An Analysis of Foreclosure Activity in the United States

Stella K Hofrenning
Augsburg College

D. Stephen Drott
Target Corporation

Greg Kugler
U.S. Bank

This paper has analyzed the foreclosure rates of US counties over an 18 month period beginning in January 2007 and ending in June 2008. Population, home and loan attributes of counties all had a statistically significant bearing on foreclosure rates. Specifically, subprime loans, the employment rate and housing price declines were correlated with higher foreclosure rates. Racial and ethnic differences did have a negative effect on foreclosure rates but were much smaller in magnitude.

INTRODUCTION

The financial economic crisis of 2007-2008 brought the difficult experiences of foreclosure to many households in the United States. In the second quarter of 2009, 2.34 percent of all single family residential mortgages have been charged off by banks, this is the highest percent in the modern era (Federal Reserve 2009). According to First American Core Logic the US House Price Index fell 32 percent from its all-time high in April 2006. RealtyTrac calculated that in 2008 there were 3.2 million foreclosures on 2.3 million properties; this was an increase of 81 percent on 2007. Foreclosure rates across the US were not evenly distributed by state or county. Florida, Nevada and California were the most heavily hit by the current wave of foreclosures. What caused this predicament? To understand this one must look at the root causes of the foreclosure epidemic.

This crisis has been blamed on the explosion in subprime mortgages, the subsequent foreclosures of those mortgages and the opaqueness of the securitization of them. In order to understand this subprime effect one first needs to define what subprime means and then quantify the scope of subprime mortgages. There are multiple definitions of subprime. One can define it through attributes of the borrower, lender or the loan (Foote et al, 295). The borrower is considered subprime if their FICO score is below 620 and Alt-A if in-between 620 and 659. FICO scores are the credit industry standard for measuring credit quality of individuals. It is based off of current and historical debt activities of individuals as reported by lending institutions. Alt-A is sometimes included in subprime statistics, however, the differentiation between the two is not consistent across reporting methodologies. A prime borrower has a score of 660 and above

(Office of the Comptroller of the Currency, 9). The US Department of Housing and Urban Development identifies subprime lenders as those that originate a smaller number of loans, have a high proportion of re-financing loans and sell less of those to government-sponsored entities [GSE's] such as Fannie Mae and Freddie Mac (Sengupta and Emmons, 1). This is a good indicator of a subprime lender as GSE's are barred from purchasing risky subprime loans. The final possible definition of subprime is looking at the loan characteristics. If a loan has been packaged into a subprime mortgage backed security [MBS] then it is subprime by definition. A MBS is a security that is sold to investors who are entitled to a share of the principal and interest payments from the assets that are pooled, i.e. the mortgages. If the rate is significantly higher than the average mortgage rate it also can be considered subprime. According to the Office of the Comptroller of Currency subprime conservatively accounts for 9 percent of mortgages in their survey. The International Monetary Fund estimates that 25 percent of outstanding mortgages are subprime, which equates to $2.75 trillion. Subprime loans are the most likely to become serious delinquent, which is the final stage before foreclosure, thus foreclosure rates are associated with sub-prime mortgages.

LITERATURE REVIEW

The subprime crisis and subsequent spillover into the greater economy has been one of significant interest with a wealth of academic literature on it. This paper builds off of these notable scholarly articles and a number of these are reviewed and discussed.

Gene Amromin and Anna L. Paulson are both financial economists at the Federal Reserve Bank of Chicago. In their 2009 paper they analyzed both subprime and prime mortgage originations each year from 2004 to 2007 to understand trends in foreclosures with respect to loan, borrower and economic attributes. They were able to base their analysis of off individual US loan level data from a private database. They used a one percent random sample for 68,000 prime loans and a ten percent random sample for 32,000 subprime loans.

When analyzing trends in subprime borrowers Amromin and Paulson noted that there was a significant decline in credit scores from 2004 to 2007. The same effect was not found with prime borrowers (21). A borrower's motive was also reviewed, a mortgage used for purchasing a house was considered more risky than one used to refinance. However Amromin and Paulson point out that one needed to differentiate between refinancing to take advantage of lower rates versus refinancing to cash-out increased equity. The latter scenario amplified risk in existing home's new loans as the cash-out option increased leverage and was tied to a belief in eternally increasing house prices. Subprime mortgages also had a higher percent that were cash out refinancing. Cash-out increased from 35 percent in 2004 to 57 percent of all subprime mortgages in 2007 (24).

Another data set reviewed by Amromin and Paulson was the prevalence of adjustable rate mortgages [ARM] in subprime lending. With ARMs a low initial interest rate is set for a period of one to seven years and then resets to another rate that typically is higher than the initial rate. An increase in rates means an increase in payments which could increase the chance of default. Also the potential of default increases if the ability to refinance before the reset date is not available, such as a lack of general credit. Amromin and Paulson found ARMs decreased from 73 percent of subprime loans to 39 percent in 2007 (24) as general credit supply was tightened.

