

Three Essays in House Foreclosures, Quasi-Experiment Model and Irrigation

by

Lina Cui

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
August 4, 2012

Keywords: housing sales price, foreclosure, REO, REO sales, irrigation, income inequality

Copyright 2012 by Lina Cui

Approved by

Diane Hite, Chair, Professor of Agricultural Economics and Rural Sociology
Joseph J. Molnar, Co-chair, Professor of Agricultural Economics and Rural Sociology
Valentina Hartarska, Associate Professor of Agricultural Economics and Rural Sociology
Denis Nadolnyak, Assistant Professor of Agricultural Economics and Rural Sociology

Abstract

In recent years much governmental and public attention has been focused on house foreclosures as they related to the recent recession. Housing spillovers can degrade neighborhood quality and depress property tax revenues, which are an important source of funding of local public goods such as public schools. It is my aim to study these spillover effects and to assess how the timing and number of foreclosures affect surrounding house values, and ultimately erode the tax base in Atlanta, Georgia.

Chapter 1 examines the effect of foreclosures on subsequent home sales prices by employing a general spatial model and a generalized spatial two stage least squares (GS2SLS) model. Potential endogeneity of foreclosures is also explored using spatial system models.

Chapter 2 employs difference-in-differences (DID) model and propensity score matching (PSM) to study the effect of foreclosures on neighborhood property values in the city of Atlanta from 2000-2010. A difference-in-differences model not only removes biases from comparisons between the treatment and control group that could be the result from systematic differences, but also removes biases from comparisons over time in the treatment group that could be the result of trends. Like difference-in-differences method, propensity score matching (PSM) removes the differences between treatment and control groups by matching treatment and control units based on a set of covariates.

Chapter 3 examines the impact of irrigation adoption on farmers' cropping income and the total profit of agricultural products sold. It also examines income inequality using agriculture

products sales value. This paper is the first attempt to use U.S. county level data to examine the 9 Southeast states' irrigation impacts. Irrigation is often promoted as a technology that can increase crop production, improve agriculture income and alleviate poverty. However, irrigation is a relatively expensive technology for small-scale and poor farmers, which impedes their opportunities to adopt irrigation technology. Income inequality may increase due to adoption barriers.

Acknowledgements

The year 2006 was my turning point in life. In the summer of that year I came to Auburn, it is the place where I started my graduate study, where I met my husband, Tong Liu, where my children, Lawrence and Caroline were born. I appreciate all of the great gifts from god and all of the great people I have ever met. First, I would thank Dr. Joseph J. Molnar, Dr. Conner Bailey, and Dr. Henry Thompson. Without them, I would not have had a chance to study at Auburn and would not have had a chance to learn quantitative research methods. Without Dr. Molnar and Dr. Thompson, I might have not even started my PhD study. Second, I would like to thank Dr. Diane Hite, who provided me with many good research ideas and insightful comments. I also appreciate the opportunities she gave to me for oral presentations at conferences and teaching at Auburn. Without her, I could not bring my work to successful completion. I also like to thank my committee members, Dr. Valentina Hartarska, Dr. Denis Nadolnyak, Dr. Henry Kinnucan (who was initially on my committee) and Dr. Tannista Banerjee for their invaluable comments and all the help they provided. Last but not least, I thank my great parents and mother-in-law for their unconditional support and love during the past six years.

Table of Contents

Abstract.....	ii
Acknowledgements.....	iii
List of Tables	ii
List of Figures	ii
List of Abbreviations	ii
CHAPTER 1	1
CHAPTER 2	3
The Effect of House Foreclosures on Neighborhood Property Value	3
1. Introduction	3
2. Literature Review	5
3. Data.....	8
4. Model.....	11
4.1. Hedonic Model.....	11
4.2. Spatial Hedonic Model	12
4.3. Endogeneity Testing	15
4.4. Zero-Inflated Negative Binomial (ZINB) Model	15
5. Results	19
5.1. Effects of Characteristics on Number of Foreclosures	28
5.2. Foreclosure Effects on House Values	35

6. Tax Loss Estimates	43
7. Conclusion and Policy Implication.....	45
CHAPTER 3	48
The Contagion Effect of Foreclosures: A Quasi-Experiment Method	48
1. Introduction	48
2. Literature Review	50
3. Data.....	53
4. Model.....	55
4.1. Difference-In-Differences (DID).....	55
4.2. Propensity Score Matching (PSM)	57
5. Results	59
5.1. OLS Regression	66
5.2. Difference-In-Differences.....	72
5.3. Propensity Score Matching	75
6. Conclusion.....	77
CHAPTER 4	80
Irrigation and Income Inequality in the Southeast United States	80
1. Introduction	80
2. Literature Review	82
3. Data.....	84
4. Model.....	86
5. Results	89
6. Discussion and Conclusion.....	96

CHAPTER 5	97
REFERENCES	100

List of Tables

Table 2.1 Descriptive Statistics for One to Four Unit Family Sales House Characteristics and Neighborhood Characteristics, Atlanta, 2008 (N=10,121)	21
Table 2.2 Descriptive Statistics for Dummy Variables, Atlanta, 2008 (N=10,121).....	25
Table 2.3 Zero-Inflated Negative Binomial Regression Analysis for Factors Affecting Foreclosures	30
Table 2.4 Regression Coefficients for Heteroskedasticity – Corrected OLS, Spatial Autoregressive Model, General Spatial Model and GS2SLS Model	39
Table 3.1 Descriptive Statistics for One to Four Unit Family Sales House Characteristics and Neighborhood Characteristic, Atlanta, 2000-2010 (N=26,352)	61
Table 3.2 Descriptive Statistics for Dummy Variables, Atlanta, 2000-2010 (N=26.352).....	64
Table 3.3 The Effect of Foreclosures within Different Buffers, Regression Coefficients for Heteroskedasticity – Corrected OLS	68
Table 3.4 The Effect of REO and REO sales on Neighborhood Property Sales Value within Different Buffers.....	71
Table 3.5 The Effect of Foreclosures within Different Buffers, Difference-In-Differences Model	73
Table 3.6 Baseline Characteristics (Treatment: DIS300>0)	76
Table 3.7 Propensity Score Matching, Caliper ($1 * E^{-4}$) method (Treatment: DIS300>0).....	77
Table 4.1 Descriptive Statistics for Study Variables, 9 Southeast States, 1997 and 2002 (N=568)	90
Table 4.2 First Stage OLS Regression Results for Irrigation	91
Table 4.3 OLS Regression Results for Determinants of Income.....	93
Table 4.4 Two Stage Least Squares (2SLS) Regression Results for Determinants of Income	94

Table 4.5 Regression Results for Farm Sale Value Inequality 95

List of Figures

Figure 2.1 Historic Quarterly Home Sales Price Index, Atlanta MSA, Seasonally Adjusted (Data Source: Federal Housing Finance Agency)	17
Figure 2.2 Spatial Relationship between Foreclosures and Percent Black Residence, Atlanta, 2008.....	32
Figure 2.3 Spatial Relationship between Foreclosures and Per Capita Income, Atlanta, 2008....	33
Figure 2.4 Spatial Relationship between Foreclosures and Percent Owner Occupied Homes, Atlanta, 2008.....	34
Figure 2.5 Spatial Relationship between Property Tax Loss and Per Capita Income, Atlanta, 2008	44
Figure 3.1 Historic Quarterly Home Sales Price Index, Atlanta MSA, Seasonally Adjusted (Data Source: Federal Housing Finance Agency)	48
Figure 4.1 Lorenz Curve	81

List of Abbreviations

GS2SLS	Generalized Spatial Two State Least Squares
OLS	Ordinary Least Squares
GMM	Generalized Method of Moments
ML	Maximum Likelihood
MCMC	Markov Chain Monte Carlo
HUD	Housing and Urban Development
ZINB	Zero-Inflated Negative Binomial
NB	Negative Binomial
ZIP	Zero-Inflated Poisson
HOPEA	Home Ownership and Equity Protection Act
LTV	Loan-To-Value
HMDA	Home Mortgage Disclosure Act
HSI	Hazardous Site Inventory
LTV	Loan-To-Value
GEPD	Georgia Environmental Protection Division
CBG	Census Block Group
MSA	Metropolitan Statistical Area
AIC	Akaike Information Criterion
REO	Real Estate Owned

DID	Difference-In-Differences
PSM	Propensity Score Matching
CPI	Consumer Price Index
2SLS	Two Stage Least Squares
CCC	Commodity Credit Corporation
USGS	U.S. Geological Survey
LSDV	Least Squares Dummy Variable

CHAPTER 1

In recent years much governmental and public attention has been focused on house foreclosures as they related to the recent recession. Housing spillovers can degrade neighborhood quality and depress property tax revenues, which are an important source of funding of local public goods such as public schools. It is my aim to study these spillover effects and to assess how the timing and number of foreclosures affect surrounding house values, and ultimately erode the tax base in Atlanta, Georgia.

Chapter 2 and Chapter 3 employ the same dataset to study the effects of house foreclosures. Chapter 2 uses cross-sectional data to examine the effect of foreclosures on subsequent home sales prices in 2008 by employing spatial models. Chapter 3 uses panel data (2000-2010) to study the effect of foreclosures on neighborhood property values in the city of Atlanta. The quasi-experiment methods, difference-in-differences (DID) model and propensity score matching (PSM) are employed and compared. A difference-in-differences model not only removes biases from comparisons between the treatment and control groups that could be the result from systematic differences, but also removes biases from comparisons over time in the treatment group that could be the result of trends. Like difference-in-differences method, propensity score matching removes the differences between treatment and control groups by matching treatment and control units based on a set of covariates.

Compared to the cross-sectional data, panel data has an advantage to reduce omitted variable problems by subtracting constant unobserved variables. However, using cross-sectional

data, spatial models also help avoid omitted variable problems by controlling spatially correlated housing prices and spatially correlated errors.

Chapter 3 examines the impact of irrigation adoption on farmers' cropping income, agricultural income and the total profit of agricultural products sold. It also examines income inequality using agriculture products sales value. This paper is the first attempt to use U.S. county level data to examine the 9 Southeast states' irrigation impacts. Irrigation is often promoted as a technology that can increase crop production, improve agriculture income and alleviate poverty. However, irrigation is a relatively expensive technology for small-scale and poor farmers, which impedes their opportunities to adopt irrigation technology. Income inequality may increase due to adoption barriers. Thus, irrigation is suspected endogenous to farmers' cropping income.

In this dissertation, addressing endogeneity problem is one interest. The endogeneity problems in Chapter 1 and Chapter 3 are caused by reverse causality. Because neighborhood house values depreciated by foreclosures may lead to more foreclosures, foreclosures may thus be endogenous to the sales price. Previous studies argue that it is hard to find an instrumental variable which is correlated with foreclosures but not correlated with the residuals of the hedonic price equation. The contributions of Chapter 1 include creating an innovative way to examine endogeneity through accounting for foreclosure timing and it also addresses the endogeneity of the spatially lagged dependent variable by using GS2SLS procedures. Chapter 3 deals with endogeneity with 2SLS regression. Because irrigation is a relatively expensive technology for small-scale farmers and poor farmers, it impedes their opportunities to adopt irrigation technology. Thus, irrigation is potentially endogenous to agricultural sales income.

CHAPTER 2

The Effect of House Foreclosures on Neighborhood Property Value

1. Introduction

In recent years much governmental and public attention has been focused on house foreclosures as they related to the recent recession. Housing price spillovers can degrade neighborhood quality and depress property tax revenues, which are an important source of funding of local public goods such as public schools. It is my aim to study these spillover effects and to assess how the timing and number of foreclosures affect surrounding house values, and ultimately erodes the tax base in Atlanta, Georgia.

Atlanta typifies the new South, and Georgia ranks as the state with the fourth highest default rate in the nation (RealtyTrac, 2011). By the end of 2006, the state foreclosure rate reached 2.7%, which was up from 1.1% in 2000. Foreclosures in Georgia are concentrated in Fulton County and DeKalb County, which are part of the same metropolitan area as the capital city, Atlanta. One of the reasons that the number of foreclosures is high in Georgia is that Georgia law permits creditors to foreclose on homes more quickly than in other states¹. Lenders can declare a borrower in default and reclaim a house in as few as 60 days (Bajaj, 2007), thus increasing the speed with which foreclosures occur. Individuals who file Chapter 13 bankruptcy due to foreclosures in Georgia do so at three times the national rate (Uchoa, 2008).

Foreclosures not only deteriorate the appeal of neighborhood and lead to disordered communities, they also depress their own and surrounding house values (Immergluck and Smith,

¹ “In New York State, the time between the filing of a *lis pendens* and the auction of the property is typically about 18 months” (Schuetz et al., 2008). In Ohio, it usually takes 150-180 days to foreclose a property (foreclosure.com).

2006; Leonard and Murdoch, 2009; Lin et al., 2009; Schuetz et al., 2008; Skogan, 1990; Towe and Lawley, 2010). This is because house prices are usually set by comparables in the neighborhood, so there is a direct price effect; further there may well be a negative spillover effect if neighborhood is reduced by foreclosures. In addition, foreclosures increase housing supply, so if demand remains the same or decreases, prices will be driven down further. Cities, counties, and school districts thus may lose tax revenue due to foreclosed homes (Immergluck and Smith, 2006). The fact that more foreclosures occur in poor neighborhoods exacerbates inequality in school quality.

Since the housing market is idiosyncratic due to differences in socio-economic characteristics and state laws on foreclosures, research on a specific market may provide information that is more valuable to local policy makers. We examine foreclosure effects in Atlanta by incorporating spatial effects and foreclosure timing. Generalized spatial two stage least squares (GS2SLS) procedures are used to examine foreclosure effects. GS2SLS is an advanced methodology that addresses endogeneity of a spatially lagged dependent variable when the model contains both spatial lags in the endogenous variables and spatial autocorrelation in the disturbances. Results from ordinary least square (OLS) regression, spatial autoregressive regression and general spatial regression are also analyzed and compared.

In addition, neighborhood house values depreciated by foreclosures may lead to more foreclosures, thus signaling potential endogeneity. We thus examine the endogeneity problem with a systems approach that incorporates surrounding house characteristics, neighborhood characteristics and loan characteristics. Because no evidence of endogeneity was found in the systems model, we use parcel level data to explore how surrounding house characteristics, neighborhood characteristics and loan characteristics affect foreclosures within a certain buffer

using a zero-inflated negative binomial (ZINB) regression, thus furthering the foreclosure literature.

2. Literature Review

There are a few recent studies addressing foreclosure effects using the hedonic price model. Nearly all the research focuses on metropolitan areas and uses publicly available datasets to study foreclosure effects in some specific time period(s). However, most of the existing literature is based on data from before the subprime crisis, and no empirical analysis using complete parcel information has been conducted for Atlanta. As an African-American dominated city, Atlanta ranks fairly low among southern cities in the dissimilarity measure of segregation; in addition, Atlanta's local government is well represented by African Americans. Since the housing market is idiosyncratic due to differences in socio-economic characteristics and state laws on foreclosures, research on a specific market may provide more valuable information for local policy makers.

Immergluck and Smith (2006) combine foreclosure data from 1997 and 1998 with neighborhood characteristics data and more than 9,600 single family property transactions in Chicago in 1999. After controlling for 40 property and neighborhood characteristics, they find that foreclosures of conventional single-family loans have a significant impact on nearby property values. Each conventional foreclosure within 1/8 mile (about 660 feet) of a single-family home results in a decline of 0.9% in value. Cumulatively, for the entire city of Chicago, the 3,750 foreclosures that occurred in 1997 and 1998 are estimated to have reduced nearby property values by more than \$598 million. However, this study covers a relatively short period of sales (one year) and foreclosures (two years), failing to address endogeneity in the model, since it cannot control for previous years' sales prices. In order to minimize the reverse causation

problem, Immergluck and Smith (2006) add the median home value at the census tract level as a control for sales price in the model. However, the census tract level control is much larger than the targeted research distance interval, which is within 1/4 mile of foreclosures, so the estimation may be biased due to measurement error. Thus, the endogeneity problem is not well addressed in their study.

Lin et al. (2009) use 20% of mortgages made in the United States from 1990 to 2006 to examine the spillover effects of foreclosures on neighborhood property values in Chicago metropolitan area in 2003 and 2006. Besides analyzing house characteristics, quarterly dummies are added to control for seasonal effects, and county and zip dummies are added to control for community level characteristics. They also apply the Heckman (1979) two-step model to correct for sample selection bias. The researchers use both loan characteristics and borrower characteristics to examine foreclosure status. Though there is a statistically significant bias, the effects on the hedonic model are quite small. The results show that spillover effects occur within ten blocks and up to five years from the foreclosure date. The effect decreases as time passes and as space between the foreclosure and the subject property increases. In addition, foreclosures reduced the surrounding house values by half in the boom period in 2003 as much as those in the downside market period in 2006. However, Lin et al. (2009) do not distinguish between the direct foreclosures effects and the spatially dependent home prices.

Schuetz et al. (2008) use property sales and foreclosure filings data in New York City from 2000 to 2005 to examine foreclosure effects on neighborhood property values. As it uses panel data, this study has the advantage of addressing the effect of previous years' sales prices on current year sales price. In New York City, the foreclosure process between the filing of a *lis pendens* and the auction of the property usually takes eighteen months. Nine time and distance

intervals are created to measure the foreclosure effects according to the foreclosure filing timeline. Regression results show that properties close to the foreclosures sell at a discount, and the magnitude of price discount increases with the number of nearby foreclosures, but not in a linear fashion.

Rogers and Winters (2009) apply a hedonic price model to study foreclosure impacts on nearby property values in St. Louis County, Missouri, by using single-family sales data from 2000-2007 and foreclosure data from 1998-2007. They adjust the model to account for spatial autocorrelation. This study supports the hypothesis that foreclosure impacts decrease as distance and time between the house sale and foreclosure increase. The results show a similar magnitude of foreclosure impact compared to Immergluck and Smith's (2006) study, but a much smaller foreclosure impact compared to Lin et al. (2009) study. Roger and Winter (2009) compared marginal foreclosure impacts in two different periods: 2003-2005 and 2006-2007. Consistent with Lin et al. (2009) study, both ordinary least squares (OLS) and generalized method of moments (GMM) estimates suggest a greater impact in 2006-2007 than in 2003-2005. In 2006-2007, a foreclosure that happened within the previous six months and within a 200 yard (600 feet) radius reduces house sales price by 1.6%, but just 0.6% in 2003-2005.

Leonard and Murdoch (2009) use four models to examine foreclosure impacts on single-family homes values in and around Dallas County, Texas, in 2006. OLS regression is used to estimate the model without controlling for spatial dependency and neighborhood pricing trends. Maximum likelihood (ML) procedures are used to examine the spatial autoregressive model and the general spatial model, and GMM procedures are used to examine the general spatial model. Regression results suggest that foreclosures within 250 feet, between 500 and 1000 feet and between 1000 and 1500 feet of a sale depreciate sales prices.

One limitation of these studies is that most of them do not fully address the potential for endogeneity of the foreclosure variable. Since depressed house prices may lead to more foreclosures, foreclosures may thus be endogenous to the sales price. These studies argue that it is hard to find an instrumental variable which is correlated with foreclosures but not correlated with the residuals of the hedonic price equation. The contributions of this study include creating an innovative way to examine endogeneity through accounting for foreclosure timing and by using GS2SLS procedures to address the endogeneity of the spatially lagged dependent variable. This study also employs zero-inflated negative binomial (ZINB) regression to explore the reasons for foreclosures.

There exists a large literature that identifies the causes of foreclosures (Baxter and Lauria, 2000; Chan et al., 2010; Gerardi et al., 2007; Immergluck, 2009; Immergluck and Smith 2005). These include borrower characteristics, loan characteristics and socio-economic characteristics. People with subprime loans are more likely to default than those with government insured loans, and subprime loans are concentrated in low-income and African-American neighborhoods (Bunce et al., 2000; Calem et al., 2004). Shlay (2006) also finds that housing abandonment is associated with poor neighborhoods.

