



## Thy Neighbor's Mortgage: Does Living in a Subprime Neighborhood Affect One's Probability of Default?

Sumit Agarwal\*, Brent W. Ambrose\*\*, Souphala Chomsisengphet\*\*\* and Anthony B. Sanders\*\*\*\*

This article focuses on the potential externalities associated with subprime mortgage origination activity. Specifically, we examine whether negative spillover effects from subprime mortgage originations result in higher default rates in the surrounding area. Our empirical analysis controls for loan characteristics, house price changes and alternative loan products. Our results indicate that, after controlling for these characteristics, the concentration of subprime lending in a neighborhood does not lead to greater default risks for surrounding borrowers. However, we do find that more aggressive mortgage products (such as hybrid adjustable rate mortgages and low/no-documentation loans) had significant negative spillovers on other borrowers. Stated differently, the aggressive alternative mortgage designs were more toxic to the housing and mortgage market than previously believed.

During the previous decade, the U.S. housing market experienced two interrelated events. First, the United States experienced a housing market bubble that began in the early 2000s, flattened in 2006 and finally burst in the latter half of 2007.<sup>1</sup> Second, during this same period, the use of alternative (or hybrid) mortgage products escalated.<sup>2</sup> These products, such as pay-option adjustable-rate mortgages (ARMs) and low-documentation and ALT-A mortgages were designed to help borrowers acquire housing in markets experiencing significant

\*Federal Reserve Bank of Chicago, 230 South LaSalle Street, Chicago, IL 60604 or [sagarwal@frbchi.org](mailto:sagarwal@frbchi.org).

\*\*Smeal College of Business, Pennsylvania State University, University Park, PA 16802 or [bwa10@psu.edu](mailto:bwa10@psu.edu).

\*\*\*Office of the Comptroller of the Currency, 250 E Street SW Washington, DC 20219 or [souphala.chomsisengphet@occ.treas.gov](mailto:souphala.chomsisengphet@occ.treas.gov).

\*\*\*\*George Mason University, 233 Enterprise Hall, Fairfax, VA 22030 or [asander7@gmu.edu](mailto:asander7@gmu.edu).

<sup>1</sup>See for example, Glaeser, Gyourk and Saks (2005) and Leamer (2007) for discussions of house price bubbles.

<sup>2</sup>See for example, Mian and Sufi (2008) and Leamer (2007) for a discussion of the role of credit expansion and the mortgage default crisis.

price appreciation. However, these mortgages were often marketed to borrowers with relatively poor credit histories as well. As a result, these mortgages became known as subprime mortgages because the borrowers for these products did not meet the underwriting criteria of the housing government sponsored enterprises (GSEs). Instead, they were securitized in the private label securitization market.

Because these mortgages were designed to provide borrowers with payment affordability during a period of rapidly rising housing prices, the most common subprime mortgages had adjustable-rate features and many had provisions for negative amortization of principal, providing borrowers with low initial payments. The general belief was that rapidly rising home values would allow borrowers to refinance prior to the impact of the negative amortization feature. Of course, many did not foresee the softening of the U.S. housing market eliminating the ability to refinance. Thus, the default rate on subprime mortgages has increased dramatically, and current estimates indicate that rising subprime defaults may add over 500,000 homes to the housing supply.<sup>3</sup> We also know that subprime mortgages appear to concentrate themselves in neighborhoods rather than being evenly spaced throughout urban areas.

The question we address is whether subprime mortgages cluster together and, if so, did their performance decay cause other defaults in the same neighborhood. In addition, we examine whether the more aggressive mortgage products (such as hybrid ARMs and low/no-documentation loans) have significant negative spillovers on other borrowers. That is, are subprime mortgages or aggressive alternative mortgage products the culprit in clusters of defaults?

The article is organized as follows. We discuss housing prices, subprime concentration and mortgage defaults in the next section. The theoretical setup is presented in the third section and the data are presented in the fourth section. The empirical method is presented in the fifth section and the results are presented in the sixth section. Our conclusions are presented in the seventh section.

## **Housing Prices, Subprime Concentration and Default**

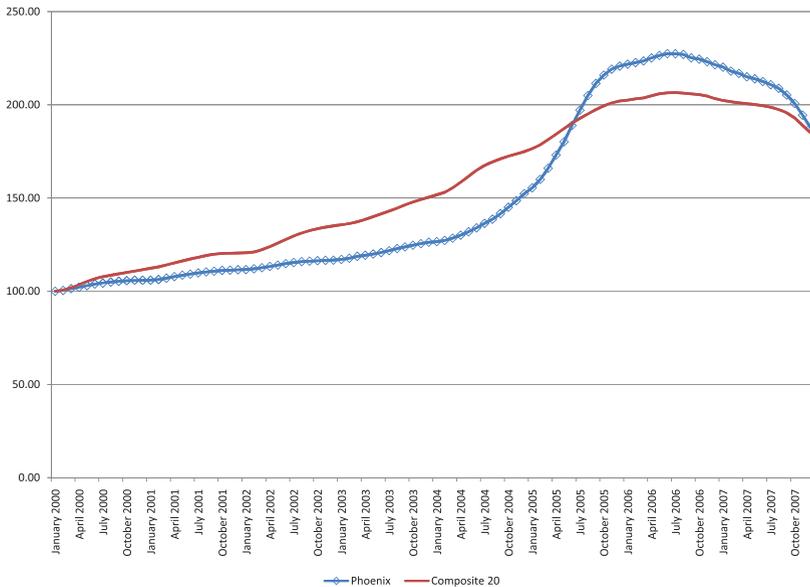
In the academic literature, the linkage between property value and mortgage default is well understood.<sup>4</sup> Holding all else constant, the boundary conditions

---

<sup>3</sup>See B. Louis, "Rising Subprime Mortgage Defaults Add to Unsold Homes Inventory," Bloomberg.com (accessed at <http://www.bloomberg.com/apps/news?pid=20601087&refer=home&sid=aC9LdDev4.Wc>).

<sup>4</sup>See for example, Kau, Keenan and Kim (1994), Foster and Van Order (1984), Jackson and Kasserian (1980), Ambrose and Buttimer (2000) and, more recently, Foote, Gerardi and Willen (2008) and Gerardi, Shapiro and Willen (2008), among many others.