In their empirical study, Amromin and Paulson used a regression of the probability of default within the first 12 months against loan and borrower characteristics as discussed above, with Metropolitan Statistical Area [MSA] level data for unemployment, Federal Housing Finance Agency [FHFA] home price changes and income and dummy variables that describe the loan regulatory and economic framework at a state level (25). Their results showed that subprime loan defaults decreased much more quickly than prime defaults for the same increase in FICO score. Loan to value ratio (LTV) was also found to be significant and again subprime loans were more sensitive with a higher LTV increasing the probability of default (26). LTV is ratio of how much a mortgage is worth relative to the appraised value of the property. Thus, higher LTVs are considered riskier, as there is less of an equity cushion.

The regression analysis also indicated that ARMs did not increase the likelihood of subprime mortgages going into default. However in another metric, Amromin and Paulson concluded that house prices were a very large reason for foreclosure increases, especially in the subprime market. They noted that this was very noticeable in 2005 and beyond. They found that the prime default relationship was not correlated with house price changes until 2007.

In 2009, Kelly D. Edmiston the senior economist at the Federal Reserve District of Kansas City reviewed foreclosure rates in that district from June 2006 through July 2008. Across that region Edmiston noted differences in foreclosure rates with high concentrations in low income areas and more recent clusters in higher income neighborhoods. At a neighborhood level Edmiston saw variance in the default rate from zero percent to 17 percent (66). Regression analysis was used to understand these variances.

In the analysis the dependent variable was the rate of foreclosure by neighborhood. Numerous variables were included in the analysis; of note were vacancy rate, violent crime, unemployment rate, self-employment rate, owner-occupied percent, low credit score percent and house price appreciation. There was a correlation between foreclosure rates and low income neighborhoods but only where subprime mortgages were present. This relationship was in fact true across all income level neighborhoods.

Edmiston's regression analysis confirmed a strong correlation between rising house prices and falling foreclosure rates, a five percent drop in home values led to a 33 percent rise in foreclosures (60). It also confirmed that not surprisingly vacancy rates were found to be positively correlated with foreclosures. An interesting finding was that as the owner-occupancy rate increased it was positively correlated with the foreclosure rate, albeit the absolute level was small (63). A positive relationship between defaults and unemployment was also not a shock; although Edmiston also confirmed that self-employment rates positively correlated with foreclosure rates at the same absolute level as unemployment (57). Of note was the correlation between a one percent increase in violent crime and a one percent decrease in foreclosure rates.

Perhaps Edmiston's most interesting data is the regression analysis of racial and ethnic demographics. His African-American variable was statistically significant at the 0.01 level for a negative correlation to foreclosures. When analyzed, the African-American subprime relationship with foreclosures turned positive (70). Looking at Hispanic demographics they correlated positively with default rates and reviewing Hispanic subprime the relationship was not significant (71).

The role of race in the subprime mortgage market is further addressed by Andrew Haughwout, Christopher Mayer and Joseph Tracy. In their 2009 paper they do not attempt to analyze foreclosure rates; instead Haughwout, Mayer and Tracy look at the initial pricing of subprime loans in relation to race, ethnicity and gender.

The New York economists narrow their study to one type of mortgage, the 2/28 adjustable rate. They point out that this 2/28 mortgage accounted for 53 percent of all subprime loans that were being serviced as of the end of 2007 (35). This type of mortgage had a set rate for two years and then adjusted every 6 months based on a margin in excess of a short term base rate which was the six-month LIBOR rate (41). They chose one month's worth of originations for that mortgage type in August 2005. Their logic behind that choice was that August 2005 was the height of issuance for subprime loans (35).

They used loan level data supplied by a private data firm that contained location, lending terms and risk factors and conjoined it with census tract and other data from the Home Mortgage Disclosure Act [HMDA]. This process enhanced the metric set for each individual loan (37).

When digging into the data Haughwout, Mayer and Tracy noted that African-American's took out mortgages on properties that experienced lower levels of price increase in 2004. They also tended to have smaller loan amounts by about 12 percent and had lower credit scores. In addition they provided full documentation for their loans 70 percent of the time. This contrasted with Asian and Hispanic borrowers who failed to provide full documentation more than half the time. Also Asian and Hispanic borrowers tended to have mortgages in faster house price appreciating locations and the vast majority of these were not for refinance but for home purchase. Haughwout, Mayer and Tracy found no difference in pricing between genders in the loans reviewed for their paper (40).

In their empirical analysis, Haughwout, Mayer and Tracy used a regression with the dependent variable of interest rate against dummy variables for Asian, Hispanic, African-American and gender. They used dummy variables for varying levels of credit score, LTV and added variables to control for geographic location such as percent of population that was African-American, homeownership rates, house price change and unemployment.

Their results were surprising in that the data showed there was no detrimental pricing effect by race, ethnicity of gender accounting for other risk variables. The regression in fact showed correlation in a slight price advantage especially in refinancing mortgages; that is on average minority borrowers received a lower interest rate. In addition geographic areas with higher rates of unemployment or African-American or Hispanic population showed negative correlation with interest rates (50). Haughwout, Mayer and Tracy's analysis also showed that for areas that had a history of high house price increase there was a significant correlation with lower interest rates (52).

