment, or electing to sell the house. However, we find that the probability of success is very sensitive to the macroeconomic environment, and specifically to house price appreciation. We estimate that the probability of default for subprime borrowers, as well as for prime borrowers, increases significantly in periods with low or negative house price appreciation. In contrast, if we were to use only data that span the existence of the subprime mortgage market, from 1994 to 2007, when Massachusetts house prices grew at an exceptional rate, we would find a significantly higher probability of success for a subprime ownership. This is an important finding because of the favorable economic environment that has largely characterized the existence of the subprime mortgage market from its emergence in 1993 up until the past few years.

This paper has largely focused on the subprime lending channel, or lenders who make mostly high-cost loans to risky borrowers. We believe that this is appropriate given that the recent foreclosure closure crisis has centered around subprime lending. However, it would also be interesting to study specific components of the subprime lending channel, and their roles in the foreclosure crisis. For example, it may be important from a policy standpoint to pinpoint the exact types of borrower and mortgage characteristics that most often lead to default and foreclosure. However, this would require more detailed data, and specifically more information regarding borrower characteristics and mortgage characteristics. We are currently exploring other sources of data that could be combined with the data used in this study to allow for a more detailed analysis of the subprime mortgage market. These sources of data include HMDA data, which contains information regarding the racial and ethnic makeup of a borrower, the income level of a borrower at the time of origination, and finally information regarding the contract interest rate associated with certain types of loans, and Loan Performance data, which includes very detailed information on mortgage characteristics, as well as information on credit scores and debt-to-income ratios of borrowers at the time origination.

References

[1] Apgar, W., and M. Duda, 2005. "Collateral Damage: The Municipal Impact of Today's Mortgage Foreclosure Boom." Report prepared for the Homeownership Preservation Foundation.

[2] Black, F., and M. Scholes, 1973. "The Pricing of Options and Corporate Liabilities." *Journal of Political Economy* 81(3):637–654.

[3] Campbell, T., J. Dietrich, and P. Zorn, 1983. "The Determinants of Default on Conventional Residential Mortgages." *Journal of Finance* 38(5): 1569–1581.

[4] Case, K., and R.J. Shiller, 1987. "Prices of Single Family Homes Since 1970: New Indexes For Four Cities." *NBER Working Paper* No. 2393.

[5] Chan, S., 1996. "Residential Mobility and Mortgages." *Regional Science and Urban Economics* 26:287–311.

[6] Courchane, M., B. Surette, and P. Zorn, 2004. "Subprime Borrowers: Mortgage Transitions and Outcomes." *Journal of Real Estate Finance and Economics* 29(4): 365–392.

[7] Cox, J., J. Ingersoll, and S. Ross, 1985. "An Intermporal General Equilibrium Model of Asset Prices." *Econometrica* 53(2): 363–384.

[8] Deng, Y., and J. Quigley, 2004. "Woodhead Behavior and the Pricing of Residential Mortgages." *Program on Housing and Urban Policy Working Paper Series* No. W00-004.

[9] Deng, Y., J. Quigley, and R. Van Order, 2000. "Mortgage Terminations, Heterogeneity and the Exercise of Mortgage Options." *Econometrica* 68(2): 275–307.

[10] Ferguson, N., 2007. "Maybe owning a home is not for everyone." http://www.telegraph.co.uk.

[11] Han, A., and J. Hausman, 1990. "Flexible Parametric Estimation of Duration and Competing Risk Models." *Journal of Applied Econometrics* 5(1): 1–28.

[12] He, H., and N. Pearson, 1991. "Consumption and Portfolio Policies with Incomplete Markets and Short-Sale Constraints: The Finite-Dimensional Case." *Mathematical Finance* 1(3): 1–10.

[13] Henderson, J., and Y. Ioannides, 1989. "Dynamic Aspec ts of Consumer Decisions in Housing Markets." *Journal of Urban Economics* 26: 212–230.

[14] Henly, A., 1998. "Residential Mobility, Housing Equity and the Labour Market." *The Economic Journal* 108(447): 414–427.

[15] Herzog, J., and J. Early, 1970. *Home Mortgage Delinquency and Foreclosure.* New York: National Bureau of Economic Research 2006-042A.

[16] Immergluck, D., and G. Smith 2006. "The Impact of Single-family Mortgage Foreclosures on Neighborhood Crime." *Housing Studies* 21(6): 851–866.

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

[18] Kau J., D. Keenan, and T. Kim, 1994. "Default Probabilities for Mortgages." *Journal of Urban Economics* 35: 278–296.

[19] Kau J., D. Keenan, W. Muller, and J. Epperson, 1995. "The Valuation at Origination of Fixed-Rate Mortgages with Default and Prepayment." *Journal of Real Estate Finance and Economics* 11: 5–36.

[20] Meyer, B., 1995. "Semiparametric Estimation of Hazard Models." MIT working paper.

[21] Pavlov, G., 2001. "Competing Risks of Mortgage Termination: Who Refinances, Who Moves, and Who Defaults?" *Journal of Real Estate Finance and Economics* 23(2): 185–211.

[22] Pennington-Cross, A., and G. Ho, 2006. "The Termination of Subprime Hybrid and Fixed Rate Mortgages." *Federal Reserve Bank of St. Louis Working Paper Series* 2006-042A.

[23] Quercia, R., and M. Stegman, 1992. "Residential Mortgage Default: A Review of the Literature." *Journal of Housing Research* 3(2): 341–379.

[24] Sueyoshi, G., 1992. "Semiparametric proportional hazards estimation of competing risks models with time-varying covariates." *Journal of Econometrics* 51: 25–58.

[25] Vandell K., 1995. "How Ruthless is Mortgage Default? A Review and Synthesis of the Evidence." *Journal of Housing Research* 6(2): 245–264.

[26] von Furstenberg, G. 1969. "Default Risk on FHA-Insured Home Mortgages as a Function of the Term of Financing: A Quantitative Analysis." *Journal of Finance* 24: 459–477.

[27] Williams, A., W. Beranek, and J. Kenkel, 1974. "Default Risk in Urban Mortgages: A Pittsburgh Prototype Analysis." *AREUEA Journal* 2(2): 101–102.

[28] Yu, X., 2006. "Competing Risk Analysis of Japan's Small Financial Institutions." *IMES Discussion Paper Series* No. 2006-E-7.

Appendix

A Construction of Weighted, Repeat-Sale House Price Indexes

We follow Case and Shiller (1987) and Deng, Quigley, and Van Order (2000) in constructing our house price indexes. The index for each town is based on the assumption that the house price of household i at time t is given by the following process:

$$\ln P_{it} = \ln \bar{P}_t + \mu_{it} + \eta_{it}, \tag{18}$$

where $\bar{P}_t$ is the house price level of the town, η_{it} is white noise, and μ_{it} is a Gaussian random walk with mean equal to zero, $E[\mu_{i,t+k} - \mu_{it}] = 0$, and variance proportional to the age of the loan, $E[\mu_{i,t+k} - \mu_{i,t}]^2 = k\sigma_1^2 + k^2\sigma_2^2$.

The index is calculated using a three-stage process on paired sales of one-family houses. High frequency sales—occurring six months or less apart—were dropped. Sales with appreciation greater than 50 percent within one year, greater than 100 percent within 1.5 years, and 150 percent within two years were dropped due to the high probability of renovation. Sales with depreciation larger than 25 percent within one year and 50 percent within two years were dropped because of the high probability of a subdivision or heavy deterioration reducing the property's previous value. House prices less then \$15,000 and greater then \$10 million were dropped.