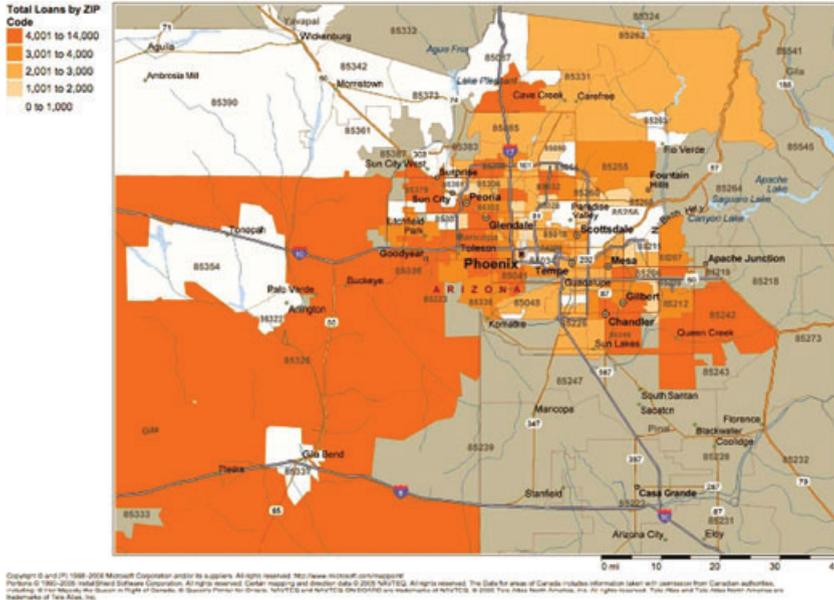
**Figure 1 ■ SP/Case-Shiller home price indices (January 2000 to January 2009 year-over-year price change).**



in option pricing models capture the idea that borrowers will rationally default when the value of the mortgaged property falls below the value of the mortgage contract. Thus, it should not be surprising that we are witnessing a wave of borrower defaults as the value of the underlying collateral declines.<sup>5</sup> The severity of housing price declines in the “sand states” of California, Arizona, Nevada and Florida has, in fact, been a primary driver of the abnormally high delinquency and foreclosure rates.

However, one interesting feature of these alternative mortgage designs, particularly subprime mortgages, that has to date not been examined is that they tend to be clustered in metropolitan areas that also experienced significant house price increases. In other words, subprime mortgages are not evenly distributed across the country. For example, Maricopa County, Arizona Phoenix, Scottsdale, Mesa and surrounding communities had one of the most explosive rates of house price growth during the 2004–2006 time period (Figure 1). Over this period, the Case-Shiller house price index grew from an index value of 100.00

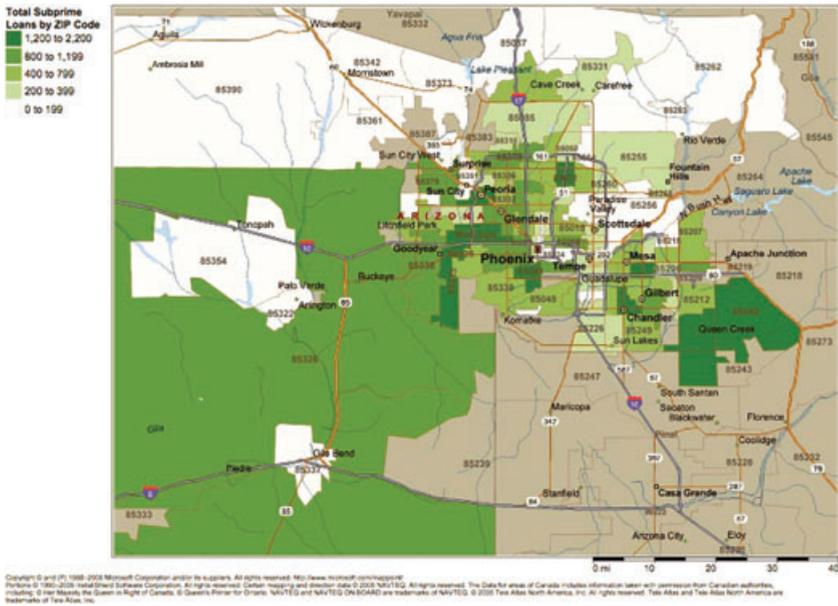
<sup>5</sup>Danis and Pennington-Cross (2008) also point out that subprime delinquency and defaults are highly correlated with loans to borrowers in markets with higher asset price volatility.

**Figure 2** ■ Total number of loans originated between 2000 and 2007 by ZIP code.

in January 2000 to a peak of 227.42 in June 2006, indicating that house values more than doubled in 6 years. At the same time, the Phoenix metropolitan area experienced an explosion in alternative mortgage origination activity. Hence, Maricopa County represents an excellent laboratory for studying the relationship between house price growth and the mortgage products used to finance home purchases.

To demonstrate the extent of subprime concentration, Figures 2 and 3 show the total mortgage origination activity and subprime origination activity for the Phoenix metropolitan area by ZIP code between 2000 and 2007. The figures clearly indicate a spatial pattern of mortgage activity. However, to gain a better perspective on subprime clustering, Figure 4 shows the concentration of subprime mortgages by ZIP code. Not surprisingly, the highest concentration of subprime activity (as a percent of all loan originations) occurs in the urban inner city as opposed to the urban-rural periphery. In fact, between 2004 and 2006, the areas with the highest volume of subprime loan origination were in new-build locations (Southeast, West and North) but, as a percentage of all loans, the lower-income neighborhoods of Phoenix (downtown, older homes along the interstates going West and North from the downtown) had the highest concentration of subprime activity. Interestingly, interest-only (IO) ARMs are

**Figure 3** ■ Total number of subprime loans originated between 2000 and 2007 by ZIP code.

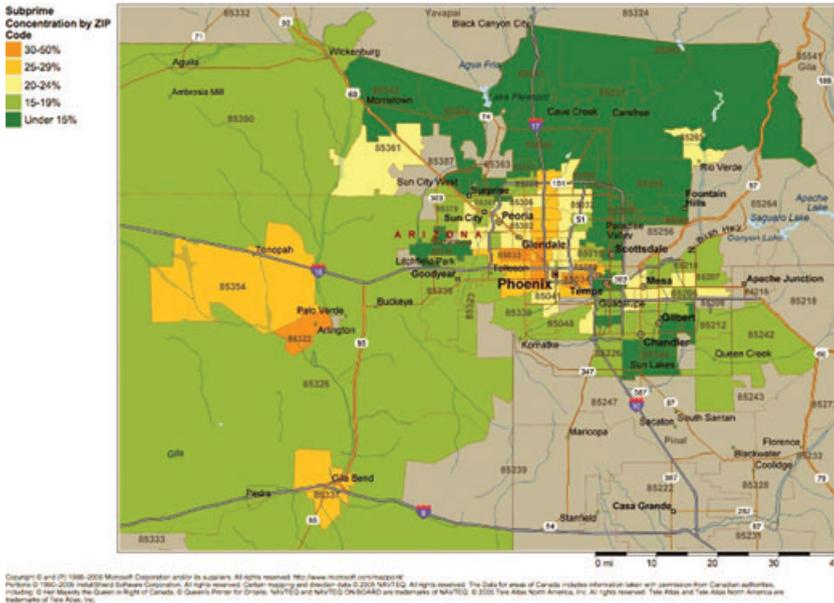


located in the highest-price areas of Maricopa County (Scottsdale, Paradise Valley and Ahwatukee), but far less than in the high-subprime-concentration ZIP codes.