Their discussion on the reason behind these enlightening findings centered on the concentrations of minority populations creating a positive supply effect for credit, in effect with elevated competition economies of scale come into play when mortgage rates were priced (Haughwout, Mayer and Tracy, 53). A second note discussed was that this regression analysis was conducted with the mean interest rate for a given variable. It could have been the case that prejudicial practices only occurred in the outer distributions of the interest rate pricing (Haughwout, Mayer and Tracy, 45).

In 2008, Jenny Schuetz, Vicki Been and Ingrid Gould Ellen used regression analysis in their paper on concentrated foreclosures' effects on surrounding neighborhoods. The three authors, from the Furman Center for Real Estate and Urban Policy, looked at foreclosures in the New York City boroughs from 2000 to 2005. They wanted to understand how a foreclosure affected the price of houses surrounding it.

In their regression they used data obtained from the New York City department of finance and a private data sourcing company to identify property attributes. Schuetz, Been and Ellen found that there is a correlation between nearness to a foreclosure and a reduced selling price. They also noted that it was not a linear relationship; dummy variables were used for distance to foreclosure and concentration of foreclosures. Within 1000 feet there was a significant effect if there were concentrated numbers of foreclosures (Schuetz et al. 316).

Schuetz, Been and Ellen also addressed the chicken or the egg question by controlling for underlying neighborhood propensity of foreclosures, thus avoiding selection bias. Given that they found that neighborhoods with lower selling prices were more likely to have higher levels of foreclosure (Schuetz et al., 317).

In Foote et al. 2008 the economists from the Federal Reserve Bank of Boston reviewed a slew of facts on subprime's role in the housing crisis. Their analysis is concentrated in Massachusetts and southern New England and goes back to 1989. They used two datasets. The first was a unique set which contained mortgage information at individual housing unit level and also contained ownership length. This allowed Foote at al. to analyze what they called "ownership experiences" (Foote et al., 294). If the mortgage was refinanced by a borrower they were able to maintain continuity in the data and not assign a different data point to the refinance.

The phenomenon of prime borrowers refinancing into a subprime mortgage was analyzed. They found that a significant number of prime borrowers refinanced into a subprime loan and subsequently defaulted. In a significant number of these cases the house had been originally purchased before 1999. Foote et al. suggested that this was evidence for cash-out refinancing. They pointed to the large house price appreciation in the 2000's, which was also calculated from this data (304). A trend that was also called out was that multi-family homes had increased in foreclosure share, 20.4 to 28.4 percent and was disproportionately large, only 10.8 percent of homes purchased were multi-family (203). This data also confirmed a correlation between house price depreciation and foreclosure rate increases (294).

The second dataset contained interest rate information on individual loans. Foote at al. used this to analyze ARMs. They found that these types of mortgages did not default in relationship to their interest rate reset timing as thought in the popular media. They reviewed the popular 2/28 ARM and found that the likelihood of default increased quickly through the first 12 months of the loan and then tapered off

(299). In further analysis Foote at al. compared subprime ARMs and FRMs to see if there was a difference in sensitivity to foreclosure risk with changes in house price appreciation. They found that in a high appreciation environment there was no statistically significant difference in default risk, however when house prices were depreciating the subprime ARMs were more likely to go into default than subprime FRMs (300).

Federal Reserve System scholars Gerardi et al. 2008 review the subprime crisis through the eyes of the investment community on Wall Street. They wanted to understand if there was data available to analysts that should have given fair warning about the impending implosion of the subprime market. Fundamentally Gerardi et al. broke down the foreclosure increase into the following equation (71):

$$\delta f / \delta t = \delta f / \delta p \times \delta p / \delta t.$$

Where f is foreclosure, t is time and p is house prices. They wanted to know if analysts underestimated the change of foreclosures with change in price or if they did not properly predict the change in prices over time. Their analysis is relevant to this paper's topic as they looked at salient factors that had an effect on foreclosure rates and then ultimately the performance of mortgage backed securities. Gerardi et al. used the same two data sets as Foote et al. and in addition used HDMA data to identify high-cost loans. High cost loans as defined by HDMA have an interest rate that is at least three percentage points above the US Treasury's 30 year bond. They ran multiple regressions with these datasets to reach some insightful conclusions.

In their research they found no evidence to support a regularly asserted belief that an increase in securitization of subprime loans led to the crisis. The argument was that since most companies that originated subprime loans quickly packaged them and sold them onto investors as MBS there was no incentive to properly screen mortgage applicants. Therefore increasingly subprime loan had little or no documentation and the borrowers were given loans that they were unable to repay. This information was not passed onto investors and the opaqueness caused the securities to be overvalued (73). Gerardi et al. found no correlation in foreclosures with this activity.

Gerardi et al. found that risk layering did increase year over year in subprime mortgages. Risk layering is where loan originators approve a mortgage for borrowers with two or more risky attributes, such a low documentation and home purchase. They analyzed the 2/28 ARM and found that one type of layering in particular had an out-weighted effect on foreclosure rates; high LTV with low documentation loans (94).

