DEPARTMENT OF ECONOMICS

Contagion Effects in Strategic Mortgage Defaults

Ryan Goodstein
Paul Hanouna
Carlos D. Ramirez
Christof W. Stahel

George Mason University
Department of Economics
Working Paper No. 13-07

Contagion effects in strategic mortgage defaults*

Ryan Goodstein, Paul Hanouna, Carlos D. Ramirez, Christof W. Stahel

January 25, 2013

* A previous version of this paper circulated under the title "Are Foreclosures Contagious?" We would like to thank, without implicating, Itzhak Ben-David, Enrica Detragiache, James Einloth, Paul Kupiec, Myron Kwast, Haluk Ünal and Erlend Nier, as well as seminar participants at the FDIC, the International Monetary Fund, the College of William & Mary, Wilfrid Laurier University, and the 2011 Ohio State Finance Alumni Conference for helpful comments and suggestions. We gratefully acknowledge financial and logistical support from the Center for Financial Research at the FDIC. The views expressed in this paper do not necessarily reflect those of the Federal Deposit Insurance Corporation. The Securities and Exchange Commission, as a matter of policy, disclaims responsibility for any private publication or statement by any of its employees. The views expressed herein are those of the author and do not necessarily reflect the views of the Commission or of the author's colleagues on the staff of the Commission.

Goodstein is at the Federal Deposit Insurance Corporation, email: *rgoodstein@fdic.gov*; Hanouna is at the Villanova School of Business, email: *paul.hanouna@villanova.edu*; Ramirez is at George Mason University, email: *cramire2@gmu.edu*; Stahel is at the U.S. Securities and Exchange Commission, email: *stahelc@sec.gov*, corresponding author.

Contagion effects in strategic mortgage defaults

Abstract

Using a large sample of U.S. mortgages, we document contagion effects in strategic mortgage defaults. These result from borrowers *choosing* to exercise their in the money default option. Our findings suggest this choice is influenced by the delinquency rate in surrounding zip codes after controlling for determinants of mortgage default. The fact that the local area delinquency rate is an important factor for strategic defaulters but not for defaults that are the result of inability to pay supports the contagion hypothesis. Our estimates suggest a 1% increase in the delinquency rate increases the probability of a strategic default up to 16.5%.

1 Introduction

The 2007-2010 real estate market collapse and ensuing financial crisis[1] has highlighted the previously little known fact outside of finance academic and banking circles that US homeowners hold the equivalent of a put option on their mortgages. That is, homeowners have the option to return their property to the lender at any time, which absent of other costs becomes valuable if the loan balance exceeds the market value of the underlying property. Exercising this option necessarily results in a mortgage default, but unlike defaults resulting from an inability to pay, these "strategic defaults" occur because homeowners recognized that the benefits of a default outweighed its costs (Das, 2012).

At the beginning of 2012, 27.8% of all residential borrowers in the United States were underwater on their mortgage loans, and therefore at risk of strategically defaulting.[2] Moreover, when compared to previous real estate downturns, more homeowners who are in a position to strategically default have done so. For example, Foote et al. (2008) find that during the 1990-1991 recession only 6.4% of homeowners with negative equity engaged in a strategic default whereas a 2011 study by Experian-Wyman estimate that in the fourth quarter of 2011 23% of all defaults were strategic.

As a result, there has been interest among academics and policymakers in the factors that compel homeowners to strategically default on their mortgages, and on the potential spillovers associated with these defaults. In particular, a number of papers in the recent academic literature have focused on the effect of foreclosures on nearby home prices. For example, Campbell et al. (2009) and Harding et al. (2009) find that foreclosures significantly

[1]See Brunnermeier (2009) and Gorton (2009) for discussions on the causes of the 2007-2010 financial crisis.

[2]CoreLogic 4^{th} Quarter 2011 Report.

reduce surrounding home prices. This effect is economically significant with both studies estimating that one foreclosure decreases nearby home prices by about 1% to 1.5%. Other studies have examined non-price externalities associated with foreclosures. For example, Immergluck and Smith (2006) and Ellen et al. (2012) find that foreclosures significantly increase local crime rates.

Using survey data, Guiso et al. (2009) and Guiso et al. (2011) identify the externality that a homeowner with negative equity is more likely to strategically default if they know someone who has already done so. We empirically investigate this channel that a mortgage default exposes other mortgages on nearby properties to an increased probability of default, controlling for all other risk factors, including recent house price changes. Consistent with terminology widespread in the extant finance literature (e.g. Bekaert et al. (2005); Boyson et al. (2010)) we use the term contagion to describe this form of externality.

In this paper we empirically investigate the externality identified in Guiso et al. (2009) and Guiso et al. (2011) that a mortgage default exposes other mortgages on nearby properties to an increased probability of default, controlling for all other risk factors, including recent house price changes. Consistent with terminology widespread in the extant finance literature (e.g. Bekaert et al., 2005; Boyson et al., 2010) we use the term *contagion* to describe this form of externality.

The academic literature provides little systematic evidence as to whether defaults are contagious. By and large, the traditional view is that defaults are primarily determined by a borrower's *inability* to pay and therefore any correlation among defaults was interpreted as originating from common economic conditions. Therefore observing clusters of defaults within a given a local area have generally been interpreted as the outcome of local economic shocks.

This view implies that default is necessarily the result of an inability to pay. However, the large pool of underwater homes that dot the US landscape has created a situation, where the option to default is so valuable that the only feature keeping entire communities from defaulting is homeowners' *willingness* to continue servicing their debt.[3] Under these circumstances, default contagion becomes possible because borrowers deciding to exercise their put option affect the willingness of other borrowers to stay current on their loans through at least three possible mechanisms.

First, abstracting from borrower morality, many if not most homeowners might simply be unaware of their option to strategically default, or if they are, may be unsure over the exact process and the costs associated with default. For instance, potential defaulters may have concerns as to whether a (strategic) default entails a bankruptcy procedure, whether other non-collateralized assets can be seized by the bank, the extent of implications of impaired credit scores, among many other considerations.[4] Therefore, strategic defaulters who have already navigated the difficult process can help, or even encourage, neighbors and friends in similar situations to default, thus generating a spiral of defaults. Moreover, learning of the cost and benefits of default through the experience of other borrowers may be more powerful and more cost effective than learning through media channels, accountants, or lawyers as it provides concrete, and locally sensitive, information about the consequences of default. Consistent with this notion Guiso et al. (2009) find that, controlling for borrower morality, "people who know someone who defaulted strategically are 82% more likely to declare their

[3]For example, an article published in the *Ventura County Star* on September 22, 2010 reports that, at the time, 70 percent of homeowners in Las Vegas were underwater.

[4]Dedicated strategic default consulting services have emerged during the crisis and common questions and answers are available at: `http://www.youwalkaway.com/faq/`.

intention to do so."

Second, insofar as moral considerations affect the willingness to default, morality may be relative. That is, as default becomes more commonplace the less morally objectionable it may become (see, for example, Guiso et al., 2009). Thus, when defaults become locally systemic, each strategic default acts to alter homeowners' moral compass towards acceptance of these actions. A similar argument can be made about reputational costs associated with default: "social stigma" (see, for example, Blume, 2010), which accompanies a strategic default, ought to decline with the proportion of borrowers strategically defaulting.

Third, communities with high foreclosure rates experience disruptions in social networks; and when strategic defaulters leave the neighborhood their departure from the community can act as a "trigger" which seals the decision of other underwater homeowners who were previously on the fence.[5] Disruptions in social networks are likely to be more pronounced in tight knit communities where members have established strong ties with each other.

Our empirical strategy for identifying the presence of foreclosure contagion is as follows. We start with a sample obtained from LPS (McDash) of over 30 million mortgages originated over the period 2000-2008 and observed from 2005 to 2009. For each month in that time period we calculate the 90+ days delinquency rates within a 5 mile radius of each zip code and call this statistic the area delinquency rate in a given month t on a given zip code. We then estimate the probability that a given loan in a zip code enters into default as a function of the 3-month lagged area delinquency rate, while controlling for economic fundamentals such as borrower and loan characteristics, changes in property values, economic and demographic conditions at the zip code level, spatial correlations, as well as time and geographic fixed

[5] Certainly, some strategic defaulters may be good riddance and therefore the decision of other homeowners may go either way.

effects. We interpret a positive and significant coefficient of the area delinquency rate on the probability of default as evidence of peer-effects or contagion.

As mentioned above, there are two reasons for households to enter default: inability to pay and unwillingness to pay. In the first case, borrowers have no choice but to default and what other borrowers choose to do is irrelevant to them. However, if foreclosure contagion exists it will be among the borrowers who retain at least some ability to pay but conclude that their option to default is too valuable to be left unexercised. We use this insight in our experimental design and find that, as expected, the area delinquency rate does not affect the probability of default in the overall population of borrowers. However, when focusing on borrowers that are most likely to be strategic defaulters (homeowners with deep negative equity in their homes and high credit scores) we find that the area delinquency rate is statistically and economically significant in determining their probability of default. Moreover, the coefficient on the area delinquency rate for borrowers in that group is significantly different from borrowers not at risk of strategically defaulting at the 1% level. We also find that the magnitude of the effect is economically important. Specifically, our estimates indicate that a one percent change in the area delinquency rate results in an 7.25% increase in the probability of default of borrowers that are potential strategic defaulters.

Given that peer-effect models are notoriously difficult to identify, we perform a battery of different specification and estimation methods. All specifications are estimated using dynamic logistic, Cox (1972) proportional hazard, and competing-risks models, as well as using different definitions of default (90+ day delinquency and foreclosure). In each case the results are qualitatively similar. Since our models are all non-linear and use panel data, they are able to circumvent the Manski (1993) "reflection problem," which we discuss in more detail further below. We are also particularly careful in addressing an omitted variable bias. These fears are assuaged by the fact that we find no evidence that the area delinquency rate affects the probability of default among the general population of borrowers, once we include an extensive array of controls. However, when we estimate the same regressions in the sub-sample of borrowers who, *ex-ante*, have a reason to strategically default, we do find evidence of contagion.