House price indexes were developed for every Massachusetts city with more than 5,000 sale observations between January 1987 and August 2007. All other towns were excluded from the sample. In the first stage, the log price of the second sale minus the log price of the first sale is regressed on a set of time dummy variables, D_t:

$$\dot{p}_i = \sum_{t=1}^{T} \beta_t D_t + \omega_i, \tag{19}$$

where $\dot{p}_i = \ln P^{2nd}_{i,t+k} - \ln P^{1st}_{i,t}$, and the dummy variables have the value +1 for the time of the second sale, and the value -1 for the time of the first sale.

In the second stage, the squared residuals, ω_i^2, are regressed on k and k^2:

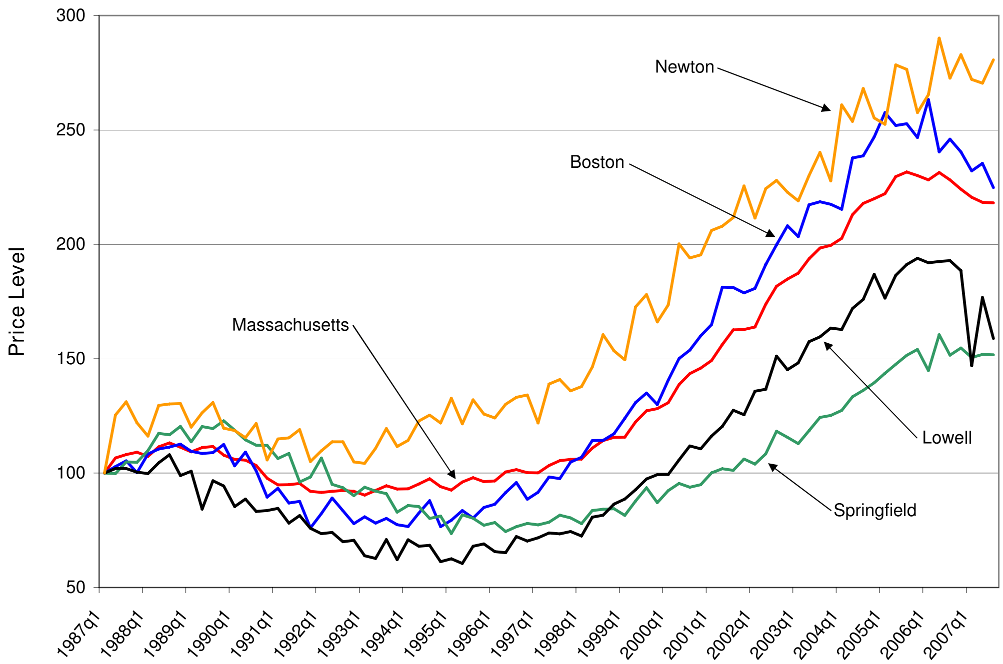
$$\omega_i^2 = A + Bk + Ck^2. \tag{20}$$

In the third stage, equation 19 is estimated by GLS, using $\hat{\omega}_i$—the square roots of the predicted values of equation 20—as weights.

The house price index is then constructed from the estimates β_t:

$$P_t^{Index} = 100 * \exp(\beta_t). \tag{21}$$

In the figure below, we plot P_t^{Index} for the entire state of Massachusetts, as well as four cities—Boston, Springfield, Lowell, and Newton—which are indicative of the state's various geographic regions, as well as different ethnic and income pools.



B Robustness Check for Subprime Lender Indicator

This appendix provides a detailed robustness check of our subprime purchase indicator variable. As we discussed above, this variable is based on HUD's annual list of lenders who originate predominantly subprime loans. This list is certainly an imperfect proxy for the subprime market, as the mortgage lenders on the HUD list do not do business in the subprime market exclusively. Moreover, there are lenders who are not on the HUD list, but who do provide loans to subprime borrowers. Therefore, by using the HUD subprime list, we are introducing two potential mistakes into our analysis . First, we are falsely labeling some mortgages as subprime that were really made to prime borrowers, a misidentification that we call "false positives." Second, we are falsely labeling some mortgages as prime that are really issued to subprime borrowers, a misidentification that we call "false negatives." If our results are to be taken seriously, it is important to estimate the magnitude of both of these identification errors.

Fortunately, we have a limited amount of information regarding mortgage interest rates in The Warren Group data that makes it possible to check the magnitude of these errors. We have detailed interest rate information on approximately 5 percent of mortgages originated between 2004 and 2006, including the initial interest rate, the date when the interest is first allowed to change, often referred to as the "hybrid" term, the index that the rate is tied to once it is allowed to fluctuate, and the difference between the mortgage rate and the index, commonly referred to as the "margin." However, this information is available only for adjustable-rate mortgages.[37]

With these variables we construct our own definition of a "high-cost" mortgage and use this to calculate subprime mortgage indicators to compare with the indicators taken from the HUD list. First, we took all of the adjustable-rate mortgages with initial interest rate information, and divided these into two categories: those with initial fixed-rate terms (hybrid terms) less than three years inclusive, and those with initial fixed-rate terms over three years. Then, we took the Federal Home Loan Mortgage Corporation (FHLMC), prime, 1-year adjustable interest rate series, and subtracted it from the initial interest rates for mortgages with initial fixed-rate terms less than three years. Similarly, we took the Freddie Mac, prime, 5-year adjustable interest rate series and subtracted it from the initial interest

[37]The information is taken directly from the mortgage documents in the Registries of Deeds. Adjustable-rate mortgage documents contain "riders" that state all of the information necessary for a borrower to calculate her interest at each stage of the mortgage. Fixed-rate mortgage documents do not contain any interest rate information. Interest rate information for fixed-rate mortgages is located in the mortgage note, which is not publicly available.

rates for mortgages with initial fixed-rate terms over three years. We labeled as subprime all mortgages for which this difference was greater than 200 basis points, *or* for which the margin was greater than 350 basis points. This method is a somewhat arbitrary way to define a subprime mortgage; however, the term “subprime” itself is somewhat arbitrary, and a universal definition of a subprime mortgage does not currently exist. Some analysts use a secondary market definition and define subprime to mean a mortgage that is part of a pool of securitized mortgages that has been labeled “subprime.” However, in many cases subprime mortgage pools also contain a non-negligible number of mortgages from credit-worthy borrowers that many analysts would consider to be prime borrowers.[38] Other analysts, including those who work with HMDA data, define a subprime mortgage as we do, based on the spread between the contracted interest rate and a market index; however, there is no general consensus on the magnitude of the spread. We believe that our subprime definition is conservative, as a 200 basis point spread between the initial interest rate and the corresponding FHLMC prime series is large, and a 350 basis point margin is also quite large. For example, a two-year hybrid, adjustable-rate mortgage originated in July 2007, with an initial interest rate of 8 percent, and a margin of 350 basis points with respect to the 6-month LIBOR, which would translate into an interest rate of approximately 9 percent (assuming that interest rates remain constant over the next two years), would barely qualify as a subprime mortgage using our definition.

The table below displays a cross-tabulation of the subprime variable constructed with the HUD list, versus the subprime variable constructed using the interest rate information. The most important statistics for our purposes are the two that are highlighted in bold font. The interpretation is that 7.85 percent of mortgages from a subprime lender on the HUD list are considered to be prime by our definition, using the spread between the contract rate and the appropriate FHLMC series. Similarly, 26.04 percent of mortgages from a non-subprime lender on the HUD list are considered to be subprime.