Turning to house price appreciation [HPA] Gerardi et al. found that foreclosure rates had a large negative correlation with HPA (134). They also found that Wall-Street analysts fully understood the key relationship of $\delta f / \delta p$. The fault with predicting the current issues was that virtually all analysts did not accurately predict $\delta p / \delta t$. The common belief was that the HPA rate of change would decrease but only to positive normal historical levels (139).

In reviewing the current articles on this subject there are a number of reoccurring themes that are worth noting. The majority of studies used a measurement of foreclosure as their dependent variable when running regression analysis. All papers that addressed it found significant negative correlation between house price appreciation and the foreclosure rates. The northeast was a definite favorite for datasets and most of the authors were able to use individual loan level data that was purchased by large institutions. An interesting finding that was repeated over multiple articles was that the reset date for ARMs had no real effect on foreclosure rates. When one looks at the role of race and ethnicity in foreclosures the findings are somewhat mixed. Anecdotal evidence would lead one to expect significant positive correlation between minorities and foreclosure. This was not the case. The findings suggest that the relationship is much more muted and in certain circumstances negatively correlated.

Whilst each of the papers reviewed provide more insight into the complex issue of foreclosure, they were not without fault. Amromin and Paulson, Gerardi et al. and Foote et al. could have improved upon their analysis by including race and ethnicity to broaden the understanding and improve their regressions as their R^2 ranged from 0.09 to 0.25 [where reported]. In Edmiston's paper one felt that detailed descriptions of what level data was being analyzed was definitely lacking. It was hard to understand if the

data being discussed was at a county or neighborhood level, an important distinction on understanding the scale of correlation between metrics.

In Haughwout, Mayer and Tracy it would have been more insightful to analyze a larger period than only one month's loans, especially given the findings on race and reduced interest pricing. In addition to pricing of subprime loans it would have been even more useful to expand the analysis to foreclosure rates on these loans. One has yet to find an analysis of the relationship of individual loan level pricing given racial and ethnic variables and then future expected and actual foreclosures. Their paper would have been ideal for this analysis.

For Schuetz, Been and Ellen their analysis on neighborhood effects only reviewed New York City. Since this city is one of the most expensive locations in the country, the data should be considered significantly skewed and application of findings to larger geographies may not fully translate. They could have improved upon it by opening up their geographical area.

THEORETICAL MODEL

This paper will construct a model that builds on the prior work and enhances understanding of foreclosure rates. The theory that helps build this framework is from Foote, Gerardi et al. 2008. The theory is that foreclosures are a function of house price change. As prices fall more borrowers owe more on their mortgage than the house is worth, this coupled with a negative life event such as being unemployed causes the foreclosure. If the borrower had positive equity in the house they could sell it and pay off the mortgage. The price decline forces the hand of the borrower when coupled with cessation or reduction in cash. Subprime mortgages are more prone to this as many have higher loan to value ratios.

The theoretical model includes loan and demographic variables of housing and population. It differs from prior models in that it increases demographic attributes. The hypothesis is that level of ethnic and racial characteristics change the rates of foreclosure. African Americans drive rates down and Hispanics drive rates up. Other hypotheses are that as income levels rise foreclosure rates fall and that crime rates drive foreclosure rates down given previous analysis' conclusions such as Edmiston 2009. The final hypothesis is that as the rate of owner occupied houses increases the foreclosure rate will increase. This is because house purchase mortgages are riskier than refinance mortgages.

The conceptual framework is shown below:

$$\text{Foreclosure Rate} = f(\text{HPA, UR, SUB, VR, CR, PCI, AAP, HLP, OO})$$
$$\text{Expected sign} = \quad -\downarrow \quad +\downarrow \quad +\downarrow \quad +\downarrow \quad \downarrow \quad \downarrow \quad \downarrow \quad +\downarrow \quad \downarrow$$

where HPA is house price appreciation; UR is unemployment rate; SUB is percent of mortgages that are subprime; VR is vacancy rate; CR is crime rate; PCI is per capita income; AAP is population percent African-American; HLP is population percent Hispanic; OO is owner occupied percent.

DATA AND EMPIRICAL MODEL

Prior analysis on foreclosures has tended to focus on certain geographic areas or random samples of loan level data. Whilst this presents useful inferences for the entire country it is not a comprehensive geographical analysis. In this paper's model county level data will be analyzed for all United States counties to enhance understanding of foreclosures. The county level data used for this paper was obtained from two sources and melded to create a hybrid data set. The first set was from the Department of Housing and Urban Development [HUD]. HUD estimated foreclosure starts in all 3122 counties in the United States for 18 months starting at the beginning of 2007. The county foreclosure estimate was a function of three factors: a measure of house price declines from their peak in the Federal Housing Finance Agency [FHFA] metropolitan area house price index. This house price index was found to be used in numerous other regression analyses on the subject including Amromin and Paulson, 2009. The second factor was the Bureau of Labor Statistics [BLS] unemployment rate in June 2008. Differing

variants of this BLS statistic are used as a variable in multiple academic discussions including Foote et al. 2009. Thirdly the percent of HDMA defined high cost loans made from 2004 to 2007 was used as a variable. This too has been widely used in other studies such as Haughwout, Mayer and Tracy 2009 and Gerardi et al. 2008. All three of these variables were included in the HDMA data along with the estimated number of foreclosures and the foreclosure rate. The HDMA also contained useful metrics such as the USPS residential 90 day vacancy rate. This data set had 12 metrics for each of the 3122 counties.