3. Data

The house transaction and foreclosure data were obtained from the Board of Tax Assessors in Fulton County, Georgia. The dataset contains all parcel information for both sold and unsold properties in city of Atlanta. It should be mentioned that a small part of Atlanta is situated in DeKalb County. Because Fulton County and DeKalb County use different variables and codes to record the sales properties, it is difficult to combine the dataset. Therefore, this study only includes the Atlanta sales in Fulton County. The dataset includes property sales price, sales date,

address, property class, and the seller and buyer names, which help to identify foreclosures and foreclosure sales.

Using transaction data from 2003 to 2008, a basic dataset that includes property characteristics, sales price and sales date is constructed. Markov Chain Monte Carlo (MCMC) is used to impute the missing data on small percentage of observations². Before analyzing the data, duplicate records are deleted. Only duplication of a parcel ID with identical date and sales price are deleted, but houses with legitimate repeat sales are maintained. This dataset contains 64,613 one-to-four unit family sales from 2003 to 2008 and there are 10,121 one to four unit family sales records in 2008 after cleaning the dataset. Properties sold directly to banks are coded as foreclosures; however, the same property may sell at depressed prices subsequently, but the subsequent sales are not considered to be foreclosures but are coded as foreclosure sales.³ The dataset contains 7,209 foreclosures and foreclosure sales at Atlanta in 2008.

Each record is geocoded to get the longitude and latitude using ArcGIS according to the property address and then intersected (overlaid) with census block group to identify which census block group the property belongs to. In this study, census block group level data is used to proxy for neighborhood characteristics. Thus, the sales records can be merged with neighborhood characteristics according to the census block group ID number. The neighborhood characteristics are sourced from the 1990 and 2000 Census data.

Since foreclosures are expected to have fewer impacts on sales at a greater distance, buffer rings for each sales property are created by ArcGIS to measure the foreclosure effects on

² The missing data in housing datasets is a common problem, usually a result of coding errors by different data entry clerks. In this dataset, the maximum missing number of observations is for the variable *number of bedrooms* with just 40 (about 0.4% of number of observations) or about. They are imputed by Markov Chain Monte Carlo (MCMC) procedure. The regression is also tested with and without imputation, and the imputation impact on the results is very small.

³ The Board of Tax Assessors records properties both sold to the bank and bank sales as foreclosures.

sales price. Schuetz et al. (2008) measure foreclosure effects by creating 0-250 feet, 250-500 feet and 500-1000 feet intervals. Leonard and Murdoch (2009) find the foreclosure effects extend to 1500 feet in Dallas County. Immergluck and Smith (2006) study foreclosures effects within ¼ miles (about 1300 feet). Rogers and Winter (2009) examine foreclosures in St. Louis County and find foreclosure effects take effect within 600 yards (1800 feet). Following the methods used in the previous literature, five foreclosures intervals are created: DIS300 is the number of foreclosures within 300 feet of sales, DIS600 is the number of foreclosures between 300 feet and 600 feet of sales, DIS1200 is the number of foreclosures between 600 and 1200 feet of sales, DIS1500 is the number of foreclosures between 1200 feet and 1500 feet, and DIS2000 is the number of foreclosures between 1500 feet and 2000 feet. According to the sales date and foreclosure date of houses, the number of foreclosures within each buffer for each property is calculated by using those foreclosures occurring before the sale of the house. For example, if a house was sold on May 10th, 2008, we use foreclosures occurred before that date to calculate the number of foreclosures within buffers.⁴

The percentage of subprime loans is a critical determinant affecting the number of foreclosures within the buffers. The percentage of low-cost/high-leverage mortgage, high-cost/low-leverage mortgage, and high-cost/high-leverage mortgage made 2004 to 2007 in census tract are obtained from Home Mortgage Disclosure Act (HMDA) data of the U.S. Department of Housing and Urban Development (HUD). The school district in which a property is located is another important variable affecting house sales price, because parents usually would like to pay

⁴ Leonard and Murdoch (2009) used cumulative foreclosures in the whole year to calculate the number of foreclosures within buffers, the endogeneity problem may occur. Also, in Immergluck and Smith (2006), foreclosures occurred in 1997 and 1998 are used to test the foreclosure effects on sales price in 1999, foreclosures occurred before 1997 (and not resolved in 1999) and occurred in 1999 are not considered, so foreclosure effects may be overestimated in the study. Lin et al. (2009) combined foreclosures from a mortgage database with information provided by outsider vendors, but the coverage of foreclosures is not very ideal.

more to be located in a good school zone. The school zone boundary for each elementary school is established by the Atlanta Board of Education. There are 55 elementary school zones in city of Atlanta, and houses in this study sample are located in 50 school zones. Thus, 50 dummies are created not only to capture the school characteristics but also the fixed effects such as property tax rate.

4. Model

4.1. Hedonic Model

The hedonic model is used to investigate the cross-sectional relationship between house sales price and foreclosures. The hedonic price model is a commonly used method for studying housing values (Hanna, 2007; Hite et al., 2001; Portney, 1981). The price of a house is usually affected by its own physical characteristics and its location. Structural characteristics, socio-demographic factors, and location to specific amenities or disamenities are included in the hedonic model to capture positive or negative effects on house prices. Properties surrounded by foreclosures suffer potential sales discount impacts. The hedonic sales price in equation (2.1) is assumed to be a function of house, neighborhood, environmental, and foreclosure characteristics

$$P = f(H, N, E, F) + \varepsilon \quad (2.1)$$

where P is house sales price, H is a vector of the house characteristics, N is a vector of census block group socio-demographic characteristics, E is a vector of the environmental disamenities, and F is a vector of foreclosure counts within certain buffers.

The log-linear hedonic price model is given by

$$\begin{aligned} \ln P_{08} = & \beta_0 + \beta_1 HouseChars + \beta_2 NeighChars + \beta_3 Enviro + \beta_4 Dis300 + \beta_5 Dis600 \\ & + \beta_6 Dis1200 + \beta_7 Dis1500 + \beta_8 Dis2000 + \beta_9 Sales_dummy \\ & + \beta_{10} Quarter_dummy + \beta_{11} School_Dummy + \varepsilon_{08} \end{aligned} \quad (2.2)$$

where P_{08} is a vector of house sales price in 2008 expressed as the natural log form; $HouseChars$ is a vector of house characteristics; $NeighChars$ is a vector of neighborhood characteristics at census block group level; $Enviro$ is a vector of the environmental disamenities; $Dis300$ is a vector of number of foreclosures within 300 feet of sales house; $Dis600$ is a vector of number of foreclosures between 300 feet and 600 feet of sales house; $Dis1200$ is a vector of number of foreclosures between 600 feet and 1200 feet of sales house; $Dis1500$ is a vector of number of foreclosures between 1200 feet and 1500 feet; $Dis2000$ is a vector of number of foreclosures between 1500 feet and 2000 feet; $Sales_dummy$ is a vector of dummies for sales type; $Quarter_dummy$ is a vector of dummies for quarters to control seasonality effects; $School_dummy$ is a vector of dummies for school zones.

4.2. Spatial Hedonic Model

The effect of potential spatial correlation on sales price needs to be addressed in the hedonic analysis. Spillover effect theory (Lin et al., 2009; Vandell, 1991) states that the price of the subject property is determined by selected comparable properties. Comparables are properties that are recently sold and those that are close to the subject property by distance. The spatial autoregressive model introduces surrounding average houses prices as a variable to explain their effects on the subject property value.

$$\begin{aligned}
 P_{08} = & \beta_0 + \beta_1 HouseChars + \beta_2 NeighChars + \beta_3 Enviro + \beta_4 Dis300 + \beta_5 Dis600 \\
 & + \beta_6 Dis1200 + \beta_7 Dis1500 + \beta_8 Dis2000 + \beta_9 Sales_dummy \\
 & + \beta_{10} Quarter_dummy + \beta_{11} School_Dummy + \rho_1 W_{08} P_{08} + \varepsilon_{08}
 \end{aligned} \tag{2.3}$$

where W_{08} is the spatial weight matrix for sales properties; ρ_1 is the parameter for spatial lag.

The spatial weight matrix can be constructed in various ways. Choosing a proper weight structure is important for the model specification. A common method to construct the weight matrix in the literature is to find the k nearest neighbor sales in terms of distance. Five to ten

nearest neighbors are the most frequent selection criterion in previous studies. This study chooses eight nearest neighbors for each sales property to establish the spatial weight matrices, where the eight nearest neighbors receive value 1 in the matrix, others receive value 0 in the matrix. The matrix is row standardized, so that the spatial average price can be controlled for each sales property⁵.

Using only one year's price may be not sufficient to control for the sales trend. Following Leonard and Murdoch's (2009) work, sales prices from year 2003 to 2007 are added into the regression,

$$\begin{aligned}
 P_{08} = & \beta_0 + \beta_1 HouseChars + \beta_2 NeighChars + \beta_3 Enviro + \beta_4 Dis300 + \beta_5 Dis600 \\
 & + \beta_6 Dis1200 + \beta_7 Dis1500 + \beta_8 Dis2000 + \beta_9 Sales_dummy \\
 & + \beta_{10} Quarter_dummy + \beta_{11} School_Dummy + \rho_1 W_{08} P_{08} + \rho_2 W_{07} P_{07} \\
 & + \rho_3 W_{06} P_{06} + \rho_4 W_{05} P_{05} + \rho_5 W_{04} P_{04} + \rho_6 W_{03} P_{03} + \varepsilon_{08}
 \end{aligned} \tag{2.4}$$

where P_{yy} is a vector of sales price in year yy ; W_{yy} is the spatial weight matrix establishing the spatial relationship between sales in 2008 and sales in year yy . For matrices W_{07} , W_{06} , W_{05} , W_{04} , and W_{03} , if there is a house sold in year 2007, 2006, 2005, 2004, or 2003 within 2000 feet to each sale in 2008, then the matrices get a non-zero entry, otherwise, the entries get zero values.

In addition to the spatial lag, this paper also controls for spatially correlated errors to control for unobserved heterogeneity. The *lmsar* function in MATLAB can be used to test for spatial correlation in the residuals of a spatial autoregressive model. If the marginal probability is less than 0.05, there is evidence that spatial dependence exists in the error structure. Thus, the general spatial model that includes both a spatial lag and spatial error is appropriate for modeling this type of dependence in the errors.

⁵ Five and ten nearest neighbors are also tested respectively, and the estimated coefficients differ less than 2% compared to results in the paper, and the significance levels do not change.

$$\begin{aligned}
P_{08} = & \beta_0 + \beta_1 HouseChars + \beta_2 NeighChars + \beta_3 Enviro + \beta_4 Dis300 + \beta_5 Dis600 \\
& + \beta_6 Dis1200 + \beta_7 Dis1500 + \beta_8 Dis2000 + \beta_9 Sales_dummy \\
& + \beta_{10} Quarter_dummy + \beta_{11} School_Dummy + \rho_1 W_{08} P_{08} + \rho_2 W_{07} P_{07} \\
& + \rho_3 W_{06} P_{06} + \rho_4 W_{05} P_{05} + \rho_5 W_{04} P_{04} + \rho_6 W_{03} P_{03} + \varepsilon_{08} \\
\varepsilon_{08} = & \lambda W_{08} \varepsilon_{08} + \mu
\end{aligned} \tag{2.5}$$

The generalized spatial two stage least squares (GS2SLS) procedure introduced by Kelejian and Prucha (1998) is also used to examine the general spatial model in this paper. GS2SLS procedures can correct for the endogeneity of spatial lag dependent variable and produce consistent estimators when the model contains both spatial lags in the endogenous variables and spatial autocorrelation in the disturbances. The spatially lagged price is instrumented by spatially weighted lagged independent variables in the price equation.

GS2SLS procedures are developed based on the GMM procedures, and have several advantages over the ML method which is often used to estimate spatial models. The ML estimation produces consistent and efficient results only if two assumptions are met. First, the disturbance should be normally distributed. Second, the variance of the disturbances should be homoskedastic. ML procedures use a numerical hessian calculation, but in the presence of outliers or non-constant variance the numerical hessian approach may not be valid because normality and homodasticity in the disturbance generating process might be violated (Lesage, 1998). Further, ML estimation is often computationally challenging when the sample size is large (Kelejian and Prucha, 1998). GS2SLS estimation relaxes the normality assumption and allows for heteroskedatic errors in the spatial model, also it is computationally simple compared to the ML estimation, especially for large sample sizes (Keijian and Prucha, 1998, 1999). In addition, the GS2SLS estimation can be used to incorporate a high degree of flexibility in the specification of the spatial weight matrix. In traditional spatial models, the selection of spatial

weight matrix is usually an *ad hoc* process. Sometimes small changes in the spatial weight matrix can result in changes to the model result. GS2SLS model can incorporate flexible spatial weight matrix specification and get consistent results (Bucholtz, 2004).

4.3. Endogeneity Testing

We hypothesize that there are endogeneities in the sales equation. Surrounding sales prices usually work as a signal that affect neighbors' decision to default. When the housing market falls, the balance of the mortgage exceeds the house value, and borrowers may choose to default in reaction. Therefore, depressed sales prices may induce more foreclosures. Most previous studies just ignore the potential endogeneity problem, only a few put forward the endogeneity problem as studying the foreclosure effects on house value (Immergluck and Smith, 2005; Rogers and Winters, 2009), but no previous study has sufficiently controlled for endogeneity in cross-sectional data.

4.4. Zero-Inflated Negative Binomial (ZINB) Model

After testing for endogeneity, a zero-inflated negative binomial (ZINB) regression can be used to examine what factors affect foreclosures. A ZINB model is a modified Poisson regression which accommodates both overdispersion (distribution variance is larger than its mean) and excess zeros found in count data. Because the distribution of foreclosures is generally skewed to the right and contains a considerable proportion of zeros in five buffers (17% within 300 feet, 28% between 300 feet and 600 feet, 15% between 600 feet and 1200 feet, 19% between 1200 feet and 1500 feet, and 13% between 1500 feet and 2000 feet), zero-inflated negative binomial (ZINB) regression model is proper to be used. Overdispersion is often caused by unobservable individual heterogeneity and/or excess zeros of the data (Sheu et al., 2004). Vuong test is used to test excess zeros and compare zero-inflated model and non-zero-inflated model.

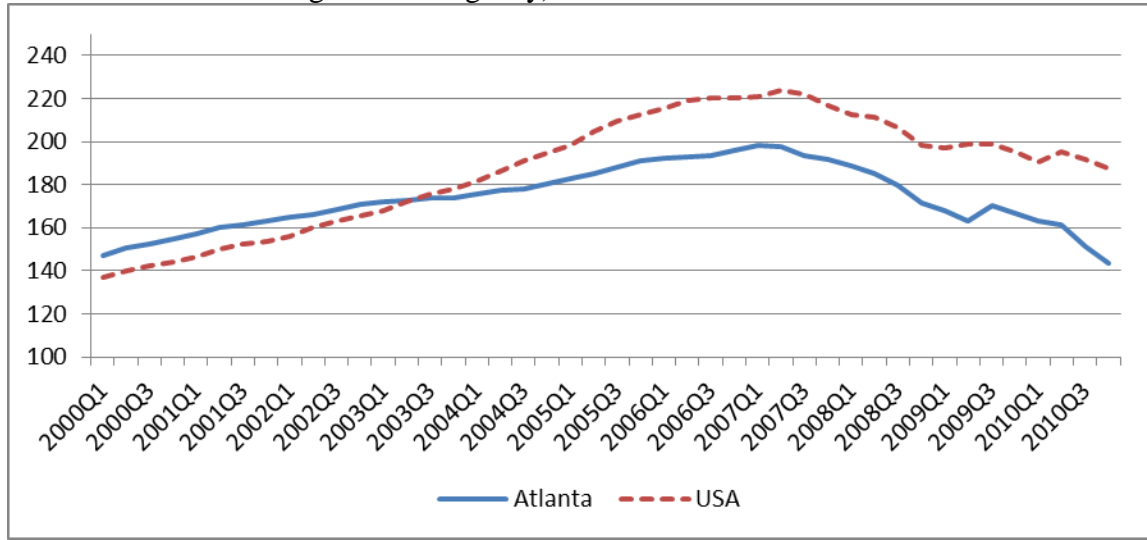
We hypothesize that for each sales property, the number of foreclosures within each buffer is influenced by its surrounding house characteristics, neighborhood characteristics and loan characteristics in the buffer. The advantage of this dataset is that it includes all properties in city of Atlanta, both sold and unsold. Thus, it is possible to measure surrounding house conditions for each sales house. The average dwelling condition, number of rooms, living area, and house age in the buffer are chosen to explain the number of foreclosures. Percentage of black residents, average household size, percentage of home ownership, per capital income, percentage of people over 65 years old are added as neighborhood characteristics to explain the number of foreclosures.

Loan characteristics at census tract level are critical to explain foreclosures, since increasing foreclosures are the direct results of growing subprime loans. Subprime loans in this paper are defined as high-cost loans, as their fees and interest rates are usually significantly above those charged to typical borrowers (Gerardi et al., 2008). The Home Ownership and Equity Protection Act (HOEPA) defines high-cost loans as loans with interest rates more than eight percentage point of loan balance. High cost loans are usually high leverage loans, which is related to the high loan-to-value (LTV) concept, i.e. borrowers have smaller down payments than typical borrowers. High-cost, high-leverage loan borrowers usually carry a higher risk of default than low-cost, low leverage loan borrowers. The census tract percentages of low-cost/high-leverage loans, high-cost/low-leverage loans, and high-cost/high-leverage loans from HMDA made from 2004 to 2007 are used to explain the number of foreclosures.

Figure 2.1 shows the historical quarterly home sales price index for Atlanta-Sandy Springs-Marietta area and the national average. From the third quarter of 2003, the housing price index in Atlanta is lower than the national average, and the price difference became larger in the

following years. In the Atlanta area, the home sales price peaked historically in mid-2007, then decreased dramatically since the second quarter of 2007 (Federal Housing Finance Agency). The sales prices in 2008 are much lower than those in 2007, so the change of sales prices between 2007 and 2008 in the buffer for each sales property in 2008 is also added as an explanatory variable ($\Delta price$) since the depreciating market is hypothesized to induce more foreclosures.

