A REGIONAL EXAMINATION OF FORECLOSURES IN WISCONSIN

Russ Kashian and David M. Welsch
Department of Economics
University of Wisconsin – Whitewater
Whitewater, WI 53190
kashianr@uww.edu, welschd@uww.edu

ABSTRACT

Foreclosures and their causes and remedies, are being discussed and fiercely debated across our nation. Although there has been some examination of the causes of foreclosures, the current research has devoted most of its attention to examining what factors change the probability that an individual will go into foreclosure. By examining Wisconsin counties over eight years this paper makes a contribution to the literature by taking a regional approach to examining the causes of foreclosures. This regional approach has greater policy applications since policy is often based on regional not individual factors. A spatial econometric model along with several other empirical models are estimated and there are consistent results that greater unemployment, a lower median age, larger families in rental units, and a smaller percentage of Asian or Native Americans leads to more foreclosures in a county. There is also evidence that education seems to affect foreclosures in a non-monotonic way; higher percentage of the population with a high school degree vs. non high school graduates increases foreclosures, but there is a negative impact or no significant difference between bachelors degree and no high school degree.

Key Words: Foreclosures, Regional Economics, Housing, Spatial Econometrics

The Industrial Geographer, 2010, Volume 7, Issue 1, Pages 19-38.

INTRODUCTION

Foreclosures in general and the causes and remedies of foreclosures in particular are currently being widely discussed. Most of the research to this point has examined the causes of foreclosures at the individual level, examining what affects an individual's probability of going into foreclosure; there are few, if any, papers that examine foreclosures at a regional level. This paper examines an eight year panel of Wisconsin counties. This regional approach to foreclosures expands beyond the current literature; it offers insight to policymakers and analysts regarding those regions in need and how to address the need. Thus this paper takes a different approach, with possibly greater policy applications, to foreclosure analysis than the historic research on mortgage foreclosure that evaluated what causes individuals to go into foreclosure.

There are regional differences in the rate of foreclosures (Ambrose, et al. 2001), therefore it is important not to ignore (as many studies of individual foreclosures do) local variables that could be important contributing factors. Our paper uses panel data from Wisconsin counties to explore these local variables and their effects on foreclosures. In Wisconsin, quite large differences in the foreclosures for single-family housing exist between counties. This paper differs from the general option-based model; these micro based studies look at individual factors where the risk-of-foreclosure is a function of such things as housing prices and incomes, as well as market-wide housing density and housing price volatility. We modify this model by using regional variables in place of the micro data on the individual homes. The basic descriptive statistics indicate that some of the factors of the option-based model explanations of the variation in foreclosure rates are consistent on an aggregate level. Specifically, price level, income, and employment should explain some of the total regional variation. However, when the structural models are estimated some of these variables do not remain significant. This analysis serves the public interest since government action and policy is often based on regional factors, not individual characteristics.

BACKGROUND SIGNIFICANCE

As much of the early literature is covered in Quercia and Stegman (1992) and Vandell (1995), it is critical to focus on the several studies that are devoted to various determinants of foreclosure and regional variables of foreclosure. The initial papers (Jung, 1962, Page, 1964, Von Furtstenberg, 1969; Von Furtstenberg, 1970a; Von Furtstenberg, 1970b; Von Furtstenberg, 1974) found that there was a relationship between foreclosures and the loan to value ratio, the interest rate, and individual microeconomic (individual loan level data) variables. This attention to microeconomic characteristics continues in many other papers, with the gradual addition of regional characteristics. For example, the issue of location emerges in one early article (Von Furtstenberg, 1974) where a dummy variable was included to differentiate between properties located in Allegheny County and those outside the county. It is also present in Sandor and Sosin (1975) with their inclusion of neighborhood rating system.

Over time, research has expanded as various demographic, family, and regional characteristics have been included in foreclosure papers. These papers use various statistical techniques and sample sizes and have been conducted in regions throughout the United States. Herzog and Earley (1970) examine 12,581 FHA and VA loans, finding that a higher loan to value ratio decreases default and delinquency rates. The applicant's occupation is also included as an independent variable, however, it is not found to have a significant impact on foreclosure rates. There was some relationship between loan delinquency rates and occupation type.

Williams, Beranek, and Kenkel (1974) examine a sample of 1,405 current loans and 125 foreclosed loans in Pittsburgh from 1962 to 1972. They use a dichotomous dependent variable which took on one of two values. These values were unity if the event occurs, and zero otherwise. As with Von Furstenberg (1974), regional characteristics (in the form of the local unemployment rate) were included in the model. The likelihood of default increased with the neighborhood unemployment rate. In addition, it increased if the borrower belonged to certain occupational groups, had a refinanced loan, had a loan to value ratio over 90 percent, had a payment to income ratio of over 30 percent, was party to junior financing, and was in the over-50 age group. The probability of default fell when the property was more expensive, when there was the presence of FHA insurance, and when the tenure of employment with the current employer increased.

Examining 545 loans from 24 banks in Connecticut, Morton (1975) found that those with a higher loan to value ratio and those secured by three-family apartment buildings have increased rates of default and delinquency; family size and borrowers job type were also studied and it was found that some borrowers with job types (such as a salesman) had a higher incidence of foreclosure. In addition, larger family size increased the probability of foreclosure. Vandell (1978) uses a logarithmic regression on a sample of 2,500 FHA loans that originated in 1970-71. This paper finds that a higher loan to value ratio, a longer age of mortgage, and a higher initial mortgage interest rate all increase the default rate.

Jackson and Kasserman (1980) apply an OLS regression to examine FHA loans. They find that higher loan to value ratio, a higher interest rate, and a longer term of mortgage all lead to a higher default probability. Webb (1982) examines a panel of 500 families whose income fluctuated. A wide range of economic and demographic characteristics on the probability of delinquency are also included. Using a Tobit regression, they found that the household head's personal characteristics (sex, race, and age) were not significant in predicting the probability of potential delinquency. Employment type did prove to be significant, with certain groups (such as farmers) having a higher probability of delinquency. There was also an increase in the probability of delinquency when the mortgage payment to income ratio was volatile over time or if the head of household held more than one job. Using a multinomial logit model and a sample of 4899 single family mortgages, Campbell and Dietrich (1983) examined what affects three different dependent variables: probability of default, probability of delinquency, and the probability of prepayment. In the initial regression they find that loans with a lower initial loan to value ratio have a negative coefficient. However, when broken into a variety of subsets (initial loan to value ratios of 80% vs. 85% vs. 90%), they find that the coefficients are not consistent. They attribute this to an adverse selection issue in lending: greater scrutiny is exerted on loans with a smaller down payment. They also find that there is a statistically significant relationship between a higher regional unemployment rate and default.