Whilst the HDMA had numerous variables, the theoretical model states that there is a relationship between foreclosures and ethnic minorities, income levels and owner occupied percent. The HUD data did not take this into account when estimating foreclosures. Therefore to augment the regression a second set of county level data was obtained from FedStats.gov. This set had 82 mostly population demographic metrics for every county in the United States. The data was collected from the US Census in 2000 and had updated metrics for 2007. There was no data found on crime rates at a county level, so this metric which was in the theoretical model was unable to be measured empirically. The final hybrid data set combined to include 94 metrics with a sample size of 3122.

The dependent variable used in this analysis was the estimated county foreclosure rate over 18 months from January 2007 through June 2008. This was obtained from the HUD data set. The independent variables were an attempt to classify three types of attributes. The first described the population characteristics of each county, such as unemployment rate and percent African American. The second type described the housing stock in a county; these metrics included owner-occupied percent and price decline in home value. The third measured the loan characteristics with a metrics of high cost loan percent. There were no dummy variables used. The variables used the in the regression analysis are defined in Table 1.

TABLE 1
DEFINITION OF VARIABLES

Variable	Type	Definition
Foreclosure rate	Dependent	HUD model, estimated count of foreclosure starts over 18 months through June 2008 divided by number of mortgages times 100.
Bachelor's degree %	Population	Educational attainment - persons 25 years and over - percent bachelor's degree or higher 2000
BLS unemployment rate	Population	June 2008 county unemployment rate
Civilian labor force unemployment rate 2007	Population	Civilian labor force unemployment rate 2007
High school graduate %	Population	Educational attainment - persons 25 years and over - percent high school graduate or higher 2000
Per capita income 2006	Population	Per capita personal income 2006
Population - 2 or more races %	Population	Resident population: Two or more races, percent (July 1 - estimate) 2007
Population - African American %	Population	Resident population: African American alone, percent (July 1 - estimate) 2007
Population - Asian %	Population	Resident population: Asian alone, percent (July 1 - estimate) 2007
Population - Latino %	Population	Resident population: Hispanic or Latino Origin, percent (July 1 - estimate) 2007
Population - Not Hispanic, White %	Population	Resident population: Not Hispanic, White alone, percent (July 1 - estimate) 2007
Population % Chg	Population	Resident total population estimate, percent change - April 1, 2000 to July 1, 2008
Population per square mile	Population	Population per square mile 2000
Poverty %	Population	People of all ages in poverty - percent 2007
Persons per households	Population	Persons per household 2000 (complete count)
90 day vacancy rate	Housing	United States Postal Service data from June 2008 on residential addresses vacant 90-days or longer divided by total addresses times 100.
Housing Units Chg %	Housing	Housing unit estimates - percent change, April 1, 2000 (base) to July 1, 2007
Multiple housing units	Housing	Housing units by units in structure - multiple units 2000 (sample)
New private housing units authorized by building permits	Housing	New private housing units authorized by building permits - total 2007 (20,000-place universe)
Owner-occupied %	Housing	Owner-occupied housing units - percent of total occupied housing units 2000

Price decline in home values	Housing	A measure of price decline in home values that uses data from the Office of Federal Housing Enterprise Oversight (OFHEO) Housing Price Index (HPI) to calculate price decline from peak value in the second quarter of any year between 2000 and 2008 and the second quarter home price in 2008
High cost loan rate	Loan	Percent of loans made between 2004 and 2006 shown to be high cost according to HMDA data.

Source: HUD http://www.huduser.org/Datasets/nsp_foreclosure_data.html & Fed Stats
http://www.fedstats.gov/qf/download_data.html

The population variables of ethnicity and race were chosen as they related back to similar metrics in Mayer and Tracy 2009 and Edmiston 2009. Poverty and per capita income were used to measure wealth as Edmiston 2009 had noted disparate rates of foreclosures between lower and higher income neighborhoods. High school diploma and bachelor degree metrics were included to further differentiate the population; historically people with bachelor degrees have had higher earnings and lower unemployment. Population density and persons per household were included to further enhance measurement of condominiums and high density housing. There is historical precedent for this as Edmiston 2009 used the same metrics in his analysis.

The unemployment variables were included as Foote at al. 2008 along with Haughwout, Mayer and Tracy 2009 had found high significance with the metric. There were two measurements of unemployment included in the combined data set. The HUD data contained a June 2008 measurement; this metric was used in estimating the county level foreclosure rate metric in the HUD data. To provide independent verification of the relationship the average 2007 unemployment level from Fed Stats was included as well.