Figure 2.1 Historic Quarterly Home Sales Price Index, Atlanta MSA, Seasonally Adjusted (Data Source: Federal Housing Finance Agency)



Equations (2.6)-(2.10) are used to explain the number of foreclosures within 300 feet, between 300 feet and 600 feet, between 600 feet and 1200 feet, between 1200 feet and 1500 feet, and between 1500 and 2000 feet as functions of house characteristics, neighborhood characteristics, loan characteristics and sales price change in the buffer.

$$\begin{aligned}
 Dis300 = & \eta_0 + \eta_1 Cdu300 + \eta_2 Lvar ea300 + \eta_3 Age300 + \eta_4 Rmbed300 + \eta_5 Black \\
 & + \eta_6 Hsize + \eta_7 Own + \eta_8 Income + \eta_9 Old + \eta_{10} Pct_lchl + \eta_{11} Pct_hcll \\
 & + \eta_{12} Pct_hchl + \eta_{13} \Delta price300 + \varepsilon_{300}
 \end{aligned} \tag{2.6}$$

$$\begin{aligned}
 Dis600 = & \eta_0 + \eta_1 Cdu600 + \eta_2 Lvar ea600 + \eta_3 Age600 + \eta_4 Rmbed600 + \eta_5 Black \\
 & + \eta_6 Hsize + \eta_7 Own + \eta_8 Income + \eta_9 Old + \eta_{10} Pct_lchl + \eta_{11} Pct_hcll \\
 & + \eta_{12} Pct_hchl + \eta_{13} \Delta price600 + \varepsilon_{600}
 \end{aligned} \tag{2.7}$$

$$\begin{aligned}
Dis1200 = & \eta_0 + \eta_1 Cdu1200 + \eta_2 Lvar ea1200 + \eta_3 Age1200 + \eta_4 Rmbed1200 + \eta_5 Black \\
& + \eta_6 Hsize + \eta_7 Own + \eta_8 Income + \eta_9 Old + \eta_{10} Pct_lchl + \eta_{11} Pct_hcll \\
& + \eta_{12} Pct_hchl + \eta_{13} \Delta price1200 + \varepsilon_{1200}
\end{aligned} \tag{2.8}$$

$$\begin{aligned}
Dis1500 = & \eta_0 + \eta_1 Cdu1500 + \eta_2 Lvar ea1500 + \eta_3 Age1500 + \eta_4 Rmbed1500 + \eta_5 Black \\
& + \eta_6 Hsize + \eta_7 Own + \eta_8 Income + \eta_9 Old + \eta_{10} Pct_lchl + \eta_{11} Pct_hcll \\
& + \eta_{12} Pct_hchl + \eta_{13} \Delta price1500 + \varepsilon_{1500}
\end{aligned} \tag{2.9}$$

$$\begin{aligned}
Dis2000 = & \eta_0 + \eta_1 Cdu2000 + \eta_2 Lvar ea2000 + \eta_3 Age2000 + \eta_4 Rmbed2000 + \eta_5 Black \\
& + \eta_6 Hsize + \eta_7 Own + \eta_8 Income + \eta_9 Old + \eta_{10} Pct_lchl + \eta_{11} Pct_hcll \\
& + \eta_{12} Pct_hchl + \eta_{13} \Delta price2000 + \varepsilon_{2000}
\end{aligned} \tag{2.10}$$

where $Cdu300$ is the average house condition within 300 feet for each sales property; $Lvarea300$ is the average living area within 300 feet for each property; $Age300$ is the average house age within 300 feet for each property; $Rmbed300$ is the average number of rooms within 300 feet for each property; Pct_lchl is the percentage of HMDA mortgage made 2004 to 2007 that are low-cost/high-leverage in the census tract; Pct_hcll is the percentage of HMDA mortgage made 2004 to 2007 that are high-cost/low-leverage in the census tract; Pct_hchl is the percentage of HMDA mortgage made 2004 to 2007 that are high-cost/high-leverage in the census tract; $\Delta price300$ is the average sales price change between year 2007 and 2008 within 300 feet; $Cdu600$ is the average house condition between 300 feet and 600 feet for each sales property; $Cdu1200$ is the average house condition between 600 feet and 1200 feet for each sales property; $Cdu1500$ is the average house condition between 1200 feet and 1500 feet; $Cdu2000$ is the average house condition between 1500 feet and 2000 feet. It is tested that Pct_hchl is correlated with number of foreclosures, but not correlated with the residuals of price equations, which indicates that it is a valid instrumental variable. A Hausman test is used to test equations (2.2) and (2.6)-(2.10) systematically for endogeneity.

5. Results

Table 2.1 reports the descriptive statistics for the variables in the model. The house characteristics include lot size in square feet, living area in square feet, number of stories, age of the house, number of bedrooms, number of full and half bathrooms, basement and attic condition, heat type, overall dwelling condition, and the street condition in the parcel. The squares of number of rooms and age are added because these variables may influence a house value in a nonlinear way. In addition to structural house characteristics, many neighborhood characteristics may affect house value. Percentage of black residents, average household size, percentage of home ownership, and percentage of people over 65 years old in each census block group (CBG) are chosen to represent neighborhood quality.

Environmental economic theory suggests that house value depends on environmental quality. Since DeKalb County is located next to Fulton County, the properties located in the east Atlanta may be affected by the hazard sites in DeKalb County. Thus, the location of point-specific Hazardous Site Inventory (HSI) in Fulton County and DeKalb County are used to calculate their effects on sales price. Data on polluting sites is obtained from the Georgia Environmental Protection Division (GEPD). Using ArcGIS, the distance of sales property to the nearest HSI can be calculated. It is hypothesized that holding other variables constant, the greater the distance to the nearest HSI, the higher the house price is expected to be. The mean distance to the nearest HSI in the sample is 2.14 kilometers. The minimum is 16 meters, and the maximum is 6,281 meters.

The count of foreclosures within a specific distance for each property is the focus of this section. Five distance intervals are created for each property. If the number of foreclosures is statistically significant for *DIS1500* but not for *DIS2000*, then spillovers from a foreclosure

affect other properties within 1500 feet of a sale, but not beyond 1500 feet. The average number of foreclosures is 2.59 within 300 feet of sales house, 4.12 between 300 and 600 feet of sales house, 12.74 between 600 feet and 1200 feet of sales house, 8.49 between 1200 feet and 1500 feet of sales house, and 16.62 between 1500 feet and 2000 feet of sales house in 2008.

Table 2.1 Descriptive Statistics for One to Four Unit Family Sales House Characteristics and Neighborhood Characteristics, Atlanta, 2008 (N=10,121)

Variable	Description	Mean	SD	Minimum	Maximum
Price	Sales price (*\$1,000)	185.87	351.11	10	10058.52
Street1	The street condition in the parcel is "paved"	0.99	0.09	0	1
Street2	The street condition is "semi-improved"	0.003	0.05	0	1
Street3	The street condition is "dirt"	0.004	0.06	0	1
Lotarea	Lot area sqft (*1000)	11.44	11.69	0.52	588.06
Lvarea	Living area sqft (*1000)	1.64	0.94	0.22	14.80
Stories	Number of stories	1.20	0.40	1	3
Rmbed	Number of bedrooms	3.01	0.93	1	12
Fixbath	Number of full bathrooms	1.70	0.88	1	9
Fixhalf	Number of half bathrooms	0.26	0.49	0	8
Bsmt1	No basement	0.08	0.28	0	1
Bsmt2	Crawl basement	0.57	0.50	0	1
Bsmt3	Part basement	0.16	0.37	0	1
Bsmt4	Full basement	0.18	0.39	0	1
Heat1	No heat	0.02	0.15	0	1
Heat2	Central heat	0.07	0.25	0	1
Heat3	Central air condition	0.21	0.40	0	1
Heat4	Heat pump	0.70	0.46	0	1
Attic1	No attic	0.87	0.33	0	1
Attic2	Unfinished attic	0.05	0.21	0	1
Attic3	Part finished attic	0.04	0.19	0	1
Attic4	Fully finished attic	0.03	0.18	0	1
Attic5	Fully finished/wall height attic	0.01	0.09	0	1
Age	Age of sales house	50.01	29.95	0	138
Cdu1	Dwelling condition is excellent	0.09	0.29	0	1
Cdu2	Dwelling condition is very good	0.19	0.39	0	1
Cdu3	Dwelling condition is good	0.20	0.40	0	1
Cdu4	Dwelling condition is average	0.40	0.49	0	1
Cdu5	Dwelling condition is fair	0.07	0.25	0	1
Cdu6	Dwelling condition is unsound	0.03	0.16	0	1
Cdu7	Dwelling condition is poor	0.02	0.12	0	1
Cdu8	Dwelling condition is very poor	0.01	0.08	0	1
Black	Percentage of black residents in CBG	0.79	0.32	0	1
Hsize	Average household size in CBG	2.68	0.45	1.34	4.19
Own	Percentage of residents own the house in CBG	0.49	0.21	0	0.98
Income	Per capital income in 1999 in CBG (*\$1,000)	19.40	18	2.76	120.93
Old	Percentage of people over 65 years old in CBG	0.11	0.06	0	0.51
HSI	The distance from sales house to the nearest hazardous site inventory (*1000m)	2.14	1.23	0.02	6.28
DIS300	Number of foreclosures within 300 feet of sales house	2.59	2.48	0	19

DIS600	Number of foreclosures between 300 and 600 yards of sales house	4.12	5.13	0	35
DIS1200	Number of foreclosures between 600 and 1200 yards of sales house	12.74	14.46	0	121
Dis1500	Number of foreclosures between 1200 and 1500 yards of sales house	8.49	10	0	76
Dis2000	Number of foreclosures between 1500 and 2000 yards of sales house	16.62	18.23	0	123
W ₀₇ P ₀₇	Spatial lag of 2007 log sales price within 2000 feet of 2008 price	11.83	0.89	0	14.95
W ₀₆ P ₀₆	Spatial lag of 2006 log sales price within 2000 feet of 2008 price	11.91	0.86	0	14.98
W ₀₅ P ₀₅	Spatial lag of 2005 log sales price within 2000 feet of 2008 price	11.83	0.94	0	14.90
W ₀₄ P ₀₄	Spatial lag of 2004 log sales price within 2000 feet of 2008 price	11.63	1.23	0	14.90
W ₀₃ P ₀₃	Spatial lag of 2003 log sales price within 2000 feet of 2008 price	11.55	1.11	0	14.93
Pct_lchl	Percentage of HMDA mortgage made 2004 to 2007 that are low-cost and high-leverage in census tract	0.08	0.04	0.01	0.22
Pct_hcll	Percentage of HMDA mortgage made 2004 to 2007 that are high-cost and low-leverage in census tract	0.27	0.12	0.02	0.49
Pct_hchl	Percentage of HMDA mortgage made 2004 to 2007 that are high-cost and high-leverage in census tract	0.16	0.07	0.01	0.28
Cdu300	Average house condition within 300 feet	3.34	1.08	1	8
Lvarea300	Average living area (*1000) within 300 feet	1.64	0.86	0.30	14.80
Age300	Average house age within 300 feet	50.04	23.41	0	132
Rmbed300	Average number of room within 300 feet	3.01	0.70	1	10
ΔPrice300	Average sales price change (*10,000 dollar) between 2007 and 2008 within 300 feet	-6.69	21.57	-429.11	262.50
Cdu600	Average house condition between 300 feet and 600 feet	3.34	1.03	1	8
Lvarea600	Average living area (*1000) between 300 feet and 600 feet	1.63	0.83	0.32	14.80
Age600	Average house age between 300 feet and 600 feet	50.26	21.26	0	123
Rmbed600	Average number of room between 300 feet and 600 feet	3.00	0.63	1	8
ΔPrice600	Average sales price change (*10,000 dollar) between 2007 and 2008 between 300 feet and 600 feet	-7.77	18.61	-431.20	152.50
Cdu1200	Average house condition between 600 feet and 1200 feet	3.54	0.84	1	7

Lvarea1200	Average living area (*1000) between 600 feet and 1200 feet	1.54	0.58	0.91	7.80
Age1200	Average house age between 600 feet and 1200 feet	55.72	12.65	1	98
Rmbed1200	Average number of room between 600 feet and 1200 feet	2.91	0.33	1.84	5.24
ΔPrice1200	Average sales price change (*10,000 dollar) between 2007 and 2008 between 600 feet and 1200 feet	-9.32	19.76	-413.02	375.27
Cdu1500	Average house condition between 1200 feet and 1500 feet	3.54	0.83	1	8
Lvarea1500	Average living area (*1000) between 1200 feet and 1500 feet	1.54	0.58	0.60	8.72
Age1500	Average house age between 1200 feet and 1500 feet	56.01	12.10	2.85	108
Rmbed1500	Average number of room between 1200 feet and 1500 feet	2.90	0.38	2	6
ΔPrice1500	Average sales price change (*10,000 dollar) between 2007 and 2008 between 1200 feet and 1500 feet	-8.54	19.06	-413.02	352.97
Cdu2000	Average house condition between 1500 feet and 2000 feet	3.54	0.81	1	6.12
Lvarea2000	Average living area (*1000) between 1500 feet and 2000 feet	1.55	0.57	0.73	6.44
Age2000	Average house age between 1500 feet and 2000 feet	56.06	11.28	4	118
Rmbed2000	Average number of room between 1500 feet and 2000 feet	2.90	0.31	2	5
ΔPrice2000	Average sales price change (*10,000 dollar) between 2007 and 2008 between 1500 feet and 2000 feet	-9.67	20.36	-431.20	648.26

Table 2.2 reports the descriptive statistics for dummy variables. Quarterly dummies are created to control for seasonal sales effects. It is hypothesized that houses are usually sold at a lower price in the winter than in the summer. The types of sale vary in the sample, and the sales types affect the sales price directly, so sales type dummies are included in the regression. Sale1 represents a valid sale, Sale2 is the sale to or from an exempt or utility, Sale3 represents properties remodeled or changed after sale, Sale4 represents sales between individual and corporation, Sale5 represents a liquidation or foreclosure sale, Sale6 represents a land contract or unusual financing sale, and Sale7 is a sale that includes additional interest. School district dummies are also important variables that affect house sales price. There are 55 elementary school zones in city of Atlanta, and houses in this sample are located in 50 school zones.

A Hausman test is used to test for endogeneity by estimating equations (2.2) and (2.6)-(2.10) simultaneously, but there is no evidence that endogeneity exists in the model, thus the OLS model is consistent. The reason is that maybe we use foreclosures occurred before sales to calculate the number of foreclosures within a buffer so that controls the potential endogeneity problem. However, equations (2.6)-(2.10) can be used to unravel the factors affect foreclosure in each buffer. Table 2.3 reports that the Vuong statistic values are larger than 1.96 and statistically significant, indicating that the ZINB model is superior to the negative binomial (NB) model. In this study, unobservable heterogeneity is likely to be another problem. The number of foreclosures varies with house characteristics, socio-demographic factors and loan characteristics. However, it also may be affected by social interaction and personal moral issues (Towe and Lawley, 2010). A ZINB model is preferred to a zero-inflated Poisson (ZIP) model because the results show the estimated alphas are statistically significant, indicating that heterogeneity also causes overdispersion even after the excess zero issue is addressed.

Table 2.2 Descriptive Statistics for Dummy Variables, Atlanta, 2008 (N=10,121)

Variable	Description	Mean	SD	Minimum	Maximum
Sale1	Valid sale	0.13	0.34	0	1
Sale2	To/from exempt or utility	0.02	0.12	0	1
Sale3	Remodeled/Changed after sale	0.01	0.12	0	1
Sale4	Related individuals or corporation	0.06	0.24	0	1
Sale5	Liquidation/Foreclosure	0.72	0.45	0	1
Sale6	Land contract/Unusual financing	0.01	0.09	0	1
Sale7	Includes Additional interest	0.00	0.02	0	1
Quarter1	Sold in the first quarter	0.25	0.43	0	1
Quarter2	Sold in the second quarter	0.26	0.44	0	1
Quarter3	Sold in the third quarter	0.26	0.44	0	1
Quarter4	Sold in the fourth quarter	0.22	0.42	0	1
SD1	Adamsville Elementary School	0.00	0.08	0	1
SD2	Benteen Elementary School	0.01	0.11	0	1
SD3	Mary Mcleod Bethune Elementary School	0.02	0.14	0	1
SD5	Capitol View Elementary School	0.02	0.13	0	1
SD6	Cascade Elementary School	0.00	0.06	0	1
SD7	Cleveland Avenue Elementary School	0.01	0.11	0	1
SD8	William M. Boyd Elementary School	0.02	0.14	0	1
SD9	Warren T. Jackson Elementary School	0.02	0.13	0	1
SD10	Morris Brandon Elementary School	0.02	0.14	0	1
SD11	Garden Hills Elementary School	0.01	0.11	0	1
SD12	E. Rivers Elementary School	0.02	0.13	0	1
SD13	Bolton Academy	0.01	0.12	0	1
SD14	Beecher Hills Elementary School	0.01	0.12	0	1
SD15	Daniel H. Stanton Elementary School	0.03	0.17	0	1

SD16	John Wesley Dobbs Elementary School	0.05	0.21	0	1
SD17	Hill-Hope Elementary School	0.01	0.10	0	1
SD18	Connally Elementary School	0.05	0.22	0	1
SD19	Centennial Place Elementary School	0.01	0.07	0	1
SD20	Continental Colony Elementary School	0.00	0.08	0	1
SD21	Ed S. Cook Elementary School	0.03	0.18	0	1
SD22	Deerwood Academy	0.02	0.15	0	1
SD23	Paul L. Dunbar Elementary School	0.01	0.11	0	1
SD25	Margaret Fain Elementary School	0.01	0.09	0	1
SD26	Fickett Elementary School	0.01	0.12	0	1
SD27	William Finch Elementary School	0.06	0.25	0	1
SD28	Charles L. Gideons Elementary School	0.06	0.24	0	1
SD29	Grove Park Elementary School	0.03	0.16	0	1
SD30	Heritage Academy Elementary School	0.01	0.12	0	1
SD31	Alonzo F. Herndon Elementary School	0.02	0.15	0	1
SD32	Joseph Humphries Elementary School	0.01	0.09	0	1
SD33	Emma Hutchinson Elementary School	0.01	0.11	0	1
SD34	M. Agnes Jones Elementary School	0.06	0.23	0	1
SD35	L. O. Kimberly Elementary School	0.01	0.10	0	1
SD36	Mary Lin Elementary School	0.00	0.06	0	1
SD37	Leonora P. Miles Elementary School	0.01	0.11	0	1
SD38	Morningside Elementary School	0.02	0.14	0	1
SD39	Parkside Elementary School	0.02	0.15	0	1
SD40	Perkerson Elementary School	0.03	0.16	0	1

SD41	Peyton Forest Elementary School	0.01	0.07	0	1
SD42	William J Scott Elementary School	0.02	0.13	0	1
SD43	Thomas Heathe Slater Elementary School	0.03	0.18	0	1
SD44	Sarah Rawson Smith Elementary School	0.02	0.13	0	1
SD45	Springdale Park Elementary School	0.02	0.13	0	1
SD46	F. L. Stanton Elementary School	0.03	0.16	0	1
SD47	Thomasville Heights Elementary School	0.01	0.10	0	1
SD49	George A. Towns Elementary School	0.02	0.12	0	1
SD50	Bazoline Usher Elem School	0.02	0.14	0	1
SD52	West Manor Elementary School	0.00	0.06	0	1
SD53	Walter F. White Elementary School	0.02	0.15	0	1
SD55	Carter G. Woodson Elementary School	0.01	0.11	0	1

5.1. Effects of Characteristics on Number of Foreclosures

Table 2.3 presents the results of the ZINB regression. Between 300 feet and 2000 feet, a property surrounded by houses in good condition has a greater number of foreclosures within the buffer than a property surrounded by fair or poor condition houses. The reason is potentially that houses in good condition have higher values and usually suffer more than those moderate condition houses if the housing market drops. So the loan balance of good condition house is more likely to exceed the house value, and it is more likely to be foreclosed. The number of foreclosures within a buffer increases if the average living area of surrounding houses decreases. The number of foreclosures within a buffer increases if the average age of surrounding houses is older. Between 300 feet and 600 feet, and between 1200 feet and 1500 feet, the number of foreclosures increases if the average number of rooms decreases.

The percentage of black residents has a statistically significant effect on the number of foreclosures. Every 10% increase in the percentage black residents increases the number of foreclosures by 0.05 within 300 feet, increases the number of foreclosures by 0.04 between 600 feet and 1200 feet, 0.05 between 1200 feet and 1500 feet, and 0.04 between 1500 feet and 2000 feet. Larger household size decreases the risk of foreclosures. Every additional household member decreases the number of foreclosures by 0.08 between 300 feet and 600 feet, by 0.1 between 600 feet and 1200 feet, by 0.11 between 1200 feet and 1500 feet, and by 0.11 between 1500 feet and 2000 feet. A higher percentage of home ownership reduces the number of foreclosures within 300 feet, but increases that number beyond 600 feet. On average, every 10% increase in home ownership decreases the number of foreclosures by 0.05 within 300 feet, and by 0.02 between 300 feet and 600 feet. But between 1500 feet and 2000 feet, every 10% increase in home ownership increases foreclosures by 0.02. Income has a significant effect on

foreclosures in all buffers. Every 1000 dollar increase in per capital income in a CBG leads to a decrease in the number of foreclosures by 0.02 unit within 300 feet, 0.04 between 300 feet and 600 feet, 0.03 between 600 feet and 1200 feet, 0.04 between 1200 feet and 1500 feet, and 0.04 between 1500 feet and 2000 feet. The percentage of people over 65 years old has no consistent effect on foreclosures within different buffers.