Evidence of foreclosure contagion has important financial and economic ramifications. First, with an outstanding amount approaching $13.56 trillion[6], the U.S. residential mortgage debt market is economically as significant as the U.S. corporate debt market[7] and, as discussed above, has been a root cause of the current U.S. financial crisis. Yet, while there has been much work on identifying and explaining contagion effects in corporate credit markets the same question has largely remained unaddressed in the residential debt markets.[8] Second, contagion has obvious implications for the pricing and design of mortgage contingent securities such as Mortgage-Back Securities (MBS) and its Collateralized Mortgage Obligations (CMO) sub-type. If properly understood, contagion effects among loans composing the MBS pool might be mitigated. Third, understanding the nature of the correlation among mortgage defaults can be a significant input in determining the loan portfolio risk of banks as required by the Basel III accords and is therefore directly relevant to banking regulatory agencies. In particular, foreclosure contagion can be an important element in the current debate on the size of individual banking units. Indeed, in the presence of foreclosure contagion an argument can be made for banks to be sufficiently large and geographically diversified

[6] 3^{rd} Quarter 2011, Mortgage debt outstanding, Board of Governors of the Federal Reserve System.

[7] The Bureau of International Settlements reports U.S. Corporate Debt at approximately $13.5 trillion in the 4^{th} Quarter of 2011.

[8] Das et al. (2007) and Jorion and Zhang (2007) for example find evidence of contagion in corporates.

so as to withstand that effect. Finally, the proper understanding of homeowners' decision to default is critical in designing home-loan modification programs and solving real estate based financial crises.

The rest of this paper is organized as follows. The following section 2 delineates the theoretical framework underlying our model specification. Section 3 discusses the data used to test our model. Section 4 discusses the details of the model, while section 5 reports the results. Section 6 concludes.

2 Framework

In the standard "option model" of mortgage default, borrowers are handed the equivalent of an American put option.[9] That is, at any time mortgage holders can give back the house posted as collateral in exchange for the loan's termination.[10] In this framework, the option to default becomes valuable if the mortgage value exceeds the home value (see, for example, Dunn and McConnell, 1981; Foster and Van Order, 1984; Deng et al., 2000; Das, 2012, among many others.). In the extreme case where transaction costs are absent and there is no memory, borrowers can exercise their "in-the-money" put option by returning their home to the bank, annulling their current mortgage and then buying it right back at the new market value. In this case, the gains from exercising the put option, which is also called a "strategic default" are:

[9] In practice, mortgage default usually occurs after the loan is 90+ day delinquent and therefore the option is more precisely a Bermudan or Canary option.

[10] As we discuss further below other non-collateralized assets may be seized by the banks in case of a deficiency judgment which can be pronounced in certain US states. This affects the option model by increasing the strike price of the option or default barrier.

$$\Pi = MAX[M - H, 0] \tag{1}$$

where M and H are the value of the mortgage and house value, respectively.

In reality, borrowers have to balance many other considerations. In particular, borrowers are handed another valuable option which is to prepay or refinance the mortgage when interest rates drop. Exercising the option to default means losing both the option to default and the option to prepay at a later date (e.g. Deng et al., 2000). Another consideration is that borrower credit scores may be sufficiently impaired as to limit further borrowing. For instance, Brevoort and Cooper (2010) find that following foreclosure, credit scores on almost all borrowers decline to subprime levels regardless of their pre-delinquency levels. Interestingly, they also find that subprime borrowers may face increases in credit scores subsequent to foreclosure and that 60% and 94% of subprime borrowers recovery their initial credit score after 2 years and 8 years, respectively. Taken together these transaction costs increase the default barrier so that the gains to exercising borrowers put option is:

$$\Pi = MAX[M - H - T, 0] \tag{2}$$

where T are the transaction costs.

Thus, having negative equity in the home may be a necessary but certainly not sufficient condition for default as transaction costs can be substantial (e.g. Foster and Van Order, 1984; Vandell, 1995). However, even accounting for transaction costs, it remains a puzzle as to why so few homeowners with deep negative equity do not default. In particular, Deng et al. (2000) point out that: "presumably a substantial number of homeowners are less likely to exercise put and call options on mortgages in the fully rational way predicted by finance theory. Accounting for this group is potentially important in understanding market behavior and in pricing seasoned mortgages." In response, the literature has focused on additional "triggers" that prompt borrowers to exercise their options such as job loss or divorce (e.g. Vandell, 1995; Gerardi et al., 2007; Foote et al., 2008; Elul et al., 2010).

In this paper, we hypothesize that homeowners who strategically default on their mortgages acts as a "trigger event" for other homeowners who also have negative equity in their homes to default. In other words, one strategic default begets further strategic defaults, which conceptually is similar to the epidemiological notion of contagion. In the finance literature, an accepted definition of contagion is the notion of excess correlation, or co-movement beyond what economic fundamentals can explain (Bekaert et al., 2005; Boyson et al., 2010). In the current context, this definition implies that other delinquencies in the neighborhood have an effect on the probability of a strategic default over and above what other factors, for example a price impact, would predict.

Therefore, to investigate whether foreclosure contagion exists, we estimate a model of individual mortgage defaults that contains the delinquency rate in the area surrounding each individual and an extensive array of controls.

Before presenting the formal model, we first discuss in more detail the data underlying this study.

3 Data

We start our analysis by obtaining a sample of mortgage loans from Lender Processing Services Inc. (LPS), formerly known as McDash, which claims to be "the market's deepest and broadest mortgage dataset." LPS consists of loan-level information provided by 16 participating mortgage servicing firms, including nine of the ten largest firms, reporting on over 30 million active loans. The sample contains a rich set of loan characteristics at origination including loan amount, property value, loan terms, interest rate, and amortization terms. Additionally, LPS maintains for each loan a monthly payment record that can be used to measure payment performance and foreclosure status. A potential drawback of LPS data is the well documented under-representation of subprime mortgages.[11] However, it is unlikely that this feature of the data affects our results because the under-represented category of subprime mortgages is, generally speaking, more likely to be underwater and also more likely to exhibit default for non-strategic reasons. Therefore, the probability of finding foreclosure contagion in this group should be lower than for the general population. But, as we elaborate in more detail further below, we already find no evidence of contagion among the general population of borrowers. Hence, the inclusion of a larger sample of subprime borrowers is likely to be redundant.

Our sample consists of loans originated between 2000 and 2008, and we examine payment and default outcomes from January 2005 to December 2009.[12] For computational reasons, we follow the literature by drawing a five-percent random sample of loans in LPS (e.g. Foote

[11]Immergluck (2008) reports that, as of year-end 2008, LPS covered roughly 58 percent of the total prime/near prime market and 32 percent of the subprime market.

[12]In 2010, major U.S. lenders such as JP Morgan Chase, Ally Financial, and Bank of America under political pressure instituted a moratorium in 23 "judicial" US States on foreclosures over the potentially fraudulent practice of robo-signing. Bank of America in particular instituted a moratorium in all 50 States. Therefore, we end our sample in 2009.

et al., 2009). Specifically, we draw the sample as follows. First, we restrict our sample of loans to owner-occupied, first-lien home purchase mortgages on 1-4 family homes. Second, we purge the sample of loans that were not active in at least one month over the period January 2005-March 2009. Third, we remove loans for which no payment records are available within the first 12 months following origination, to address a potential left-censoring bias. Finally, we drop loans on properties not located in the 4,110 zip codes covered by the Case-Shiller zip-code level House Price Index. From these remaining loans we draw a random five-percent sample, and for each loan i we track the monthly performance in a variable D_{it} that is equal to 1 in the first month in which the loan enters 90+ day delinquency and 0 otherwise. All loan-month observations following the month in which a loan first becomes delinquent are excluded from our sample.

From the purged LPS dataset we also compute for each zip code and month from January 2005 to December 2009 the "local area" delinquency rates. The local area delinquency rate, $DQ_{00.05}$, is calculated using the following procedure. First, we compute for each zip code and month pairs a delinquency rate, where the delinquency rate is defined as the number of mortgage loans that are 90 or more days overdue, irrespective of when that event occurred, divided by all active loans. Then, for each target zip code-month, the area delinquency rate is calculated as the average delinquency rate, weighted by the number of active loans, across all zip code-months with zip codes lying within a 5 mile radius of the target.[13] The zip code in which the property is located is excluded when computing $DQ_{00.05}$ for reasons explained further below.

Next, we include variables that proxy for local economic conditions. Specifically, we

[13]Distances are computed based on the geographical coordinates of zip code area centroids.

include monthly county-level unemployment rates and state-level mean unemployment duration from the Bureau of Labor Statistics' Local Area Unemployment Statistics, quarterly county-level credit card delinquency rates from TrenData, monthly county-level housing permit applications from the US Census Bureau regional permit series, and finally the number of monthly zip code level mortgage loan applications calculated from the Home Mortgage Disclosure Act (HMDA) data.[14]

Additionally, to control for the potential price spillover effects in foreclosures documented in Campbell et al. (2009) and Harding et al. (2009), we include the two-year change in house prices at the zip code level based on the Case-Shiller house price indices obtained on Moody's Analytics. We also use Case-Shiller data to compute the current loan-to-value ratio (LTV) for every loan-month observation in our sample. Since LPS does not include information on the presence of junior liens, we are unable to compute the combined LTV ratio for the loans in our sample. Thus, we may be understating the true LTV for some loans in our sample.[15] We control for this possibility by including a dummy variable, which takes the value of 1 if LTV at origination is equal to 80%, and 0 otherwise. This variable serves as a proxy for the presence of a second lien (see Foote et al., 2008).