Vandell and Thibodeau (1985) use a logit regression on a sample of 450 mortgages in Dallas. They find that a higher loan to value ratio and a lower payment to income ratio increase the default rate; their paper also includes a ratio component by including neighborhood rating. Evans, Maris, Weinstein (1985) use regression analysis on a dataset of FHA loans. They find a small loan value, small loan amount, and being white all reduce the default rate; this paper also included some regional components. Zorn and Lea (1989) use a multinomial logit on a Canadian lender loan portfolio from two cities and find that older borrowers and those with larger net equity have a lower default rate. However, larger number of dependents cited on the loan application increases the probability of both delinquency and default.

Using a multinomial logit model, Cunningham and Capone (1990) examined mortgages in Texas from 1982 to 1985 that were terminated. They find that as the mortgage ages in months and the borrower's age at inception increases the probability of foreclosure increases. However the square of this variable is negative, implying that this probability increases at a decreasing rate. They also include the regional unemployment rate and find that a higher employment rate decreases the default rate. Cunningham and Capone acknowledge that the unemployment variables are significant with the incorrect sign. They offer limited explanation for this result, limiting the attribution to underwriting guidelines. Ambrose and Capone (1998, 2000) examine samples of 406,986 and 5650 defaults respectively. They use state unemployment rate in both of these studies. In the 1998 study they find that it has a positive and significant effect on the default rate. Mian and Sufi (2009) examine a basic relationship between high subprime zip codes and defaults, although they do not include regional variables such as unemployment.

With the default rate as his dependent variable, Capozza et al. (1997) examine the effects of three mortgage default trigger events: unemployment, moving rates, and divorce. He looked at the default rate as the percentage of mortgages in default in the region, finding that as unemployment increases, the region experiences higher default rates. A similar result occurs with the region's divorce rate.

Given 51 units of regional analysis (the 50 states and the District of Columbia), Clauretie (1987) uses as dependent variable the log of a state's foreclosure rate to evaluate statewide foreclosure rates. Given 51 units of regional analysis and quarterly data for 10 years, there are 960 observations. Claurite finds that foreclosures rise over time if interest rates rise, thus creating an incentive to opt out of lower interest loans. In addition, he finds that in markets with rising real estate values, the foreclosure rates are smaller and, in one of his regressions, the percentage of foreclosures increases with the unemployment rate.

Baxter and Lauria (2000) also create a regional analysis by employing a sample from New Orleans to focus on the relationship between race and neighborhood transition to foreclosure. By breaking the city up into block groups, mortgage foreclosure data are merged with 1980 and 1990 census data aggregated at the block group level. The final data set contains 4,174 residential mortgage foreclosures sold at judicial auctions between 1985 and 1990. Baxter and Lauria determined, through the use of a structural model that examined the demographic changes in New Orleans neighborhoods, that housing foreclosure is associated with a process of rapid racial residential succession and is high in predominantly black neighborhoods. Foreclosures also rose with unemployment as this foreshadowed racial transition.

These models continue as Merry and Wilson (2006) included 51 regions in their fixed effects regression, which estimated a constant state-specific effect for each state and the District of Columbia. Three years of data (2003-2005) were used in this analysis. Rather than using unemployment in the fixed effects regression, this paper used the growth rate in employment which it found was negative and significant to the dependent variable (share of loans 60+ days past due or in foreclosure). They also found that more recent loans were less likely to be delinquent or in foreclosure and as the share of counties in the state with significant hurricane damage increased, so did the mortgage delinquency rate.

Schuetz, Been, and Ellen (2008) use New York City Data from 2000-2005 to create a hedonic fixed effects model to assess the negative spillover effect foreclosures have on adjacent property sales prices. Using a ordinary least squares regression and data from Akron Ohio, Kaplan and Sommer (2009) find that there is a higher incidence of foreclosure in neighborhoods with higher percentages of minorities. A second positive and significant variable was as neighborhoods have higher rates of new residents, the foreclosure rate rises. Controlling for regional characteristics such as unemployment, the supply of housing, and housing price appreciation, Keifer and Keifer (2009) have found that house prices and foreclosure rates are clearly correlated. Using quarterly state level data for the 48 contiguous U.S. states between 1982 Q2 and 2009 Q2, they found that there is a large and significant relationship between foreclosure rates and house price appreciation. We will use much of the above literature to explain the variable choices; this can be found later in this paper in the section on variable selection.

DATA

Foreclosures are widely discussed but rarely formally defined. This paper uses the traditional definition of foreclosures; a foreclosure is defined as the occurrence of a foreclosure filing. This court filing is a legal procedure in which a mortgagee (or a lien holder) attempts to obtain a court ordered termination of a mortgagor's equitable right of redemption. This redemption would occur as a borrower attempts to bring their delinquent account current.

We must take this one step further and adjust this foreclosure variable. Foreclosures must be adjusted because it is common that debtors, by the time they have arrived at a state of foreclosure, have acquired more than one mortgage, all of which may have the option or opportunity to foreclose. In addition, they may have delinquent property tax bills leading the government to foreclose. For example in Wisconsin one property was foreclosed on eight times in a given year. While a property may be facing two, three, or in rare cases even more legal actions, it is only one property. By reporting these, without correcting for the repeated foreclosure, the impact of the situation is overstated. In effect, only one property is at risk. Those properties foreclosed on multiple times in a single year have been corrected to reflect only one foreclosure. This leads to the use of what we will refer to (and use as a dependent variable in the rest of the paper) as “adjusted foreclosures”.

Our observations are at the county level; data are drawn from 71 counties in the state of Wisconsin over 8 years (2000-2007). Only one county from Wisconsin has been excluded due to their lack of participation in the statewide Circuit Court database. We will examine how the following county characteristics affect the adjusted foreclosure variable discussed above: the unemployment rate in the county, fair market rent (a proxy for home value)[1], log of the number of housing units, log of population density, median age of the county, average household size of owner-occupied units, average household size of renter-occupied units, per capita income (lagged one period),[2] percentage of the population that is black, Native American, Asian, or other race (percentage white is the reference group for these four variables), percentage of the population that is Hispanic (non Hispanic is the reference group for this variable),[3] percentage of the population that has a high school degree but no four year degree, percentage of the population that has a Bachelor degree or higher (percentage of the population that did not obtain a high school degree is the reference group for the previous two groups), and percentage of the population that lives in an urban location (percentage of the population living in rural areas is the reference group). We discuss the variable choices in the next section.