Percentage of subprime loans made from 2004 to 2007 significantly affects foreclosures in all buffers. Within 300 feet, every 10% increase in low-cost and high-leverage loans increases foreclosures by 0.15, a 10% high-cost and low-leverage loans increase foreclosures by 0.16, and 10% more of high-cost and high-leverage loans increase foreclosures by 0.28. The price difference between 2007 and 2008 ($\Delta Price$) also affects foreclosures as a feedback effect. Since most of the sales prices in 2008 are lower than those in 2007, a large proportion of price change values are negative. Thus, every 10,000 dollars increase in price difference in absolute value increases foreclosures by 0.04 units within 300 feet, 0.01 between 300 feet and 600, 0.008 between 600 feet and 1200 feet, 0.008 between 1200 feet and 1500 feet, and 0.006 between 1500 feet and 2000 feet. Figure 2.2 shows the spatial relationship between foreclosures and percentage of black residents in the Atlanta area in 2008. Foreclosures are concentrated in the central city and areas with a high percentage of African American residents, especially concentrated in the CBGs where African American residents are more than 80%. Figure 2.3 shows the spatial relationship between foreclosures and per capita income. Foreclosures are more likely to be concentrated in low income districts. In Figure 2.4, the relationship between foreclosures and percentage of home ownership suggests that foreclosures are concentrated in areas with relatively moderate percentage of home ownership, not in the lowest home ownership area as expected.

Table 2.3 Zero-Inflated Negative Binomial Regression Analysis for Factors Affecting Foreclosures

Variable	Coefficient (t statistics)				
	Dis300	Dis600	Dis1200	Dis1500	Dis2000
Intercept	0.07 (0.40)	0.50** (2.19)	0.88*** (4.35)	0.25 (1.13)	0.56*** (2.59)
Cdu300	-0.02 (-1.19)				
Lvarea300	-0.14*** (-3.29)				
Age300	0.0025*** (4.24)				
Rmbed300	0.04 (1.35)				
Δ Price300	-0.04*** (-6.74)				
Cdu600		-0.07*** (-2.77)			
Lvarea600		-0.01 (-0.14)			
Age600		0.01*** (7.12)			
Rmbed600		-0.08* (-1.79)			
Δ Price600		-0.01*** (-11.73)			
Cdu1200			-0.18*** (-6.58)		
Lvarea1200			-0.23*** (-6.58)		
Age1200			0.02*** (17.05)		
Rmbed1200			-0.06 (-1.27)		
Δ Price1200			-0.01*** (-9.08)		
Cdu1500				-0.23*** (-7.58)	
Lvarea1500				0.02 (0.25)	
Age1500				0.02*** (17.15)	
Rmbed1500				-0.15*** (-2.76)	
Δ Price1500				-0.01***	

				(-8.66)	
Cdu2000					-0.24***
					(-8.13)
Lvarea2000					-0.04
					(-0.65)
Age2000					0.03***
					(25.13)
Rmbed2000					-0.09
					(-1.50)
ΔPrice2000					-0.01***
					(-6.60)
Black	0.54***	0.14	0.38***	0.50***	0.39***
	(6.77)	(1.45)	(4.61)	(5.34)	(4.75)
Hsize	-0.04	-0.08**	-0.10***	-0.11***	-0.11***
	(-1.48)	(-2.04)	(-3.04)	(-3.25)	(-3.58)
Own	-0.48***	-0.20***	0.01	0.00	0.24***
	(-8.57)	(-2.85)	(0.19)	(0.07)	(4.09)
Income	-0.02***	-0.04***	-0.03***	-0.04***	-0.04***
	(-9.04)	(-15.12)	(-16.06)	(-16.29)	(-18.57)
Old	-0.03	0.08	-0.59***	0.12	-0.25
	(-0.17)	(0.35)	(-2.89)	(0.61)	(-1.38)
Pct_lchl	1.50***	2.67***	2.82***	3.32***	1.90***
	(2.64)	(3.82)	(5.05)	(5.16)	(3.40)
Pct_hcll	1.61***	2.27***	3.16***	3.42***	2.88***
	(7.29)	(8.04)	(13.58)	(13.35)	(12.64)
Pct_hchl	2.84***	5.38***	6.47***	6.41***	6.87***
	(12.27)	(18.47)	(25.81)	(23.72)	(27.95)
Alpha	0.18***	0.59***	0.60***	0.62***	0.61***
	(22.61)	(43.44)	(56.75)	(52.24)	(58.72)
Vuong test of ZINB versus NB	5.09***	6.17***	7.50***	5.47**	6.22***

Note: ***Statistically significant at 1%; **Statistically significant at 5%; *Statistically significant at 10%.

The asymptotic t statistics are in parentheses.

Figure 2.1 Spatial Relationship between Foreclosures and Percent Black Residence, Atlanta, 2008

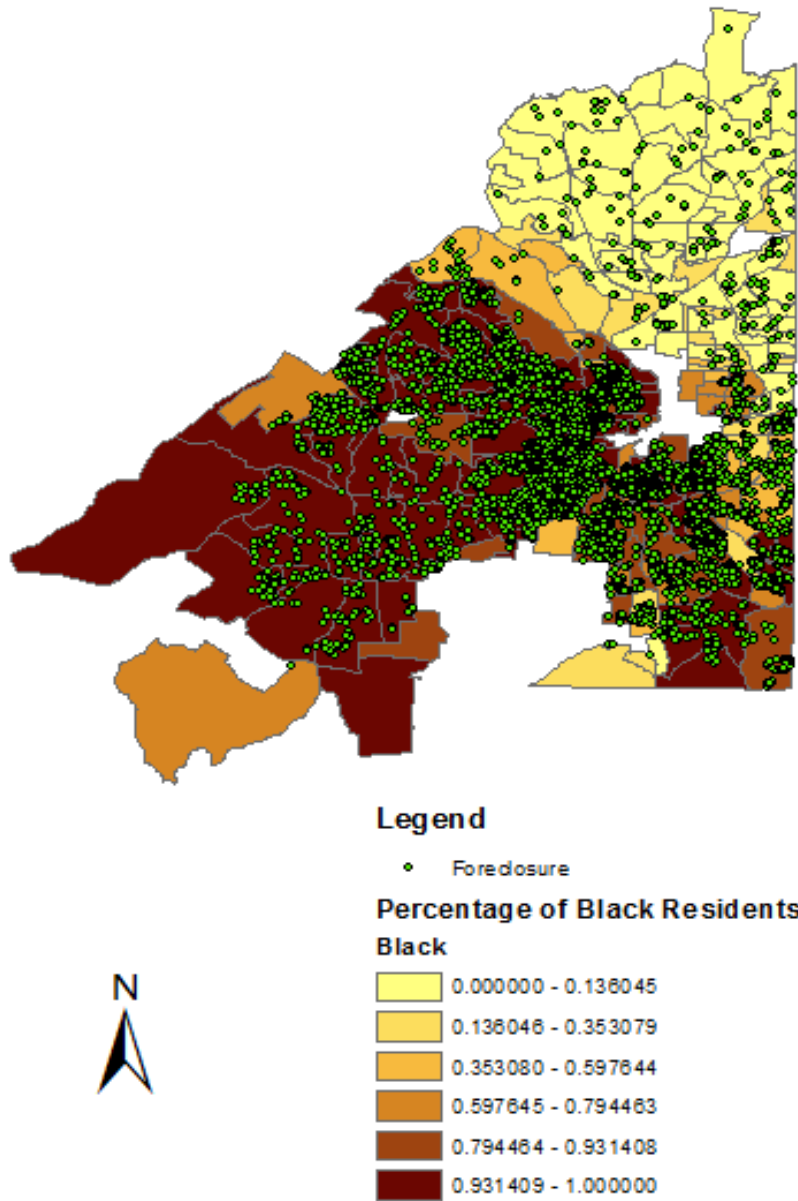


Figure 2.2 Spatial Relationship between Foreclosures and Per Capita Income, Atlanta, 2008

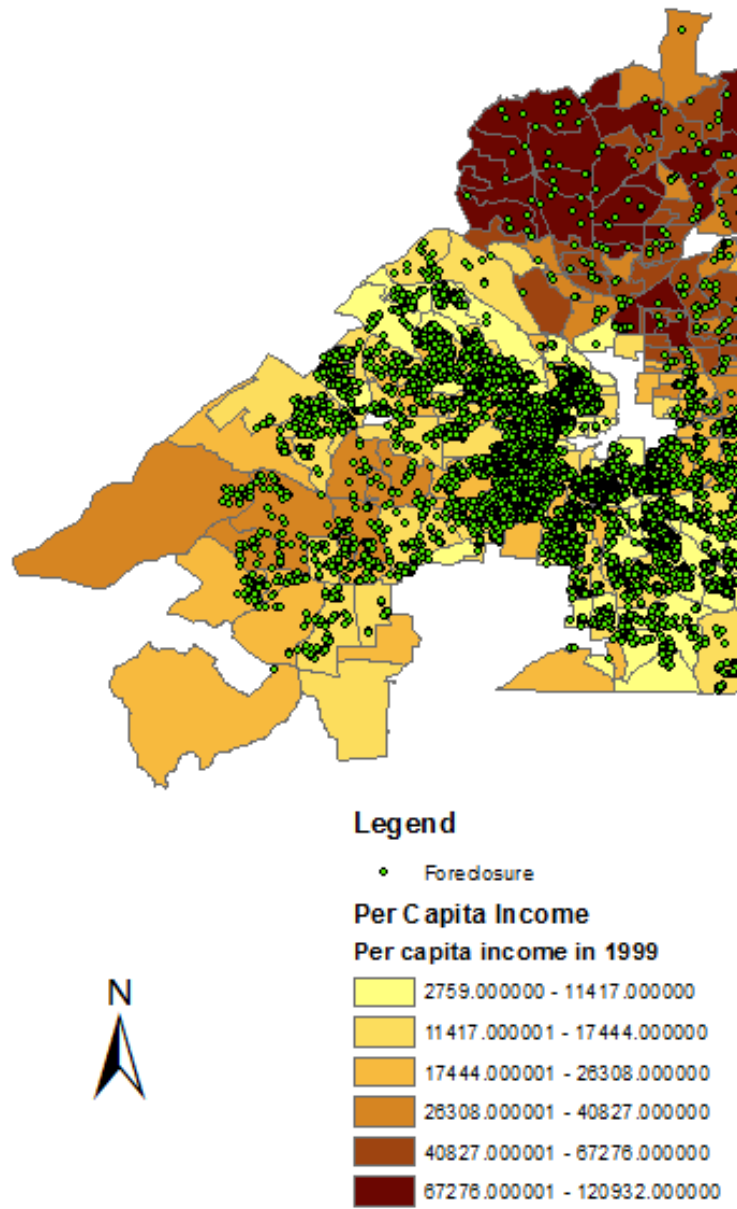
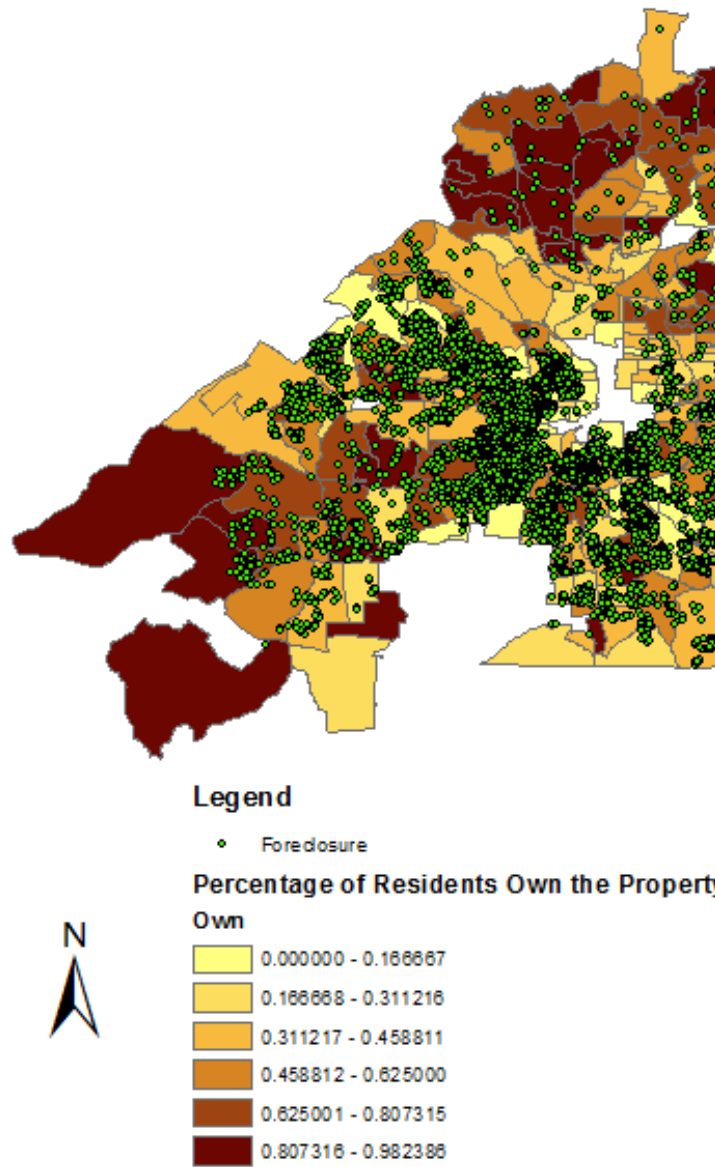


Figure 2.3 Spatial Relationship between Foreclosures and Percent Owner Occupied Homes, Atlanta, 2008



5.2. Foreclosure Effects on House Values

Foreclosure effects are the focus of this study. Table 2.4 reports the estimation results for different regression models. Two models are estimated by OLS, the spatial autoregressive regression, the general spatial regression, and the GS2SLS regression respectively. The reported coefficients of OLS regression are corrected for heterogeneity, and the reported t-statistics are based on robust errors. Model 1 serves as the base for all models, including the foreclosure variables *DIS300*, *DIS600*, *DIS1200* and *DIS1500* as well as house characteristics and neighborhood characteristics variables. It is found that foreclosures take effects within 1500 feet, so another interval between 1500 and 2000 feet, *DIS2000* is added in Model 2.

In Model 1, OLS regression suggests that foreclosure spillover effects extend to 1500 feet. One more foreclosure within 300 feet of a property depresses its sales price by 1.59%, one more foreclosure between 300 feet and 600 feet decreases sales price by 0.66%, one more foreclosure between 600 feet and 1200 feet decreases sales price by 0.35%, and one more foreclosure between 1200 feet and 1500 feet decreases sales price by 0.41%. The spatial autoregressive regression, general spatial regression and the GS2SLS regression also indicate that foreclosure effects extend to 1500 feet, although suggesting a slightly smaller effect than the OLS regression, which makes sense after controlling for the spatial dependency and correlation between spatial errors. The GS2SLS regression presents that one more foreclosure within 300 feet depresses the sales price by 1.57%, one more foreclosure between 300 feet and 600 feet depresses the sales price by 0.54%, one more foreclosure between 600 feet and 1200 feet reduces the sales price by 0.3%, and one more foreclosure between 1200 feet and 1500 feet decreases the sales price by 0.37%.

Model 2 reports smaller coefficient estimates than Model 1 after including 1500 feet to 2000 feet buffer. Both the OLS and GS2SLS regressions find that the foreclosure effect extends only to 1200 feet, which suggests a relatively local spillover effects when compared to the estimates from Model 1. The GS2SLS regression shows sales price decreases by 1.54% when there is one more foreclosure within 300 feet, 0.53% if there is one more foreclosure between 300 feet and 600 feet, and 0.25% if there is one more foreclosure between 600 feet and 1200 feet. Foreclosures do not take effect beyond 1200 feet. The spatial autoregressive regression and general spatial regression report that foreclosures take effect within 2000 feet, but not between 1200 feet and 1500 feet.

In general, the spatial models control for spatial lags and spatial autocorrelation, which produce a better fit than OLS regression, as indicated by higher adjusted R^2 value. Comparing the three spatial models, the spatial autoregressive regression deals with spatial dependency, and the positive spatial lag coefficient ρ indicates that a higher sales price in neighboring buffers exerts a positive influence on average selling price across the entire Atlanta one to four unit family neighborhood sample. The general spatial regression not only controls for spatial dependency, but also deals with spatial autocorrelation in the residuals. The spatial autocorrelation coefficient is statistically significant, which indicates there may be unobserved heterogeneity, which the spatial error can mitigate. In addition, the general spatial regression produces a better fit to the sample data indicated by higher log likelihood values and lower Akaike Information Criterion (AIC). Thus, the general spatial model outperforms the OLS model and the spatial autoregressive model. The GS2SLS estimates are obtained by introducing a set of instruments in a two stage least squares (2SLS) procedure (Keijian and Prucha, 1998). Although the coefficient differences between the general spatial regression estimated by the ML method and GS2SLS method are not

considerable, the GS2SLS procedures can address endogenous spatially lagged dependent variable and produce consistent coefficients by relaxing normality assumption and homogeneity assumption, so GS2SLS regression is preferred to the other spatial regressions.

All the structural variables and neighborhood variables have their expected signs based on previous findings. For example, the GS2SLS results suggest that one more bedroom results in a 10.58% ($13.23\% - 2 * 1.33\%$) house price increase. An additional year of house age decreases the sales price by 1.7% ($-1.73\% + 2 * 0.014\%$), but when the house is older than 62 years, the age begins to have positive effect on the sales price. The *Cdu* is the general dwelling condition for each property. Eight dummy variables are created to distinguish different dwelling conditions, with *Cdu1* being the excellent condition and *Cdu8* being the very poor condition. The coefficients reflect a nonlinear relationship between the dwelling conditions and the sales price. In the regression, the dummy *Cdu8* is deleted, so the coefficients of *Cdu1* to *Cdu7* should be interpreted as the differences from the coefficient of *Cdu8*. In the GS2SLS Model 1, a property in excellent condition (*Cdu1*) on average sells for 22% higher price than one in very poor condition (*Cdu8*) and the difference is statistically significant. A property in very good condition (*Cdu2*) on average sells for 14.5% more than one in very poor condition (*Cdu8*), a property in good condition (*Cdu3*) on average sells for 11.3% more than one in very poor condition (*Cdu8*), the property with average condition (*Cdu4*) sells 4.83% higher price and the property with fair condition (*Cdu5*) sells 0.61% higher price than that with very poor condition (*Cdu8*) although the statistics values are not significant. However, the property with unsound (*Cdu6*) and poor condition (*Cdu7*) sells less than that with very poor condition (*Cdu8*). As the neighborhood characteristics, all the regression models report significant effects of percentage of black residents, household size and per capita income on house sales price. The GS2SLS regression

results report that every 1% increase in the percentage of black residents in a CBG reduces sales prices by 0.16%; one more household member on average in a CBG reduces sales price by 7.8%; and every \$1000 increase in per capita income increases house sales price by 0.29%.

One interesting point is that the coefficient for percentage of home ownership is statistically significant in the OLS model, but not in any of the spatial models. Increasing homeownership does not necessarily increase house price. From the early 1990s to 2005, national homeownership rate increased dramatically due to emergence of innovative subprime loans. Subprime loans helped borrowers who were not qualified for prime mortgages to afford a house. However, due to a subprime loan borrower's low income, impaired credit scores, and a subprime loan's high interest, most borrowers choose to default. Defaults increase the housing supply and decrease house price as a result. Thus, higher homeownership rates within a census block group do not lead to a significantly higher house price.