We obtain zip code level demographic characteristics from the US Census Bureau. The demographic characteristics were compiled in the 2000 decennial census and we include the proportion of the zip-code population self-identified as a minority and as Hispanic. We also

[14]Geographic information in HMDA data are published at the Census Tract level. We convert to zip level data using a tract to zip code crosswalk file produced using the Missouri Census Data Centers MABLE program, weighting by the share of the census tract population that resides in each zip code. We de-seasonalize and de-mean the zip-level loan application data by using the 12 month rolling average, and dividing by the average annual count of loan applications over the full 2005 to 2009 period.

[15]Based on a sample of mortgages from LPS of first-lien, fixed rate originations in 2005-06 matched with credit bureau records, Elul et al. (2010) find that for the 26% of borrowers with a second mortgage, combined LTV is approximately 15 percentage points higher than LTV.

include the proportion of the zip code population under the age of 30 and over the age of 60.

Finally, we arrive at the final sample which consists of 191,533 unique loans for a total of 5,067,258 loan-month observations by removing loan-month observations with missing unemployment data and local area delinquency rates calculated with less than 100 active loans.

Descriptive statistics are presented in table 1. The raw hazard rate into delinquency across all loan-month observations in our sample, that is the time-series and cross-sectional average of D_{it}, is 0.3%, and 6.8% of the loans in our sample become delinquent at some point during our period of observation.

Figure 1 presents baseline hazard rates, where loan-month observations are stratified into four groups based on current LTV ("High" if current LTV is greater than 100, and "Low" otherwise) and the area delinquency rate ("High" if $DQ_{00.05}$ is greater than or equal to 2.1, the 75^{th} percentile value of $DQ_{00.05}$, and "Low" otherwise). The figure shows that the probability of becoming delinquent in a given month conditional on not being delinquent up to the prior month is substantially larger for the two high current LTV groups compared to the low current LTV groups. However, the baseline hazard rate of high current LTV loans is larger for those in high $DQ_{00.05}$ areas compared to low $DQ_{00.05}$ areas, suggesting that $DQ_{00.05}$ may affect foreclosure decisions. Of course, an alternative explanation is that this difference in hazard rates is driven by other factors correlated with the $DQ_{00.05}$. We control for such factors in our econometric specification described in the next section.

4 Model Specification

As indicated above, in order to investigate whether or not foreclosure contagion exists, we test whether the neighborhood delinquency rate affects an individual's probability of default, after controlling for a comprehensive set of variables. Specifically, in our most exhaustive specification, we estimate the probability of default as follows:

$$
\begin{aligned}
Pr(D_{it} = 1) = F[& DQ_{00.05,t-3}, \textit{loan and borrower characteristics at origination}, \\
& \textit{current loan characteristics, housing price changes,} \\
& \textit{demographic characteristics, local economic activity,} \\
& \textit{spatial correlations, county fixed effects, quarterly fixed effects}]
\end{aligned}
$$

where it indexes an individual loan i in month t. The dependent variable D_{it} is equal to 1 in the first month in which the loan enters 90+ day delinquency, and 0 otherwise.[16] Our variable of interest is $DQ_{00.05,t-3}$ which is the 3-month lagged local area delinquency rate. We use a 3-month lag to coincide with the time borrowers would initiate the decision to become 90+ day delinquent themselves.[17] After including all control variables, we interpret a positive coefficient on DQ as evidence in support of foreclosure contagion (see Boyson et al., 2010, for a similar approach). As previously mentioned, we define the local area as zip codes with centroids lying within a 5 mile radius from the centroid of the zip code where the

[16] In an alternative specification we use the first incidence of the loan entering foreclosure. The baseline estimates from this model – not shown – are qualitatively similar.

[17] In an alternative specification we use the local area delinquency rate with a one-month lag, with qualitatively similar results.

loan is located. We choose a 5 mile radius as it balances obtaining a precise local area with having too few surrounding zip codes. In this sense, our choice of 5 miles is educated by Mok et al. (2007), who find that prior to the use of the internet face-to-face and telephone contacts decrease substantially after a distance of 5 miles. Additionally, Mok et al. (2010) find that this distance remains an inflection point even after the popularization of internet usage.

A possible concern is the definition of what constitutes a default. We use the 90+ day delinquency as our measurement of mortgage default because it best reflects the homeowners *intent* to default. However, foreclosures are public events and are therefore more informative to other homeowners than delinquencies. We therefore re-estimate all our models using an actual foreclosure as our definition of default as opposed to 90+ delinquencies and find qualitatively similar results.

We discuss our control variables in detail below.

4.1 Control variables

4.1.1 Loan and borrower characteristics at origination

We start by controlling for loan and borrower characteristics at origination that are known default risk factors. For loan characteristics we include dummy variables that account for whether the loan has a fixed or variable rate, whether the loan is a hybrid Adjustable Rate Mortgage (ARM) with an initial fixed period of less than 5 years, whether the loan has low or no documentation, whether the loan is interest only, whether the loan had a negative amortization option, whether the loan has a prepayment penalty, whether the loan is a jumbo loan, and whether the loan has an original term of more than 30 years. In addition, we include a dummy variable for whether the loan is categorized as subprime by the servicer ("B" or "C" type loans). We also include a dummy variable for whether or not the loan-to-value (LTV) ratio at origination was exactly equal to 80.[18] Foote et al. (2008) note that these loans are likely to have been accompanied by a second lien. Thus the combined LTV on these loans is likely higher than what we are able to compute from LPS data. For borrower characteristics, we include dummy variables to indicate whether their credit score at origination was less than 640 (low), between 640 and 719 (middle), or 720 or more (high), and whether the Debt to Income (DTI) ratio was between 28% inclusive and 36%, 36% inclusive to 42%, and equal to or higher than 42%. If $FICO$ or DTI information at origination is missing we account for it in separate dummy variables.

4.1.2 Current loan characteristics

We also include the current loan characteristics that enter into borrowers decision to strategically default. These include dummy variables for whether the current LTV (LTV_t) is between 90% and 100%, whether LTV_t is between 100% and 120% and LTV_t is greater than 120%. As described above in section 3, LTV_t is calculated by taking the loan balance in a given month and updating the historical property value (purchase price) using the Case-Shiller zip-code level index. Since property value estimates are imprecise in this construction because the zip code level index measures the average change in value, we opt for broad categories of LTV_t. We also include the interest rate on the loan which determines the value of the prepayment option. Moreover, we also include a cubic function in age to control for the age of the loan.

[18]Nearly 25 percent of loans in our data had an LTV of 80 percent at origination.

4.1.3 Local house price changes

As noted earlier, Campbell et al. (2009) and Harding et al. (2009) find that foreclosure has a negative impact on surrounding home prices. We refer to this foreclosure externality as *price* contagion. In this paper we focus on the *non-price* contagion effect of foreclosures, i.e. the extent to which mortgage defaults are responsible for *directly* causing nearby homeowners to default, above and beyond the price contagion channel. Therefore, we control for home price changes by including the Case-Shiller zip code level house price changes over the prior 2-years. This variable also reflects local macro-economic conditions and is a possible proxy for borrower expectations of future house prices (see Nadauld and Sherlund, 2009).

4.1.4 Local economic activity

Local economic activity is an important determinant of mortgage default because it reflects average differences across areas in homeowners' ability to repay [Deng et al. (2000)]. As such, foreclosures may appear to be spatially correlated only because economic factors common to all borrowers are causing them to default. Therefore, we use several measures of economic performance as controls in addition to the zip-code level house price index changes mentioned above. Specifically, we include the monthly county-level unemployment rate, monthly county-level unemployment duration, monthly county-level credit card delinquency rates, monthly county-level housing permit applications, and finally monthly zip code level loan applications. Also, as is elaborated in the robustness section as an alternative specification we also use county *times* quarterly fixed effects that absorb all quarterly time-varying county level changes in economic activity.

4.1.5 Demographic characteristics

Our controls include the racial and age zip code demographic variables. Our racial characteristics include the proportion of minorities and the proportion of Hispanics at the zip code level. Guiso et al. (2009) in a survey find that if faced with a $100,000 equity shortfall in their homes "Hispanics are much more likely to default than black or white." Also, we include the proportion of the zip code population aged 60 and over and the proportion aged 30 or under. Guiso et al. (2009) find that homeowners aged 65 or above are more likely to strategically default if faced with a $100,000 equity shortfall in their homes.

4.1.6 Spatial correlations

One concern is that the probability of default on individual mortgages in a zip code is correlated with the area delinquency rate because the loan and borrower characteristics in the local area are similar to those in the target zip code. We control explicitly for this possibility by including as control variables local area loan and borrower characteristics. We include, the average $FICO$ at origination, the average LTV at origination, share of loans that have a fixed rate, the shares of loans that have full documentation, and the share of loans that are interest only, for borrowers located in zip codes lying within a 0 to 5 mile radius from the target zip code (omitting the target zip code). The calculation of these local area loan and borrower characteristics mimics the calculation of the area local delinquency rates.

4.1.7 Fixed Effects

Finally, we include county fixed-effects to control for persistent differences in observable and unobservable factors across counties, and quarterly fixed-effects to control for changes in the national macro-economic conditions. Notably, county fixed-effects control for differences in State laws such as provision that allow for "deficiency judgments" to be pronounced. As mentioned above, as a robustness we also include county interacted with quarter fixed-effects, a much stronger requirement. We also include zip-code fixed-effects as a robustness which requires non-varying zip-code level variables (such as demographic characteristics) to be withdrawn from the regression specification.