If subprime lending is correlated with poor underwriting standards, then the clustering of subprime mortgages may cause a spillover effect in terms of default. A number of studies have documented that a common outcome of default (foreclosure) is a negative spillover onto the value of surrounding properties and neighborhoods.<sup>6</sup> For example, recent studies indicate that the impact of foreclosures on surrounding property prices ranges from 8.7% (Lin, Rosenblatt and Yao 2009) to approximately 1% (Immergluck and Smith 2006, Campbell, Giglio and Pathak 2009). In addition, Schuetz, Been and Ellen (2008) in a study of foreclosures in New York City document that proximity to a foreclosure does result in a price discount.

<sup>6</sup>See Lauria and Baxter (1999) for evidence on the impact on communities, Moreno (1995) for direct cost estimates of foreclosures on cities and neighborhoods, Immergluck and Smith (2006), Lin, Rosenblatt and Yao (2009), Clauterie and Daneshvary (2009) and Lee (2008) for evidence of the effect on property prices. Frame (2010) provides an excellent survey of the extant evidence regarding the effect of foreclosures on house prices.

Figure 4 ■ Subprime mortgage concentration by ZIP code.

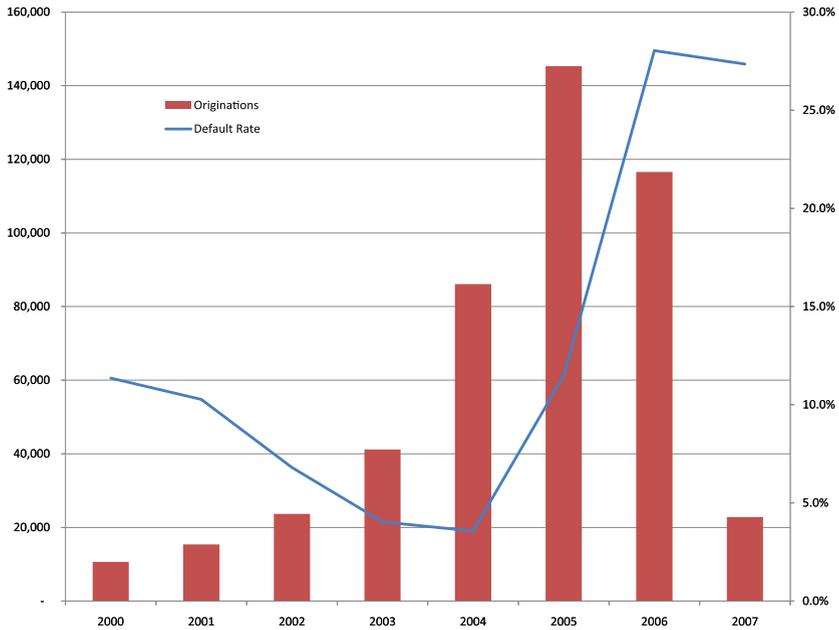


But do the spillover effects from subprime defaults imply that borrowers in neighborhoods (or ZIP codes) that are clustered together have a higher probability of default? That is, once we control for loan characteristics, house price changes and alternative loan products, do borrowers in neighborhoods with higher concentrations of subprime borrowers have a greater likelihood of default? That is the question we explore in this article.

### Theoretical Setup

The theoretical setup for a spillover effect causing a default cascade is straight forward and is similar to the simple model of observational learning presented in Bikhchandani, Hirshleifer and Welch (1998). First, we assume that homeowners follow the wealth-maximizing decisions underlying modern mortgage option pricing models. That is, we assume that borrowers only default when the value of the underlying property is less than the present value of the mortgage debt. Second, we assume that homeowners observe noisy private signals about the value of their property. The noisy signal comes in two forms: high ( $H$ ) or low ( $L$ ). A high signal implies that the property market may be appreciating and the homeowner updates her property valuation accordingly. Examples of high signals include frequent sales in the neighborhood, short sale times on the

**Figure 5 ■** Subprime mortgage origination volume and default rates for Phoenix, AZ (January 2000 to December 2007).



market, favorable news reports about the neighborhood, *etc.* Conversely, a low signal implies that the property market may be depreciating. Examples of low signals include longer observed time on the market, more property for sale with fewer actual sales, foreclosure sales, evidence that houses are being abandoned, news reports about crime in the neighborhood, *etc.* As noisy low signals are observed, the homeowner updates her property valuation estimate downward. As the frequency of noisy low signals increase, the lower the homeowner's estimate of property value becomes. Assuming the homeowner rationally applies the default boundary condition prior to each mortgage payment due date, the perceived decline in property value may result in an optimal default situation.

The problem is that the individual homeowner's default decision depends upon her individual loan-to-value (LTV) ratio, which is private information. However, if she defaults and the lender sells the property at foreclosure, the foreclosure sale becomes a public signal of a declining property market. That is, the remaining homeowners must assume that property values have declined from the time that the homeowner originated her mortgage, otherwise she would not have defaulted.

Because mortgage default decisions convey signals to neighboring homeowners about the direction of changes in property values, one homeowner's decision to default may start a default cascade by causing the remaining homeowners to reevaluate their property values downward, perhaps to a level triggering an optimal default decision on their part. However, an initial default does not imply that a default cascade will occur. Recall that each homeowner evaluates the property value signal in light of the present value of his or her mortgage debt. Thus, a default cascade will most likely occur in neighborhoods where the majority of the homeowners have high LTV ratios.

To illustrate, consider a neighborhood with four households ( $a$ ,  $b$ ,  $c$  and  $d$ ) in a two-period model. In each period, the households receive a private signal regarding the property market.<sup>7</sup> All else being equal, we assume a low property value signal is sufficient to cause a borrower with a high LTV ratio to believe that she no longer has any positive equity in the house and thus default is optimal. Consistent with the lag between default and foreclosure, we assume that a borrower default is only observed by the other households at the following period when the house is sold at foreclosure.