Table 2.4 Regression Coefficients for Heteroskedasticity – Corrected OLS, Spatial Autoregressive Model, General Spatial Model and GS2SLS Model^a

Variable	OLS		Spatial Autoregressive Model		General Spatial Model		GS2SLS Model	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Intercept	10.50*** (40.08)	10.50*** (40.09)	8.86*** (26.12)	8.88*** (26.15)	9.40*** (61.15)	9.38*** (27.95)	8.54*** (18.02)	8.5394*** (18.03)
DIS300	-0.0159*** (-2.88)	-0.0156*** (-2.82)	-0.0159*** (-3.12)	-0.0156*** (-3.06)	-0.0161*** (-3.06)	-0.0157*** (-2.99)	-0.0157*** (-3.15)	-0.0154*** (-3.10)
DIS600	-0.0066** (-1.99)	-0.0065** (-1.96)	-0.0058* (-1.93)	-0.0057* (-1.90)	-0.006* (-1.94)	-0.0059* (-1.89)	-0.0054* (-1.84)	-0.0053* (-1.82)
DIS1200	-0.0035** (-2.36)	-0.0029* (-1.93)	-0.0033** (-2.38)	-0.0028** (-1.96)	-0.0037** (-2.56)	-0.0031** (-2.09)	-0.003** (-2.20)	-0.0025* (-1.82)
DIS1500	-0.0041** (-2.02)	-0.0024 (-1.14)	-0.0038** (-2.14)	-0.0023 (-1.12)	-0.0038** (-2.06)	-0.0022 (-1.06)	-0.0037** (-2.16)	-0.0023 (-1.15)
DIS2000		-0.002 (-1.47)		-0.0019* (-1.70)		-0.002* (-1.76)		-0.0018 (-1.62)
Street2	-0.34* (-1.70)	-0.34* (-1.69)	-0.34* (-1.96)	-0.34** (-1.96)	-0.33* (-1.92)	-0.33* (-1.92)	-0.34** (-1.97)	-0.34** (-1.97)
Street3	-0.23* (-1.90)	-0.23* (-1.91)	-0.20 (-1.45)	-0.20 (-1.46)	-0.20 (-1.44)	-0.20 (-1.45)	-0.19 (-1.39)	-0.02 (-1.40)
Lotarea	0.002** (2.30)	0.002** (2.29)	0.002** (2.41)	0.002** (2.40)	0.002** (2.38)	0.002** (2.36)	0.002** (2.45)	0.002** (2.44)
Lvarea	0.13*** (8.03)	0.14*** (8.06)	0.13*** (6.94)	0.13*** (6.97)	0.13*** (6.94)	0.13*** (6.94)	0.13*** (6.83)	0.13*** (6.85)
Stories	0.13*** (4.43)	0.13*** (4.37)	0.13*** (3.53)	0.12*** (3.49)	0.13*** (3.89)	0.13*** (3.64)	0.12*** (3.48)	0.12*** (3.44)
Age	-0.02*** (-13.21)	-0.02*** (-13.25)	-0.02*** (-8.75)	-0.02*** (-8.79)	-0.02*** (-16.90)	-0.02*** (-9.44)	-0.02*** (-11.34)	-0.02*** (-11.39)
Age2	0.00015*** (10.34)	0.00015*** (10.40)	0.00014*** (6.69)	0.00014*** (6.73)	0.00014*** (12.73)	0.00014*** (7.20)	0.00014*** (8.66)	0.00014*** (8.72)
Rmbed	0.15*** (3.56)	0.15*** (3.55)	0.13*** (3.48)	0.13*** (3.48)	0.13*** (3.54)	0.13*** (3.44)	0.13*** (3.47)	0.13*** (3.47)

Rmbed2	-0.015*** (-2.76)	-0.015*** (-2.76)	-0.014*** (-2.64)	-0.014*** (-2.64)	-0.014*** (-2.69)	-0.014*** (-2.61)	-0.014*** (-2.62)	-0.013*** (-2.62)
Fixbath	0.06*** (3.34)	0.06*** (3.36)	0.05*** (2.72)	0.05*** (2.74)	0.05*** (2.79)	0.05*** (2.74)	0.05*** (2.70)	0.05*** (2.72)
Fixhalf	0.04* (1.71)	0.04* (1.70)	0.03 (1.33)	0.03 (1.32)	0.03 (1.30)	0.03 (1.29)	0.03 (1.31)	0.03 (1.30)
Bsmt2	0.05 (1.37)	0.05 (1.41)	0.06 (1.62)	0.06* (1.66)	0.05 (1.41)	0.05 (1.45)	0.06* (1.85)	0.06* (1.89)
Bsmt3	0.13*** (3.16)	0.13*** (3.17)	0.12*** (3.05)	0.12*** (3.07)	0.11*** (2.80)	0.12*** (2.83)	0.13*** (3.22)	0.13*** (3.24)
Bsmt4	0.14*** (3.68)	0.14*** (3.70)	0.14*** (3.50)	0.14*** (3.53)	0.13*** (3.28)	0.13*** (3.30)	0.14*** (3.66)	0.14*** (3.68)
Heat2	-0.08 (-0.97)	-0.08 (-0.95)	-0.09 (-1.35)	-0.09 (-1.33)	-0.09 (-1.41)	-0.09 (-1.38)	-0.09 (-1.29)	-0.08 (-1.27)
Heat3	-0.03 (-0.36)	-0.03 (-0.34)	-0.03 (-0.56)	-0.03 (-0.54)	-0.04 (-0.62)	-0.04 (-0.60)	-0.03 (-0.48)	-0.03 (-0.47)
Heat4	0.05 (0.64)	0.05 (0.66)	0.04 (0.61)	0.04 (0.63)	0.03 (0.56)	0.03 (0.58)	0.04 (0.65)	0.04 (0.67)
Attic2	0.08** (2.03)	0.08** (2.06)	0.08* (1.84)	0.08* (1.86)	0.08* (1.86)	0.08* (1.88)	0.08* (1.80)	0.08* (1.82)
Attic3	0.01 (0.27)	0.01 (0.27)	0.01 (0.26)	0.01 (0.26)	0.02 (0.36)	0.02 (0.35)	0.01 (0.18)	0.01 (0.18)
Attic4	0.07 (1.51)	0.07 (1.49)	0.07 (1.45)	0.07 (1.43)	0.07 (1.49)	0.07 (1.46)	0.07 (1.40)	0.07 (1.38)
Attic5	0.22*** (3.60)	0.22*** (3.55)	0.20** (2.14)	0.20** (2.11)	0.21** (2.18)	0.21** (2.14)	0.19** (2.05)	0.19** (2.01)
Cdu1	0.28** (2.33)	0.28** (2.34)	0.23** (2.07)	0.23** (2.08)	0.23** (2.21)	0.23** (2.12)	0.22** (2.05)	0.22** (2.06)
Cdu2	0.20* (1.78)	0.20* (1.80)	0.15 (1.46)	0.15 (1.47)	0.16 (1.58)	0.16 (1.52)	0.14 (1.43)	0.15 (1.44)
Cdu3	0.16 (1.45)	0.16 (1.45)	0.12 (1.14)	0.19 (1.14)	0.12 (1.23)	0.12 (1.17)	0.11 (1.12)	0.11 (1.12)
Cdu4	0.09	0.09	0.05	0.05	0.05	0.05	0.05	0.05

	(0.78)	(0.81)	(0.50)	(0.51)	(0.54)	(0.54)	(0.49)	(0.50)
Cdu5	0.05	0.05	0.01	0.01	0.01	0.01	0.01	0.01
	(0.42)	(0.43)	(0.09)	(0.10)	(0.12)	(0.12)	(0.06)	(0.07)
Cdu6	-0.25**	-0.25**	-0.29**	-0.29**	-0.29***	-0.29**	-0.29***	-0.29***
	(-2.03)	(-2.03)	(-2.54)	(-2.54)	(-2.69)	(-2.57)	(-2.58)	(-2.58)
Cdu7	-0.07	-0.07*	-0.11	-0.11	-0.11	-0.11	-0.11	-0.11
	(-0.50)	(-0.49)	(-0.89)	(-0.89)	(-0.94)	(-0.89)	(-0.92)	(-0.91)
Black	-0.25**	-0.24**	-0.18*	-0.18*	-0.19**	-0.19*	-0.16*	-0.15*
	(-2.50)	(-2.43)	(-1.95)	(-1.89)	(-2.01)	(-1.91)	(-1.75)	(-1.69)
Hsize	-0.10***	-0.10***	-0.08**	-0.08**	-0.09**	-0.09**	-0.08**	-0.08**
	(-2.79)	(-2.77)	(-2.25)	(-2.24)	(-2.33)	(-2.19)	(-2.25)	(-2.23)
Own	0.10*	0.11*	0.09	0.09	0.09	0.10	0.08	0.08
	(1.72)	(1.76)	(1.32)	(1.36)	(1.35)	(1.38)	(1.21)	(1.25)
Income	0.0046***	0.0046***	0.0035**	0.0035**	0.0039**	0.0038**	0.0029**	0.0029**
	(3.98)	(3.92)	(2.41)	(2.37)	(2.55)	(2.49)	(2.07)	(2.02)
Old	-0.21	-0.21	-0.18	-0.18	-0.20	-0.20	-0.15	-0.15
	(-1.06)	(-1.06)	(-0.85)	(-0.86)	(-0.91)	(-0.91)	(-0.77)	(-0.77)
HSI	0.003	0.002	-0.0004	-0.001	-0.0005	-0.0016	-0.0004	-0.0014
	(0.24)	(0.15)	(-0.03)	(-0.07)	(-0.04)	(-0.11)	(-0.03)	(-0.10)
W ₀₇ P ₀₇	0.009	0.009	0.003	0.003	0.006	0.005	-0.0005	-0.0006
	(0.49)	(0.48)	(0.16)	(0.16)	(0.31)	(0.29)	(-0.03)	(-0.03)
W ₀₆ P ₀₆	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
	(0.58)	(0.62)	(0.46)	(0.49)	(0.53)	(0.55)	(0.38)	(0.41)
W ₀₅ P ₀₅	0.03*	0.03*	0.03	0.03	0.03	0.03	0.02	0.02
	(1.68)	(1.67)	(1.26)	(1.25)	(1.24)	(1.22)	(1.14)	(1.13)
W ₀₄ P ₀₄	0.008	0.008	0.01	0.01	0.01	0.01	0.016	0.016
	(1.15)	(1.13)	(1.30)	(1.28)	(1.09)	(1.08)	(1.60)	(1.58)
W ₀₃ P ₀₃	-0.01	-0.001	-0.003	-0.003	-0.006	-0.006	0.001	0.001
	(-1.35)	(-1.35)	(-0.26)	(-0.26)	(-0.46)	(-0.45)	(0.08)	(0.08)
Sale1	0.24***	0.24***	0.23***	0.23***	0.23***	0.23***	0.22***	0.22***
	(4.79)	(4.81)	(4.43)	(4.45)	(4.56)	(4.54)	(4.30)	(4.32)
Sale2	-0.82***	-0.82***	-0.83***	-0.83***	-0.84***	-0.83***	-0.83***	-0.83***
	(-4.11)	(-4.09)	(-10.11)	(-10.08)	(-10.16)	(-10.10)	(-10.06)	(-10.03)

Sale3	0.03 (0.45)	0.03 (0.48)	0.01 (0.14)	0.015 (0.16)	0.016 (0.17)	0.018 (0.20)	0.006 (0.06)	0.008 (0.08)
Sale4	-0.52*** (-8.19)	-0.52*** (-8.19)	-0.52*** (-8.98)	-0.52*** (-8.96)	-0.52*** (-8.94)	-0.52*** (-8.91)	-0.52*** (-9.00)	-0.52*** (-8.98)
Sale5	-0.30*** (-5.73)	-0.30*** (-5.72)	-0.30*** (-6.23)	-0.30*** (-6.23)	-0.30*** (-6.16)	-0.30*** (-6.15)	-0.30*** (-6.23)	-0.30*** (-6.22)
Sale6	-0.24 (-1.60)	-0.24 (-1.60)	-0.24** (-2.22)	-0.24** (-2.22)	-0.24** (-2.22)	-0.24** (-2.32)	-0.24** (-2.21)	-0.24** (-2.21)
Sale7	0.48* (1.85)	0.46* (1.89)	0.49 (0.95)	0.47 (0.93)	0.52 (1.01)	0.51 (0.99)	0.46 (0.90)	0.45 (0.87)
Quarter1	0.32*** (10.34)	0.31*** (9.61)	0.33*** (10.86)	0.32*** (10.18)	0.33** (10.56)	0.31*** (9.88)	0.34*** (11.10)	0.32*** (10.43)
Quarter2	0.19*** (7.07)	0.18*** (6.68)	0.19*** (7.17)	0.19*** (6.82)	0.19*** (7.05)	0.19*** (6.68)	0.20*** (7.28)	0.19*** (6.94)
Quarter3	0.19*** (7.21)	0.18*** (7.04)	0.19*** (7.39)	0.18*** (7.22)	0.19*** (7.38)	0.18*** (7.20)	0.19*** (7.39)	0.19*** (7.23)
ρ			0.15***	0.15***	0.10***	0.11***	0.23***	0.22***
λ					0.08***	0.07***	-0.07**	-0.06**
Adj. R ²	0.58	0.58	0.58	0.58	0.58	0.58	0.58	0.58
Log-likelihood			-9537.99	-9536.53	-9534.76	-9533.18		
AIC			19287.98	19287.06	19281.52	19280.36		

^a School district dummy variable parameter estimates not reported here for sake of space.

Note: ***Statistically significant at 1%; **Statistically significant at 5%; *Statistically significant at 10%.

The asymptotic t statistics are in parentheses.

6. Tax Loss Estimates

One of the direct effects of foreclosures is a reduction of house values depresses local tax revenues. Because property taxes fund local public goods, property taxes losses would depreciate welfare of local economy. This section conducts an analysis of how tax collections will be impacted by changes in marginal implicit prices. Property taxes for Atlanta are calculated by subtracting the homestead exemption from 40% of the purchase price of the home, dividing that by 1000 and multiplying that amount by the county millage rate. The homestead exemption in Fulton county is \$15,000, so the property tax is calculated as follows,

$$\text{Property Tax} = \frac{40\% * \text{sales price} - 15000}{1000} * 50.91 \quad (2.11)$$

The tax loss is thus calculated as

$$\text{Property Tax Loss} = \frac{40\% * \text{sales price reduction} - 15000}{1000} * 50.91 \quad (2.12)$$

The marginal effect of foreclosures on surrounding house prices is computed using the parameters of the GS2SLS regression Model 1 from Table 2.4.

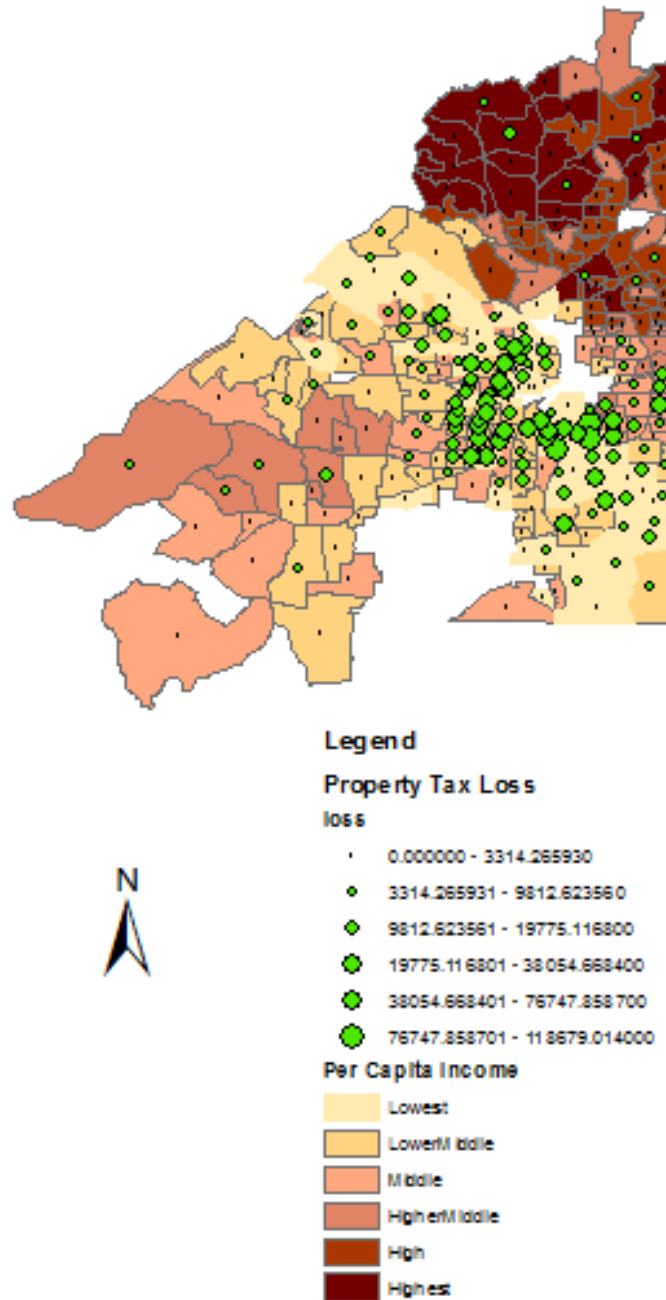
$$\frac{\partial \text{ sales price}}{\partial \text{ foreclosure}} = \beta_4 * \text{sales price} + \beta_5 * \text{sales price} + \beta_6 * \text{sales price} + \beta_7 * \text{sales price} \quad (2.13)$$

The estimated property tax loss for 10,121 one-to-four unit family houses in Atlanta is about \$2.2 million in 2008. It is hypothesized that the biggest losers will likely be in the poorest neighborhoods where foreclosures tend to be concentrated. Figure 2.5 shows the spatial relationship between property tax loss and per capita income, confirming that the biggest tax losers are those census block groups with lower per capita income.

However, this analysis underestimates the foreclosure effects significantly, as it only includes one to four unit family houses but does not include larger multi-family houses, commercial or industrial properties. In addition, relocation costs in the event of a foreclosure and

other transaction costs are not considered. The benefits from foreclosure reduction would be expected to be higher if a full sample of spectrum is included.

Figure 2.5 Spatial Relationship between Property Tax Loss and Per Capita Income, Atlanta, 2008



7. Conclusion and Policy Implication

By using a unique dataset, this study examines foreclosure impacts on neighborhood property values in Atlanta using one to four unit family houses. The spatial distribution of the foreclosure pattern is also analyzed. The OLS results, spatial autoregressive regression, general spatial regression and GS2SLS regression are analyzed and compared. The general spatial model controls for spatial lags and controls for spatial error to avoid omitted variable problem. The GS2SLS regression is more appealing when the residuals are heteroskedastic and when the finite samples do not meet the normality requirement. The foreclosure effects extend up to 1500 feet of a property. The results present a slight larger spillover effects when compared to other studies. The marginal foreclosure impact is -1.57% within 300 feet, - 0.54% between 300 feet and 600 feet, -0.3% between 600 and 1200 feet, and -0.37% between 1200 feet and 1500 feet. Immergluck and Smith (2006) find that the marginal foreclosure impact in Chicago City in 1999 is -1.14% within 1/8 mile (about 600 feet) and -0.33% between 1/8 and 1/4 mile (about between 600 feet and 1200 feet). Roger and Wither (2009) find that the marginal foreclosure impact on single-family sales in Louis County, Missouri, is about -1% of sales price within 200 yards (about 600 feet) from 2000-2007. Leonard and Murdoch (2009) find that foreclosures within 250 feet of a sale depreciate selling price by 0.5%, and foreclosures between 250 feet and 500 feet decrease sales price by 0.1% in Dallas County, Texas, in year 2006. The larger effect of foreclosures may be due to the faster speed of foreclosures in Atlanta caused the higher percentage of foreclosures.

Although endogeneity problems are not found in this study after testing our equations systems with the Hausman test, the zero-inflated negative binomial regression results explain the

reasons for number of foreclosures and ArcGIS program presents the foreclosure patterns through maps of foreclosure activity.

Foreclosures are more likely to be concentrated in low-income and minority districts, which is consistent with results of previous studies finding low-income and minority borrowers are more likely to have subprime loans. A sales property surrounded by houses in good condition has a greater number of foreclosures within the buffer than a sales property surrounded by houses in fair or poor condition. The number of foreclosures within a buffer increases if the surrounding houses' living area decreases and if the average age of surrounding houses is greater. Larger household size decreases the risk of foreclosures. Higher percentage of home ownership decreases the number of foreclosures within 300 feet, but increases the number of foreclosures beyond 600 feet. Towe and Lawley (2010) also found that an increase in the percentage of owner-occupied units in a neighborhood has a positive impact on the hazard rate of foreclosure. The reason is maybe because there are more houses have potential to be foreclosed on with higher home ownership while renters probably do not foreclose that fast.