4.2 Estimation

We estimate the probability of default model using a dynamic logistic framework, allowing the associated hazard function to vary non-parametrically by including a cubic spline in age. However, for robustness we also estimate the model using the Cox (1972) proportional hazard framework and the competing risk model preferred by Deng et al. (2000). The competing risk model accounts for the possibility that a loan may have gone into default but that outcome was not observed because the loan was refinanced beforehand. Econometrically, the dynamic logistic framework is closely related to the Cox (1972) proportional hazard model (see, for example, Sueyoshi, 1995; Shumway, 2001, for a discussion) and overall all three estimation methods provide qualitatively very similar results. Therefore, we focus our attention on the dynamic logistic framework which is computationally less onerous than the other two estimation methodologies. Robust standard errors are clustered at the zip code level.

4.3 Omitted variable bias

In theory, if the area delinquency rate remains statistically significant even after the inclusion of this extensive array of controls, we would have evidence in favor of contagion. There is, however, an important challenge in interpreting the coefficient of the area delinquency rate as an indication of contagion. The challenge lies on the list of control variables included. Because there is no a-priori and agreed upon list of "controls", there is always a danger that, despite the inclusion of a large set of controls, any remaining explanatory power of the area delinquency rate variable could be attributable to an omitted variable. Thus, if the area delinquency rate remains significant after the set of controls is included, it still possible that the set is "insufficient" and that, in reality, the area delinquency rate could simply be capturing the effect of the omitted variable.

Given this concern, the only reliable way of confidently ruling out the possibility of an omitted variable bias is to add sufficient controls to render the variable of interest (area foreclosure rate) statistically insignificant in the full sample regression. The fact that the area delinquency rate completely loses its explanatory power when this set of controls is included clearly suggests that there is no longer an omitted variable bias.

Once we establish that our specification is indeed "complete" (i.e. we have enough controls), we then proceed to estimate this specification for the sub-sample of borrowers who are most likely to be strategic defaulters–those who are underwater (i.e. their LTV is above 120%), but that do not appear to be financially distress (i.e. have FICO scores above 720). If the area delinquency rate remains statistically significant for this subset of borrowers, we would be confident in concluding that there is evidence of contagion for this group.

4.4 Manski (1993) reflection problem

Before discussing the empirical results, it is important to address another possible concern that arises in models of social interaction such as ours. The concern is the "reflection problem" of Manski (1993), which arises when a researcher tries to infer whether the average behavior in some group influences the behavior of the members of that same group. Since the behavior appears on both the right and left hand side of the regression equation, the resulting specifications are unidentified. Our methodology specifically addresses this issue in several important ways.

First, as pointed out in Brock and Darlauf (2001) the reflection problem applies to *linear models* of social interactions whereas the logistic models that we use do not suffer from identification issues. Second, Sirakaya (2006) demonstrates that the Cox proportional hazard model which we also use as an alternative specification to corroborate our results is also identified in this context. Third, as Brock and Darlauf (2007) show the panel data specifications we use allow for peer-effects to be identified regardless of whether the model is linear or not. Fourth, Lee (2007) shows that variation in group sizes (in our case the number of borrowers in the 0-5 mile radius) alleviates the reflection problem. Fifth, our explanatory variable $DQ_{00.05}$ is a cumulative aggregate, a stock variable, whereas the left hand side variable, D_{it}, is the corresponding flow variable. Lastly, as a final precaution, the area delinquency rate we calculate as a peer-effect measure excludes the zip code where the individual borrowers are located.

5 Results

5.1 Baseline

We first estimate a baseline model, reported in table 2, which estimate the probability of default on an individual loan as a function of the nearby area delinquency rate ($DQ_{00.05}$) and control variables using the dynamic logistic model. In our baseline estimation we start with the general set of borrowers regardless of whether these borrowers are underwater on their mortgages.

We also estimate all baseline specifications using the Cox (1972) proportional hazard rate model which provides similar results. Furthermore, we estimate all baseline specifications using the competing risk model advocated by Deng et al. (2000), which accounts for the possibility that a loan may terminate due to prepayment prior to a default being observed. These results are also qualitatively similar to the ones being reported.[19] Therefore, our findings are robust to alternative estimation methodologies.

Model 1 in table 2 indicates that $DQ_{00.05}$ is an important factor driving the probability of mortgage default after controlling for loan characteristics at origination, loan characteristics at the time of the decision to default, local area demographic characteristics, as well as quarter and county fixed effects. The coefficient of 1.0441 is statistically significant at the 1% level with robust standard errors clustered at the zip code level. However, as indicated above, it is possible that its significance is simply capturing the effect of an omitted variable. The results in model 2 suggest that this may actually be the case. After including controls for recent house price changes and other local economic conditions (unemployment rate

[19]Since the results for the Cox (1972) proportional hazard and competing risk models are qualitatively identical to the dynamic logistic regression we do not report them.

and duration, credit card delinquencies, housing permit and loan applications) the $DQ_{00.05}$ coefficient, while still statistically significant at the 1% level, drops by nearly 45% (from 1.0441 to 0.581). Evidently, inclusion of these controls captures a great deal of the variation in the local delinquency rate.

While the set of additional controls included in model 2 help to control for an omitted variable, they may not be "sufficient" as it is still possible that $DQ_{00.05}$ is capturing the effect of local common factors, not included in our set of local economic conditions. In model 3, we include additional controls to capture the effect of local area loan and borrower characteristics. That is we may observe the area delinquency rate to be an important determinant of the probability of default in an individual loan only in so far as the area delinquency rate proxies for spatial similarities in borrowers. We specifically address this issue by including the average borrower characteristics on mortgage loans in the area that corresponds to $DQ_{00.05}$. These variables include the area's average credit score at origination, average LTV at origination, the share of loans that have a fixed interest rate mortgage, the share of loans that have full documentation, and the share of loans that have an interest only option. The inclusion of these controls capture more cleanly the effect of the average "quality" of borrowers in a given area, which may influence the probability of default.

The results in model 3 suggest that, after including this last set of controls, the $DQ_{00.05}$ coefficient is no longer statistically distinguishable from 0 at any conventional level of significance. Moreover, the $DQ_{00.05}$ drops substantially from 0.5810 to 0.0349. We gauge the economic significance of $DQ_{00.05}$ in all three models by reporting in table 3 the marginal effect and the elasticity of $DQ_{00.05}$. Focusing on the elasticity a 1% change in $DQ_{00.05}$ results in an economically small and statistically insignificant 0.16% change in the probability of default. We confirm that $DQ_{00.05}$ is not significant by estimating model 4 which does not include $DQ_{00.05}$ to see if the other parameters remain similar. We find that the pseudo-R^2 on model 4 is the same as in model 3 (15.10%) and all coefficient estimates are nearly identical.

Hence, we can conclude that after including an extensive array of controls the probability of default on an individual loan is not affected by the area delinquency rate. However, as our previous discussion emphasized peer-effects are important only in so far as they "trigger" the decision to strategically default when that decision is beneficial to homeowners. Nonetheless, the results in table 2 identify the proper set of controls in our regressions. At face value, they indicate that, in the overall population of borrowers, there is no evidence of contagion, once a large set of controls is included.

5.2 Strategic defaulters

As Guiso et al. (2011) point out it is difficult to identify strategic defaulters as these defaulters have all the "incentives to disguise themselves as people who cannot afford to pay." However, we can identify a set of conditions that increase the likelihood of a strategic default.

First, we concentrate on the subset of borrowers with high LTV_t ratios because a strategic default makes sense *only* for underwater borrowers. However, while loans with LTV_t above 100% are technically underwater, existing evidence suggests that in practice LTV must be higher before a borrower exercises his option to strategically default. Guiso et al. (2009) find, using survey data, that strategic foreclosures are unlikely to take place for properties that are less than 10% under water. Bhutta et al. (2010) find that for subprime borrowers in Arizona, California, Florida, and Nevada, the median strategic defaulter has an LTV_t of 162% and 90% of strategic defaulters have an LTV_t of 120% or more. Based on this evidence, we choose 120% as the threshold for "high" LTV_t over which a borrower might choose to strategically default.[20] We thus create a dichotomous variable equal to 1 if LTV_t is 120% or more, 0 otherwise.

Second, being significantly underwater alone does not necessarily ensure that we are capturing borrowers who have the choice to default. We need to further identify borrowers that are less likely to experience a cash flow shock or, if they do, are less likely to be liquidity constrained after suffering such a shock. We use the borrower's $FICO$ score at origination of the loan as a proxy for a borrower's financial constraints, classifying borrowers with a score of 720 or higher into the "high" credit score bucket. Using credit score to identify borrowers that are more likely to be strategic defaulters is consistent with evidence from a 2011 study by Experian-Wyman finding that strategic defaults as a share of overall defaults are roughly three times higher among prime borrowers compared to non-prime borrowers.

Table 4 presents descriptive statistics of D_{it} and $DQ_{00.05}$ within cells defined by LTV_t and $FICO$. As expected, the default rates are higher within the "high" LTV_t cells and higher within the "low" $FICO$ cells. However, for a given LTV_t stratum the stock of loans in delinquency, $DQ_{00.05}$, are comparable across $FICO$ levels. Specifically, for the "high" ("low") LTV_t stratum, $DQ_{00.05}$ are 0.1462 and 0.1436 (0.0387 and 0.0405).

To investigate whether there is evidence of contagion among potential strategic defaulters, we examine whether the *impact* of the area delinquency rate variable, $DQ_{00.05}$, varies by LTV_t and $FICO$ categories. Specifically, we add interaction terms of $DQ_{00.05}$ with indicators for "high" LTV_t and "high" $FICO$ to the empirical specification. Our focus is on the "high" LTV_t *and* "high" $FICO$ group. After controlling for the "full" set of controls included in

[20]Our results (not shown) are similar when alternatively defining "high LTV_t" as greater than 130%.

model 3 of table 2, one can reasonably make the claim that default is more of a choice for this group of borrowers. Therefore, we expect our contagion measure, $DQ_{00.05}$, to be important for this group. In contrast, we expect default contagion to be much weaker, or even statistically insignificant for other groups, in particular, those with "low" LTV_t ratios and "low" $FICO$ scores.