In case 1, we consider the scenario where the neighborhood has only one risky (or subprime) borrower (household  $a$ )—represented by having a high LTV. We assume all the other homeowners have low LTV ratios. At  $t = 0$  each household observes a noisy private signal regarding the value of its house [ $a = L$ ,  $b = H$ ,  $c = L$ ,  $d = H$ ]. Because homeowner  $a$  with a high LTV ratio received a low signal, she evaluates her position and recognizes that she is in a negative equity situation. Thus, she defaults on her mortgages and the lender sells the house at foreclosure at  $t = 1$ . At  $t = 1$  the remaining households receive a second private signal and observe the consequence of  $a$ 's default at  $t = 0$ . Thus, the remaining homeowners correctly assume that  $a$  received a low signal at  $t = 0$ . Although the remaining households observe the low signal resulting from  $a$ 's default, none of the remaining homeowners default at  $t = 1$  because they have low LTV ratios and the payoff from defaulting is negative (even if their signals were [ $L$ ,  $L$ ]). Thus, a default cascade never materializes.

Now, consider a second neighborhood where all the homeowners have high (but not equal) LTV ratios such that ( $LTV_a > LTV_b > LTV_c > LTV_d$ ). Again, we assume that at each period the homeowners receive a private noisy signal of the change in property value. Again, at  $t = 0$  one of the four households ( $a$ ) receives a low signal and determines that default is optimal. At  $t = 1$  the remaining households receive a new private signal plus they observe the

---

<sup>7</sup>For ease of exposition, we assume an equal probability of a high or low signal.

outcome of  $a$ 's default. Thus, the remaining households now have three signals to consider: the initial signal from  $t = 0$ , the new signal and the observed default.

First, consider household  $b$  who received the following private signals:  $[H, H]$ . This household has two private signals indicating property values are appreciating and thus discounts the observed default signal. Thus,  $b$  does not default at  $t = 1$ .

Now consider household  $c$ , whose private signals were  $[H, L]$ . In this case, the two private signals should cancel out; however, the observed default causes  $c$  to place greater weight on the second signal and thus believe that property values are falling. Therefore,  $c$  defaults at  $t = 1$ .

Lastly,  $d$ 's private signals were  $[L, L]$ . Although the first  $L$  signal was insufficient to cause a default at  $t = 0$ , the combination of  $[L, L]$  plus the observed default reinforce the perception of falling property values, and thus  $d$  defaults. In this case, we observe a default cascade as the default at  $t = 1$  reinforces the  $L$  signals received by households  $c$  and  $d$  at  $t = 1$ .

Based on this simplistic example, we address the following research question: Do borrowers in neighborhoods with higher concentrations of risky mortgages (as a percentage of total mortgage origination volume) experience larger-than-average default rates?<sup>8</sup>

## Data

### *Mortgage Data*

In order to answer the question of whether neighborhood risk impacts individual borrower default probability, we collect data from the asset-backed securities (ABS) data series of the LoanPerformance Corporation (LPC), Incorporated. This data series contains a large set of loan-level information describing the characteristics of the subprime loans that were securitized in the private label market. LPC indicates that the data cover 61% of the outstanding subprime

---

<sup>8</sup>Previous empirical work in finance and economics has found evidence supporting the insights obtained from this simple information cascade type model. For example, in a classic experimental setting Anderson and Holt (1997) demonstrate how an information cascade forms, leading individuals to select against their private signal and follow the actions of others. In finance, Chang, Chaudhuri and Jayaratne (1997) find support for information cascades by demonstrating that a bank's decision to open a branch in a particular location depends upon the number of existing bank branches in that area.

market. We focus on the 461,729 mortgages contained in the LPC database that were originated from January 2000 through December 2007 in Maricopa County, Arizona.

The LPC data contain complete information on mortgage types. For example, LPC classifies mortgages as Subprime, Alt-A or Prime and identifies whether the loan was originated with full documentation (Full-Doc), partial or low documentation (Low-Doc) or no documentation (No-Doc) of borrower income and assets. In addition, LPC identifies whether the mortgage was a fixed-rate mortgage (FRM) or ARM. Furthermore, for ARMs, LPC notes whether the mortgage is a traditional ARM or a hybrid ARM. In terms of borrower characteristics, the LPC data indicate whether the mortgage was originated as a refinance and whether the borrower also cashed out equity at refinancing (cashout refinance). We also make use of information concerning the presence of prepayment penalties on the mortgage and whether the loan was originated for a condominium or to an investor.

Table 1 provides an overview of the characteristics of the securitized mortgages originated in Phoenix between 2000 and 2007. Consistent with the booming housing market over this period, we see the number of mortgage originations increases dramatically from 10,653 in 2000 to a peak of 145,333 in 2005. In the second section of Table 1 we note that, overall, 67.3% of loans were classified as Subprime by the originator and 32.4% were classified as Alt-A mortgages.

It is important to remember that Subprime and Alt-A are simply labels of convenience applied by the originating lender and that no standard definition exists. Thus, the third section of Table 1 shows the breakdown of mortgage type based on “hard” information describing the level of documentation required at origination, the type of origination (purchase or refinance), the presence of prepayment penalties, payment type (fixed or adjustable), type of property (investor, single-family or condominium) and whether the mortgage is a first-lien. One of the most important risk characteristics is the level of documentation provided by the borrower at origination. We see that in 2000 75% of borrowers provided full documentation. However, by 2006 only 41% of borrowers were providing full documentation of assets and income, while over 55% were providing only limited (or low) documentation and 3% were providing no documentation to support their mortgage application. As the subprime market grew over this period, the proportion of FRMs declined from over 50% of origination volume in 2000 to 34% in 2004 (and continued to stay in the 30% to low 40% range). While the market share of FRMs declined, the proportion of ARMs increased from 46% in 2000 to 65% in 2004. Traditional ARM market share declined from 46% in 2000 to 15% in 2007. Finally, Table 1 shows that mortgage refinance activity generally tracked changes in mortgage interest rates

Table 1 ■ Descriptive characteristics.