The price difference between 2007 and 2008 also affects the number of foreclosures. Falling house prices between 2007 and 2008 caused by previous foreclosures creates a climate in which even more foreclosures occur, since subprime loans borrowers may lose confidence to the market and choose to default.

The percentage of subprime loans made from 2004 to 2007 significantly increases foreclosures in all buffers. High-cost and high-leverage loans have the greatest effects on the number of foreclosures. The sales price regressions show that higher home ownership in a neighborhood does not significantly improve the house price. Thus, the increasing home ownership resulting from access to subprime loans in recent years has a deleterious effect on

sustainability in the housing market. On the contrary, a portion of houses were foreclosed on as a result of excess speculation in the housing market. From the results, policy makers should consider programs to make home ownership consistent with a borrower's economic situation and credit history, as well as program to quickly resolve the large inventory of foreclosed houses on the market. Banks should be more careful about the subprime loans lending standards, but should also avoid refusing responsible borrowers in low-income and high-minority neighborhoods. Hartarska and Gonzalez-Vega (2006) find that the credit counseling program, which helps low-income borrowers to estimate the amount of debt they will be able to service, can help decrease mortgage loan default rate.

The Georgia law increases foreclosures quickly, which speeds up the crisis. Although foreclosures can be put back on the market sooner than in other states, most foreclosures are sold at a greatly discounted price. It may also reduce surrounding house sales prices and thus affect property tax collection. Because property taxes fund local public goods, losses in the property taxes revenues would have a multiplier impact in degrading provision of local public goods. The estimated property tax loss for 10,121 one-to-four unit family houses at Atlanta is about \$2.2 million in 2008. If the full spectrum of houses types and foreclosures were considered, reducing foreclosures would result in an even higher social benefit. The result also confirms that the biggest tax losers are those census block groups with lower per capita income where more foreclosures are likely to occur.

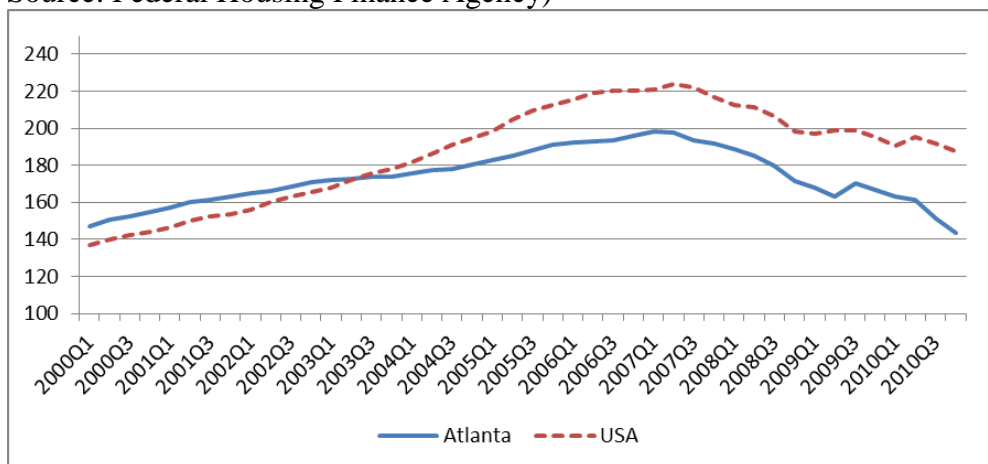
CHAPTER 3

The Contagion Effect of Foreclosures: A Quasi-Experiment Method

1. Introduction

The financial crisis caused by subprime loans started in 2006 and became apparent in 2007. The number of foreclosures in 2007 was 79% higher than in previous years nation-wide (Veiga 2008; Rogers and Winter 2009). Defaults depress housing market, and foreclosures reduce surrounding house values (Immergluck and Smith 2006; Schuetz, Been and Ellen 2008; Lin, Rosenblatt and Yao 2009; Leonard and Murdoch 2009). Figure 2.1 shows the historical quarterly home sales price index for Atlanta-Sandy Springs-Marietta area versus the national average. From the third quarter of 2003, the housing price index in Atlanta is lower than the national average, and the price difference became larger in the following years. In the Atlanta area, the home sales price peaked historically in mid-2007, then decreased dramatically since the second quarter of 2007 (Federal Housing Finance Agency).

Figure 3.1 Historic Quarterly Home Sales Price Index, Atlanta MSA, Seasonally Adjusted (Data Source: Federal Housing Finance Agency)



Previous studies indicate that foreclosures depreciate neighborhood house sales price. However, there is difficulty over time in establishing how much of the depreciation is caused by foreclosures, and how much is caused by macroeconomic or regional trends. The problem is that there could be a trend over time that the prices of all the houses in a neighborhood, not only houses in foreclosure infestation areas, are declining due to external economic conditions. The question is that whether the foreclosures cause a decline in sales price through spillover effects or all the houses in a neighborhood experience price decline due to external economic conditions. Thus, it is important to identify whether the foreclosures or the time trend caused depreciation in housing sales price.

Difference-in-differences model can not only remove biases from comparisons between the treatment and control group that could be the result from permanent differences, but also can remove biases from comparisons over time in the treatment group that could be the result of trends (Wooldridge 2007). Like the difference-in-differences method, propensity score matching (PSM) can also eliminate selection bias between treatment and control groups by matching treatment and control units based on a set of covariates. This study employs both a difference-in-differences model and propensity score matching to study the effects of foreclosures on the neighborhood property values in the city of Atlanta from 2000-2010.

In addition, this study also distinguishes between real estate owned (REO) property and REO sales. REO occurs when the borrower misses mortgage payments and the property becomes owned by banks, government agencies or mortgage institutions; REO sale occurs when bank sells the foreclosed property to individuals or investment corporations. There are very few literatures distinguish REO from REO sales and they did not incorporate both variables in the model to study the foreclosure effects. The hypothesis is that both REO and REO sales will

decrease surrounding houses' sales price. REO reduces surrounding houses' sales price by introducing disordered communities. REO sale reduces surrounding houses' sales price because house prices are usually set by comparables in the neighborhood. In order to quickly resolve REOs, Banks usually sell houses to individual or companies at a great discount, so there is a direct negative spillover effect.

2. Literature Review

There are a few recent studies addressing the effects of foreclosure on housing values using the hedonic price model. Most studies are done by employing cross-sectional data, while a few of them use panel data or repeat sales data.

Immergluck and Smith (2006) combine foreclosure data from 1997 and 1998 with neighborhood demographic characteristics data and more than 9,600 single family property transactions in Chicago in 1999. After controlling for forty characteristics of properties and their respective neighborhoods, they find that foreclosure of conventional single-family loans have a significant impact on nearby property values. Each conventional foreclosure within a 1/8 mile (about 660 feet) of a single-family home results in a decline of 0.9% in value.

Leonard and Murdoch (2009) use four models to examine the foreclosure impacts on single-family home values in and around Dallas County, Texas, in 2006. Maximum likelihood estimation and GMM estimation are used to examine and compare the spatial lag model and general spatial model. Regression results suggest that foreclosures within 250 feet, between 500 and 1000 feet and between 1000 and 1500 feet of a sale depreciate surrounding neighborhood selling prices.

Lin, Rosenblatt and Yao (2009) use 20% of the mortgages made in the United States from 1990 to 2006 to examine the spillover effects of foreclosures on neighborhood property

values in Chicago metropolitan area in 2003 and 2006. They also apply the Heckman (1979) two-step model to correct for sample selection bias. The researchers use both loan characteristics and borrower characteristics to examine foreclosure status. Though there is a statistically significant bias, the effects on the hedonic model are quite small. The results show that spillovers take effect within ten blocks and for up to five years from the foreclosure. The effect decreases as time passes and as space between the foreclosure and the subject property increases. In addition, foreclosures reduced surrounding 2006 house values by half compared to prices during the boom period in 2003.

Schuetz, Been and Ellen (2008) use property sales and foreclosure filings data in New York City from 2000 to 2005 to examine foreclosure effects on neighborhood property values. Nine time and distance intervals were created to measure foreclosure effects according to the foreclosure filing timeline. Regression results show that properties close to the foreclosures sell at a discount, and the magnitude of price discount increases with the number of nearby foreclosures, but not linearly.

Rogers and Winters (2009) apply a hedonic price model to study foreclosure impacts on nearby property values in St. Louis County, Missouri, using single-family sales data from 2000-2007 and foreclosure data from 1998-2007. They adjust the model to account for spatial autocorrelation, since foreclosures are usually spatially clustered. The ten nearest neighboring sales are used to construct the spatial weight matrix. This study supports the hypotheses that foreclosure impacts decrease as distance and time between the house sale and foreclosure increase. The results show a similar magnitude of foreclosure impacts compared to Immergluck and Smith's (2006) study, but a much smaller foreclosure impact compared to Lin, Rosenblatt and Yao's (2009) study.

Campbell, Giglio and Pathak (2009) use housing transaction data from Massachusetts over the last twenty years to examine the effect of foreclosures on house prices at zip code level. The results show that the average discount resulting from a forced sale is 28% of house value. Unforced sales take place at efficient prices, while forced sale prices reflect time-varying illiquidity in neighborhood housing markets. It also shows that the house price is reduced by 1% when there is a foreclosure at a distance of 0.05 mile.

One limitation of these studies is that they do not distinguish whether foreclosures themselves lead to depreciated surrounding housing price, or the external economic conditions reduce the sales price around the whole neighborhood. There is only one paper mentioning this issue. Harding, Rosenblatt and Yao (2009) use repeat sales model to examine the effect of foreclosures on property values. However, for each sales property they arbitrarily pick just two repeated sales during study period 1989-2007, the first sale had to occur in 1990 or later and the resale had to be completed in 2007 or earlier. Because housing is sold randomly, some houses are sold once, while others may be sold more than 3 times during the study period. In their study, a number of observations are deleted in the analysis, and not every market transaction is available to estimate foreclosure effects. In addition, the authors do not mention how the two repeat sales are selected for each sales property. For example, if a house is sold three times in 2006, 2007 and 2008 respectively, why do they choose sales in 2006 and 2007 as repeat sales but not those in 2006 and 2008 as repeat sales?

In contrast, this paper includes every valid sale that sold twice or more during the study period. Every house is transacted at a different time, so this dataset comprises an unbalanced panel.

3. Data

The transaction and foreclosure data are from the Board of Tax Assessors in Fulton County, Georgia. It should be mentioned that a small part of Atlanta belongs to DeKalb County. Because Fulton County and Dekalb County use different variables and codes to record properties sales, it is difficult to combine the datasets. This study only includes the Atlanta sales in Fulton County. The sales data from 2000 and 2010 include all the transaction information we need for the analysis, including sales price, sales type, sales date, address, and the seller and buyer name, which helps to identify real estate owned (REO) property and REO sales. Because the sales data and the housing characteristics data are in different files, I combine the datasets according to housing parcel id to get all the information needed for analysis. There are 74,424 transactions in total from 2000 to 2010, including valid sales, REOs and REO sales. The REOs are either transacted at a price of 0 when the borrower has less equity in the property than the amount owed to the bank or is transacted at price great than 0 when the borrower has more equity in the property than the amount owed to the bank. The REOs should be deleted from the transaction data because they are not real sales and do not reflect the housing values. After deleting the REOs and observations with missing variables, there are 26,352 one to four unit family house sales records and all of them are sold twice or more.

The REO occurs if the individual sells the property to the bank, government agency or mortgage institutions when the borrower cannot pay the loans. Then the bank usually sells the property to other investment companies or individuals, which is REO sale. For some properties, the REO sale transacted more than once during the years. Also, the REO sale could become the REO again in subsequent years. To calculate the number of foreclosures nearby, each property's sales date and foreclosure date are used to determine how many REOs and REO sales in each

buffer. For example, a REO occurs on June 20th, 2008, it is then sold to an individual on July 30th, 2008. If there is a house within 300 feet sold on August 2nd, 2008 then I identify there is one REO sale nearby. However, if the house sold on June 25th, then I identify there is one REO nearby.

Each record is geocoded to get the longitude and latitude using ArcGIS according to the property address, then intersected (overlaid) with census block group to identify which census block group the property belongs to, and intersected with Atlanta school zones to identify which school zone the property belongs to. In this study, census block group level data is used as a proxy for neighborhood characteristics. Thus, the sales records can be merged with neighborhood characteristics according to the census block group id number. The neighborhood characteristics data are from the 1990 and 2000 Census Bureau. Neighborhood characteristics include the average per capita income, percentage of black residents, average household size, percentage of residents who own the property, and percentage of residents over 65 years old.

This study uses the difference-in-differences methodology to estimate the effects of foreclosures on property values. Since foreclosures are expected to have fewer impacts on sales at a greater distance, buffer rings for each sales record were created. *DIS300* is the number of foreclosures within 300 feet of sales occurred before sales. *DIS600* is the number of foreclosures within 600 feet of sales occurred before sales. *DIS900* is the number of foreclosures within 900 feet of sales occurred before sales. *DIS1200* is the number of foreclosures within 1200 feet of sales occurred before sales.

The Georgia law permits lenders to declare a borrower in default and reclaim a house in as little as sixty days, which speeds up the foreclosure process and puts foreclosures back on the market quickly. The dataset shows that most of the REOs are sold by bank or mortgage

institutions within one year and usually sell at a discounted price, which is much lower than surrounding house values. If the REOs are bought by an investment companies, the investment companies usually can sell the house to the market at a much higher price than the REO sales price made by the banks.

4. Model

The hedonic price model is a commonly used method for studying housing values. House price is usually affected by its own physical characteristics and its location. Structural characteristics, socio-demographic factors, and location to specific amenities or disamenities are included in the hedonic model to attribute positive or negative effects on house prices. The hedonic model estimated by OLS regression serves as the baseline for analysis. Difference-in-differences and propensity score matching methods will then be used for identifying the effects of foreclosures on property values.

4.1. Difference-In-Differences (DID)

The difference-in-differences (DID) model can not only remove biases from comparisons between the treatment and control group that could be the result from permanent differences, but also can remove biases from comparisons over time in the treatment group that could be the result of trends (Wooldridge 2007). Although previous studies indicate that foreclosures depreciate neighborhood house sales prices, there is a problem in that there could be a trend over time that the prices of all the houses in a neighborhood, not only houses in foreclosure infestation areas, are declining due to external economic conditions. The question is that whether the foreclosures cause a decline in sales price through spillover effects or all the houses in a neighborhood experience price decline due to external economic conditions.

Repeat sales data are powerful for estimating the foreclosure effect. Because number of foreclosures is the interest of this study, and the treatment effect varies with different number of foreclosures. With many time periods and arbitrary treatment patterns, we can use

$$P_{it} = \lambda_t + \tau NF_{it} + \gamma X_{it} + c_i + d_i + u_{it}, t = 1, 2, \dots, T \quad (3.1)$$

where P_{it} is a vector of property sales price from year 2000 to 2010 deflated by year consumer price index (CPI) expressed as the natural log form; λ_t is a full set of time effects, representing the overall market price level at time t ; NF_{it} is the number of foreclosures within certain buffers; X_{it} are variables that change over time to affect sales price, which include vectors of property characteristics⁶, sales type, and sales quarter dummy variables; c_i is observed characteristics that do not change over time; and d_i is unobserved characteristics that do not change over time. Estimation by fixed effects can absorb observed and unobserved invariant characteristics c_i and d_i across time, provided the number of foreclosures, NF_{it} , is strictly exogenous.

Bertrand, Duflo and Mullainathan (2004) point out that conventional DID standard errors are understated due to a serial correlation problem. One factor causing serial correlation is that the treatment variable itself changes very little within a state over time. However, they find that the serial correlation problem can be eliminated by randomly choosing ten treatment dates between study years, instead of just choosing one date after which all the states in the treatment group are affected by the treatment. If the observation relates to a state that belongs to the treatment group at one of these ten dates, the law is defined as 1, 0 otherwise. In other words, the intervention variable is now repeatedly turned on and off, so its value in one year tells us nothing about its value the next year. In this study, the treatment is the number of foreclosures within certain buffers. Because the number of foreclosures changes over time, a property not

⁶ In Harding et al. (2008)'s paper, they assume that the property characteristics are fixed between sales in the standard repeat sales methodology. However, the characteristics of some houses may change due to remodeling or innovation. This study incorporates dummy variables to control these changes in the model.

surrounded by foreclosures at initial time may experience foreclosures in its neighborhood next year, but when the foreclosures are resolved by a bank or third party, the property would not be exposed to the nearby foreclosures. In other words, the treatment is repeatedly turned on and off. Thus, the serial correlation problem can be avoided in this model.

4.2. Propensity Score Matching (PSM)

Propensity score matching (PSM) is often used in observational studies where subjects are not randomly assigned to treatment and control groups. Randomization may assure that treatment and control groups have identical characteristics, so that the differences between the groups after applying a treatment can be attributed to the treatment effect. However, when the subjects are not randomly assigned to groups, it causes causal inference complicated because we do not know whether the differences of outcome come from the treatment itself or is a product of differences among treatment group and control group. The propensity score matching method can achieve randomized experiment effect by making the characteristics of subjects in treatment and control groups close to identical. The idea is to estimate the probability (the propensity score) that a subject would be assigned to the treatment given certain characteristics. If a treated subject has the same propensity score as a control subject, then the difference between them is the result of the treatment effect itself.

Previous studies indicate that foreclosures are more likely to occur in low-income and minority districts. In other words, houses located in foreclosure infestation areas and non-foreclosure infestation areas have systematically different characteristics, with respect to housing characteristics and neighborhood characteristics. We could match the propensity score for subjects in treatment and control groups to make sure they have similar probability of being surrounded by foreclosures.

The first step is to estimate the propensity score for the sales property surrounded by foreclosures. It is estimated by logistic regression in which the dependent variable is *DIS300*, indicating whether there are foreclosures within 300 feet. The treatment *DIS300* equals to 1 if there are there are 1 or more than 1 foreclosures within 300 feet. In particular, 68% of the sales properties in my sample from year 2000 to 2010 were affected with foreclosures.

$$DIS300 = f(M, H, N, Y) + \varepsilon \quad (3.2)$$

where *M* is a vector of mortgage characteristics, including percentage of low-cost/high-leverage mortgage, high-cost/low-leverage mortgage, and high-cost/high-leverage mortgage in census tract; *H* is a vector of the house characteristics, including lot size, living areas, and house age; *N* is a vector of census block group socio-demographic characteristics, including the average per capita income, percentage of black residents, average household size, percentage of residents who own the property, and percentage of residents over 65 years old; *Y* is a vector of year dummies.

After estimating the propensity score, there are various methods for matching the scores between treatment and control groups. The most commonly used matching methods include the nearest available neighbor and caliper matching. In the nearest available neighbor method, the treatment unit is selected to find the closest control match if the absolute value of the difference between their propensity scores are the smallest. The procedure is repeated for all the treated units. If there is a replacement, then the matched control unit can be selected again to match other treatment units. Otherwise, once it is matched, it will not be considered for matching against other treatment units. The nearest available neighbor method guarantees that all the treated units can find their control matches even if their propensity scores are not close enough.

Caliper matching is similar to the nearest available neighbor matching method but it adds an additional restriction (Coca-Peraillon, 2006). The treated unit is selected to find its closest control match based on the propensity score but only if the control unit's propensity score is within a certain radius. Thus, it is possible that not all the treated units will be matched to control units, but the method can avoid bad matches.

5. Results

Table 3.1 reports the descriptive statistics for the variables in the model. The house characteristics include lot size in square feet, living area in square feet, number of stories, age of the house, number of bedrooms, number of full and half bathrooms, basement and attic condition, heat type, overall dwelling condition, and the street condition for each parcel.

In addition to structural house characteristics, many neighborhood characteristics may affect house value. Percentage of black residents, average household size, percentage of residents who own a house, and percentage of people over 65 years old in each CBG are chosen to represent neighborhood quality.