For housing attributes the house unit change and new permits were included as similar variables were in Edmiston 2009. These metrics approximate new housing activity and as the literature found that new housing loans were more risky than refinancing it was important to include. The owner occupied and multi-housing metrics were incorporated as both of these had been found by Gerardi et al. 2008 and Foote et al. 2008 to be correlated with foreclosure rates.

Ninety day vacancy rate was included as a measure because of the analysis noted of Edmiston 2009 and Schuetz, Been and Ellen 2008. As vacancy rates increase, prices are reduced as there is an increase in supply for the same level of demand. In addition vacancy rates increase neighborhood blight and neighborhood flight. The last housing metric was the most important one, all literature reviewed was in agreement that price change was one of the most significant variables to correlate with foreclosures. The metric used is a measure of price decline not appreciation, so signs will be reversed from the theoretical model. The only loan attribute used was also a very important metric, high cost loans measured the proportion of higher than average loans for the county. This allowed us to measure a rate of subprime activity

The following regression contains the variables and signs that, given the theoretical base, will estimate the relative strength of association between foreclosure rate and observable attributes of population, housing and loans in US counties:

Foreclosure Rate = β_1 - β_2 Price decline in home values % + β_3 BLS unemployment rate

+ β_4 High cost loan rate + β_5 90 day vacancy rate

+ β_6 Per capita income 2006 + β_7 High school graduate %

- β_8 Bachelor's degree % - β_9 Population African American %

+ β_{10} Population Asian % + β_{11} Population Latino %

+ β_{12} Multiple housing units % + β_{13} Owner-occupied %

The most important hypothesis behind this formula is that the foreclosure rate is negatively related to price decline in home value. The second most important is that as high cost loan proportion goes up so does the foreclosure rate.

RESULTS

The descriptive statistics are presented below in Table 2 and estimations of the regressions are shown in Table 3.

TABLE 2
SUMMARY STATISTICS OF DATA

Variable	Type	Mean	Std. Dev.	Median	Mode	Min	Max
Foreclosure rate	Dependent	4.94	2.14	4.78	0	0	33.33
Bachelor's degree %	Population	16.53	7.79	14.5	12.1	4.9	63.7
BLS unemployment rate	Population	5.76	2.03	5.54	3.29	1.32	22.73
Civilian labor force unemployment rate 2007	Population	4.86	1.7	4.6	4.6	1.5	18
High school graduate %	Population	77.39	8.73	79.2	81.2	34.7	97
Per capita income 2006	Population	28330.51	7713.82	27372.5	0	0	110292
Persons per households	Population	2.54	0.2	2.5	2.45	2	4.36
Population - 2 or more races %	Population	1.19	1.11	1	0.7	0	25.7
Population - African American %	Population	9.16	14.47	2.3	0.2	0	85.8
Population - Asian %	Population	1.1	2.33	0.5	0.3	0	45
Population - Latino %	Population	7.45	12.64	2.6	1.3	0.1	97.3
Population - not Hispanic, White %	Population	79.71	19.07	87.1	96.5	2.1	99.4
Population % chg	Population	3.11	11.77	1.4	-0.1	-43.9	89.6
Population per square mile	Population	244.52	1670.91	43.1	2.4	0	66834.6
Poverty %	Population	15.09	6.25	14.1	16.4	2.4	55.9
90 day vacancy rate	Housing	3.79	3.22	3.13	0	0	27.78
Housing units chg %	Housing	7.92	9.03	5	2.8	-68.7	98.2
Multiple housing units	Housing	12.07	9.44	9.3	5.6	0	99.1
New private housing units authorized by building permits	Housing	446.9	1547.47	69	0	0	40238
Owner-occupied %	Housing	74.05	7.55	75.4	77.9	19.6	89.9
Price decline in home values	Housing	-0.75	3.11	0	0	-40.55	0
High cost loan rate	Loan	30.08	10.33	28.83	33.33	0	100

TABLE 3
REGRESSION RESULTS

VARIABLE	REGRESSION A	REGRESSION B
Intercept	-5.885*	-5.053*
	(0.456)	(0.482)
Price decline in home values	-0.165*	-0.184*
	(0.007)	(0.007)
BLS unemployment rate	0.397*	
	(0.012)	
Civilian labor force unemployment rate 2007		0.369*
		(0.014)
High cost loan rate	0.151*	0.147*
	(0.003)	(0.003)
90 day vacancy rate	0.036*	0.043*
	(0.006)	(0.007)
Per capita income 2006	-3.54E-06	-9.16E-06*
	(3.49E-06)	(3.69E-06)
High school graduate %	0.042*	0.029*
	(0.004)	(0.004)
Bachelor's degree %	-0.044*	-0.043*
	(0.005)	(0.005)
Population - African American %	-0.014*	-0.007*
	(0.002)	(0.002)
Population - Asian %	-0.01	0.001
	(0.011)	0.011)
Population - Latino %	-0.008*	-0.011*
	(0.002)	(0.002)
Multiple housing units	0.03*	0.040*
	(0.004)	(0.004)
Owner-occupied %	0.015*	0.025*
	(-0.004)	(0.005)
R^2	0.759	0.730
R^2ADJ	0.758	0.728
F	815.11	698.61