The data come from several sources. Adjusted foreclosures come from Wisconsin Circuit Court Documents, unemployment rate from the United States Bureau of Labor Statistics, number of housing units from the Wisconsin Department of Administration, Demographic Servicers Center, Annual Housing Survey for years following the 2000 Census, per capita income from the Wisconsin Department of Workforce Development, and Population Density comes from Maponics.[4] The remaining variables come from the 2000 Census.

VARIABLE CHOICES

While no study is able to incorporate all of the variables that one would want our study is able to obtain the crucial variables. Our paper includes most variables that were found to be important in previous studies both at the micro and regional level. Below we discuss the variables we include in comparison to previous studies (discussed in the literature review) and we discuss the variables we were not able to include.

First we discuss the variables that we include or proxy for. Many previous studies have found the unemployment rate of the area to be crucial; even many studies that examine this at the micro level have included regional unemployment. Some micro studies have found a more valuable mortgage to be

[1] While rental properties and single family houses are not perfect substitutes, census data for fair market rent is used due to the unavailability of consistent county by county single family home value data throughout Wisconsin.

[2] We are forced to use per capita income lagged one period since data on income from 2007 was not yet available when we collected the data, but it probably does not change our results since if the models are rerun with the years 2000-2006 with current per capita income as an explanatory variable the results change little. These results are available from the authors.

[3] Note that Hispanic is from census data so the reference group is non-Hispanic. That is in our racial descriptors each of the following sum to 100%, (% black)+(%native American)+(%Asian)+(%Other Race)+(%White) and (%Hispanic +%Non-Hispanic)

[4] More information on Maponics data collection service can be found on their website: http://www.maponics.com/index.html

important. We were unable to obtain average home value although we were able to include fair market rent (commonly used as a proxy for housing value). Including the log of the number of housing units allows us to control for the number of housing units in the area. Ideally we would include the number of mortgages in an area since more mortgages would lead to more foreclosures; since we cannot obtain this number we proxy it with number of housing units. Many micro level studies have included occupation. It is impossible for us to obtain this data at the regional level and it would be difficult to interpret the meaning of these even if they were available at the regional level. More appropriate measures that capture similar ideas at the regional level are education levels and per capita income of the area, both of which we include. Many previous studies have included age; we control for median age which will most likely have a slightly different and less precise interpretation at the regional level. Some previous micro level studies have included demographics of the individual holding the mortgage; we include the percentages of the respective racial groups in the area. At least one study has included the family size of those holding the mortgage; we include both the average household size of owner-occupied and renter occupied units.

Our study is able to examine an issue that would be impossible to study with most micro studies. Most micro studies are unable to examine whether foreclosures are more of an issue in urban or rural areas or areas with a more dense population. Most micro studies examine only one area so all of the houses are in either an urban or a rural area or the data does not describe the location of the mortgage; for example Baxtor and Lauria (2000) use data from New Orleans and Schuetz, J., et al. (2008) examine New York City. Our study is able to examine this and includes both variables to see which, if either, has a greater foreclosure problem.

Now we discuss the main variables that were used in previous studies that we were unable to include or even find a suitable proxy for. The key variable included in many micro studies that we were not able to fully include was loan to value ratio, although fair market rent captures part of this variable; this will complicate the interpretation of fair market rent. The other variable omitted from our study that some studies have included and found important is age of the mortgage (although many studies do not include this variable). Other variables included in a few previous studies that we do not control for are: refinancing of loans, interest rate, and prepayment penalties, although these variables were included as controls in relatively few previous studies.

Our study includes many more regional controls than the vast majority of the other studies and we are able to include most of the variables found important in previous studies. It is important to note the major issue with omitting a variable is only to the extent of how correlated it is with the variable we are interpreting.

DESCRIPTIVE STATISTICS

Table 1 shows that Wisconsin counties have great variation in adjusted foreclosures and other county characteristics. The standard deviation of adjusted foreclosures is about twice the mean, showing that there is a wide variation in adjusted foreclosures across counties/time. Table 2 reexamines the means and standard deviations of the independent variables, but it also divides the counties into two groups: those that are above the mean number of adjusted foreclosures per housing units and those that are below the mean number of adjusted foreclosures per housing units.

Two interesting findings are that in counties with a large number of adjusted foreclosures the fair market rent and the per capita income are higher, showing that a larger number of adjusted foreclosures occur in counties where houses are worth more and the residents earn more money. Another interesting finding is that counties with a large number of foreclosures also have a larger percentage of residents that are blacks or Hispanics. An additional finding related to race is that standard deviation is higher for counties with a large number of foreclosures, indicating that counties with a large number of foreclosures differ more from each other in minority make up than counties with few adjusted foreclosures. Yet one must interpret these results with caution, since these results are not from a full structural equation that holds other characteristics of the counties constant.

Table 1: Descriptive Statistics

	N	Mean	Median	Standard Deviation	Min	Max
ADJUSTED FORECLOSURE VARIABLES						
Adjusted foreclosures	568	162.99	78.00	349.21	0.00	4,815.00
Log of Adjusted foreclosures	567	4.39	4.36	1.11	0.69	8.48
Log of 1+ Adjusted foreclosures	568	4.40	4.37	1.10	0	8.48
Adjusted foreclosures as a percentage of housing units	568	0.43	040	0.20	0.00	1.48
INDEPENDENT VARIABLES						
Unemployment Rate	568	5.17	5.00	1.31	2.30	12.40
Fair Market Rent	568	527.67	518.00	105.75	392	951
log of Number of Housing Units	568	9.92	9.89	0.92	7.65	12.93
Log of Population Density	568	4.06	3.82	1.19	2.20	8.26
Median Age	568	38.01	38.00	3.29	27.70	45.80
Average household size of owner-occupied units	568	2.62	2.63	0.14	2.30	3.13
Average household size of renter-occupied units	568	2.15	2.11	0.25	1.74	3.99
Per capita median family income (lagged 1 period)	568	26,884	26156.50	5,300	15,883	56,816
% Black (White is the reference group)	568	1.14	0.28	3.20	0.06	24.59
% Native American	568	2.70	0.42	10.55	0.11	87.26
% Asian	568	0.78	0.33	0.94	0.00	4.54
% Other Race	568	1.55	1.18	1.10	0.53	6.50
% Hispanic	568	1.70	0.95	1.72	0.33	8.77
high school degree but no 4 year degree [ŧ]	568	66.64	67.45	3.64	51.52	72.61
Bachelor degree or higher [ŧ]	568	17.03	15.47	6.16	9.97	40.64
% urban (Rural is the reference group)	568	0.38	0.37	0.29	0.00	1.00