The effect of foreclosures is the focus of this study. The DID specification is used to estimate the effect of foreclosures on property values. The distance interval is created for each property. *DIS300* is a vector of the number of foreclosures (including both REO and REO sales) within 300 feet of sales occurring before sale, *DIS600* is a vector of the number of foreclosures within 600 feet of sales occurring before sale, *DIS900* is a vector of the number of foreclosures within 900 feet of sales occurring before sale, and *DIS1200* is a vector of the number of foreclosures within 1200 feet of sales occurring before sale. The average number of foreclosures is 3.12 within 300 feet of a sales property, 9.91 within 600 feet of a sales property, 19.04 within 900 feet of a sales property, and 30.6 within 1200 feet of a property.

In this study, I also distinguish between REO and REO sales. The hypothesis is that both REO and REO sales will depreciate neighborhood house sales price. REO reduces surrounding house sales price because it causes disordered community, the vacant houses usually attract criminals and rodent animals, which may cause both house physical appearance deterioration and lead to mental stress for surrounding neighbors. REO sales could also depreciate its surrounding neighborhood house sales price, because the price of the subject property is determined by its comparable properties that are recently sold and close to the subject property (Lin et al., 2009; Vandell, 1991). Banks usually sell REO to individuals or investment companies with a great discount. The discount price thus has a spillover effect to surrounding sales price. The average number of REOs is 1.1 within 300 feet, 3.49 within 600 feet, 6.71 within 900 feet, and 10.82 within 1200 feet. The average number of REO sales is 2.02 within 300 feet, 6.43 within 600 feet, 12.33 within 900 feet, and 19.78 within 1200 feet.

Table 3.1 Descriptive Statistics for One to Four Unit Family Sales House Characteristics and Neighborhood Characteristic, Atlanta, 2000-2010 (N=26,352)

Variable	Description	Mean	SD	Minimum	Maximum
Price	Sales price (*\$1000)	180.99	289.51	1	6,499
Street1	The street condition in the parcel is "paved"	0.99	0.08	0	1
Street2	The street condition is "semi-improved"	0.002	0.05	0	1
Street3	The street condition is "dirt"	0.004	0.06	0	1
Lotarea	Lot area sqft (*1000)	10.91	10.05	0.06	387.68
Lvarea	Living area sqft (*1000)	1.61	0.99	0.12	24.55
Stories	Number of stories	1.19	0.39	1	3
Rmbed	Number of bedrooms	2.98	0.96	0	14
Fixbath	Number of full bathrooms	1.70	0.90	0	11
Fixhalf	Number of half bathrooms	0.25	0.48	0	7
Bsmt1	No basement	0.06	0.24	0	1
Bsmt2	Crawl basement	0.60	0.49	0	1
Bsmt3	Part basement	0.16	0.37	0	1
Bsmt4	Full basement	0.17	0.38	0	1
Heat1	No heat	0.02	0.14	0	1
Heat2	Central heat	0.05	0.22	0	1
Heat3	Central air condition	0.17	0.38	0	1
Heat4	Heat pump	0.76	0.43	0	1
Attic1	No attic	0.88	0.33	0	1
Attic2	Unfinished attic	0.05	0.22	0	1
Attic3	Part finished attic	0.03	0.18	0	1
Attic4	Fully finished attic	0.03	0.18	0	1
Attic5	Fully finished/wall height attic	0.01	0.09	0	1
Age	Age of sales house	51.00	29.13	0	140
Cdu1	Dwelling condition is excellent	0.10	0.30	0	1
Cdu2	Dwelling condition is very good	0.21	0.41	0	1
Cdu3	Dwelling condition is good	0.20	0.40	0	1
Cdu4	Dwelling condition is average	0.39	0.49	0	1
Cdu5	Dwelling condition is fair	0.06	0.23	0	1
Cdu6	Dwelling condition is unsound	0.03	0.16	0	1
Cdu7	Dwelling condition is poor	0.01	0.11	0	1
Cdu8	Dwelling condition is very poor	0.003	0.05	0	1
Black	Percentage of black residents in CBG	0.79	0.32	0	1
Hsize	Average household size in CBG	2.65	0.44	1.28	4.19
Own	Percentage of residents own the house in CBG	0.49	0.21	0	0.98
Income	Per capital income in 1999 in CBG (*\$1000)	19.37	17.74	2.76	120.93
Old	Percentage of people over 65 years old in CBG	0.11	0.06	0	0.51
DIS300	Number of total foreclosures (including REO and REO sales) within 300 feet of sales house	3.12	3.90	0	37

DIS300REO	Number of foreclosures within 300 feet of sales house	1.10	1.64	0	21
DIS300REOS	Number of foreclosure sales within 300 feet of sales house	2.02	2.71	0	26
DIS600	Number of total foreclosures (including REO and REO sales) within 600 feet of sales house	9.91	11.75	0	91
DIS600REO	Number of foreclosures within 600 feet of sales house	3.49	4.54	0	41
DIS600REOS	Number of foreclosure sales within 600 feet of sales house	6.43	7.88	0	57
DIS900	Number of total foreclosures (including REO and REO sales) within 900 feet of sales house	19.04	22.26	0	153
DIS900REO	Number of foreclosures within 900 feet of sales house	6.71	8.42	0	77
Dis900REOS	Number of foreclosure sales within 900 feet of sales house	12.33	14.72	0	94
DIS1200	Number of total foreclosures (including REO and REO sales) within 1200 feet of sales house	30.60	35.45	0	268
DIS1200REO	Number of foreclosures within 1200 feet of sales house	10.82	13.33	0	132
DIS1200REOS	Number of foreclosure sales within 1200 feet of sales house	19.78	23.30	0	152

Table 3.2 reports the descriptive statistics for dummy variables. Quarter dummies are created to control for sales seasonal effects. It is hypothesized that houses are usually sold at a lower price in the winter than in the summer. There are many types of sales within the sample, and the sales type affects the sales price directly. Thus, we add sales type dummies into the regression. D1 is valid sale, D2 is the sale to or from exempt or utility, D3 is remodeled or changed after sale, D4 is related to individual or corporation sale, D5 is liquidation or foreclosure sale, D6 is land contract or unusual financing sale, and D7 is the sale that includes additional interest. School zone variable is another important factor determining house sales price. Good quality schools usually attract good quality teachers who receive higher salaries. Parents fund the schools through buying houses located in good school zones. Houses located in good school zones are usually more expensive because parents thus can fund the school by paying more property tax. Fifty elementary school zone dummies were added in the model.

Table 3.2 Descriptive Statistics for Dummy Variables, Atlanta, 2000-2010 (N=26.352)

Variable	Description	Mean	SD	Minimum	Maximum
Sale1	Valid sale	0.35	0.48	0	1
Sale2	To/from exempt or utility	0.03	0.17	0	1
Sale3	Remodeled/Changed after sale	0.06	0.24	0	1
Sale4	Related individuals or corporation	0.08	0.27	0	1
Sale5	Liquidation/Foreclosure	0.43	0.49	0	1
Sale6	Land contract/Unusual financing	0.05	0.21	0	1
Sale7	Includes Additional interest	0.002	0.05	0	1
Quarter1	Sold in the first quarter	0.25	0.43	0	1
Quarter2	Sold in the second quarter	0.28	0.45	0	1
Quarter3	Sold in the third quarter	0.26	0.44	0	1
Quarter4	Sold in the fourth quarter	0.21	0.41	0	1
SD1	Adamsville Elementary School	0.004	0.06	0	1
SD2	Benteen Elementary School	0.01	0.10	0	1
SD3	Mary Mcleod Bethune Elementary School	0.02	0.15	0	1
SD5	Capitol View Elementary School	0.02	0.15	0	1
SD6	Cascade Elementary School	0.002	0.05	0	1
SD7	Cleveland Avenue Elementary School	0.01	0.08	0	1
SD8	William M. Boyd Elementary School	0.01	0.12	0	1
SD9	Warren T. Jackson Elementary School	0.02	0.14	0	1
SD10	Morris Brandon Elementary School	0.02	0.14	0	1
SD11	Garden Hills Elementary School	0.01	0.11	0	1
SD12	E. Rivers Elementary School	0.02	0.14	0	1
SD13	Bolton Academy	0.01	0.11	0	1
SD14	Beecher Hills Elementary School	0.02	0.12	0	1
SD15	Daniel H. Stanton Elementary School	0.03	0.17	0	1

SD16	John Wesley Dobbs Elementary School	0.03	0.18	0	1
SD17	Hill-Hope Elementary School	0.01	0.11	0	1
SD18	Connally Elementary School	0.06	0.24	0	1
SD19	Centennial Place Elementary School	0.01	0.09	0	1
SD20	Continental Colony Elementary School	0.004	0.07	0	1
SD21	Ed S. Cook Elementary School	0.04	0.20	0	1
SD22	Deerwood Academy	0.01	0.11	0	1
SD23	Paul L. Dunbar Elementary School	0.01	0.11	0	1
SD25	Margaret Fain Elementary School	0.01	0.10	0	1
SD26	Fickett Elementary School	0.01	0.08	0	1
SD27	William Finch Elementary School	0.05	0.22	0	1
SD28	Charles L. Gideons Elementary School	0.07	0.25	0	1
SD29	Grove Park Elementary School	0.04	0.19	0	1
SD30	Heritage Academy Elementary School	0.01	0.09	0	1
SD31	Alonzo F. Herndon Elementary School	0.03	0.17	0	1
SD32	Joseph Humphries Elementary School	0.01	0.07	0	1
SD33	Emma Hutchinson Elementary School	0.01	0.10	0	1
SD34	M. Agnes Jones Elementary School	0.06	0.25	0	1
SD35	L. O. Kimberly Elementary School	0.004	0.07	0	1
SD36	Mary Lin Elementary School	0.003	0.06	0	1
SD37	Leonora P. Miles Elementary School	0.006	0.08	0	1
SD38	Morningside Elementary School	0.02	0.14	0	1
SD39	Parkside Elementary School	0.03	0.16	0	1
SD40	Perkerson Elementary School	0.03	0.17	0	1

SD41	Peyton Forest Elementary School	0.006	0.08	0	1
SD42	William J Scott Elementary School	0.02	0.13	0	1
SD43	Thomas Heathe Slater Elementary School	0.03	0.16	0	1
SD44	Sarah Rawson Smith Elementary School	0.02	0.12	0	1
SD45	Springdale Park Elementary School	0.01	0.12	0	1
SD46	F. L. Stanton Elementary School	0.03	0.18	0	1
SD47	Thomasville Heights Elementary School	0.01	0.09	0	1
SD49	George A. Towns Elementary School	0.02	0.12	0	1
SD50	Bazoline Usher Elem School	0.03	0.16	0	1
SD52	West Manor Elementary School	0.01	0.09	0	1
SD53	Walter F. White Elementary School	0.03	0.18	0	1
SD55	Carter G. Woodson Elementary School	0.01	0.12	0	1

5.1. OLS Regression

Table 3.3 reports the foreclosure effects within different buffers estimated by OLS regressions. The standard deviations are heteroskedastic-corrected, so they are robust. All the structure variables and neighborhood variables have their expected signs. For example, 1000 more square feet living areas increases the sales price by 13%. The *Cdu* is the general dwelling condition for each property. Eight dummy variables are created to distinguish among different dwelling conditions, with *Cdu1* being excellent condition and *Cdu8* being very poor condition. In the regression, we delete one dummy *Cdu8*, so the coefficients of *Cdu1* to *Cdu7* should be interpreted as the differences from the coefficient of *Cdu8*. Within 300 feet, properties with excellent condition (*Cdu1*) sells at a 39% higher price on average than houses with very poor

condition (*Cdu8*), while properties with fair condition (*Cdu5*) sells at only 18% higher price on average than houses with very poor condition (*Cdu8*).

In the OLS regression, one more foreclosure within 300 feet reduces the property sales price by 1.4%, one more foreclosure within 600 feet reduces the property sales price by 0.6%, one more foreclosure within 900 feet reduces the property sales price by 0.3%, and one more foreclosure within 1200 feet reduces the sales price by 0.2%.

Table 3.3 The Effect of Foreclosures within Different Buffers, Regression Coefficients for Heteroskedasticity – Corrected OLS^a

Variable	Within 300 Feet	Within 600 Feet	Within 900 Feet	Within 1200 Feet
Intercept	8.10*** (38.73)	8.06*** (38.39)	8.05*** (38.28)	8.04*** (38.29)
DIS300	-0.014*** (-7.65)			
DIS600		-0.006*** (-8.82)		
DIS900			-0.003*** (-8.22)	
DIS1200				-0.002*** (-8.84)
Street2	-0.10 (-0.74)	-0.10 (-0.75)	-0.10 (-0.72)	-0.10 (-0.72)
Street3	-0.11 (-1.22)	-0.10 (-1.21)	-0.10 (-1.15)	-0.10 (-1.16)
Lotarea	0.006*** (8.59)	0.006*** (8.42)	0.006*** (8.51)	0.006*** (8.56)
Lvarea	0.13*** (9.98)	0.13*** (10.05)	0.13*** (10.09)	0.13*** (10.10)
Stories	0.07*** (3.13)	0.07*** (3.27)	0.07*** (3.27)	0.08*** (3.30)
Age	0.003*** (10.48)	0.003*** (10.43)	0.003*** (10.45)	0.003*** (10.46)
Rmbed	0.03*** (4.73)	0.03*** (4.66)	0.03*** (4.67)	0.03*** (4.67)
Fixbath	0.05*** (4.78)	0.05*** (4.79)	0.04*** (4.72)	0.04*** (4.73)
Fixhalf	0.01 (0.77)	0.01 (0.66)	0.01 (0.64)	0.01 (0.67)
Bsmt2	0.02 (0.95)	0.02 (1.11)	0.02 (1.13)	0.03 (1.23)
Bsmt3	0.08*** (3.45)	0.09*** (3.57)	0.09*** (3.63)	0.09*** (3.77)
Bsmt4	0.09*** (3.54)	0.09*** (3.62)	0.09*** (3.71)	0.09*** (3.81)
Heat2	0.07* (1.74)	0.07* (1.81)	0.07* (1.84)	0.07* (1.85)
Heat3	0.13*** (3.52)	0.13*** (3.64)	0.13*** (3.65)	0.13*** (3.63)
Heat4	0.20*** (5.70)	0.20*** (5.80)	0.20*** (5.78)	0.20*** (5.78)
Attic2	0.07*** (3.20)	0.07*** (3.30)	0.07*** (3.35)	0.07*** (3.38)
Attic3	0.04	0.01	0.01	0.01

	(0.13)	(0.18)	(0.20)	(0.19)
Attic4	0.01	0.01	0.01	0.01
	(0.38)	(0.43)	(0.51)	(0.50)
Attic5	0.10*	0.10**	0.10**	0.11**
	(1.92)	(1.99)	(2.04)	(2.07)
Cdu1	0.39***	0.38***	0.39***	0.39***
	(3.75)	(3.73)	(3.79)	(3.80)
Cdu2	0.31***	0.31***	0.32***	0.32***
	(3.12)	(3.10)	(3.17)	(3.18)
Cdu3	0.27***	0.27***	0.28***	0.28***
	(2.70)	(2.69)	(2.75)	(2.76)
Cdu4	0.21**	0.21**	0.21**	0.22**
	(2.08)	(2.06)	(2.12)	(2.15)
Cdu5	0.18*	0.17*	0.17*	0.18*
	(1.66)	(1.65)	(1.71)	(1.73)
Cdu6	0.07	0.07	0.07	0.08
	(0.65)	(0.65)	(0.71)	(0.75)
Cdu7	0.11	0.12	0.12	0.13
	(1.04)	(1.11)	(1.12)	(1.16)
remodel	0.13***	0.14***	0.14***	0.14***
	(3.72)	(3.80)	(3.82)	(3.81)
Black	-0.11**	-0.09	-0.09	-0.08
	(-2.10)	(-1.57)	(-1.60)	(-1.48)
Hsize	-0.01	-0.01	-0.01	-0.01
	(-0.64)	(-0.41)	(-0.42)	(-0.35)
Own	0.09**	0.10***	0.10***	0.11***
	(2.51)	(2.78)	(2.86)	(3.04)
Income	0.005***	0.005***	0.005***	0.005***
	(6.49)	(6.42)	(6.32)	(6.18)
Old	-0.01	-0.02	-0.01	-0.002
	(-0.09)	(-0.20)	(0.08)	(-0.01)
Quarter1	0.04**	0.03**	0.03**	0.03**
	(2.51)	(2.30)	(2.30)	(2.21)
Quarter2	0.02*	0.02	0.02	0.02
	(1.67)	(1.51)	(1.52)	(1.47)
Quarter3	0.04***	0.04***	0.04***	0.04***
	(3.03)	(2.97)	(3.00)	(2.98)
Year Dummy	Yes	Yes	Yes	Yes
Adj. R ²	0.57	0.57	0.57	0.57

^a School district dummy, sales type dummy and year dummy variable parameter estimates not reported here for sake of space.

Note: ***Statistically significant at 1%; **Statistically significant at 5%; *Statistically significant at 10%.

The asymptotic t statistics are in parentheses.

I also distinguish between REO and REO sales to examine their effects on neighborhood property sales price respectively. Table 3.4 reports that one more REO within 300 feet reduces surrounding sales price by 4.3% after controlling housing characteristics and neighborhood characteristics, and one more REO sale reduces surrounding sale price by 0.8%. If both REO and REO sales are included in the regression model, Model 3 reports that one more REO within 600 feet reduces surrounding sales price by 2.7%, while one more REO sales increases surrounding sales price by 0.6%. It makes sense because REO sales would resolve REO. Once a REO is sold by the bank, it is not considered as a REO, and instead it becomes a REO sale if it occurs before the subject property sale. Although REO sales themselves may affect surrounding house sales price in a negative way, REO sales help eliminate vacant houses and rebuild community stability, and one more REO sale indicates that there is one less REO, REO sales thus help raise surrounding houses' sales price *per se* after controlling REOs. The effect of REO and REO sales are consistent with different buffers although the magnitudes are different.

Table 3.4 The Effect of REO and REO sales on Neighborhood Property Sales Value within Different Buffers⁷

Variable	Within 300 Feet			Within 600 Feet			Within 900 Feet			Within 1200 Feet		
	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3	Model 1	Model 2	Model 3
DIS300REO	-0.043*** (-11.24)		-0.045*** (-11.05)									
DIS300REOS		-0.008*** (-3.37)	0.003 (1.35)									
DIS600REO				-0.022*** (-13.49)		-0.027*** (-14.02)						
DIS600REOS					-0.004*** (-4.54)	0.006*** (5.10)						
DIS900REO							-0.012*** (-12.79)		-0.019*** (-14.84)			
DIS900REOS								-0.003*** (-4.45)	0.006*** (7.89)			
DIS1200REO										-0.009*** (-14.18)		-0.015*** (-17.62)
DIS1200REOS											-0.002*** (-4.75)	0.005*** (10.53)

Note: ***Statistically significant at 1%; **Statistically significant at 5%; *Statistically significant at 10%.
The asymptotic t statistics are in parentheses.

⁷ Regression Coefficients for Heteroskedasticity – Corrected OLS

5.2. Difference-In-Differences

One disadvantage of OLS regression is that the omitted variable problem may exist. Although we have controlled for housing characteristics, neighborhood characteristics, and school districts, there are still some unobserved variables may affect housing sales price, including some location variables which are unchanged over time but cannot be measured with in the dataset. Difference-in-differences method could help eliminate those invariant observed and unobserved variables, and thus avoid the omitted variable problem. Table 3.5 reports the difference-in-differences regression results. Model 1 serves as the base model, which includes the foreclosure variable *DIS300*, *DIS600*, *DIS900*, and *DIS1200* (including both REO and REO sales) respectively and variant house characteristics, including house age, sales type, sales quarter and whether it is remodeled or not. On average, one more foreclosure within 300 feet reduces its surrounding sales price by 2.5%, one more foreclosure within 600 feet reduces its surrounding sales price by 1.1%, one more foreclosure within 900 feet reduces its surrounding sales price by 0.6%, and one more foreclosure within 1200 feet reduces its surrounding sales price by 0.4%. The coefficients estimated by the difference-in-differences method are larger than those estimated by OLS regression. In Model 2, after separating REO and REO sales, the estimated coefficients are also larger than those estimated by OLS regression. For example, one more REO within 600 feet reduces surrounding sales price by 3.7%, while one more REO sales within 600 feet increases sales price by 0.6%. In general, the difference-in-differences models produce a better fit than OLS models indicated by higher adjusted R^2 value.