Table 5 presents the results of the $DQ_{00.05}$ interacted with a dummy variable equal to 1 if FICO is high (a score of 720 or above) and a dummy variable equal to 1 if the LTV_t is greater than 120%. Also, as is standard, we include the variables themselves as well as the two-way interaction terms in the regressions. Mimicking the baseline results we report three specifications and focus on model 3 as it is the most comprehensive. In table 5, model 3 shows that all the interaction terms of interest are statistically significant at the 1% level of significance.

To ease the interpretation, table 6, panel A, presents the estimated nearby area delinquency rate marginal effects, broken down by LTV_t (high or not) and $FICO$ scores (high or not) based on the dynamic logistic model (model 3) discussed above. The table shows that the marginal effect of $DQ_{00.05}$ for the high $FICO$ *and* high LTV_t bucket is positive and statistically significant at the 5% level. This is precisely the group where we expect peer-effects or contagion to take place. Comparing the marginal effects of the High LTV_t and High $FICO$ group with the Low LTV_t and Low $FICO$ (the group with the lowest likelihood of being strategic defaulters) the difference in the coefficients is highly significant with a p-value on the χ^2 test of 1.19%. Thus, the evidence indicates the existence of contagion among those who are underwater *and* have the choice to default. This evidence supports the findings in Guiso et al. (2011) who, using survey data, find results consistent with an information spillover hypothesis (contagion) and inconsistent with a clustering hypothesis (common economic conditions driving the correlations).

The estimated marginal effect of 0.0112 for the High LTV_t and High $FICO$ bucket reported in panel A, translates to an elasticity of 0.0725. [21] This figure implies that a 1% increase in the local area delinquency rate increases the probability of a strategic default by 7.25%, a figure of considerable economic magnitude.

Given that (a) we control for current and past loan and borrower characteristics, a large set of local economic variables, local demographic characteristics, local area loan and borrower characteristics, as well as geographical and time fixed-effects; and that (b), we do not find the area delinquency rate to be a significant variable in the general population of borrowers, it is unlikely these results are driven by an omitted variable bias. Nonetheless, we further investigate the robustness of our findings in the next subsection.

5.3 Robustness

As mentioned above, our chief concern is that local economic conditions that lead homeowners to default are not properly measured or account for thereby biasing our results in favor of a spillover hypothesis as opposed to a local economic shock to the area (clustering) hypothesis. We therefore conduct several robustness checks to rule out this possibility. First, we estimate an additional model which includes county fixed effects *interacted* with quarter fixed effects. This model absorbs all local (county) and time (quarter) variation and it therefore represents our most complete model.[22] Table 6, Panel B, presents the marginal

[21]To arrive at this figure we multiply the marginal elasticity times the ratio of the mean area foreclosure rate to the mean delinquency rate for the the High LTV_t and High $FICO$ subsample.

[22]We are forced to remove zip code level demographic information as it is nearly co-linear with the interacted fixed effects.

effects of the area delinquency rate $DQ_{00.05}$ for each of our LTV_t and $FICO$ buckets. The results appear even stronger than before with the "high" LTV_t *and* "high" $FICO$ marginal effect being positive and statistically significant at the 1% level.

The implied elasticity for the "high" LTV_t, "high" $FICO$ bucket under this new specification is 0.165, which is even larger than the 0.0725 estimated with the previous model.

As an additional robustness check we re-estimate the same model using zip code fixed effects and removing non-varying zip code level variables. The results (not reported) are qualitatively similar with an elasticity in the same order of magnitude as reported in table 6. Finally, we re-estimate our models using up to 6-month lags of local economic activity variables with no effect on the statistical significance $DQ_{00.05}$.

6 Conclusion

In a large sample of U.S. mortgages observed between 2005 and 2009 we document the presence of contagion effects. We define contagion as the direct impact that nearby defaults (as measured by 90+ day delinquencies or foreclosures) have on the conditional probability of a mortgage default after controlling for all other known determinants, including house price appreciation, which Harding et al. (2009) and Campbell et al. (2009) highlight as being important drivers of default.

Thus, our contagion hypothesis is that borrowers are influenced by other borrowers who are in default controlling for all other economic mechanisms. This hypothesis can only hold for strategic defaulters: borrowers who stand to benefit from exercising their default option resulting from the value of the mortgage exceeding the property market value, but maintain the ability to repay their monthly mortgage payments. As many economists have pointed out, these borrowers stand to benefit from defaulting but do not "ruthlessly" do so [e.g. Deng et al. (2000), Vandell (1995)]. Instead, a set of "triggers" appear to be necessary that either force or influence homeowners in these circumstances to default. While there is an established literature examining "hard triggers" such as divorce or job loss, there are few studies examining the role of "soft triggers" such as peer-effects where homeowners are being influenced by other defaulters.[23]

Consistent with the contagion hypothesis, we find that the area delinquency rate surrounding a mortgage is an important determinant of default for potential strategic defaulters. Moreover, we find that it is not for borrowers who are not at risk of strategically defaulting. Specifically, we find that holding everything else constant, a 1% increase in the delinquency rate of surrounding zip codes (within a radius of 5 miles), increases the likelihood of an individual strategic mortgage default by at least 7.25%. These results are not only statistically significant but economically important.

Our results are robust to a battery of controls including housing price changes, county unemployment rate, state unemployment duration, county housing permit applications, county credit card delinquency rates, and the number of loan applications in the given zip code. Additionally, we use alternative specifications that include county and quarter fixed effects, county interacted with quarter fixed effects and zip code fixed effects. We use the dynamic logistic regression as our estimation procedure of choice but we also estimate a Cox (1972) proportional hazard model and a competing risk model with similar results.

This study contributes to the literature in several dimensions. First, the results empiri-

[23] A notable exception is Guiso et al. (2011) who use evidence from surveys to establish the presence of information spillovers and social stigma in influencing strategic defaulters.

cally support the findings in Guiso et al. (2011) that the decision to strategically default is not taken in a vacuum and itself creates significant spillovers to other households. Second, our study furthers the notion of strategic default "triggers" to include the information and perceptions that other borrowers possess as additional "soft triggers." Third, evidence that household decisions have community effects has the potential to explain other household-level financial decision-making which has an impact on consumption, debt and investment allocation. Moreover, the presence of contagion effects provides insights for home loan modification programs. Indeed, these programs need to view home loans in a community as a group. Finally, in addition to our findings on contagion effects, our baseline results provide further evidence on the determinants of mortgage default. In particular, we find that spatial correlations are important determinants of default in the general population of borrowers.

We also discuss three non-mutually exclusive mechanisms by which contagion in strategic mortgage defaults can take place: (a) information spillovers, (b) reputational spillovers, and (c) disruptions in social networks. The commonality among all three mechanisms is that defaults either provide underwater homeowners informational updates as to the consequences of strategically defaulting or influence their *willingness* to stay current. Disentangling the relative importance of each mechanism is a difficult, yet interesting task, which we leave this to future research.

References

Bekaert, G., C. R. Harvey, and A. Ng (2005). Market integration and contagion. *The Journal of Business 78*(1), 39–69.

Bhutta, N., J. Dokko, and H. Shan (2010). The depth of negative equity and mortgage default decisions. *Finance and Economics Discussion Series, Federal Reserve Board, No. 2010-35*.

Blume, L. E. (2010). Stigma and social control. *Cornell University Working Paper*.

Boyson, N., C. Stahel, and R. Stulz (2010). Hedge fund contagion and liquidity shocks. *Journal of Finance 65*, 1789–1816.

Brevoort, K. P. and C. R. Cooper (2010). Foreclosure's wake: The credit experiences of individuals following foreclosure. *Federal Reserve Board Working Paper*.

Brock, W. A. and S. N. Darlauf (2001). Discrete choice with social interactions. *Review of Economic Studies 68*, 235–260.

Brock, W. A. and S. N. Darlauf (2007). Identification of binary choice models with social interactions. *Journal of Econometrics 140*, 52–77.

Brunnermeier, M. K. (2009). Deciphering the liquidity and credit crunch 2007-2008. *Journal of Economic Perspectives 23*, 77–100.

Campbell, J. Y., S. Giglio, and P. Pathak (2009). Forced sales and house prices. *NBER Working Paper No. 14866*.

Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society, Series B 34*, 187–220.

Das, S. (2012). The principal principle. *Journal of Finance and Quantitative Analysis Forthcoming.*

Das, S., D. Duffie, N. Kapadia, and L. Saita (2007). Common failings: How corporate defaults are correlated. *Journal of Finance 62*, 93–118.

Deng, Y., J. M. Quigley, and R. Van Order (2000). Mortgage terminations, heterogeneity, and the exercise of mortgage options. *Econometrica 68*, 275–307.

Dunn, K. B. and J. J. McConnell (1981). The valuation of mortgage-backed securities. *Journal of Finance 36*, 599–617.

Ellen, I., J. Lacoe, and C. Sharygin (2012). Do foreclosures cause crime? *Journal of Urban Economics 74*, 59–70.

Elul, R., N. S. Souleles, S. Chomsisengphet, D. Glennon, and R. Hunt. (2010). What triggers mortgage default? *Working Paper 10-13, Research Department, Federal Reserve Bank of Philadelphia.*.

Foote, C. L., K. Gerardi, L. Goette, and P. S. Willen (2009). Reducing foreclosures: No easy answers. *Federal Reserve Bank of Atlanta Working Paper 2009-15*.