[38]For example, we found using data from Loan Performance Inc. for Middlesex County, MA, that 20 percent of mortgages in subprime pools are made to borrowers with FICO scores above 700.

	HUD subprime indicator		
	prime	*subprime*	Total
prime	43,485	1,361	44,846
(< 125 basis point difference)	96.97%	3.03%	
	73.96%	**7.85%**	
subprime	15,308	15,981	31,289
(≥ 125 basis point difference)	48.92%	51.08%	
	26.04%	92.15%	
Total	58,793	17,342	

With these two statistics we can develop a simple model to estimate the effect of these errors on foreclosure rates. If we let D^*_{sub} be the observed foreclosure rate for subprime ownerships, D^*_{prime} the observed rate for prime ownerships, and D_{sub} and D_{prime} the respective true rates, we can write the observed rates as linear functions of the true rates and the errors:

$$D^*_{sub} = D_{sub} * (1 - \alpha) + D_{prime} * \alpha \tag{22}$$

and

$$D^*_{prime} = D_{sub} * (1 - \beta) + D_{prime} * \beta, \tag{23}$$

In equation 22, α is the fraction of mortgages identified as subprime from the HUD list that are really prime, and β is the fraction of mortgages identified from the HUD list as prime that are really subprime. Since we have estimates of both α and β from the table above, the system of equations 22 and 23 contains two unknowns (the true, unbiased foreclosure rates), and thus it is easy to solve. If we assume a spread of 200 basis points and 350 basis points as our subprime definition criteria, our result that approximately 18 percent of subprime ownership experiences end in default after 12 years, then changes to approximately 19.5 percent ending in default. It is important to note that the unbiased foreclosure rates are sensitive to our estimates of α and β, which in turn are a function of the spread between the contracted mortgage rate and the rate on upon which we based our subprime definition. Furthermore, the key assumption on which the above relationship depends is that the sample of subprime mortgages that we capture with the HUD list, and the entire population of subprime mortgages, are not systematically different.

Table 1: Number of Foreclosures and Sales by Year

Year	**Foreclosures**	**Sales**	Year	**Foreclosures**	**Sales**
1990	1,641	53,686	1999	2,003	100,380
1991	5,432	54,170	2000	1,431	94,666
1992	9,101	62,409	2001	1,060	92,240
1993	8,044	69,231	2002	940	96,046
1994	6,990	76,675	2003	572	99,468
1995	4,617	72,518	2004	615	109,816
1996	4,156	82,274	2005	873	106,504
1997	3,780	88,107	2006	2,731	91,075
1998	2,712	98,239	2007	4,798	80,436

Note: Sale and foreclosure numbers come from data compiled by The Warren Group, and reflect only residential properties. Data from 2007 are available through August.

Table 2: Initial Loan-to-Value Ratios

	All ownerships			Ownerships that Default		
	#	mean	median	#	mean	median
1988	70,010	0.76	0.80	5,767	0.83	0.80
1989	58,192	0.76	0.80	4,429	0.85	0.88
1990	47,478	0.79	0.80	2,549	0.87	0.90
1991	49,257	0.79	0.80	1,243	0.90	0.95
1992	58,104	0.80	0.80	932	0.91	0.95
1993	64,374	0.82	0.85	912	0.92	0.95
1994	70,976	0.82	0.87	924	0.92	0.95
1995	66,153	0.83	0.88	828	0.94	0.98
1996	75,208	0.83	0.87	811	0.94	0.98
1997	80,450	0.83	0.85	786	0.93	0.97
1998	90,441	0.83	0.85	666	0.93	0.95
1999	91,734	0.82	0.85	685	0.93	0.95
2000	86,167	0.81	0.82	659	0.91	0.95
2001	83,958	0.82	0.85	570	0.92	0.95
2002	87,308	0.81	0.82	630	0.92	0.95
2003	89,525	0.81	0.85	737	0.93	0.95
2004	98,142	0.82	0.86	1,294	0.94	0.99
2005	95,586	0.83	0.90	1,429	0.95	1.00
2006	80,993	0.84	0.90	529	0.96	1.00
2007	51,512	0.84	0.90	7	0.96	1.00

Note: Each year corresponds to a specific ownership cohort. Loan-to-value ratios are cumulative in the sense that they include all mortgages obtained at the time of purchase, including second and third mortgages where applicable.

Table 3: Subprime Lenders

	# loans	% of subprime purchase mortgages	status
Option One Mtg. Corp.	11,243	18.6	operating
New Century Financial Corp.	5,951	9.9	shutdown
Freemont Investment & Loan	5,550	9.2	shutdown
Argent Mtg. Co.	3,599	6.0	shutdown
Summit Mtg. Co.	3,067	5.1	shutdown
Mortgage Lender Net	2,798	4.6	shutdown
Long Beach Mtg. Co.	2,520	4.2	shutdown
WMC Mtg. Corp.	2,316	3.8	shutdown
Accredited Home Lenders	2,174	3.6	shutdown
First Franklin Financial	1,896	3.1	operating
Total	41,114	68.1	-

Note: This is a list of the top ten subprime lenders in terms of number of purchase mortgage originations in Massachusetts from 1993 to 2007. The status of each lender is updated through November 2007.

Table 4: Subprime Foreclosure Statistics

	Subprime Ownerships Started	Foreclosures from ownerships with		
		subprime purchase	last mortgage subprime	at least one subprime
1993	0.09%	0.00%	0.00%	0.00%
1994	0.37%	0.00%	0.00%	0.00%
1995	0.42%	0.00%	0.15%	0.17%
1996	0.89%	0.05%	0.50%	0.64%
1997	1.88%	0.16%	1.57%	1.80%
1998	2.58%	0.98%	4.08%	4.73%
1999	2.45%	3.34%	8.90%	11.16%
2000	2.47%	4.40%	12.44%	15.60%
2001	2.90%	4.98%	13.93%	19.37%
2002	3.92%	7.60%	17.59%	24.04%
2003	6.92%	9.98%	23.06%	29.60%
2004	10.06%	10.78%	26.15%	35.34%
2005	14.81%	17.54%	37.17%	46.60%
2006	13.05%	28.65%	45.60%	59.24%
2007	4.03%	30.40%	43.94%	58.05%

Note: Subprime foreclosure numbers are calculated as percentages of the total number of foreclosures in a given year.

Table 5: Subprime Market Shares

	Subprime Market Share (%)	
	Massachusetts	California
2001	2.3	24.5
2002	2.8	31.3
2003	3.0	32.3
2004	2.3	19.7
2005	2.3	25.9

Note: Statistics are based on HMDA Data, as recorded in The 2007 Mortgage Market Statistical Annual.