Variable	Origination Year Cohort									
	2000	2001	2002	2003	2004	2005	2006	2007		
# Mortgages	461,729	15,388	23,690	41,186	86,105	145,333	116,569	22,805		
Subprime versus Alt-A										
Subprime (0,1)	67.63%	79.74%	77.58%	71.00%	64.29%	65.14%	69.92%	53.00%		
Alt A (0,1)	32.37%	20.26%	22.42%	29.00%	35.71%	34.86%	30.08%	47.00%		
Loan Characteristics										
Full Documentation (0,1)	51.90%	70.67%	66.44%	58.70%	58.13%	50.41%	41.38%	41.66%		
Low Documentation (0,1)	45.01%	26.75%	30.27%	38.42%	38.83%	46.27%	55.47%	55.57%		
No Documentation (0,1)	3.09%	2.57%	3.29%	2.88%	3.04%	3.32%	3.16%	2.77%		
Refinance (0,1)	46.33%	56.01%	55.91%	52.94%	39.01%	43.76%	46.29%	58.68%		
Cashout Refinance (0,1)	37.92%	44.92%	41.38%	36.65%	29.17%	38.28%	40.57%	46.28%		
Prepayment Penalty (0,1)	63.39%	66.82%	65.50%	66.85%	65.16%	61.94%	62.94%	57.26%		
All Fixed-Rate Mortgage (0,1)	39.57%	52.57%	48.44%	46.23%	34.29%	35.43%	40.98%	42.21%		
All Adjustable-Rate Mortgage (0,1)	60.43%	47.43%	51.56%	53.77%	65.71%	64.57%	59.02%	57.79%		
Traditional ARM (0,1)	29.94%	47.18%	49.87%	45.32%	35.28%	27.07%	19.27%	15.07%		
Hybrid ARM (0,1)	46.01%	45.78%	48.28%	45.54%	49.49%	48.57%	43.20%	30.14%		
Condominium (0,1)	5.69%	4.43%	3.98%	4.31%	5.10%	5.80%	6.88%	6.99%		
Investor Occupancy (0,1)	13.12%	7.56%	10.32%	13.83%	15.12%	15.11%	10.71%	12.40%		
Lien > 1 (0,1)	19.85%	19.50%	17.29%	16.08%	18.85%	18.40%	25.39%	15.42%		

with a sharp decline in 2004 coinciding with an increase in interest rates during that year. Thus, the changing average loan characteristics between 2000 and 2007 clearly paint a picture of increasing penetration of higher-risk mortgage origination activity in Phoenix.

Because the LPC data cover primarily nonprime mortgages, we merge the LPC data with the Home Mortgage Disclosure Act (HMDA) database to determine the overall volume of mortgage origination activity in Maricopa County. Thus, using HMDA to determine the number of mortgages originated in ZIP code  $i$  at month  $t$ , we calculate concentration measures of outstanding loans by product type for each ZIP code and month. Furthermore, based on the loan-level payment performance behavior of these loans, we calculate average default rates for each of the 109 individual ZIP codes from January 2000 to December 2008. We define defaults as 90+ days past due in foreclosure, real estate owned or in bankruptcy and alive in the prior time period (current or 89 days or less delinquent). Table 2 shows the average monthly default rate and average concentration of loans by neighborhood (ZIP code) based on mortgage characteristics. To gain a better perspective, Figure 5 juxtaposes the average annualized default rate against the annual subprime origination activity. As we noted above, the number of subprime originations in Phoenix climbed from 10,653 in 2000 to a peak in 2005 of 145,333; while after 2006 loan origination activity fell dramatically, and by 2008 no new subprime mortgages were originated in Phoenix. Over this same interval, the Phoenix housing market experienced a significant increase in house values. For example, the Case-Shiller Index for Phoenix rises from a 100.00 in January 2000 to a peak of 227.42 in June 2006 and then declines (Figure 1). In fact, the index growth for Phoenix was far faster than the rest of the country (as measured by the Case-Shiller Composite Index of 20 cities during the 2004–2005 period).

Consistent with the option pricing view that mortgage default results from declines in house values relative to mortgage value, Table 2 shows the dramatic increase in the monthly average default rate starting in 2006. We note that during the period between 2000 and 2005 the average monthly default rate was less than 1%. However, as the house prices peaked and then started to decline in 2006 and 2007, the average monthly default rate skyrocketed to 2.34% and 2.28%, respectively.

Table 2 also reports the overall and yearly average ZIP code concentration by mortgage classification. For example, we see the rise and fall of subprime activity between 2000 and 2007, noting that the average concentration of subprime origination activity rose from 4.5% in 2000 to 12% in 2005 and then declined to 5.3% in 2007. However, the real growth in alternative mortgages occurred in the use of Alt-A and low/no-documentation mortgages. For example,

**Table 2 ■ Mean neighborhood characteristics by origination year cohort.**

	All	2000	2001	2002	2003	2004	2005	2006	2007
Default Rate (Monthly)	1.13%	0.95%	0.86%	0.57%	0.34%	0.30%	0.96%	2.34%	2.28%
ZIP Code Concentration Measures									
Percent of Subprime Loans	10.96	4.58	4.12	5.21	15.35	9.96	12.53	11.98	5.46
Percent of ALT A Loans	5.35	0.98	1.09	1.53	7.06	5.50	6.54	5.10	4.48
Percent of ARM Loans	10.04	2.60	2.49	3.39	12.75	10.16	12.32	10.08	5.76
Percent of Hybrid ARMs	4.78	1.36	1.27	1.71	5.94	5.07	6.09	4.50	1.88
Percent of Investor Occupancy	2.18	0.55	0.37	0.82	2.82	2.45	2.85	1.80	1.19
Percent of No-Doc loans	0.54	0.08	0.15	0.24	0.97	0.46	0.65	0.54	0.26
Percent of Low-Doc Loans	7.55	1.25	1.37	2.00	8.67	6.00	8.87	9.61	5.77
Percent of Cashout Refinance	5.75	2.35	2.27	2.68	6.68	4.35	6.99	6.65	4.22
Percent of Loans with Prepayment Penalty	10.29	3.62	3.50	4.39	14.70	10.06	11.78	10.73	5.53
Percent of Foreclosed Homes	0.16	0.07	0.11	0.24	0.30	0.12	0.10	0.18	0.32

between 2000 and 2005 the concentration of Alt-A mortgages increased about six times while the concentration of no-documentation loans increased by eight times. In addition, we also see the dramatic increase in the use of prepayment penalties between 2003 and 2006, with the percentage of loans containing a prepayment penalty more than tripling over the 2000 to 2002 period.