Foote, C. L., K. Gerardi, and P. S. Willen (2008). Negative equity and foreclosure: Theory and evidence. *Journal of Urban Economics 64*, 234–245.

Foster, C. and R. Van Order (1984). An option-based model of mortgage default. *Housing Finance Review 3*, 351–372.

Gerardi, K., A. H. Shapiro, and P. S. Willen (2007). Subprime outcomes: Risky mortgages, homeownership experiences and foreclosures. *Federal Reserve Bank of Boston Working Paper*.

Gorton, G. B. (2009). Information, liquidity, and the (ongoing) panic of 2007. *American Economic Review 99*, 567–572.

Guiso, L., P. Sapienza, and L. Zingales (2009). Moral and social constraints to strategic default on mortgages. *NBER Working Paper No. 15145*.

Guiso, L., P. Sapienza, and L. Zingales (2011). The determinants of attitudes towards strategic default on mortgages. *Journal of Finance*.

Harding, J. P., E. Rosenblatt, and V. W. Yao (2009). The contagion effect of foreclosed properties. *Journal of Urban Economics 66*, 164–178.

Immergluck, D. (2008). The accumulation of foreclosed properties: Trajectories of metropolitan reo inventories during the 2007-2008 mortgage crisis. *Community Affairs Discussion Paper No. 02-08, Federal Reserve Bank of Atlanta.*.

Immergluck, D. and G. Smith (2006). the external costs of foreclosure: The impact of single-family mortgage foreclosures on property values. *Housing Policy Debate 17*, 57–79.

Jorion, P. and G. Zhang (2007). Good and bad credit contagion: Evidence from credit default swaps. *Journal of Financial Economics 84*, 860–883.

Lee, L.-F. (2007). Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. *Journal of Econometrics 84*, 333–374.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies 60*, 531–542.

Mok, D., B. Wellman, and R. Basu (2007). Did distance matter: A pre-internet analysis. *Social Networks 29(3)*, 430–62.

Mok, D., B. Wellman, and J. A. Carrasco (2010). Does distance still matter in connected lives? a pre- and post-internet comparison. *Urban Studies 47(3)*, 2747–2784.

Nadauld, T. and S. M. Sherlund (2009). The role of the securitization process in the expansion of subprime credit. *SSRN eLibrary*.

Shumway, T. (2001). Forecasting bankruptcy more accurately: A simple hazard model. *The Journal of Business 74*(1), 101–124.

Sirakaya, S. (2006). Recidivism and social interactions. *Journal of the American Statistical Association 101*, 863–877.

Sueyoshi, G. T. (1995). A class of binary response models for grouped duration data. *Journal of Applied Econometrics 10*, 411–431.

Vandell, K. (1995). How ruthless is mortgage default? a review and synthesis of the evidence. *Journal of Housing Research 6*, 245–264.

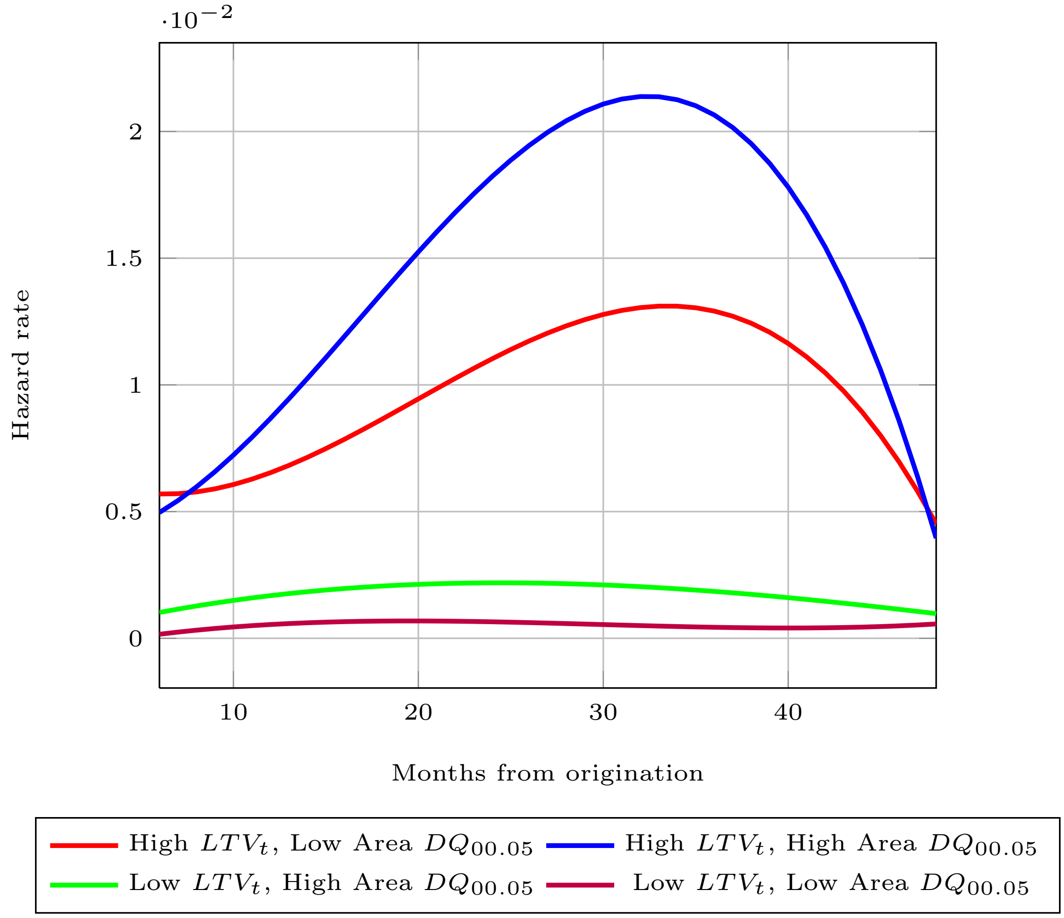

Figure 1: Baseline hazard rates into 90-day delinquency by current loan-to-value (LTV_t) and area delinquency rate ($DQ_{00.05}$).

A loan is categorized as "High LTV_t" in month t if LTV_t is greater than or equal to 120, and categorized as "Low LTV_t" otherwise. A loan is categorized as "High $DQ_{00.05}$" if the delinquency rate within 0 to 5 miles is greater than or equal to 2.6 (the value of $DQ_{00.05}$ at the 75^{th} percentile), and categorized as "Low $DQ_{00.05}$" otherwise. Each line represents the predicted value from a regression of the raw hazard rate on a cubic function of loan age. Regressions are estimated separately for each LTV_t-$DQ_{00.05}$ group.

Table 1: Descriptive Statistics

This table presents the sample descriptive statistics. The sample obtained from LPS (McDash) consists of 5,067,258 loan-month observations over the period 2005-2009. The area 90-day delinquency rate is calculated for each zip code-month pairs as the weighted average 90-day delinquency rate in zip codes lying within a radius of 5 miles (using the entire LPS database). Local area loan and borrower characteristics are calculated in the same manner. Current LTV (LTV_t) is computed from LPS data and Case-Shiller zip code level price indices. Time variant variables are reported as of the last observation for each loan in the sample.

Panel A: Time invariant variables				
Variable	Mean	Std. Dev.	Min	Max
Characteristics at origination				
Fixed rate	0.631	0.483	0	1
Hybrid $<$ 5 ARM	0.096	0.294	0	1
Hybrid $<$ 5 ARM missing	0.135	0.342	0	1
Low or no doc.	0.277	0.447	0	1
Low or no doc. missing	0.405	0.491	0	1
Low FICO score	0.117	0.322	0	1
Middle FICO score	0.313	0.464	0	1
FICO score missing	0.138	0.345	0	1
DTI $\in$ [28,36)	0.135	0.341	0	1
DTI $\in$ [36, 42)	0.145	0.352	0	1
DTI $\geq$ 42	0.222	0.416	0	1
DTI missing	0.356	0.479	0	1
$LTV_{t=0} = 80$	0.282	0.450	0	1
Interest only option (I/O)	0.147	0.354	0	1
Interest only option missing	0.008	0.091	0	1
Neg. Am option	0.046	0.211	0	1
Prepayment penalty	0.132	0.338	0	1
Type "B" or "C" loan	0.079	0.269	0	1
Jumbo loan	0.164	0.370	0	1
Original term $>$ 30 years	0.030	0.170	0	1
ZIP demographics				
Minority population share (ZIP)	24.4	11.9	3.1	72.0
Hispanic population share (ZIP)	20.7	15.2	0.9	77.3
Age $\leq$ 30 population share (ZIP)	41.5	3.4	23.8	51.6
Age $\geq$ 60 population share (ZIP)	16.5	3.2	9.0	43.2

Panel B: Time variant variables				
Variable	Mean	Std. Dev.	Min	Max
Area 90-day delinquency rate (3 month lag)				
DQ rate (0-5 miles)	0.176	0.381	0	1
Current loan characteristics				
Interest rate	0.062	0.010	0.010	0.146
$LTV_t \in (90, 100)$	0.124	0.329	0	1
$LTV_t \in [100, 120)$	0.153	0.360	0	1
$LTV_t \geq 120$	0.131	0.338	0	1
Local economic conditions				
Recent house price change (ZIP)	-9.3	29.1	-85.9	112.9
Unemployment rate (County)	7.9	3.2	2.4	31.3
Mean unemployment duration (State)	13.7	5.4	3.0	26.0
Credit card DQ rate (County)	2.5	0.8	0.8	5.6
Loan applications (ZIP)	-24.0	52.6	-92.6	181.7
House permit applications (County)	-23.4	40.0	-99.6	201.8
Local area loan and borrower characteristics (0-5 miles)				
Average FICO	711.8	18.9	622.0	781.4
Average $LTV_{t=0}$	71.8	6.9	42.0	90.9
Share of loans that are fixed rate	73.2	13.0	16.9	98.5
Share of loans that are full doc.	65.3	9.9	17.1	94.5
Share of loans that have an I/O option	10.5	4.9	0.9	59.3
No. of loans: 191,533				
No. of loan-months: 5,067,258				