Table 6: Descriptive Statistics of Estimation Sample

	Mean	**Std. dev.**	**Min**	**Max**
age	18.25	12.64	1	48
initial LTV	0.83	0.17	0.01	1.25
subprime indicator	0.05	0.22	0	1
condominium indicator	0.22	0.42	0	1
multi-family indicator	0.12	0.33	0	1
cumulative appreciation	0.23	0.33	-0.74	1.54
libor (6-month)	4.44	1.85	1.17	8.60
10-yr t-bill	5.47	1.17	3.62	8.70
unemployment rate (town level)	5.02	2.06	1.13	18.60
median family income (tract-level)	$67,176	$24,493	$12,669	$191,062
% of minority households (tract-level)	17.70	19.58	1.36	99.70

Table 7: Estimation Results

	Default				**Sale**			
	Polynomial Baseline		Non-Parametric Baseline		Polynomial Baseline		Non-Parametric Baseline	
	Coef	Std. Err.	Coef	Std. Err.	Coef	Std. Err.	Coef	Std. Err.
age	0.358	0.017			0.152	$3.36e^{-03}$		
age^2	-0.013	$8.12e^{-04}$			$-6.27e^{-03}$	$1.67e^{-04}$		
age^3	$1.46e^{-04}$	$1.16e^{-05}$			$7.70e^{-05}$	$2.39e^{-06}$		
initial LTV	2.82	0.13			0.732	0.027		
cumulative appreciation	-3.05	0.15			-0.411	0.024		
libor (6-month)	$-1.41e^{-03}$	0.017			-0.045	$3.95e^{-03}$		
10-yr t-bill	0.043	0.027			-0.080	7.03^{-03}		
unemployment rate	0.086	$9.06e^{-03}$			-0.071	3.31^{-03}		
% minority (zip)	$8.65e^{-03}$	$9.26e^{-04}$			$1.22e^{-03}$	$2.85e^{-04}$		
Median Income (2000 zip)	$-1.78e^{-05}$	$1.76e^{-06}$			$-1.47e^{-06}$	$3.05e^{-07}$		
subprime	2.10	0.07			0.323	0.027		
condo	0.512	0.041			0.549	$9.83e^{-03}$		
multi-family	0.620	0.042			0.059	0.015		
constant	-12.10	0.23			-4.69	0.05		
# observations	3,925,438				3,925,438			
Log Likelihood	-314,620				-314,620			

Note: We are currently in the process of estimating the model with non-parametric baseline hazards. Preliminary estimation results with very small subsamples (around 2%) suggest that results are largely unchanged.

Figure 1: Massachusetts House Price Growth, Foreclosures and Delinquencies, January 1989 to August 2007

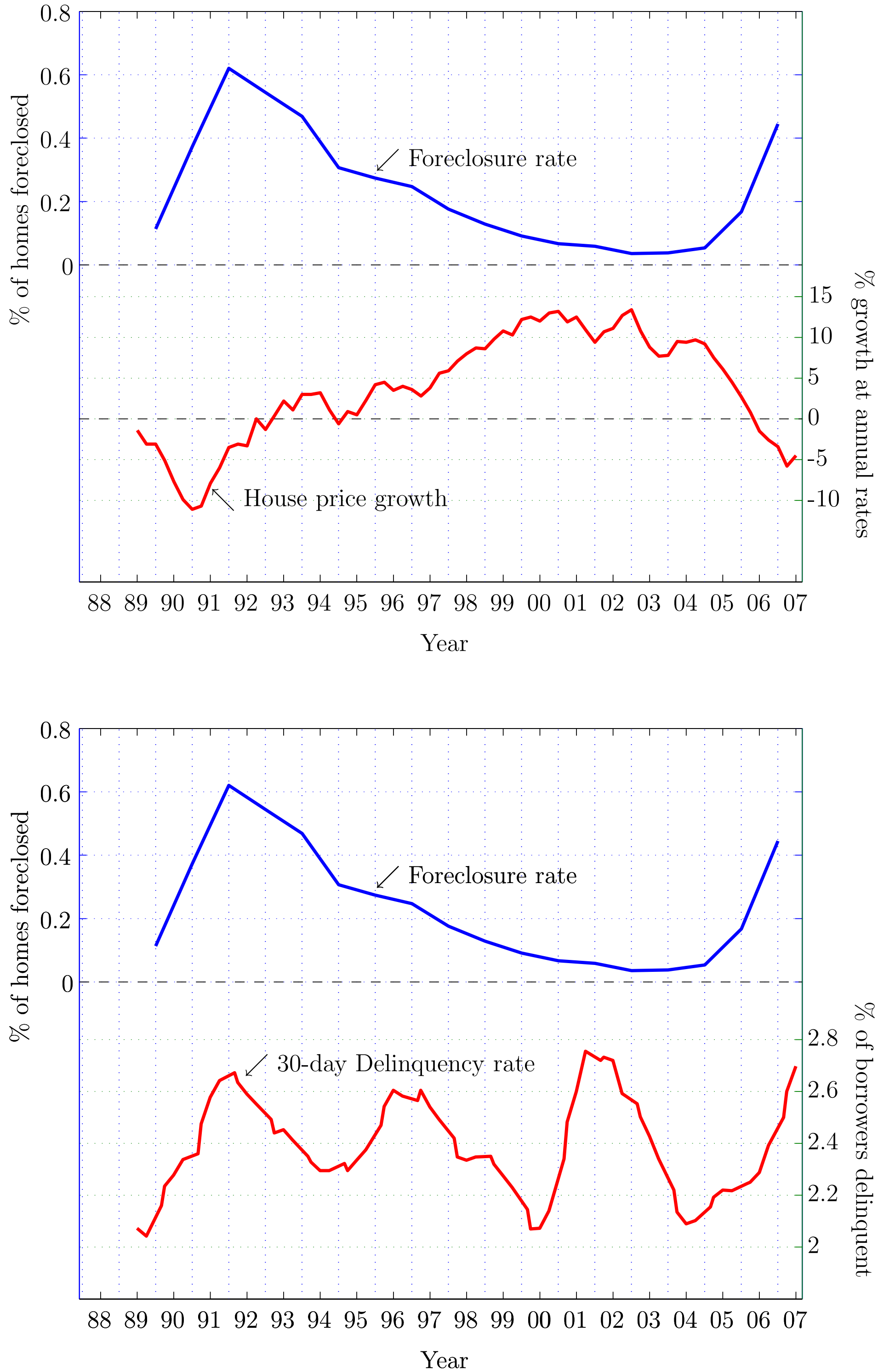
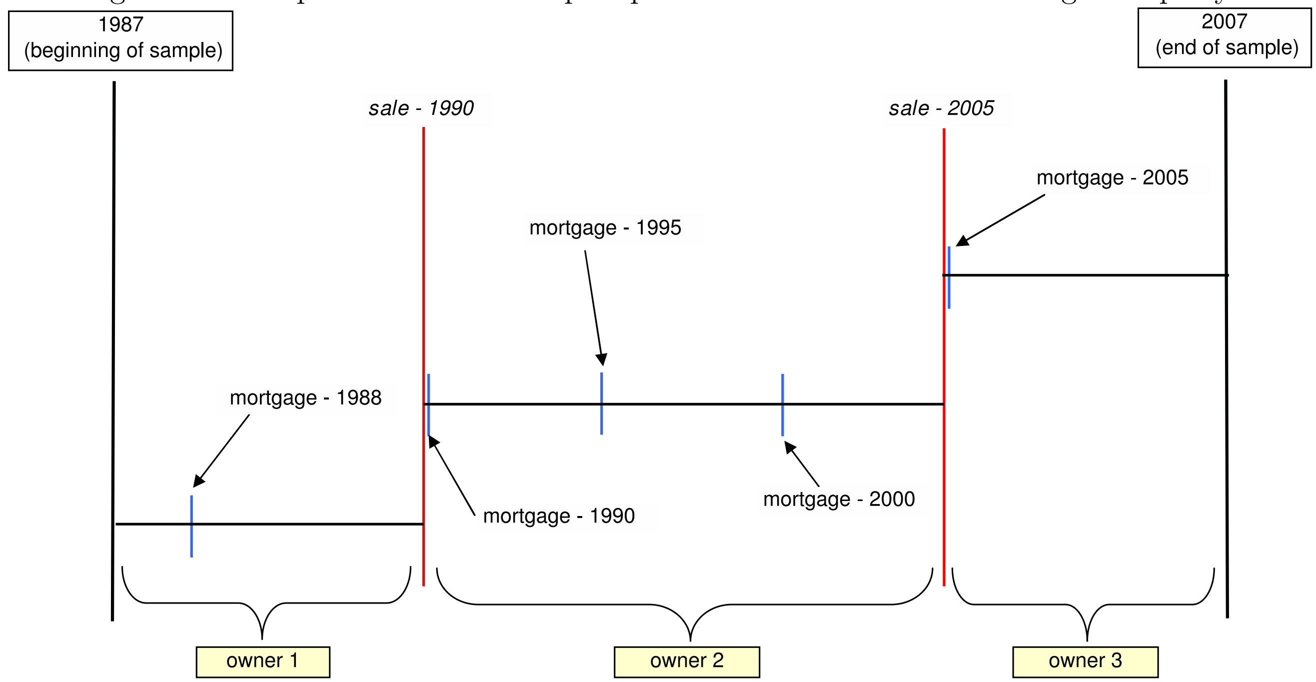
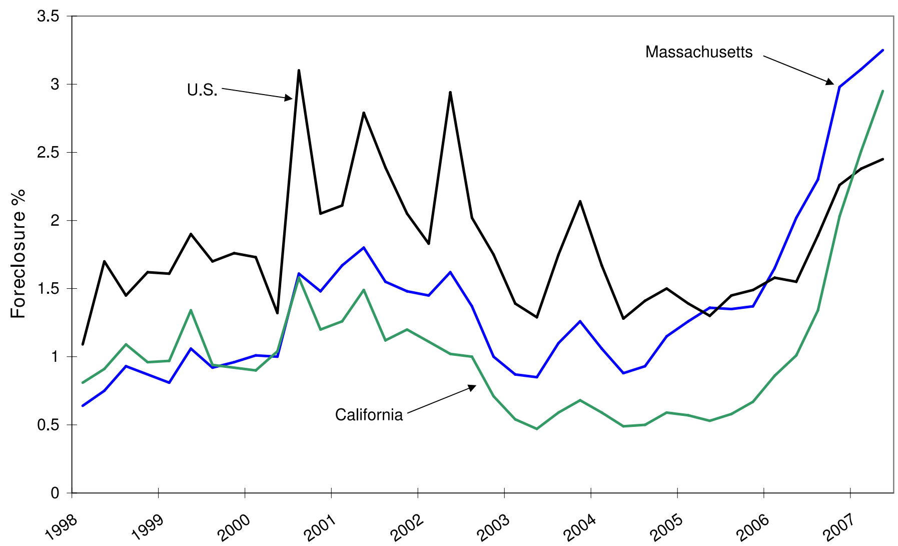