ŧ - no high school degree is the reference group

Table 2: More Characteristics of the Independent Variables by Two County Groupings

	Counties above mean Foreclosures per Housing Units		Counties below mean Foreclosures per Housing Units	
Number of Counties	247		321	
	Mean	SD	Mean	SD
Unemployment Rate	5.27	0.93	5.09	1.53
Fair Market Rent	569.62	106.15	495.40	93.53
log of Number of Housing	40,081.26	66,437.83	29,341.48	42,288.03
Log of Population Density	221.40	642.82	108.17	250.94
Median Age	37.64	2.72	38.30	3.65
Average household size of owner-occupied units	2.63	0.11	2.61	0.16
Average household size of renter-occupied units	2.16	0.11	2.14	0.32
Per capita income (lagged 1 period)	25,932.58	3,851.12	23,885.97	5,767.61
% Black (White is the reference group)	1.78	4.40	0.64	1.63
% Native American	0.86	1.63	4.12	13.80
% Asian	0.75	0.87	0.80	0.98
% Other Race	1.84	1.40	1.32	0.71
% Hispanic (non-Hispanic is reference group)	2.23	2.20	1.29	1.06
% No high school degree	16.86	3.23	15.93	3.95
% high school degree but no 4 year degree	66.94	2.94	66.40	4.09
% Bachelor degree or higher	16.20	4.57	17.67	7.08
percentage urban (Rural is the reference)	42.29	28.31	34.89	29.04

In Figure 1, it is visually apparent that counties on the Wisconsin/Minneapolis border have witnessed higher levels of foreclosures. This is possibly due to a slowdown in the Minneapolis economy that is across the border from Wisconsin Counties, such as St. Croix County, Wisconsin. In addition, the counties on the Wisconsin/Illinois state line are also witnessing high foreclosure rates, relative to the rest of Wisconsin. This may be the impact of the Chicago and Rockford markets. However, while Milwaukee and Racine County see very high levels of foreclosure, Dane County (the second largest county in the State) does not display similar levels. This may be a reflection of some insulation offered Dane County by serving as the home of Wisconsin State Government and the University of Wisconsin-Madison.

Figure 2 examines the total foreclosures in Wisconsin over time. This graph shows that while there were jumps in foreclosures in 2006 and 2007 they were already trending up; in fact these numbers of foreclosures in 2006 and 2007 do not lie even outside a 90% confidence interval line. This time trend further demonstrates why the model should include time dummies to account for this statewide rise in foreclosures.

EMPIRICAL STRATEGIES

To examine what types of variables affect foreclosures, our main specification is a Random Effects Spatial Error Model with 1 + log of adjusted foreclosures as the dependent variable. We also estimate other more basic models to show our model is robust (not sensitive) to model specification; we estimate a basic Random Effects model, as well as a Negative Binomial model with total adjusted foreclosures as the dependent variable, and a Random Effects model with adjusted foreclosures in a county as a percentage of total households as the dependent variable.[5] In the Basic Random Effects model a Breusch-Pagan Lagrange Multiplier test is estimated; the results are found at the bottom of Table 4. The Breusch-Pagan Test overwhelming rejects pooled OLS versus random effect. Note that a fixed-effects model is not estimated because many of the variables we are interested in do not vary over time in our data (since for many variables we use census data) and others vary little over time

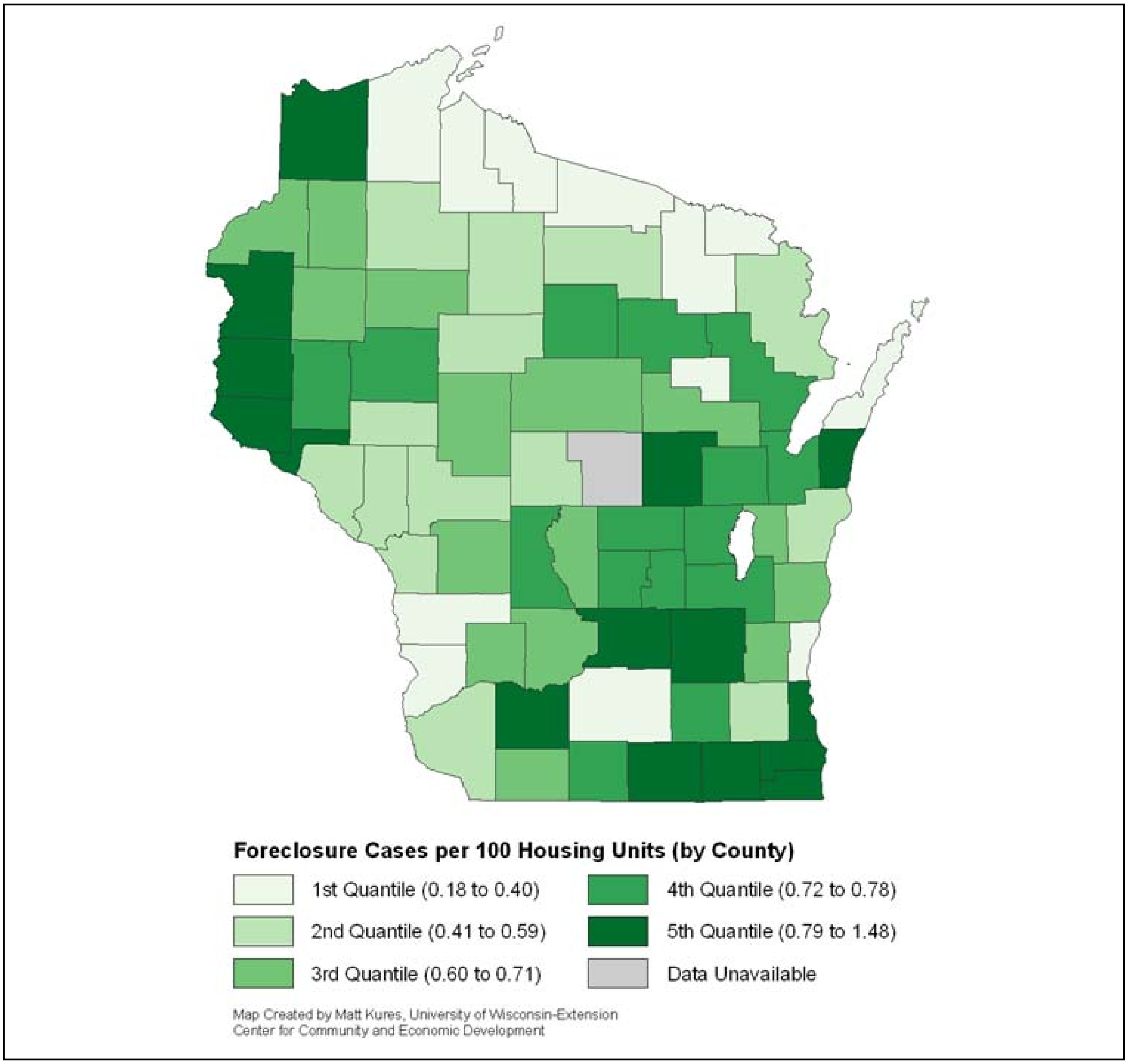
The regression model takes the form:

$$F_{it} = C_{it}'\delta + \mu_i + \varepsilon_{it} \qquad (1)$$

where F_{it} is adjusted foreclosures, which will be estimated with two variables, the first will be the natural log of 1 + adjusted foreclosures (the coefficients [times 100] in this regression may be interpreted as the effect of the independent variable on percentage of adjusted foreclosures) and the second will be the adjusted foreclosures as a percentage of housing units in the county ; C_{it} is a vector of the county characteristics discussed in the data and variable choices sections, time dummy variables (which account for any statewide time trend in foreclosures) and a constant, where some vary over time but most do not, with estimable coefficients δ and μ_i is the unobserved effect, which captures the unobserved time constant factors that affect our dependent variable (it is constant across time), which is the individual county heterogeneity; not

[5] Earlier versions of this paper also included OLS models (clustered on counties) as well as a lagged dependent variable model. Nearly all of the results from this paper are qualitatively the same as the results included in this version of the paper.

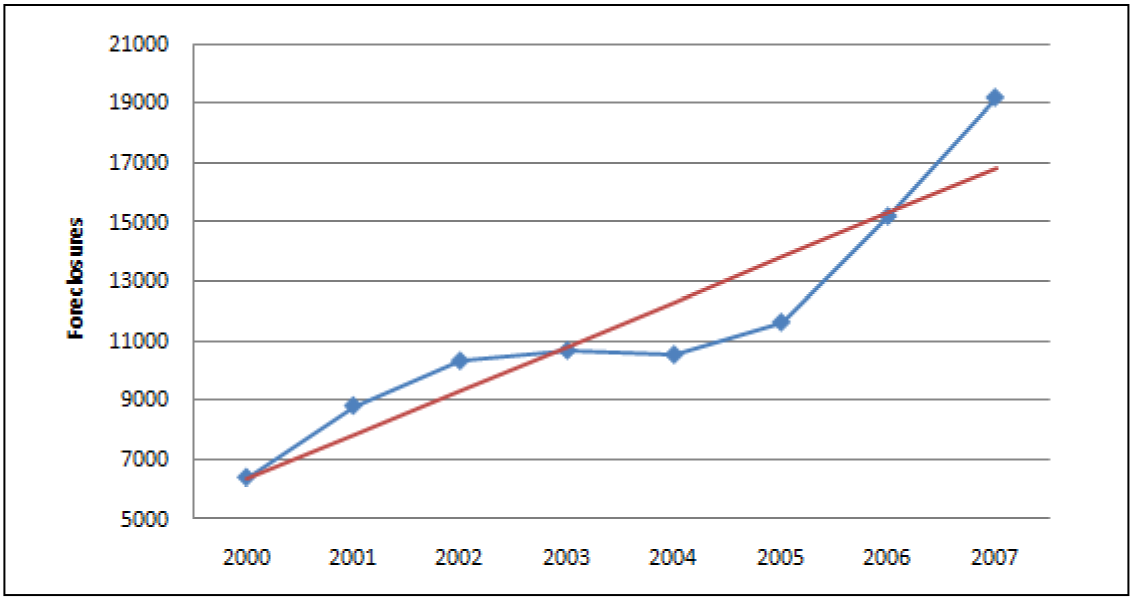
Figure 1: Foreclosure Cases per 100 Housing Units for Wisconsin Counties in 2007



accounting for this county effect could bias the estimates. In a pooled OLS model μ_i would be set equal to zero; for the random effects model to hold, it must be assumed that μ_i is uncorrelated with each explanatory variable. Finally, ε_{it} is the time-varying error or idiosyncratic error; it captures unobserved factors that affect the dependent variable which vary over time. We do not estimate a fixed effects model since many of the variables of interest do not vary over time.

Beyond greater policy applications, examining aggregate data to county level could have econometric benefits as well. It may actually be advantageous to aggregate to the county level; the use of

Figure 2: Total Foreclosures in Wisconsin



average characteristics and total foreclosures probably has fewer errors in measurement than a model examining individual characteristics and individual probability of foreclosures.[6] This aggregation may have one particular disadvantage; this disadvantage is that while (assuming linearity) the estimates of the model are unbiased, they are less precise. Since this study is primarily concerned with the ability to make policy application on the regional level we believe this regional approach leads to a greater ability to apply the results to policy.

When observations are geographic in nature, spatial issues are often a concern. In our model spatial autocorrelation could be an issue. The error term from one county may be correlated with the error term in a neighboring county (conditional on the independent variables); that is, the unobserved heterogeneity in one county may affect the unobserved heterogeneity in a neighboring county.

We start with a brief introduction of the (non panel data) spatial error model (Anselin, 1988 and Lesage, 1998) and then extend this to the random effects spatial error model. That is we assume for the time being that μ_i in equation 1 is equal to zero. For greater readability we remove the subscripts denoting county and year for the remainder of the discussion.

$$
\begin{aligned}
F &= C\delta + \varepsilon \\
\varepsilon &= \lambda W \varepsilon + \upsilon \\
\upsilon &\sim N\left(0, \sigma^2 I_n\right)
\end{aligned}
\qquad (2)
$$

where W is a symmetric spatial weight matrix where $W_{ij} = 1$ for counties that share a border and 0 otherwise (usually this matrix is standardized to have the row sum to one) and λ is the coefficient

[6] For more details on why aggregation may have fewer errors in measurement see Hanushek (1979).

on the spatially correlated errors. For information on how to estimate this model by maximum likelihood see Anselin (1988) and Lesage (1998).