Table 2: Logistic regression of area 90-day delinquency rate on the probability of mortgage default

This table reports estimates of a logistic model using a panel of 5,067,258 loan-month observations over the period 2005-2009. The dependent variable is 1 when a loan enters 90-day delinquency (default) and 0 otherwise. The key variable of interest is the area 90-day delinquency rate which is calculated for each zip code-month pairs as the weighted average 90-day delinquency rate in zip codes lying within a radius of 5 miles. Model 1 controls for loan and borrower characteristics at origination, zip code demographic characteristics and current loan characteristics where current LTV (LTV_t) is computed from LPS data and Case-Shiller zip code level price indices. Model 2, augments the model to include housing prices and measures of local economic activity. Model 3, includes local area loan and borrower characteristics which is calculated for each zip code-month pairs as the weighted average of each characteristic in zip codes lying within a radius of 5 miles. Model 4, omits the key variable of interest from model 3. All regressions include county and quarter fixed effects as well as a cubic function of loan age. Robust standard errors, clustered at the zip code level, are placed below the coefficients in parentheses. ***, **, and * denote significance at the 1%, 5% and 10% levels, respectively.

Continued on next page

	Model							
	(1)		(2)		(3)		(4)	
DQ rate (0-5 miles)	1.0441	***	0.5810	***	0.0349			
	(0.1382)		(0.1374)		(0.1567)			
Fixed rate	-0.4449	***	-0.4472	***	-0.4524	***	-0.4523	***
	(0.0187)		(0.0187)		(0.0187)		(0.0186)	
Hybrid < 5 ARM	0.2366	***	0.2367	***	0.2301	***	0.2301	***
	(0.0281)		(0.0281)		(0.0281)		(0.0281)	
Hybrid < 5 ARM missing	-0.1014	***	-0.1049	***	-0.1039	***	-0.1039	***
	(0.0269)		(0.0269)		(0.0269)		(0.0269)	
Low or no doc	0.3063	***	0.3050	***	0.3070	***	0.3070	***
	(0.0158)		(0.0159)		(0.0158)		(0.0158)	
Low or no doc. missing	-0.0883	***	-0.0897	***	-0.0865	***	-0.0865	***
	(0.0206)		(0.0205)		(0.0205)		(0.0205)	
Low FICO score	1.2619	***	1.2652	***	1.2537	***	1.2537	***
	(0.0239)		(0.0240)		(0.0239)		(0.0238)	
Middle FICO score	0.7740	***	0.7743	***	0.7693	***	0.7693	***
	(0.0163)		(0.0163)		(0.0162)		(0.0162)	
FICO score missing	0.7107	***	0.7129	***	0.7103	***	0.7104	***
	(0.0241)		(0.0241)		(0.0241)		(0.0241)	
DTI missing	0.5101	***	0.5100	***	0.5065	***	0.5065	***
	(0.0263)		(0.0263)		(0.0263)		(0.0263)	
DTI $\in$ [28,36)	0.3291	***	0.3303	***	0.3281	***	0.3280	***
	(0.0253)		(0.0252)		(0.0252)		(0.0252)	
DTI $\in$ [36, 42)	0.3789	***	0.3804	***	0.3785	***	0.3785	***
	(0.0252)		(0.0252)		(0.0252)		(0.0252)	
DTI $\geq$ 42	0.4823	***	0.4822	***	0.4793	***	0.4793	***
	(0.0237)		(0.0237)		(0.0237)		(0.0237)	
$LTV_{t=0} = 80$	0.2068	***	0.2050	***	0.2042	***	0.2042	***
	(0.0136)		(0.0136)		(0.0136)		(0.0136)	
Interest only option (I/O)	0.0585	***	0.0586	***	0.0585	***	0.0585	***
	(0.0182)		(0.0182)		(0.0181)		(0.0181)	
Interest only option missing	0.2339	***	0.2321	***	0.2316	***	0.2316	***
	(0.0850)		(0.0850)		(0.0855)		(0.0855)	
Neg. Am option	0.1244	***	0.1275	***	0.1272	***	0.1271	***
	(0.0333)		(0.0332)		(0.0332)		(0.0332)	
Prepayment penalty	0.0766	***	0.0829	***	0.0825	***	0.0826	***
	(0.0216)		(0.0216)		(0.0216)		(0.0216)	
Type "B" or "C" loan	0.2540	***	0.2519	***	0.2497	***	0.2497	***
	(0.0259)		(0.0260)		(0.0259)		(0.0259)	
Jumbo loan	0.0469	**	0.0563	***	0.0781	***	0.0779	***
	(0.0210)		(0.0209)		(0.0211)		(0.0211)	
Original term > 30 years	0.1400	***	0.1412	***	0.1424	***	0.1423	***
	(0.0268)		(0.0269)		(0.0269)		(0.0269)	
Minority population share (ZIP)	-0.0091		0.0137		0.0004		-0.0002	
	(0.0520)		(0.0505)		(0.0498)		(0.0494)	
Hispanic population share (ZIP)	0.3448	***	0.1605	***	0.1675	***	0.1675	***
	(0.0408)		(0.0481)		(0.0477)		(0.0477)	
Age $\leq$ 30 population share (ZIP)	0.0100		0.1241	***	0.1156	***	0.1159	***
	(0.0489)		(0.0495)		(0.0492)		(0.0492)	
Age $\geq$ 60 population share (ZIP)	-0.0590		0.0059		-0.0044		-0.0044	
	(0.0603)		(0.0603)		(0.0598)		(0.0598)	
Interest rate	37.2008	***	37.2124	***	36.9357	***	36.9355	***
	(0.7683)		(0.7654)		(0.7652)		(0.7651)	
$LTV_t \in (90, 100)$	0.6677	***	0.6272	***	0.6146	***	0.6144	***
	(0.0200)		(0.0199)		(0.0200)		(0.0200)	
$LTV_t \in [100, 120)$	0.9704	***	0.9048	***	0.8857	***	0.8856	***
	(0.0210)		(0.0214)		(0.0214)		(0.0214)	
$LTV_t \geq 120$	1.4548	***	1.3373	***	1.3116	***	1.3119	***
	(0.0262)		(0.0274)		(0.0274)		(0.0275)	
Recent house price change (ZIP)			-0.0095	***	-0.0093	***	-0.0093	***
			(0.0008)		(0.0008)		(0.0008)	
Unemployment rate (County)			-0.0080		-0.0025		-0.0024	
			(0.0094)		(0.0093)		(0.0093)	
Mean unemployment duration (State)			-0.0008		0.0000		0.0000	
			(0.0037)		(0.0037)		(0.0037)	
Credit card DQ rate (County)			-0.0221		0.0157		0.0169	
			(0.0356)		(0.0356)		(0.0351)	
House permit applications (County)			0.0015	***	0.0015	***	0.0015	***
			(0.0004)		(0.0004)		(0.0004)	
Loan applications (ZIP)			-0.0008		-0.0010	*	-0.0010	*
			(0.0006)		(0.0006)		(0.0006)	
Average FICO (0-5 miles)					-0.0047	***	-0.0047	***
					(0.0007)		(0.0007)	
Average $LTV_{t=0}$ (0-5 miles)					0.0090	***	0.0090	***
					(0.0025)		(0.0025)	
Share of loans that are fixed rate (0-5 miles)					0.0015			
					(0.0013)			
Share of loans that are full doc. (0-5 miles)					0.0008			
					(0.0011)			
Share of loans that have an I/O option (0-5 miles)					0.0032		0.0032	
					(0.0020)		(0.0020)	
Quarterly FE	Y		Y		Y		Y	
County FE	Y		Y		Y		Y	
N	5067258		5067093		5067093		5067093	
Pseudo R^2	0.1499		0.1505		0.1510		0.1510	

Table 3: Estimated effects of area default rate on the probability of mortgage default

This table reports the marginal effects and elasticities of the local area delinquency rate on the probability of mortgage default for the logistic model regressions reported in table 2. Robust standard errors, clustered at the zip code level, are placed below the coefficients in parentheses. ***, **, and * denote significance at the 1%, 5% and 10% levels, respectively.

	Model						
	(1)		(2)		(3)		(4)
Marginal effect of area default rate	0.0067	***	0.0037	***	0.0002		–
(dy/dx)	(0.0018)		(0.0013)		(0.0010)		–
Elasticity of area default rate w/ respect to default	0.0474	***	0.0264	***	0.0016		–
(ey/ex)	(0.0062)		(0.0062)		(0.0071)		–
Characteristics at origination	Y		Y		Y		Y
ZIP demographics	Y		Y		Y		Y
Current loan characteristics	Y		Y		Y		Y
Local economic conditions			Y		Y		Y
Local area loan/borrower characteristics					Y		Y
Quarterly FE	Y		Y		Y		Y
County FE	Y		Y		Y		Y
N	5067258		5067093		5067093		5067093
Pseudo R^2	0.1499		0.1505		0.1510		0.1510

Table 4: Default rates by current LTV and $FICO$ categories

This table presents averages of loan level delinquency probability D_{it} and local area delinquency rates $DQ_{00.05}$ for strata defined by current LTV and $FICO$. The row LTV-High contains all loans with LTV_t above 120%, and the column $FICO$-High contains all loans with $FICO$ scores of 720 or above.