Figure 2: Example of the Ownership Experiences Associated with a Single Property



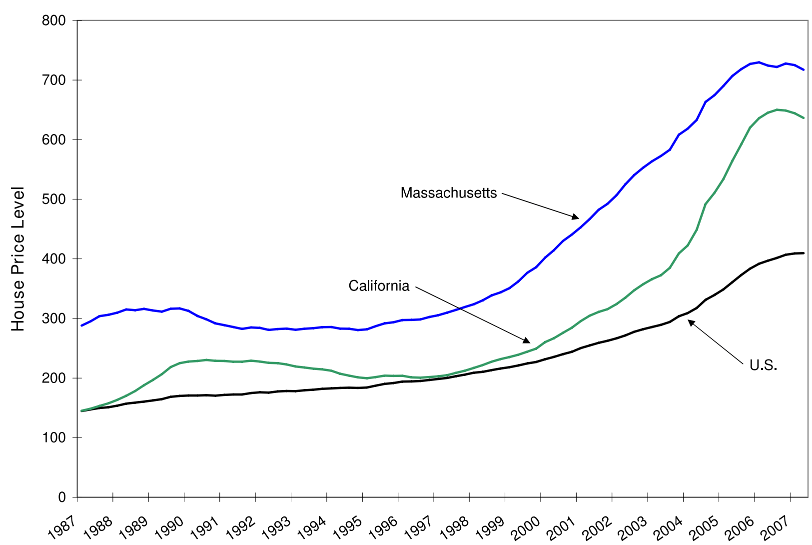
Notes: This example considers a property in which we see two sale transactions (red lines) and five mortgage transactions (blue lines). In this case we are able to identify three separate ownership experiences associated with this property. The first ownership experience comprises all transactions between the beginning of our data and the first sale of the property (1 mortgage). The second ownership experience includes all the transactions between the first and second sale of the property (3 mortgages), and finally, the third ownership experience includes all transactions between the second sale and the end of our data set (1 mortgage). The example in the figure illustrates that censoring is an issue with our data. In this particular example, the first ownership experience is left-censored, since we do not see the start of the experience, while the third ownership experience is right-censored, since we do not see the end of the experience. We address how we treat censored observations in our discussion of the estimation method in section 3.

Figure 3: Foreclosures Started on Conventional Subprime Mortgages



Notes: "Foreclosures started" refers to the percentage rate of loans for which a foreclosure has been initiated during the quarter, that is, the number of loans sent to the foreclosure process as a percentage of the total number of mortgages in the pool. Data is obtained from the Mortgage Bankers Association, and includes only conventional mortgages, which are mortgages with values below the GSE conforming loan limit.

Figure 4: OFHEO House Price Indexes



Notes: Data is from the Office of Federal Housing Enterprise Oversight (OFHEO). Only single-family residential homes are included the house price calculations.