Following Elhorst (2003) and Baltagi and Lee (2004 and 2006) we estimate an extension of this model to random effects (now we eliminate the assumption that μ_i equals zero). This starts with an equation that combines the random effects aspect of equation (1) with the spatial error autocorrelation of equation (2):

$$F = C'\delta + (\iota_T \otimes I_N)\mu + (I_T \otimes B^{-1})\varepsilon \qquad (3)$$

where ι_T is a $(T \times 1)$ vector of ones, I_N is a $(N \times N)$ identity matrix, and $B = I_N - \lambda W$; λ is again called the spatial autocorrelation (autoregressive) coefficient. The result for the spatial autocorrelation coefficient is at the bottom of Table 3 and shows that it is highly significant.[7]

Elhorst shows how this can be used to create a simplified log-likelihood function (see Figure 3). For more details on this log likelihood function, and how to estimate it, see Elhorst (2003). [8]

Note spatial models are often estimated with a neighboring region's dependent variable as an independent variable (a lagged spatial model). Intuitively this model is probably not appropriate for our estimations because foreclosures in one county are most likely not directly dependent on a neighboring county's foreclosure level; it is more likely that the unobserved heterogeneity (unobserved to the researcher) from one county will affect the foreclosure level in a neighboring county. As in most of the other specifications we use log of 1 + adjusted foreclosures as the dependent variable.

RESULTS

Table 3 displays the results for the Random Effects Spatial Error Model with log of 1 + adjusted foreclosures as the dependent variable; coefficients (times 100) should be interpreted at the effect of a one unit change of the independent variable on percentage of foreclosures.[9]

Consistent with expectations, the dummy variables for years shows an increasing trend in the amount of adjusted foreclosures over time and all the results show that there was an increase in foreclosures in 2006 and 2007; also with an increase in the percentage of housing units we would expect the percentage of adjusted foreclosures to increase. This increasing trend in the number of foreclosures may be a less obvious result than at first blanche; first it is important to note that there was an increasing trend in foreclosures prior to the most recent "foreclosures crisis"; second one must remember this increasing trend is once we control for other factors in the economy namely unemployment and housing value. Next the main results of the paper are discussed.

[7] The variance covariance matrix is: $\sigma_\mu^2(\iota_T\iota_T' \otimes I_N) + \sigma_\varepsilon^2(I_T \otimes (BB')^{-1})$

[8] To do this we use a variant of a program first written by J. Paul Elhorst.

[9] If the independent variable is entered in log form then the interpretation of the coefficient is the elasticity.

Figure 3: The Elhorst (2003) Equation

$$\log L = -\frac{NT}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2}\sum_{i=1}^{N}\log\left[1+T\theta^2\left(1-\lambda\omega_i\right)^2\right] + T\sum_{i=1}^{N}\log\left(1-\lambda\omega_i\right) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}e_t{}'e_t$$

Where ω_i are the characteristic roots of W, $e_t = F_t^* - C_t^*\delta$, $F_t^* = P\overline{F} + B\left(F_t - \overline{F}\right)$, $C_t^* = \left(I_N - \lambda W\right)C_t - \left(P - \left(I_N - \lambda W\right)\right)\overline{C}$, and P is such that $P'P = \left[T\theta^2 I_N + \left(B'B\right)^{-1}\right]^{-1}$.

Table 3: Spatial Model with Log of Foreclosures as the Dependent Variable

	RE with Spatial Autocorrelation	
	Coef.	S.E.
Unemployment Rate	0.079***	0.022
Fair Market Rent	2.00E-05	3.03E-04
Log of Number of Housing Units	1.082**	0.064
Log of Population Density	0.069	0.079
Median Age	-0.051***	0.018
Average household size of owner-occupied units	-0.258	0.388
Average household size of renter-occupied units	0.605**	0.244
Per capita income (lagged 1 period)	1.50E-05*	8.60E-06
% Black (White is the reference)	-0.008	0.013
% Native American	-0.025***	0.006
% Asian	-0.089**	0.037
% Other Race	0.004	0.088
% Hispanic	0.029	0.057
High School Degree but no 4 year Degree ŧ	0.030**	0.014
Bachelor Degree or Higher ŧ	-0.020*	0.011
Percentage Urban (Rural is the reference)	0.020	0.191
Year 2001 (2000 is the reference) ¥	0.228***	0.051
Year 2002	0.318***	0.065
Year 2003	0.287***	0.072
Year 2004	0.327***	0.070
Year 2005	0.412***	0.074
Year 2006	0.635***	0.080
Year 2007	0.767***	0.093
Constant	-8.082***	2.276
R-square	0.99	
n	568	
Sigma	0.44	
Spatial Autocorrelation Coef.	0.228***	0.075

S.E. is the standard error
*,**,***: significant at the 10,5, and 1% level, respectively.
ŧ : no high school degree is the reference group

Consistent with most previous models of foreclosures, a higher unemployment rate is associated with a larger number of adjusted foreclosures (higher unemployment leads to more foreclosures); not only is unemployment highly statistically significant it is also large in "practical significance". Specifically if a county has a 1 percentage point increase in unemployment we would expect foreclosures to increase by approximately 7.9%. Even if we take the conservative estimate of 4% (approximately the lower end of a 90% confidence interval) this effect is reasonably large. Suppose there are two counties that are exactly the same except one has an unemployment rate of 5% (approximately the mean) and the other has an unemployment rate that is 6.3% (approximately 1 standard deviation above the mean); since the second county's unemployment is 1.3 percentage points higher, this would mean that the second county would have 5.2% more adjusted foreclosures.

We find no evidence that fair market rent (the proxy for housing value), population density, percentage urban, or housing size of owner occupied units affects the number of foreclosures, indicating that once you control for other factors the fluctuation in housing values between places will not affect the number of foreclosures. The insignificance of population density and percentage urban indicates that foreclosures are most likely not only an urban nor are they only a rural issue, but a statewide (countrywide) concern.

Areas with an older population have fewer foreclosures. This could be an indication of many things: older individuals have had time to pay down their mortgages and have more equity in their homes, or the existence of a cultural difference between older and younger individuals. This is consistent with Zorn and Lea (1989) and Cunningham and Capone (1990) who found that the age of the borrower produced a negative and significant coefficient on the foreclosure rate regression.

Also of interest is that as the size of households in renter occupied units increases adjusted foreclosures increase; this variable is significant in all specifications. One possible explanation for this is that landlords that rent to large families have an increased probability of foreclosures due to possibly higher costs and or lower returns associated with renting to larger families. This result is also consistent with previous studies; Herzog and Early (1970) found that very large families (eight or more dependents) yielded high risk coefficients in all three samples, and two of these were significantly greater than zero. This is consistent with Morton (1975) who found similar results.