		$FICO$-High	$FICO$-Low
LTV_t-High	N	120129	185593
	Default rate D	0.0227	0.0377
	Local area DQ rate (0-5 miles)	0.1462	0.1436
LTV_t-Low	N	2278431	2482940
	Default rate D	0.0019	0.0079
	Local area DQ rate (0-5 miles)	0.0387	0.0405

Table 5: Logistic Regression of area 90-day delinquency rate interacted with high $FICO$ and high current LTV on the probability of mortgage default

This table reports estimates of a logistic model using a panel of 5,067,258 loan-month observations over the period 2005-2009. The dependent variable is 1 when a loan enters 90-day delinquency (default) and 0 otherwise. The variables of interest are: the area 90-day delinquency rate which is calculated for each zip code-month pairs as the weighted average 90-day delinquency rate in zip codes lying within a radius of 5 miles, high $FICO$ (a dummy variable equal to 1 if the borrower has a $FICO$ score of 720 or above, 0 otherwise), high current LTV (a dummy variable equal to 1 if current LTV (LTV_t) is higher than 120%, 0 otherwise), and the 2-way and 3-way interaction terms of these variables. Model 1 controls for loan and borrower characteristics at origination, zip code demographic characteristics and current loan characteristics where LTV_t is computed from LPS data and Case-Shiller zip code level price indices. Model 2, augments the model to include housing prices and measures of local economic activity. Model 3, includes local area loan and borrower characteristics which is calculated for each zip code-month pairs as the weighted average of each characteristic in zip codes lying within a radius of 5 miles. All regressions include county and quarter fixed effects as well as a cubic function of loan age. Robust standard errors, clustered at the zip code level, are placed below the coefficients in parentheses. ***, **, and * denote significance at the 1%, 5% and 10% levels, respectively.

Continued on next page

	Model					
	(1)		(2)		(3)	
DQ rate (0-5 miles)	0.6091	***	-0.0117		-0.6993	***
	-0.1643		-0.1703		-0.1868	
High FICO Score	-1.0716	***	-1.0849	***	-1.087	***
	-0.0296		-0.0296		-0.0299	
Current LTV 120 or More	1.3017	***	1.1492	***	1.0993	***
	-0.0381		-0.0397		-0.0403	
(High FICO) x ($LTV_t \geq 120$)	0.7513	***	0.7654	***	0.7566	***
	-0.0502		-0.0505		-0.0513	
(High FICO) x (DQ rate)	2.4749	***	2.5874	***	2.6805	***
	-0.2242		-0.2259		-0.2315	
($LTV_t \geq 120$) x (DQ rate)	0.1319		0.3572		0.5862	***
	-0.214		-0.2202		-0.226	
(High FICO) x ($LTV_t \geq 120$) x (DQ Rate)	-1.9711	***	-2.0223	***	-2.0585	***
	-0.3221		-0.3243		-0.3322	
Fixed rate	-0.4409	***	-0.4432	***	-0.4479	***
	-0.0184		-0.0185		-0.0184	
Hybrid $<$ 5 ARM	0.2334	***	0.2331	***	0.2269	***
	-0.0277		-0.0277		-0.0276	
Hybrid $<$ 5 ARM missing	-0.0949	***	-0.0989	***	-0.0976	***
	-0.0266		-0.0265		-0.0266	
Low or no doc	0.2981	***	0.296	***	0.2977	***
	-0.0157		-0.0157		-0.0156	
Low or no doc. missing	-0.0819	***	-0.0835	***	-0.0807	***
	-0.0203		-0.0202		-0.0202	
Low FICO score	0.5601	***	0.5623	***	0.5534	***
	-0.026		-0.026		-0.0259	
Middle FICO score	0.0778	***	0.0763	***	0.0738	***
	-0.0221		-0.022		-0.022	
FICO score missing	(omitted)		(omitted)		(omitted)	
DTI missing	0.4978	***	0.4978	***	0.4947	***
	-0.0262		-0.0261		-0.0261	
DTI $\in$ [28,36)	0.3261	***	0.3276	***	0.3255	***
	-0.0252		-0.0251		-0.0251	
DTI $\in$ [36, 42)	0.371	***	0.3726	***	0.3708	***
	-0.0252		-0.0251		-0.0251	
DTI $\geq$ 42	0.4751	***	0.4753	***	0.4725	***
	-0.0236		-0.0236		-0.0236	
$LTV_{t=0} = 80$	0.2013	***	0.1989	***	0.1979	***
	-0.0135		-0.0134		-0.0134	
Interest only option (I/O)	0.0531	***	0.0536	***	0.0534	***
	-0.0178		-0.0178		-0.0178	
Interest only option missing	0.23	***	0.2268	***	0.2264	***
	-0.0839		-0.0839		-0.0844	
Neg. Am option	0.1171	***	0.1202	***	0.12	***
	-0.0328		-0.0326		-0.0326	
Prepayment penalty	0.0806	***	0.0863	***	0.0858	***
	-0.0212		-0.0212		-0.0213	
Type "B" or "C" loan	0.2503	***	0.2489	***	0.2467	***
	-0.0254		-0.0254		-0.0253	
Jumbo loan	0.0501	***	0.0605	***	0.0812	***
	-0.0207		-0.0206		-0.0208	
Original term $>$ 30 years	0.1494	***	0.1504	***	0.1514	***
	-0.026		-0.0261		-0.0261	
Minority population share (ZIP)	-0.0045		0.019		0.0068	
	-0.0509		-0.0499		-0.0495	
Hispanic population share (ZIP)	0.3411	***	0.1574	***	0.1651	***
	-0.0402		-0.0476		-0.0472	
Age $\leq$ 30 population share (ZIP)	0.0024		0.1208	***	0.1109	**
	-0.0486		-0.0492		-0.0489	
Age $\geq$ 60 population share (ZIP)	-0.0676		-0.0004		-0.0094	
	-0.0597		-0.0598		-0.0595	
Interest rate	36.9897	***	36.976	***	36.7088	***
	-0.7581		-0.7551		-0.7547	
$LTV_t \in (90, 100)$	0.6475	***	0.6046	***	0.5929	***
	-0.0199		-0.0199		-0.0199	
$LTV_t \in [100, 120)$	0.9464	***	0.8779	***	0.8613	***
	-0.021		-0.0214		-0.0214	
Recent house price change (ZIP)			0.0014	***	0.0014	***
			-0.0004		-0.0004	
Unemployment rate (County)			-0.0099	***	-0.0098	***
			-0.0008		-0.0008	
Mean unemployment duration (State)			-0.0115		-0.0063	
			-0.0094		-0.0093	
Credit card DQ rate (County)			-0.0005		0.0004	
			-0.0037		-0.0037	
House permit applications (County)			-0.0426		-0.0048	
			-0.0354		-0.0353	
Loan applications (ZIP)			-0.0013	**	-0.0015	***
			-0.0006		-0.0006	
Average FICO (0-5 miles)					-0.4784	***
					-0.0691	
Average $LTV_{t=0}$ (0-5 miles)					0.835	***
					-0.2444	
Share of loans that are fixed rate (0-5 miles)					0.0013	
					-0.0013	
Share of loans that are full doc. (0-5 miles)					0.0007	
					-0.001	
Share of loans that have an I/O option (0-5 miles)					0.0031	
					-0.002	
Quarterly FE	Y		Y		Y	
County FE	Y		Y		Y	
N	5067093		5067093		5067093	
Pseudo R^2	0.1515		0.1522		0.1527	

Table 6: Estimated effects of area default rate on the probability of mortgage default categorized by high current LTV Ratio and high $FICO$ score

This table presents the marginal effects of the local area delinquency rate on the probability of mortgage default for each category of high current LTV (LTV_t above 120%) and high $FICO$ score (scores of 720 or above). The LTV_t-High, $FICO$-high category represents borrowers with the highest risk of strategically defaulting. Panel A, reports marginal effects that are computed from model 3 of table 5. Panel B, reports marginal effects from a similar model that includes Quarter interacted with County fixed effects. P-values of Chi-square tests for the difference between categories is provided below each panel. Robust standard errors, clustered at the zip code level, are placed below the coefficients in parentheses. ***, **, and * denote significance at the 1%, 5% and 10% levels, respectively.

Panel A: Quarterly and county fixed effects				
	FICO-High		FICO-Low	
LTV_t-High	0.0111959	**	-0.0039449	
	(0.0050888)		(0.007174)	
LTV_t-Low	0.0037817	***	-0.0053536	***
	(0.001019)		(0.0020433)	
Characteristics at origination				Y
ZIP demographics				Y
Current loan characteristics				Y
Local economic conditions				Y
Local area loan/borrower characteristics				Y
Quarterly FE				Y
County FE				Y
Chi-square test (High/High vs Low/Low)			P-values	
LTV_t-High/FICO-High vs. LTV_t-Low/FICO-Low:			0.0119	**
LTV_t-High/FICO-High vs. LTV_t-High/FICO-Low:			0.0273	**
LTV_t-High/FICO-High vs. LTV_t-Low/FICO-high:			0.1351	

Panel B: Quarterly interacted with county fixed effects				
	FICO-High		FICO-Low	
LTV_t-High	0.0255	***	0.0214	***
	(0.0052)		(0.0076)	
LTV_t-Low	0.0049	***	-0.0016	
	(0.0005)		(0.0015)	
Characteristics at origination				Y
ZIP demographics				N
Current loan characteristics				Y
Local economic conditions				Y
Local area loan/borrower characteristics				Y
County x Quarter FE				Y
Chi-square test (High/High vs Low/Low)			P-values	
LTV_t-High/FICO-High vs. LTV_t-Low/FICO-Low::			0.0000	***
LTV_t-High/FICO-High vs. LTV_t-High/FICO-Low:			0.5581	
LTV_t-High/FICO-High vs. LTV_t-Low/FICO-high:			0.0000	***