NOTE: Standard errors are in parentheses below each estimate.
*Statistically significant at the 0.05 level

Reviewing the results from the regressions one can see that the estimated empirical model in Regression A explains approximately 76 percent of the variation in the foreclosure rate in US counties. Population, home and loan attributes of counties all had a statistically significant bearing on foreclosure rates. As expected, an increase in the unemployment rate has a significant positive impact on foreclosure rates holding everything else constant. Both price declines in home values and increases in high cost loan rates have a significant negative relationship with foreclosures rates. Racial and ethnic differences did have a negative effect on foreclosure rates but were much smaller in magnitude than housing attributes such as the owner-occupied percent.

Regression B in Table 3 uses the civilian labor force unemployment rate in 2007 instead of the BLS unemployment rate. The signs of independent variables remain consistent between the two regressions using the BLS unemployment rate.

The results showed that holding all other explanatory variables equal an increase in the 90 day vacancy rate, high school graduate proportion of county population, multiple housing units or owner-occupied percent of county houses increased average US county foreclosure rates. The following explanatory variables showed that holding all other variables equal a decrease in the proportion of county population with bachelor degrees, African American percent of population or Latino percent of population increased average US county foreclosure rates.

Interestingly neither per capita income nor Asian population percent were statistically significant variables for foreclosure rates. This could potentially be the case because the foreclosure rate of the Asian population with county level data may not differ from the general population. Also the variable of Latino population % had a negative relationship to foreclosure. One possible reason for this could be the positive supply impact on pricing as mentioned in Haughwout, Mayer and Tracy's 2009 analysis. The lower pricing could offset other factors in Latino communities that increase foreclosure rates.

CONCLUSION

This paper has analyzed the foreclosure rates of US counties over an 18 month period beginning in January 2007 and ending in June 2008. It reviewed previous literature on foreclosures to build a theoretical model and tested hypotheses with an empirical model and regression analysis. It found that population, home and loan attributes of counties all had a statistically significant bearing on foreclosure rates. Subprime loans were correlated with higher foreclosure rates; along with unemployment rates and house price declines these affected foreclosures by one factor larger than any other variable. Racial and ethnic differences did have a negative effect on foreclosure rates but were much smaller in magnitude than housing attributes such as the owner-occupied percent.

Potential next steps are to investigate private sourcing of foreclosure rates as a thorough examination of available options showed that the estimate used was the only viable data available from public sources. This dependent variable of county foreclosure rates was a HUD estimate. The metrics of house price index, unemployment and high cost loan estimate that helped to calculate HUD's estimate were also used in this paper's regression. However all data sets have been the standard used in multiple studies and industry analysis at levels above individual loan data. Given that fact a further next step would be to obtain loan level data from counties across the US and rerun the analysis.

Other potential data improvement would include a different measure of price decline, only 400 counties had a measurement other than zero for house price decline. Although those counties accounted for over half the US population a more complete data set would be available from a private firm such as First American. Also to further enhance price decline an additional metric for negative equity could be sourced, again this would have to be purchased from a private data company.

Despite these opportunities for improvement this research still adds to the understanding of the foreclosure phenomenon. The results point to policy makers needing to take a closer look at regulation of subprime loans and macroeconomic policies which helped inflate house prices. This paper contributed to this subject by confirming unemployment and house prices as triggers for foreclosure; but also showing

that race and ethnicity does not explain increased foreclosure rates. It points to further research needing to be done to understand what other factors caused assumptions that race drove foreclosures.

REFERENCES

Aalbers, Manuel. "Geographies of the financial crisis." *Area* 41, no. 1 (March 2009): 34-42. *Academic Search Premier*, EBSCO*host* (accessed October 5, 2009).

Aalbers, Manuel B. "The Sociology and Geography of Mortgage Markets: Reflections on the Financial Crisis." *International Journal of Urban & Regional Research* 33, no. 2 (June 2009): 281-290. *Academic Search Premier*, EBSCO*host* (accessed October 5, 2009).

Aalbers, Manuel B. "Why the Community Reinvestment Act Cannot be Blamed for the Subprime Crisis." *City & Community* 8, no. 3 (September 2009): 346-350. *Academic Search Premier*, EBSCO*host* (accessed October 5, 2009).

Amromin, G., and A. Paulson. "Comparing patterns of default among prime and subprime mortgages." *Economic Perspectives* 33, no. 2 (April 1, 2009): 18. http://www.proquest.com/ (accessed October 7, 2009).

Ashton, Philip. "An appetite for yield: the anatomy of the subprime mortgage crisis." *Environment & Planning A* 41, no. 6 (June 2009): 1420-1441. *Academic Search Premier*, EBSCO*host* (accessed October 7, 2009).

Daglish, T.. "What motivates a subprime borrower to default?" *Journal of Banking & Finance* 33, no. 4 (April 1, 2009): 681. http://www.proquest.com/ (accessed October 7, 2009).