Although per capita income is positive and significant in this specification, its practical significance is small. It would take a $100,000 difference in per capita income between two counties to yield even a 1.5% change in the number of foreclosures; also this variable is not significant in any of the other specifications.

The racial results yield some important findings. We find no evidence that areas with a larger percentage of Black individuals or Hispanic individuals (versus whites) will have a larger percentage of foreclosures. Counties with a larger percentage of Native Americans or Asians have less adjusted foreclosures (versus percentage of county that is white).

When taken together the education variables suggest that education has a non-monotonic effect on foreclosures. Counties with a larger percentage of individuals with high school degrees versus individuals that have not obtained a high school degree have a greater percentage of foreclosures, but the results show that there is either a negative effect or no statistical difference between a county with a larger percentage of college graduates versus no high school degree. This implies more education leads to greater foreclosures up to a point, but when education increases even more foreclosures may begin to decrease. One possible explanation is that people with high school degrees (but not a four year degree) are more likely to have access to loans than those without high school degrees, but are also more likely not to be able to handle homeownership as well as people with bachelor degrees.

ROBUSTNESS CHECKS

Table 4 estimates three other models to see if our results are sensitive to model specification. We show the results of non-spatial models; the first is a random effects model with log of adjusted foreclosures as the dependent variable, the second is a negative binomial model with adjusted foreclosures as the dependent variable, and the third is a random effects model with adjusted foreclosures as a percentage of housing units as the dependent variable. In all of these the standard errors are clustered on counties.[10]

The interpretation of the size of the coefficients from the negative binomial results are similar to the log dependent variable model since they are approximately what the proportionate change (or percentage change if the coefficient is multiplied by 100) in the dependent variable is for a one unit change in the independent variable (if the independent variable is in level not log form).[11] The interpretation of the model with Foreclosures as a percentage of housing units is slightly different; the interpretation of the estimated coefficient is the effect of a change in the independent variables on the percentage *points* of foreclosures (as a percentage of housing units). In most cases the results are qualitatively the same as our primary model and many cases they are quantitatively very similar as well.

CONCLUSIONS

Discovering the factors that lead to a change in foreclosures has important policy implications. This paper contributes to the literature by being the first to examine a full structural model of foreclosures at a regional level. To summarize our key findings: a higher unemployment rate or larger size of households in renter occupied units leads to more foreclosures, counties with larger populations of Native Americans or Asians and higher median age have fewer foreclosures, and education appears to affect foreclosures in a non-monotonic way as percentage of the population with a high school degree versus non high school graduates increases foreclosures, but there is a negative impact or no significant difference between bachelors

[10] The "R-square" reported for the Random Effects Models in Table 4 is in quotes because it is not the typical OLS R^2 and does not have all of the properties of the OLS R^2. Rather it is a correlation squared or a R^2 from a second round regression.

[11] Technically it measures how much the difference in logs of the expected counts changes (or the log of the ratio of counts) for a one unit change in the independent variable, which should be approximately the proportion, if the proportion is small.

Table 4: Alternative Specifications

	Log of Foreclosures		Negative Binomial		Percentage of Housing Units	
	RE				RE	
	Coef.	S.E.	Coef.	S.E.	Coef.	S.E.
Unemployment Rate	0.082***	0.030	0.078***	0.021	0.029**	0.014
Fair Market Rent	-1.70E-04	2.42E-04	8.08E-04***	3.00E-04	-1.37E-04	1.87E-04
Log of Number of Housing Units	1.095***	0.070	1.112***	0.044	0.079*	0.045
Log of Population Density	0.061	0.092	0.031	0.062	0.002	0.048
Median Age	-0.053**	0.022	-0.049***	0.014	-0.012	0.008
Average household size of owner-occupied units	-0.224	0.517	-0.290	0.299	0.169	0.240
Average household size of renter-occupied units	0.624***	0.206	0.612***	0.167	0.314***	0.098
Per capita income (lagged 1 period)	0.000	0.000	0.000	0.000	0.000	0.000
% Black (White is the reference)	-0.007	0.012	-0.009	0.008	0.001	0.006
% Native American	-0.025***	0.005	-0.027***	0.005	-0.012***	0.003
% Asian	-0.093**	0.046	-0.077**	0.032	-0.056**	0.023
% Other Race	0.004	0.094	0.059	0.068	0.025	0.036
% Hispanic	0.032	0.055	-0.010	0.043	0.004	0.022
high school deg. but no 4 year deg. ŧ	0.032**	0.016	0.028***	0.010	0.019**	0.008
Bachelor degree or higher ŧ	-0.018	0.015	-0.025***	0.009	0.003	0.007
Percentage urban (Rural is the reference)	-0.007	0.200	-0.042	0.159	-0.010	0.095
Year 2001 (2000 is the reference) ¥	0.227***	0.047	0.219***	0.036	0.066***	0.018
Year 2002	0.320***	0.079	0.302***	0.059	0.111***	0.032
Year 2003	0.304**	0.120	0.315***	0.068	0.127***	0.034
Year 2004	0.341***	0.084	0.266***	0.065	0.131***	0.032
Year 2005	0.434***	0.091	0.347***	0.071	0.179***	0.034
Year 2006	0.660***	0.102	0.566***	0.082	0.302***	0.041
Year 2007	0.795***	0.122	0.674***	0.096	0.411***	0.052
Constant	-8.354***	2.998	-8.362***	1.665	-2.485*	1.438
"R-square"	0.94				0.67	
n	568		568		568	
Breusch-Pagan LM (p-value)	221.3 (0.00)				254.4 (0.00)	

S.E. is the heteroskedasticity-robust standard error clustered on counties
*,**,***: significant at the 10,5, and 1% level, respectively.
ŧ : no high school degree is the reference group

degree and no high school degree. The above results appear to be robust since they are consistent across a multitude of specifications.

If policy makers want to address the foreclosure issue, evidence from this paper suggests they should address unemployment, create policies to reduce the size of households in rental units, and increase college education. Practical advice for lenders is that they should investigate a strategy of increasing mortgages in areas (assuming a constant interest rate) where the median age is higher and the percentage of Native Americans or Asians is larger; these areas may present an arbitrage opportunity for lenders.

This study could aid policy makers in which areas to target for assistance. For example the results from fair market rent indicate that we do not need to target areas with higher (or lower) housing values for help. There is also indication that this is not an urban or a rural issue, and that areas with a large number of high school graduates but few college graduates could use more assistance.

REFERENCES

Ambrose, B. W., C.A. Capone, and Y. Deng. 2001. "Optimal Put Exercise: An Empirical Examination of Conditions for Mortgage Foreclosure." *Journal of Real Estate Finance and Economics* 23 No 2: 213-234.