### *House Price Data*

The housing data set consists of only single-family houses that sold in Maricopa County, Arizona between January 1989 and September 2007. The data were acquired from Ion Data. We use these data to create a repeat-sales index by ZIP code. In order to be included in the repeat-sales index, the following criteria had to be met: (a) all sales must be between unrelated parties, (b) sales of new houses were excluded, (c) the period between sales should be at least 6 months, (d) the price of a house must be greater than \$5,000 and (e) appreciation or depreciation must be no more than up 80% or down 60% per year.<sup>9</sup>

The repeat-sales indices were created using a three-step process:

Step 1: Qualitative variables were formed based on the starting quarter/month and the ending quarter/month and frequency. The number of qualitative variables equals the number of observations in the index. For example, the monthly index starting January 2000 and ending April 2008 has 88 qualitative variables. Thus, if a house was sold in January of 2007, then the dummy variable for that month would be a 1, the previous sale month will get a value of  $-1$  and all others receive a value of 0.

Step 2: After assigning the dummy variable, we estimate a pooled weighted ordinary least squares regression (of all the observations), weighted by the gap between the current sale and previous sale.

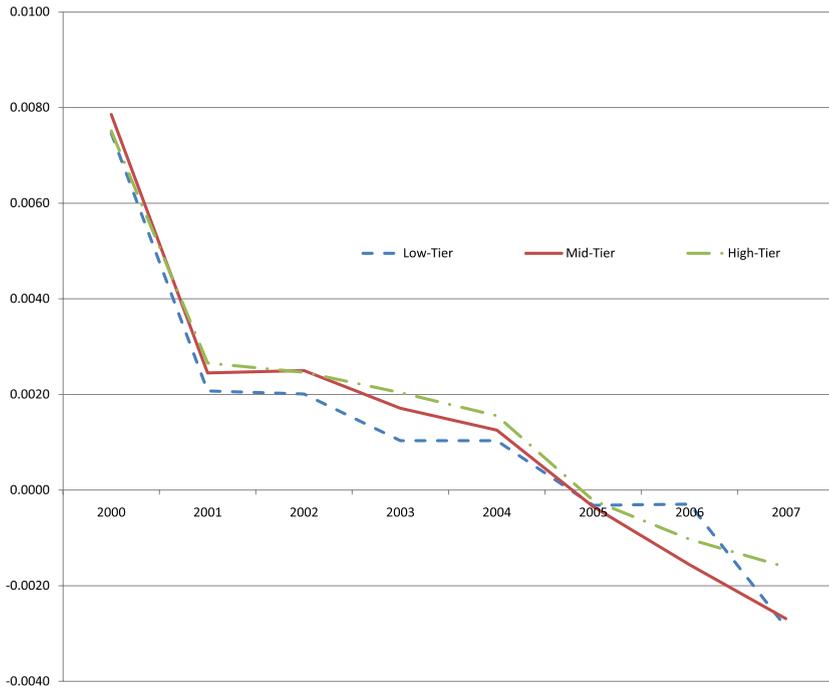
Step 3: The coefficients obtained from the regression are then based to 100 from the first period which gives the house price index (HPI).

Our repeat-sales indices are constructed following Case and Shiller (1987) in order to correct for heteroskedasticity found in the original repeat-sales indices. Within each quarter and for each ZIP code, we use our repeat-sales index to divide home sales into three groups: high, medium and low. We then select the average price within each bucket to represent higher, medium and lower price houses in that ZIP code.

---

<sup>9</sup>We used these data screens to remove obviously incorrectly coded observations.

**Figure 6 ■** Year-over-year house price index change for ZIP codes in the first quintile of subprime concentration for Phoenix, AZ (January 2000 to December 2007).



As noted above, by merging the LPC data with total origination activity reported in HMDA, we are able to calculate ZIP-code-level concentration measures of subprime activity. Figures 6 and 7 show the average yearly house price change for ZIP codes at the bottom and top of the subprime concentration. For example, in the ZIP codes with the highest concentration of subprime activity (Figure 7) we find that house price appreciation was greater in the lower priced housing market during the accelerating bubble years (2003 and 2004). In contrast, Figure 6 reveals that the lower-priced housing market in ZIP codes with the lowest concentration of subprime activity had the lowest level of house price appreciation. Thus, it appears that subprime origination activity is correlated with house price appreciation, suggesting that access to credit played a role in fueling the housing bubble in Phoenix.<sup>10</sup>

<sup>10</sup>This observation is consistent with the findings of Coleman, LaCour-Little and Vandell (2008) that subprime origination activity is correlated with house price changes during the peak years of the housing bubble (2004–2006).

**Figure 7** ■ Year-over-year house price index change for ZIP codes in the 5th quintile of subprime concentration for Phoenix, AZ (January 2000 to December 2007).



## Empirical Method

To test the default cascade hypothesis, we focus on individual mortgages to explore the impact of the concentration of subprime mortgages in a neighborhood on the probability that a specific mortgage will default. Following standard practice in the empirical mortgage performance literature, we estimate a proportional hazard model of borrower default. We begin by denoting  $T$  as the latent duration of each loan to default and  $\tau$  as the observed duration of the mortgage. Conditional on a set of explanatory variables,  $x$ , the probability density function (*pdf*) and cumulative density function *cdf* for  $T$  are

$$f(T|x; \theta) = h(T|x; \theta) \exp(-I(x; \theta)) \quad (1)$$

and

$$F(T|x; \theta) = 1 - \exp(-I(x; \theta)) \quad (2)$$

where  $I$  is given as:

$$I(T|x; \theta) = \int_0^T h(x; \theta) ds \quad (3)$$

and  $h$  is the hazard function. Thus, assuming that  $h(\tau|x; \theta) = \exp(x'\beta)$ , then the conditional probability of default is given as

$$Pr(\tau, x; \theta) = \frac{\exp(x'\beta)}{1 + \exp(x'\beta)} \quad (4)$$

and is estimated via maximum likelihood.<sup>11</sup>

Following Gross and Souleles (2002), we separate  $x$  into components reflecting various risk characteristics. These include individual borrower risk characteristics, loan characteristics, ZIP-code-level mortgage concentration measures, ZIP-code-level repeat-sales index house price changes and measures of nearby foreclosures. We measure individual borrower risk characteristics at origination as reflected by their FICO score and LTV ratio. We also include a variety of control variables that identify the type of loan originated (*i.e.*, low-documentation, no-documentation, adjustable-rate, hybrid, IO, *etc.*) and a set of ZIP-code-level concentration variables that capture the percentage of loans outstanding in the borrower's ZIP code at the time of origination to capture various high-risk characteristics (*i.e.*, the percentage of loans that are low-documentation, no-documentation, adjustable-rate, hybrid, IO, *etc.*). Thus, by examining these concentration variables, we are able to identify the impact that higher concentrations of risky loans have on the odds of borrower default.