Edmiston, Kelly D. "Characteristics of High-Foreclosure Neighborhoods in the Tenth District." *Economic Review (01612387)* 94, no. 2 (2009 2nd Quarter 2009): 51-75. *Academic Search Premier*, EBSCO*host* (accessed October 6, 2009).

Emmons, William R., and R Sengupta " What Is Subprime Lending?" *Federal Reserve Bank of St. Louis,* June 2007. http://research.stlouisfed.org/publications/mt/20070601/cover.pdf (accessed December 6, 2009).

Foote, Christopher L., et al. "Reducing Foreclosures." *Research Review* no. 11 (January 2009): 7-10. *Academic Search Premier*, EBSCO*host* (accessed October 5, 2009).

Foote, Christopher L., et al. "Just the facts: An initial analysis of subprime's role in the housing crisis." *Journal of Housing Economics* 17, no. 4 (December 2008): 291-305. *Academic Search Premier*, EBSCO*host* (accessed October 6, 2009).

Foote, Christopher L., Kristopher Gerardi, and Paul S. Willen. "Negative Equity and Foreclosure: Theory and Evidence." *Federal Reserve Bank of Boston,* June 2008. http://www.bos.frb.org/economic/ppdp/2008/ppdp0803.pdf (accessed December 6, 2009).

Gerardi, K., A. Lehnert, S. Sherlund, P. Willen, D. Lucas, and N. Souleles. "Making Sense of the Subprime Crisis/Comments and Discussion." *Brookings Papers on Economic Activity* no. 2 (October 1, 2008): 69-159. http://www.proquest.com/ (accessed October 7, 2009).

Gerardi, Kristopher S., and Paul S. Willen. "Subprime Mortgages, Foreclosures, and Urban Neighborhoods." *Research Review* no. 10 (July 2008): 6-12. *Academic Search Premier*, EBSCO*host* (accessed October 5, 2009).

Haughwout, Andrew, Christopher Mayer, and Joseph Tracy. "Subprime Mortgage Pricing: The Impact of Race, Ethnicity, and Gender on the Cost of Borrowing." *Brookings-Wharton Papers on Urban Affairs* no. 10 (July 2009): 33-56. *Academic Search Premier*, EBSCO*host* (accessed October 5, 2009).

Hernandez, Jesus. "Redlining Revisited: Mortgage Lending Patterns in Sacramento 1930–2004." *International Journal of Urban & Regional Research* 33, no. 2 (June 2009): 291-313. *Academic Search Premier*, EBSCO*host* (accessed October 6, 2009).

Immergluck, Dan. "Core of the Crisis: Deregulation, the Global Savings Glut, and Financial Innovation in the Subprime Debacle." *City & Community* 8, no. 3 (September 2009): 341-345. *Academic Search Premier*, EBSCO*host* (accessed October 6, 2009).

Marcuse, Peter. "A Critical Approach to the Subprime Mortgage Crisis in the United States: Rethinking the Public Sector in Housing." *City & Community* 8, no. 3 (September 2009): 351-356. *Academic Search Premier*, EBSCO*host* (accessed October 6, 2009).

Mills, E.. "Urban Land-Use Controls and the Subprime Mortgage Crisis." The Independent Review 13, no. 4 (April 1, 2009): 559-565. http://www.proquest.com/ (accessed October 7, 2009).

Neale, C.. "Subprime Loans and Bankruptcy: The Memphis Experience Post-BAPCPA." American Bankruptcy Institute Journal 28, no. 5 (June 1, 2009): 50-51,73. http://www.proquest.com/ (accessed October 7, 2009).

Newman, Kathe. "Post-Industrial Widgets: Capital Flows and the Production of the Urban." *International Journal of Urban & Regional Research* 33, no. 2 (June 2009): 314-331. *Academic Search Premier*, EBSCO*host* (accessed October 7, 2009).

Saegert, Susan, Desiree Fields, and Kimberly Libman. "DEFLATING THE DREAM: RADICAL RISK AND THE NEOLIBERALIZATION OF HOMEOWNERSHIP." Journal of Urban Affairs 31, no. 3 (August 2009): 297-317. *Academic Search Premier*, EBSCO*host* (accessed October 5, 2009).

Schuetz, Jenny, Vicki Been, and Ingrid Gould Ellen. "Neighborhood effects of concentrated mortgage foreclosures." *Journal of Housing Economics* 17, no. 4 (December 2008): 306-319. *Academic Search Premier*, EBSCO*host* (accessed October 5, 2009).

Seidenberg, S.. "HOMING   IN   ON   FORECLOSURE." ABA   Journal   94, no. 7 (July 1, 2008): 54-59. http://www.proquest.com/ (accessed October 7, 2009).

Strauss, Kendra. "Accumulation and Dispossession: Lifting the Veil on the Subprime Mortgage Crisis." *Antipode* 41, no. 1 (January 2009): 10-14. *Academic Search Premier*, EBSCO *host* (accessed October 7, 2009).