Figure 5: Kaplan-Meier Non-Parametric Default Hazard Rates for Massachusetts

Figure 6: Kaplan-Meier Non-Parametric Sale Hazard Rates for Massachusetts

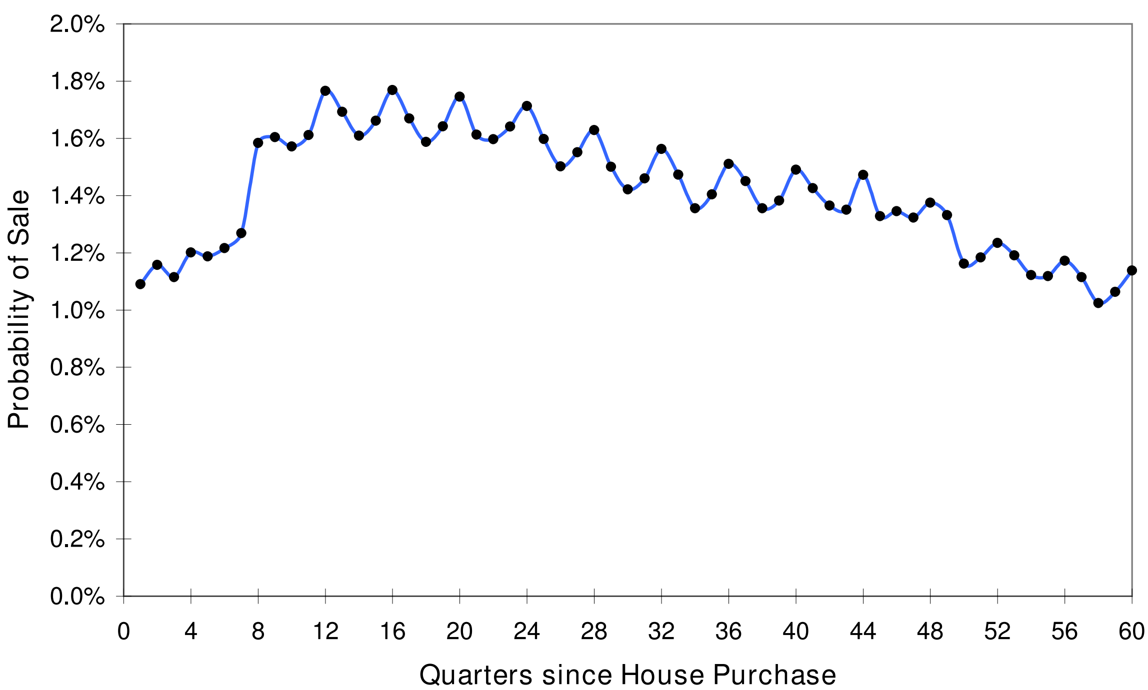


Figure 7: Kaplan-Meier Non-Parametric Default Hazard Rates by Initial CLTV Ratio Categories for Massachusetts

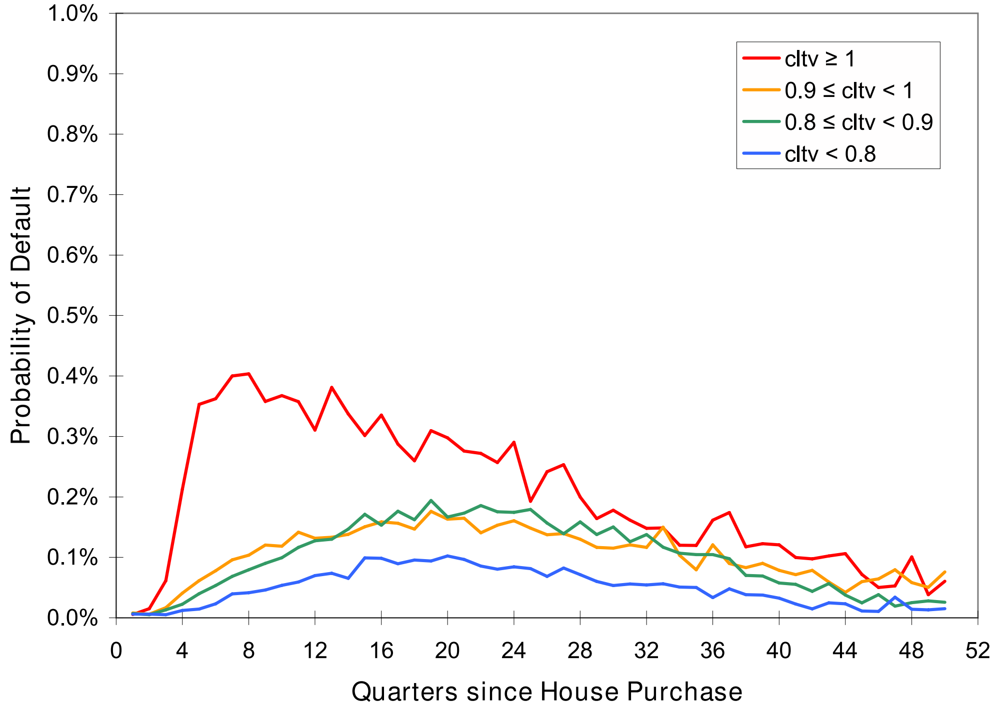


Figure 8: Kaplan-Meier Non-Parametric Default Hazard Rates by Cumulative Appreciation Categories

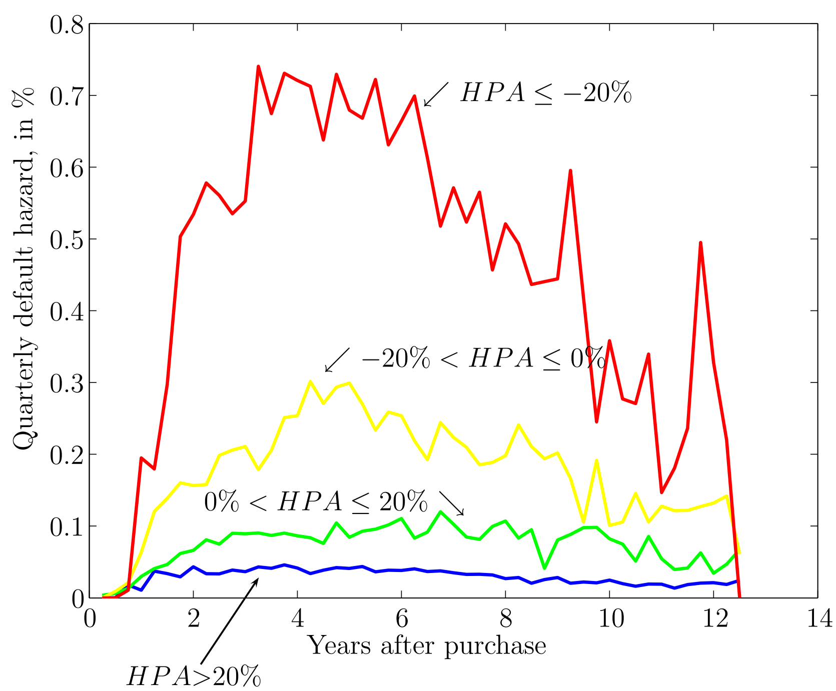


Figure 9: Kaplan-Meier Non-Parametric Default Hazard Rates by Unemployment Rate Categories

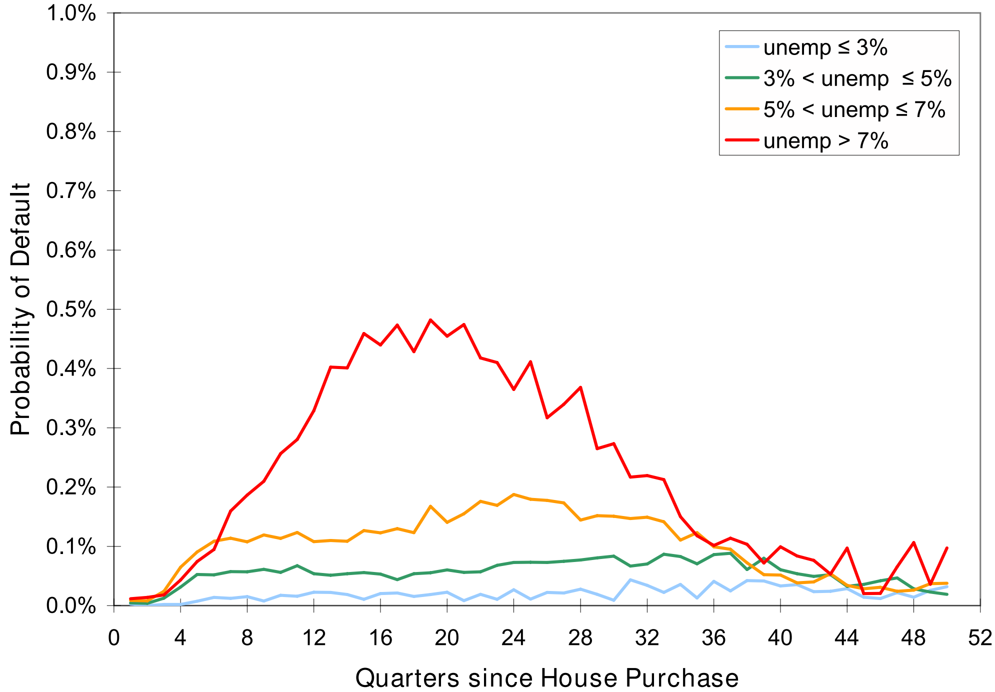


Figure 10: Kaplan-Meier Non-Parametric Default Hazard Rates by Subprime Lender Categories