## Results

This section presents the estimation results for the proportional hazard rate model for borrower default discussed above. Table 3 reports the estimated coefficients from the proportional hazard model.

Consistent with previous studies of borrower default, we find that borrower credit score at origination is inversely related to default risk. That is, higher FICO scores are correlated with lower probabilities of default. We also see that higher LTV ratios are associated with higher risk of defaults.

Turning to the impact of mortgage type, we find that subprime mortgages are 1.3 times more likely to default, all else being equal, than prime mortgages.

---

<sup>11</sup>Loan performance is tracked through December 2008. Thus, mortgages still current as of December 2008 are treated as censored.

**Table 3** ■ Hazard rate regression analysis of the probability of default.

	Parameter Estimate	Standard Error	$9 \chi^2$	$p$ value	Hazard Ratio
Age of Loan (in months)	-0.4791	0.007	4324.0	<.0001	0.619
Age Square	0.0034	0.000	2,176.2	<.0001	1.003
Borrower and Loan Characteristics					
FICO (origination)	-0.0080	0.000	3,324.0	<.0001	0.992
LTV (origination)	0.0237	0.001	774.2	<.0001	1.024
Subprime (0,1)	0.2903	0.023	164.3	<.0001	1.337
Low-Documentation (0,1)	0.5968	0.012	2,681.1	<.0001	1.816
No-Documentation (0,1)	0.8541	0.034	640.6	<.0001	2.349
Refinance (0,1)	-0.3699	0.013	768.3	<.0001	0.691
Prepayment Penalty (0,1)	0.1185	0.019	40.4	<.0001	1.126
Adjustable-Rate Mortgage (0,1)	-0.5565	0.014	1,611.9	<.0001	0.573
Hybrid ARM (0,1)	0.7576	0.018	1,776.7	<.0001	2.133
Condominium (0,1)	-0.2646	0.021	154.2	<.0001	0.767
Investor Occupancy (0,1)	0.1140	0.019	37.9	<.0001	1.121
Lien > 1 (0,1)	0.4777	0.029	278.3	<.0001	1.612
ZIP Code monthly House Price Return (lag 1 month)	-1.1864	0.039	910.5	<.0001	0.305
ZIP Code monthly House Price Return (lag 2 month)	-1.5408	0.050	936.2	<.0001	0.214
ZIP Code monthly House Price Return (lag 3 month)	-1.1852	0.045	699.6	<.0001	0.306
ZIP code Concentration Measures					
Percent of Subprime Loans concentrated in ZIPcode	-0.0376	0.004	100.9	<.0001	0.963
Percent of ARM Loans Concentrated in ZIPcode	-0.1087	0.007	247.0	<.0001	0.897
Percent of Hybrid ARMs Concentrated in ZIPcode	0.0236	0.010	5.7	0.0171	1.024
Percent of Investor Occupancy Concentrated in ZIPcode	-0.0973	0.005	327.7	<.0001	0.907
Percent of No-Doc Loans Concentrated in ZIPcode	0.0953	0.015	38.0	<.0001	1.100
Percent of Low-Doc Loans Concentrated in ZIPcode	0.0812	0.004	466.7	<.0001	1.085
Percent of Cashout Refinance Concentrated in ZIPcode	0.0716	0.005	247.6	<.0001	1.074
Percent of Loans with Prepayment Penalty Concentrated in ZIPcode	0.0318	0.005	39.6	<.0001	1.032
Percent of Foreclosed Homes Concentrated in ZIPcode	0.0288	0.001	1,204.9	<.0001	1.029
-2 Log Likelihood (Restricted)	3,004.001				
-2 Log Likelihood (Unrestricted)	2,746.188.5				
Pseudo $R^2$	8.58%				

*Note:* This table reports the maximum-likelihood parameter estimates for the proportional hazard rate model of loan default probability. The dependent variable is a dummy variable equal to one if the loan defaulted (90-days delinquent) and zero otherwise. The ZIP code concentration variables capture the percentage of loans outstanding in the loan's ZIP code at loan origination.

Furthermore, borrowers that originated loans with either low or no documentation are 1.8 and 2.4 times more likely to default than borrowers that provide documentation of their incomes and assets. Not surprisingly, we find that borrowers who originated a mortgage in order to refinance an existing mortgage are less likely to default, while the presence of a prepayment penalty raises the odds of default by 12.6%.

Much discussion in the popular press has blamed the use of ARMs for the current default crisis. However, the estimated coefficient indicating an ARM is negative and significant, indicating that ARMs have a significantly lower default rate than FRMs. However, borrowers who selected hybrid-ARMs (the product most associated with higher-risk subprime borrowers) have significantly higher default rates than FRM borrowers. In fact, the odds ratio for hybrid-ARMs indicates that these mortgages have default rates that are twice as high as FRMs. Finally, we also observe that nonowner-occupied mortgages and mortgages with junior liens have significantly higher default rates than traditional first-lien, owner-occupied mortgages.

To the examine the impact of house price changes on default, we include the lagged monthly house price return measured at the ZIP-code level. As expected, the estimated coefficients indicate that default probability is lower in periods when house prices are rising.

Turning to the measures of mortgage activity in the surrounding area, we find that default risk is highly correlated with mortgage origination activity, albeit in some surprising ways. First, we note that the negative and significant coefficient on subprime concentration indicates that borrower risk actually decreases as the percentage of subprime mortgage in the ZIP code increases. This is in stark contrast with the estimated coefficient indicating that the risk of default is highly correlated with whether the loan is a subprime mortgage. One explanation for this result is apparent in Figure 7 where we see that ZIP codes with the highest subprime concentration had the highest yearly price appreciation in 2003 and 2004 (the peak subprime boom years). This suggests that subprime origination activity was a credit supply phenomena that led to rising house prices in those areas during the periods when these mortgages were being most utilized.

We also see that the risk of default decreases as the concentration of ARMs increases. However, the concentration of hybrid ARMs is positively related to default risk with each percentage increase in hybrid ARM concentration raising the odds of default by 2.4%. Not surprisingly, the presence of low-doc and no-doc borrowers in an area does significantly increase the odds of default, with a one percent increase in no-doc concentration raising the odds of default

by 10%. Consistent with previous studies, follow that foreclosure sales impact surrounding property values, we find that a 1% increase in the percentage of foreclosed homes in a ZIP code increases the odds of default by 2.9%.<sup>12</sup>

## Conclusion

In this article, we examine the relationship between default and subprime mortgage concentration on a local rather than a national level. Subprime mortgages are not evenly distributed over urban areas (in this case, Phoenix, Arizona). Rather, we find that subprime mortgages are more highly concentrated in certain ZIP codes. In the case of Phoenix, these concentrations are found around the Central Business District and other lower-income neighborhoods.