Ambrose, B. and C. Capone. 1998. "Modeling the Conditional Probability of Foreclosure in the Context of Single Family Mortgage Default Resolutions." *Real Estate Economics* 26 No 3: 391-429.

Ambrose, B. and C. Capone. 2000. "The Hazard Rates of First and Second Defaults" Journal of *Real Estate Finance and Economics* 20: 275-293.

Anselin, L. 1988. *Spatial econometrics: Methods and models*. Dordrecht, the Netherlands: Kluwer.

Baltagi, B. H. and D. Lee 2004. "Prediction in the panel data model with spatial correlation." In *Advances in spatial econometrics: Methodology, tools, and applications,* edited by L. Anselin, R. J. G. M. Florax, and S. J. Rey. Heidelberg: Springer.

Baltagi, B. H. and D. Lee 2006. "Prediction in the Panel Data Model with Spatial Correlation: the Case of Liquor." *Spatial Economic Analysis* 1, No. 2: 175-185.

Baxter, V. and M. Lauria. 2000. "Residential Mortgage Foreclosure and Neighborhood Change." Housing Policy Debate 11 No 3: 675-699.

Campbell T.S. and J.K. Dietrich. 1983. "The Determinants of Default on Insured Conventional Residential Mortgage Loans." *The Journal of Finance* 38, No 5: 459-477.

Capozza, D.R., D. Kazarian, and T. A. Thomson. 1997. "Mortgage Default in Local Markets." *Real Estate Economics* 25 No 4: 631-656.

Clauretie, T. M. 1987. "The Impact of Interstate Foreclosure Cost Differences and the Value of Mortgages on Default Rates." *AREUEA Journal 15: 152-167.*

Cunningham, D.F. and C.A. Capone, Jr. 1990. "The Relative Termination Experience of Adjustable Fixed rate Mortgages". Journal of Finance 45 No5: 1687-1703

Elhorst, J. P. 2003. "Specification and Estimation of Spatial Panel Data Models." *International Regional Science Review* 26, No.3: 244-268.

Evans, R. D., B. A. Maris, R. I. Weinstein. 1985. "Expected loss and mortgage default risk."
Quarterly Journal of Business and Economics 24: 75–92.

Hanushek, Eric, A. 1979. Conceptual and Empirical Issues in the Estimation of Educational Production Functions. *The Journal of Human Resources.* 14, no. 3: 351-388.

Herzog P. and J. S. Earley. 1970. "Home Mortgage Delinquency and Foreclosure." National Bureau of Economic Research, ISBN: 0-87014-206-2.

Jackson, J. R. and D. L .Kaserman. 1980. " Default Risk on Home Mortgage Loans: A Test of

Competing Hypotheses." The Journal of Risk and Insurance 47: 678-690

Jung, A.F. 1962. "Terms on Conventional Mortgage Loans on Existing Houses." *Journal of Finance* 17 No 3: 432-443.

Kaplan, D. H. and Sommers, G. G. 2009. "'An Analysis of the Relationship Between Housing
Foreclosures, Lending Practices, and Neighborhood Ecology: Evidence from a Distressed County," The Professional Geographer 61 No1: 101 — 120.

Keifer, H. and L. C. Keifer. 2009. "The Co-Movement of Mortgage Foreclosure Rate and House Price Depreciation: A Spatial Simultaneous Equation System". Unpublished Manuscript. Office of the Comptroller of the Currency. Washington.

Lesage, J. P. 1998. *Spatial Econometrics*. Unpublished manuscript University of Toledo.

Merry, E. A. and M. D. Wilson. 2006. "The Geography of Mortgage Delinquency." Federal Reserve Board of Governors, Washington, DC. Unpublished Manuscript.

Mian, A. and A. Sufi. 2009. "The Consequences of Mortgage Credit Expansion: Evidence From The US Mortgage Default Crisis." *The Quarterly Journal of Economics* 124: 1449-1496.

Morton, T. G. 1975. "A Discriminate Function Analysis of Residential Mortgage Delinquency and Foreclosure." AREUEA Journal 3 No1: 73-90

Page, A. N. 1964. " The Variation of Mortgage Interest Rates." The Journal of Business 37: 280-294

Quercia. R.G. and M.A. Stegman. 1992. "Residential Mortgage Default: A Review of the Literature." *Journal of Housing Research* 3 No 2: 341-379.

Sandor, R. L. and H. B. Sosin. 1975. "The Determinants of Mortgage Risk Premiums: A Case Study of the Portfolio of a Savings and Loan Association." The Journal of Business 48: 27-38

Schuetz, J.. V Been and I. G. Ellen. 2008. "Neighborhood Effects on Concentrated Mortgage Foreclosures." Journal of Housing Economics 17: 306-319

Vandell, K.D. 1978. "Default Risk Under Alternative Mortgage Instruments The Journal of Finance." 33: 1279-1296

Vandell, K.D. 1995. "How Ruthless is Mortgage Default? A Review and Synthesis of the Evidence." Journal of Housing Research 6 No 2: 245-264

Vandell, K. D. and T. Thibodeau. 1985. "Estimation of Mortgage Defaults Using Disaggregate Loan History Data." AREUEA Journal 13: 292-316.

Von Furtstenberg, G. 1969. "Default Risk on FHA-Insured Home Mortgages as a Function of the Terms of Financing: A Quantitative Analysis." The Journal of Finance 24, No 3: 459-477.

Von Furtstenberg, G. 1970a. "Risk Structures and the Distribution of Benefits within the FHA Home Mortgage Insurance Program." The Journal of Money Credit and Banking 2, No 3: 303-322.

Von Furtstenberg, G. 1970b. . "Home Mortgage Delinquencies: A Cohort Analysis." *The Journal of Risk and Insurance* 37, No 3: 437-445.

Von Furtstenberg, G. 1974. "Home Mortgage Delinquencies: A Cohort Analysis." *The Journal of Finance* 29, No 5: 1545-1548.

Webb, B. G. 1982. " Borrower Risk Under Alternative Mortgage Instruments." The Journal of Finance 37: 169-183

Williams, A.O., W. Beranek, and J. L. Kenkel. 1974. "Default Risk in Urban Mortgages: A Pittsburgh Prototype Analysis." *American Real Estate and Urban Economics Association Journal* 2 No 2: 101-112.

Zorn, P. and M. Lea. 1989. "Mortgage Borrower Repayment Behavior: A Microeconomic Analysis with Canadian Adjustable Rate Mortgage Data." AREUA Journal 17 No1: 118-13