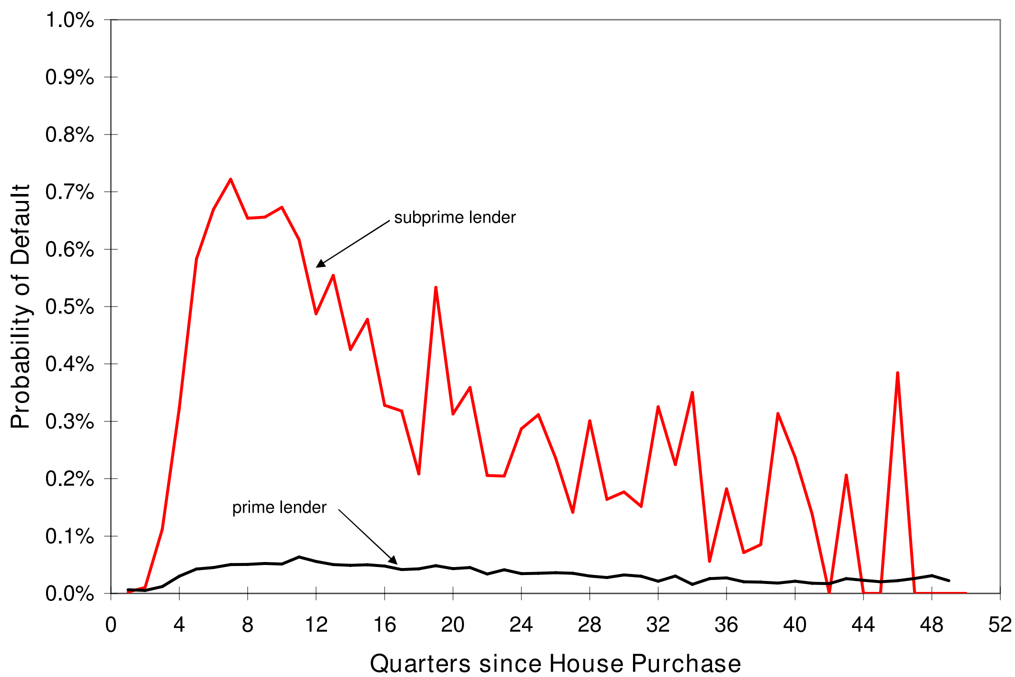


Figure 11: Kaplan-Meier Non-Parametric Cumulative Incidence (Default) - Subprime vs. Prime

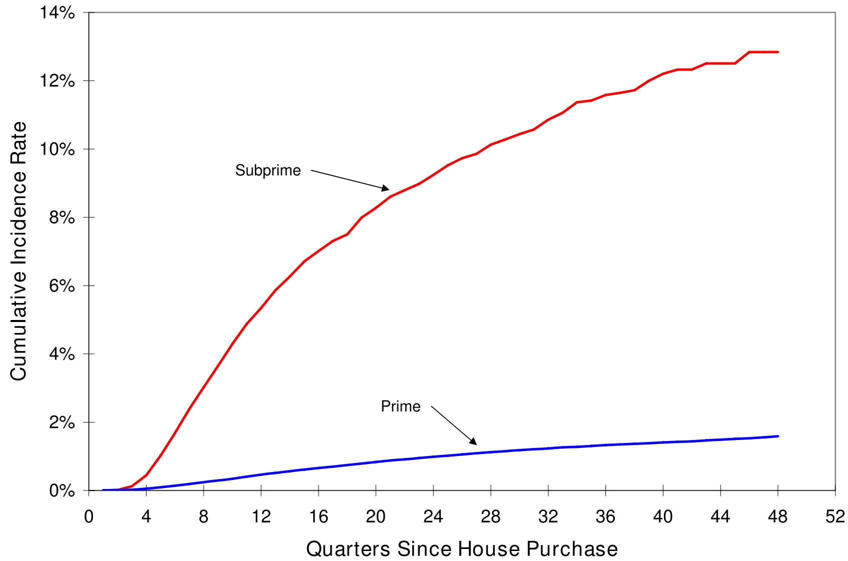


Figure 12: Predicted Hazard Rate versus Kaplan-Meier Hazard Rate (1987-2002)