As we would expect, individual borrower risk characteristics play a significant role in explaining the probability of borrower default. For example, borrower credit quality and LTV ratios are important determinants of mortgage risk. Furthermore, individual loans that were classified as Subprime or Alt-A mortgages were significantly riskier than loans to traditional, prime borrowers. Furthermore, our analysis shows that increases in the local foreclosure rate (using the concentration of foreclosures in the ZIP code as a proxy) raises the probability of borrower default. None of these results are surprising.

However, our analysis does reveal that, after controlling for individual borrower risk characteristics and foreclosures in the area, the concentration of subprime lending in the neighborhood does not increase the risk of borrower default. In fact, we find the opposite. As a result, it does not appear that extending credit to subprime borrowers in general increased the probability of borrower default. Rather, our analysis suggests that subprime lending is a credit supply effect that led to rising house prices in those areas.

We do find that higher concentrations of the more aggressive mortgage products (hybrid ARMS and no- or low-documentation loans) did increase the probability of borrower default. This finding is important given the current policy debates concerning the role of subprime lending and the formation and burst of the housing bubble.

---

<sup>12</sup>We conducted a series of robustness tests to check if the results are sensitive to specification errors, omitted variables and nonlinear explanatory variable specifications. For instance, we estimated models with FICO and LTV splines and/or dummies, and our results are robust to these alternative specifications. We also estimated the model with individual fixed effects and a series of alternative specifications for the house price return series. Again, our results are robust to these alternative specifications. Results of these tests are available upon request.

*We thank Gene Amromin, Gadi Berlevy, Scott Frame and the participants at the American Real Estate and Urban Economics Association mid-year meeting for their helpful comments and suggestions. We also thank Jacqui Barrett and Sriram Villupuram for their research assistance and the Arizona State University Center for Real Estate Theory and Practice for the use of the Maricopa County housing data. The views expressed in this research are our own and not necessarily those of the Federal Reserve System, the Federal Reserve Bank of Chicago or the Office of the Comptroller of the Currency.*

## References

- Ambrose, B.W. and R.J. Buttimer, Jr. 2000. Embedded Options in the Mortgage Contract. *The Journal of Real Estate Finance and Economics* 21(2): 95–111.
- Anderson, L. and C. Holt. 1997. Information Cascades in the Laboratory. *American Economic Review* 87: 847–862.
- Bikhchandani, S. D. Hirshleifer and I. Welch. 1998. Learning from the Behavior of Others: Conformity, Fads, and Informational Cascades. *Journal of Economic Perspectives* 12(3): 151–170.
- Chang, A., S. Chaudhuri and J. Jayaratne. 1997. Rational Herding and the Spatial Clustering of Bank Branches: An Empirical Analysis. Federal Reserve Bank of New York Research Paper No. 9724.
- Clautetie, T. and N. Daneshvary. 2009. Estimating the House Foreclosure Discount Corrected for Spatial Price Interdependence and Endogeneity of Marketing Time. *Real Estate Economics* 37(1): 43–67.
- Coleman, M. M. LaCour-Little and K.D. Vandell. 2008. Subprime Lending and the Housing Bubble. *Journal of Housing Economics* 17(4): 272–290.
- Danis, M.A. and A. Pennington-Cross. 2008. The Delinquency of Subprime Mortgages. *Journal of Economics and Business* 60(1/2): 67–90.
- Campbell, J., S. Giglio and P. Pathak. 2009. Forced Sales and House Prices. NBER Working Paper Number 14866.
- Case, K.E. and R.J. Shiller. 1987. Prices of Single Family Homes since 1970: New Indexes for Four Cities. *New England Economic Review* 45–56.
- Foote, C.L. K. Gerardi and P.S. Willen. 2008. Negative Equity and Foreclosure: Theory and Evidence. *Journal of Urban Economics* 64(2): 234–245.
- Foster, C. and R. Van Order. 1984. An Option Based Model of Mortgage Default. *Housing Finance Review* 3: 351–372.
- Frame, W.S. 2010. Estimating the Effect of Mortgage Foreclosures on Nearby Property Values: A Critical Review of the Literature. Working Paper, Federal Reserve Bank of Atlanta.
- Gerardi, K., A.H. Shapiro and P.S. Willen. 2008. Subprime Outcomes: Risky Mortgages, Homeownership Experiences, and Foreclosures. Federal Reserve Bank of Boston, No. 07.a15.
- Glaeser, E.L., J. Gyourko and R.E. Saks. 2005. Why Have Housing Prices Gone Up? *American Economic Association Papers and Proceedings* 95(2): 329–333.
- Gross, D.B. and N.S. Souleles. 2002. An Empirical Analysis of Personal Bankruptcy and Delinquency. *Review of Financial Studies* 15(1): 319–347.
- 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(1): 57–79.

- Jackson, J. and D. Kasserman. 1980. Default Risk on Home Mortgage Loans: A Test of Competing Hypotheses. *Journal of Risk and Insurance* 3: 678–690.
- Kau, J.B. D.C. Keenan and T. Kim. 1994. Default Probabilities for Mortgages. *Journal of Urban Economics* 35(3): 278–296.
- Lauria, M. and V. Baxter. 1999. Residential Mortgage Foreclosure and Racial Transition in New Orleans. *Urban Affairs Review* 34(6): 757–786.
- Leamer, E.E. 2007. Housing in the Business Cycle. National Bureau of Economic Research Working Paper 13428.
- Lee, K. 2008. Foreclosure's Price-Depressing Spillover Effects on Local Properties: A Literature Review. *Federal Reserve Bank of Boston Community Affairs Discussion Papers*, 2008-01.
- Lin, Z., E. Rosenblatt and V.W. Yao Spillover. 2009. Effects of Foreclosures on Neighborhood Property Values. *The Journal of Real Estate Finance and Economics* 38(4): 387–407.
- Mian, A. and A. Sufi. 2008. The Consequences of Mortgage Credit Expansion: Evidence from the 2007 Mortgage Default Crisis. NBER Working Paper No. W13936.
- Moreno, A. 1995. *The Cost-Effectiveness of Mortgage Foreclosure Prevention*. Family Housing Fund: Minneapolis, MN.
- Schuetz, J. V. Been and I.G. Ellen. 2008. Neighborhood Effects of Concentrated Mortgage Foreclosures. *Journal of Housing Economics* 17(4): 306–319.