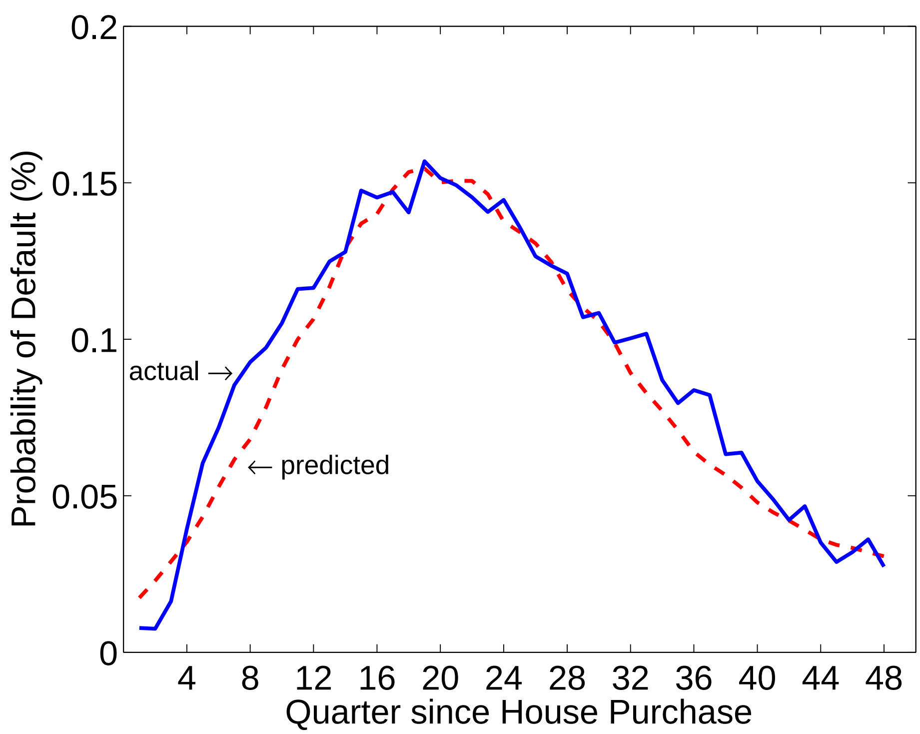


Figure 13: Predicted versus Actual Default Hazard Rates

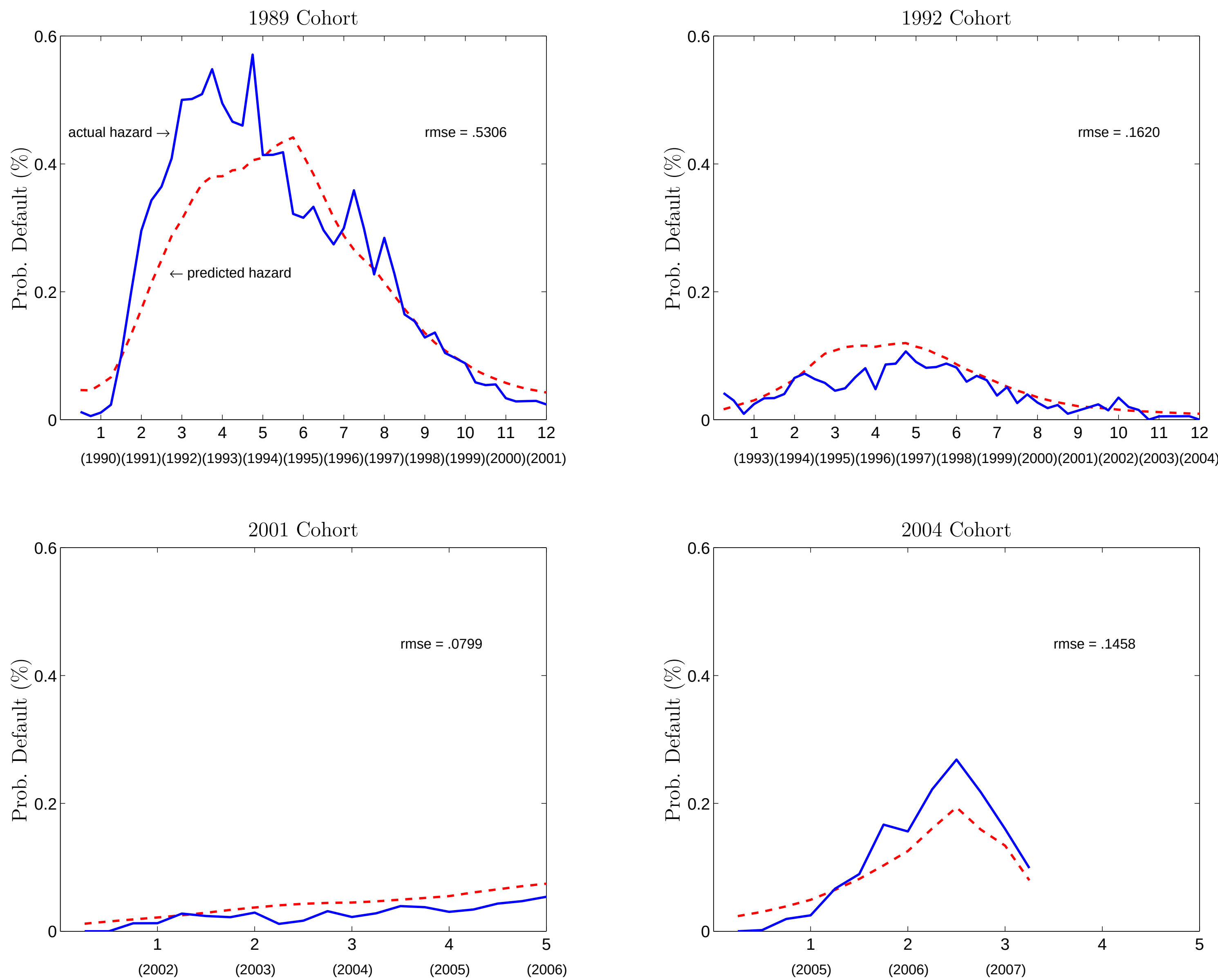


Figure 14: Predicted vs. Actual Subprime Defaults (1993–2007)

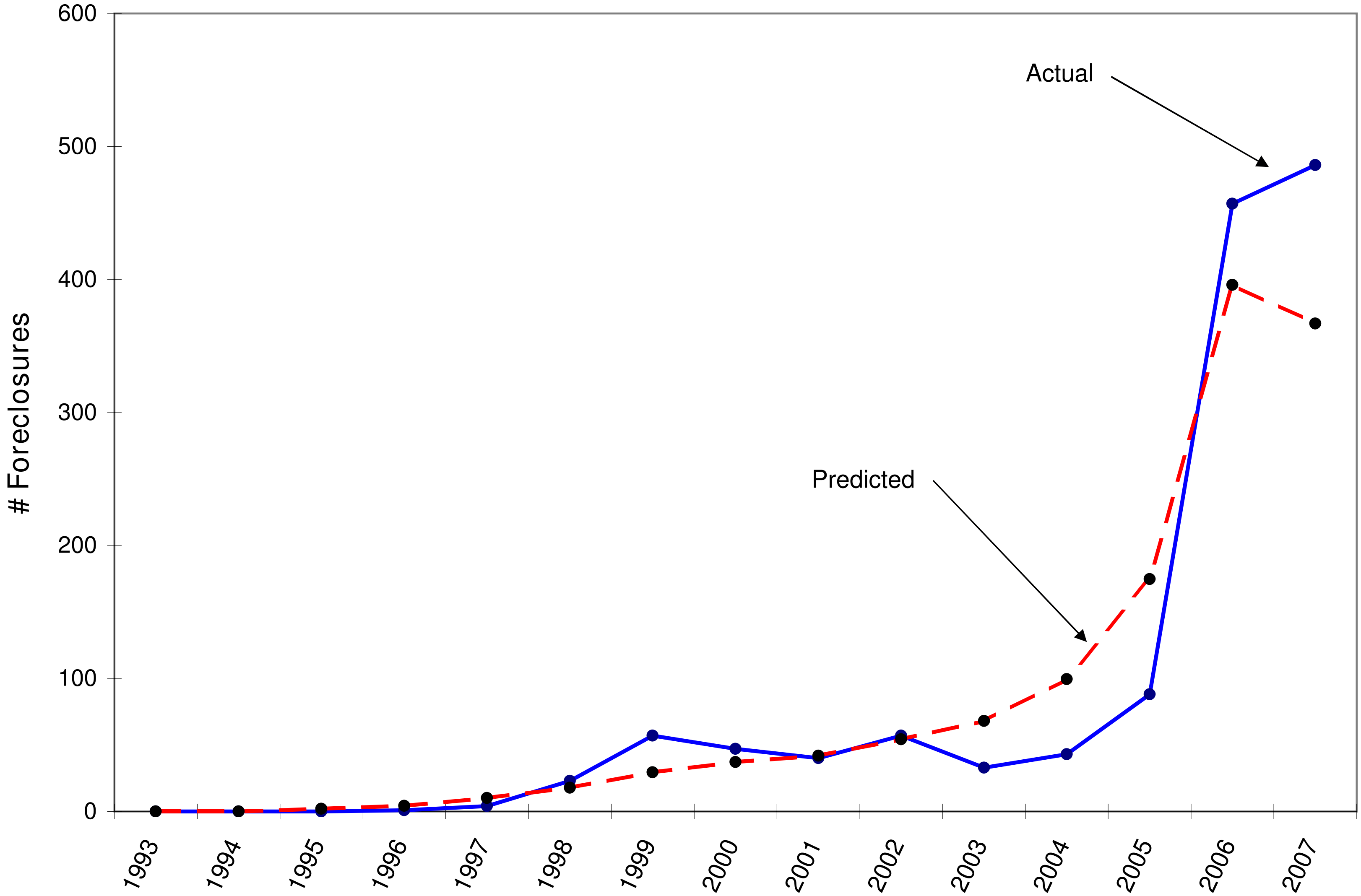


Figure 15: Estimated Cumulative Incidence of foreclosure for Prime and Subprime Ownerships