Table 3.5 The Effect of Foreclosures within Different Buffers, Difference-In-Differences Model

Variable	Within 300 Feet		Within 600 Feet		Within 900 Feet		Within 1200 Feet	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
DIS300	-0.025*** (-7.91)							
DIS300REO		-0.068*** (-11.64)						
DIS300REOS		-0.001 (-0.13)						
DIS600			-0.011*** (-9.52)					
DIS600REO				-0.037*** (-14.49)				
DIS600REOS				0.006*** (3.23)				
DIS900					-0.006*** (-9.64)			
DIS900REO						-0.024*** (-15.30)		
DIS900REOS						0.006*** (5.38)		
DIS1200							-0.004*** (-10.31)	
DIS1200REO								-0.018*** (-16.77)
DIS1200REOS								0.005*** (7.11)
Age	-0.07*** (-14.15)	-0.07*** (-14.21)	-0.07*** (-14.08)	-0.07*** (-14.17)	-0.07*** (-14.07)	-0.07*** (-14.26)	-0.07*** (-14.11)	-0.07*** (-14.26)
Sale1	1.31*** (8.13)	1.30*** (8.10)	1.32*** (8.21)	1.31*** (8.13)	1.32*** (8.21)	1.30*** (8.11)	1.33*** (8.23)	1.30*** (8.11)
Sale2	0.53*** (3.22)	0.52*** (3.18)	0.54*** (3.28)	0.53*** (3.20)	0.54*** (3.27)	0.52*** (3.17)	0.54*** (3.30)	0.52*** (3.16)

Sale3	0.70*** (4.28)	0.69*** (4.25)	0.70*** (4.33)	0.70*** (4.29)	0.70*** (4.33)	0.69*** (4.27)	0.71*** (4.35)	0.69*** (4.28)
Sale4	0.82*** (5.06)	0.81*** (5.02)	0.84*** (5.14)	0.82*** (5.06)	0.84*** (5.15)	0.82*** (5.05)	0.84*** (5.17)	0.82*** (5.06)
Sale5	0.65*** (4.00)	0.63*** (3.93)	0.66*** (4.06)	0.63*** (3.94)	0.66*** (4.06)	0.63*** (3.90)	0.66*** (4.08)	0.62*** (3.89)
Sale6	0.20 (1.22)	0.19 (1.18)	0.21 (1.29)	0.20 (1.25)	0.21 (1.29)	0.20 (1.22)	0.22 (1.32)	0.20 (1.23)
Sale7	1.28*** (6.33)	1.26*** (6.26)	1.29*** (6.38)	1.26*** (6.26)	1.29*** (6.39)	1.26*** (6.25)	1.30*** (6.42)	1.26*** (6.25)
Quarter1	0.02 (1.01)	0.02 (1.08)	0.01 (0.81)	0.02 (1.01)	0.01 (0.76)	0.02 (1.02)	0.01 (0.65)	0.02 (0.99)
Quarter2	0.004 (0.27)	0.01 (0.43)	0.002 (0.10)	0.01 (0.33)	0.001 (0.07)	0.01 (0.33)	-0.00 (-0.01)	0.01 (0.31)
Quarter3	0.04*** (2.95)	0.05*** (2.97)	0.05*** (2.85)	0.05*** (2.93)	0.05*** (2.83)	0.05*** (2.92)	0.05*** (2.78)	0.05*** (2.83)
Remodel	0.06 (0.92)	0.06 (0.85)	0.05 (0.78)	0.04 (0.62)	0.04 (0.66)	0.03 (0.44)	0.03 (0.53)	0.03 (0.41)
Adj. R ²	0.76	0.76	0.76	0.76	0.76	0.76	0.76	0.76

^a School district dummy and year dummy variable parameter estimates not reported here for sake of space.

Note: ***Statistically significant at 1%; **Statistically significant at 5%; *Statistically significant at 10%.

The asymptotic t statistics are in parentheses.

5.3. Propensity Score Matching

Table 3.6 and Table 3.7 report the propensity score matching results. Table 3.6 shows the baseline characteristics of the treatment and control groups. The treatment group (N=18,030) is very different from the control group (N=8,312) in all selected variables. Houses are more likely to be surrounded by foreclosures within 300 feet if they have smaller living areas, older, located in the census block group with higher percentage of subprime mortgages, higher percentage of black residents, larger household size, lower home ownership, lower per capita income, and lower percentage of people older than 65 years old.

Table 3.7 shows the matching results using the caliper matching method with replacement with a caliper of 1×10^{-4} . There are 2,183, i.e. 12% treated units are matched. The samples appear to be well balanced except lot size. Then, t test is used to compare the sales price mean between treated and control group. The average sales price in treatment group is \$128,712, and the average sales price in control group is \$141,052. Thus, the treatment effect of foreclosure within 300 feet is to reduce average sales price by \$4,728, which is about 8.7% less. The result is larger than the coefficient estimated by OLS and difference-in-differences regression. Propensity score matching reports whether foreclosure reduces surrounding house sales price. The treatment is a binary variable, if the number of foreclosures within 300 feet for a house is not 0, the house is considered to be treated with foreclosures. Thus, foreclosures within 300 feet reduce property sales price by 8.7% on average.

Table 3.6 Baseline Characteristics (Treatment: DIS300>0)

Characteristic	Treatment (N=18,030)		Control (N=8,312)		Analysis	
	Mean	SD	Mean	SD	T-test	P-value
Pct_lchl	0.06	0.03	0.10	0.05	t = 66.05	<0.0001
Pct_hcll	0.31	0.10	0.20	0.14	t = -75.88	<0.0001
Pct_hchl	0.19	0.06	0.12	0.08	t = -79.48	<0.0001
Lot size	8.97	6.37	15.11	14.36	t = 48.10	<0.0001
Living area	1.43	0.57	2.02	1.48	t = 46.69	<0.0001
Age	52.45	29.73	47.86	27.51	t = -11.93	<0.0001
Black	0.89	0.20	0.59	0.42	t = -77.61	<0.0001
Household size	2.74	0.39	2.46	0.48	t = -50.52	<0.0001
Own	0.46	0.19	0.56	0.23	t = 40.48	<0.0001
Income	14039.9	8507.2	30933.4	25397.2	t = 80.11	<0.0001
Old	0.11	0.06	0.12	0.07	t =4.08	<0.0001

^a School district dummy variable parameter estimates not reported here for sake of space.
 Note: ***Statistically significant at 1%; **Statistically significant at 5%; *Statistically significant at 10%.

Table 3.7 Propensity Score Matching, Caliper (1×10^{-4}) method (Treatment: DIS300>0)

Characteristic	Treatment (N=2,183)		Control (N=2,183)		Analysis	
	Mean	SD	Mean	SD	T-test	P-value
Pct_lchl	0.07	0.04	0.07	0.04	t = -0.24	0.81
Pct_hcll	0.29	0.11	0.29	0.11	t = -0.76	0.45
Pct_hchl	0.17	0.06	0.17	0.06	t = 1.21	0.23
Lot size	9.75	6.77	10.76	6.40	t = 5.05	<0.0001
Living area	1.45	0.64	1.48	0.63	t = 1.57	0.12
Age	52.44	28.81	51.21	26.90	t = -1.46	0.14
Black	0.85	0.25	0.85	0.26	t = 0.50	0.62
Household size	2.67	0.42	2.67	0.43	t = -0.51	0.61
Own	0.48	0.19	0.49	0.21	t = 0.75	0.45
Income	15810	10236.2	16375.1	10938.3	t = 1.76	0.08
Old	0.12	0.07	0.12	0.07	t = -0.36	0.72
Price	128,712	127,935	141,052	12,340.4	t = 2.72	0.01

^a School district dummy variable parameter estimates not reported here for sake of space.
 Note: ***Statistically significant at 1%; **Statistically significant at 5%; *Statistically significant at 10%.

6. Conclusion

By using repeat sales, this study applies quasi-experiment models, difference-in-differences and propensity score matching methods to analyze the impacts of foreclosures on neighborhood one to four unit residential property values in city of Atlanta. The regression results by the OLS, difference-in-differences and propensity score matching are analyzed and compared. Using repeat sales, the difference-in-differences and propensity score matching methods avoid omitted variable bias which is likely a problem in hedonic models.

In this study, I also distinguish between REO and REO sales. The difference-in-differences model reports that one more foreclosure (including REO and REO sales) reduces surrounding house sales prices by 2.5% within 300 feet, 1.1% with 600 feet, 0.6% within 900 feet, and 0.4% within 1200 feet. However, after separating REO and REO sales, the effect of REO increases dramatically, one more REO reduces surrounding sales price by 3.7%, while one more REO sales increases sales price by 0.6% within 600 feet.

Propensity score matching also indicate that after balancing treatment and control groups, foreclosures within 300 feet reduces average sales price by 8.7% less. The result is larger than the coefficient estimated by both OLS and difference-in-differences regression. Propensity score matching only reports whether foreclosure reduces surrounding house sales price. The treatment is a binary variable, if the number of foreclosures within 300 feet for a house is not 0, the house is considered to be treated with foreclosures. However, difference-in-differences reports how much surrounding house sales price will be reduces by increasing one more foreclosure within certain buffer. Compared with difference-in-differences method, propensity score matching method reports an aggregate effect of foreclosures. Thus, it makes sense that the coefficient of foreclosures estimated by propensity score matching is larger than those estimated by difference-in-difference method.

Compared to Harding, Rosenblatt and Yao's (2009) work, this study improves their model from several aspects. First, besides number of nearby foreclosures, this study controls more property characteristics that are expected to change between sales, including whether the house remodeled or not, sales quarter and sales type. Second, instead of arbitrarily picking out two repeat sales, this study includes every transaction record sold more than once during the study periods, which gives me more precise estimates due to the efficiency gain bout by more

data. Third, using transaction buyers and sellers' names, this study distinguishes between REO and REO sales. Because each REO sale resolves a REO, it helps increase surrounding house sales price as a result. The study separates effects of the price trend over time and the contagion effects of foreclosures. The results confirm negative contagion effects of foreclosures for surrounding sales properties

CHAPTER 4

Irrigation and Income Inequality in the Southeast United States

1. Introduction

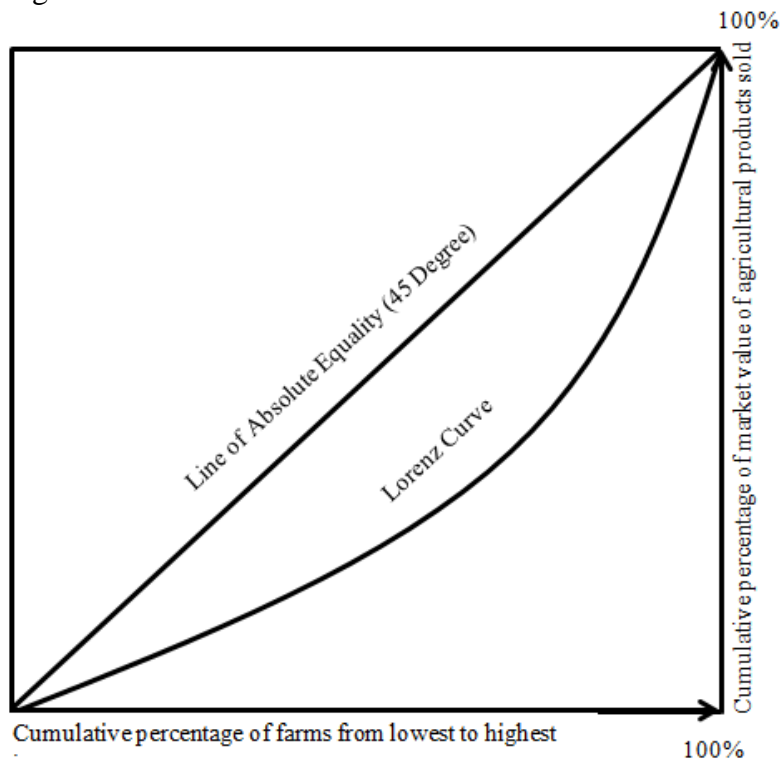
Irrigation is often promoted as a technology that can increase crop production, improve agriculture income and alleviate poverty. However, irrigation is a relatively expensive technology for small-scale farmers and poor farmers, which impedes their opportunities to adopt irrigation technology. Income inequality may increase due to the adoption barriers.

According to the agricultural treadmill theory, anyone of a number of small farms produce the same products cannot affect the commodity's price; hence farmers who initially adopt new technology and thus increase productivity are able to gain significant benefits. The income inequality increases between technology adopters and non-adopters. However, after some time, others follow and commodity prices tend to fall with the increased supply. Thus, increased efficiency in agricultural production can drive down prices. The downward pressure on crop price directly has two results: (1) those who have not yet adopted the new technology must now do so lest they lose income because of price squeeze and (2) those who are too old, sick, poor or indebted to innovate eventually have to exit from farming. Their resources are then absorbed by those who make the windfall profits or "scale enlargement" (Cochran, 1958; Bai, 2008). In effect, the consequence of the first situation decreases farmers' income inequality when more and more farms adopt irrigation technology, and the second situation may result in redistribution of natural resources and rural income and further exacerbates inequality.

Lorenz curves and Gini coefficient are used to measure the income inequality. A Lorenz curve plots the cumulative percentages of total income received against the cumulative percentages of recipients, starting with the poorest farms. With perfect equality, the number of percent of the farms would receive exactly number of percent of the income. The corresponding Lorenz curve would therefore be a straight 45 degree line. A Gini index helps to compare income inequality among farms. It is the area between a Lorenz curve and the line of absolute equality, as a percentage of the triangle under the line of absolute equality. A Gini index of 0% represents perfect equality and a Gini index of 100% implies perfect inequality.

In this study, farm is ordered as the unit of analysis to calculate the Gini index. The Lorenz curve is constructed by plotting the cumulative percentage of farms and cumulative percentage of market value of agricultural products sold. In a perfectly equal system, all farms would contribute the same share to the market value of agricultural products sold.

Figure 4.1 Lorenz Curve



There are several research studies examine the relationship between irrigation and income distribution as well as poverty in the developing countries. However, the existing literature does not adequately address the endogeneity of irrigation adoption and thus not convincingly establish a causal relationship. Because irrigation is a relatively expensive technology for small-scale farmers and poor farmers, low farm income may impede their opportunities to adopt irrigation technology. Thus, irrigation may be endogenous to agricultural income and profit. This paper corrects the endogeneity problem and uses county level data from the Census of Agriculture to examine effects of irrigation technology adoption on agriculture income and income inequality in the 9 Southeast states, including Alabama, Arkansas, Florida, Georgia, Louisiana, Mississippi, North Carolina, South Carolina, and Tennessee. Compared to the household level data in the previous studies, county level data may be more comprehensive to study the impact of irrigation in the country on average.

2. Literature Review

Huang et al. (2005) conducted survey in rural China, using household level data, they find that irrigation increases farmers' cropping income and total income. Holding other household characteristics constant, increasing irrigated land per capita by one hectare will lead to an increase of 3082 yuan in annual cropping income per capita, and an increase of 2628 yuan in annual total income per capital.

Using household data from 26 irrigation systems, Hussain (2007) states direct and indirect irrigation benefits. The direct benefits include increasing crop yields, reducing climate risk and increasing employment opportunities for irrigation system construction and maintenance work. The indirect benefits include increasing farming labor demand and activating the rural

community through various water related activities (such as fish farming). The indirect benefits could be larger than direct benefits through the multiplier effect.

Chamber (1988) cites several empirical studies across countries that irrigation increases wage rates by increasing labor demand. Irrigation raises employment for landless labors via increased working days per hectare and increased working days during a cropping season and additional employment in a second or third irrigation season (Smith, 2004).

However, irrigation adoption decisions are usually affected by many factors. Adoption of irrigation may be difficult for poor farmers because it requires capital, familiarization and is cash intensive to operate. At first it raises inequality, as only a few of farmers share the initially high income generated (Smith, 2004).

Caswell and Zilberman (1985) apply a multinomial logit framework to predict irrigation choice as a function of water cost, farm location, water source, and crops grown in the San Joaquin Valley of California. Skaggs (2000) find that the probability a grower will be a higher tech irrigator decreases as age increases. Farmers owning larger farms are more likely to install or expand drip irrigation technology. Shrestha and Gopalakrishnan (1993) use a probit model to examine the choice of drip technology as a function of differential water use and yields, plant cycle, soil type, temperatures and field gradients. They find that advanced irrigation technologies tend to be adopted first in areas with relatively low land quality and expensive water (particularly deep groundwater) areas.

Huang et al. (2006) report that a higher proportion of good quality land leads to higher cropping income, but has no effect on off-farm income and other income⁸. Good quality land

⁸ Other income includes livestock income, income from gifts (non-remittances), rental income, income from subsidies and pensions, income from interest, income from asset sales, net value of commercial agricultural (e.g. vegetable and fruit), value of crop subsidiaries (e.g. fodders), net value of processed crop products, and miscellaneous income.

with higher water-holding capacity reduced farmers' propensity to adopt irrigation technology.

Lichtenberg (1980) also find that "irrigated agriculture was restricted to areas with relatively flat topography and good soils." Irrigation on sloped land is not economics due to labor consumption and irrigation on sandy soils may cause runoff and thus waste water.

3. Data

Agricultural production and related data come from the 1997 and 2002 Census of Agriculture. For confidentiality reasons, counties are the finest geographic unit of observation in these data. In the subsequent regression analysis, the total market value of crops sold per acre, the total market value of agricultural products sold per acre and total profit of agricultural products sold per acre are the dependent variables. The market value of agricultural products sold represents the gross market value before taxes and production expenses of all agricultural products sold or removed from the place regardless of who received the payment. It is equivalent to total sales which include sales from crops, some livestock and animal specialties. The figure also includes the value of commodities placed in commodity credit corporation (CCC) loans. The market value of crops sold only includes sales from crops. The average total profit of agricultural products sold is constructed as the difference between the average market value of agricultural products sold and average production expenses in a county. The sales value may be a better measure of the economic size of farm because it represents all income resources from production operation, other than income from farm-related sources.

The soil quality data come from National Resource Inventory (NRI). The NRI is a massive survey of soil sample and land characteristics from roughly 800,000 sites which is conducted in census years. Follow Deschênes and Greenstone's (2007) work, a number of soil quality variables are selected as controls in the following regressions, including fraction of sand,

soil erosion (K-Factor), susceptibility to flood, slope length, permeability, moisture capacity, wetland, and salinity.

Number of wells in each county is received from U.S. Geological Survey (USGS) Groundwater Watch. The USGS provides wells' location, site number, site name and measurement begin date and end date. Using these measurement begin date and end date, I could calculate number of wells in 1997 and 2002 in each county. Then, number of wells could be merged with irrigation data by county fips code and year.

The Gini coefficient is calculated from Census of Agricultural total agricultural products sold values categories. The 12 sales value categories range from less than \$1000 to \$500,000 or more. In each category, the number of farms and their total sales values are provided, so I could get cumulative percentage of farms from lowest to highest agricultural sales income and cumulative percentage of market value of agricultural produce sold. Following Foster and Sen (1997), the Gini coefficient⁹ is calculated as

$$Gini = 1 - \frac{\sum G}{\sum A / \sum E} \quad (4.1)$$

where A_i is the number of farms in each sales value category, $\sum A$ is the sum of all the farms in each category in the county, E_i is the total sales value in each sales category, $\sum E$ is the sum of total sales value in each category, G_i is Gini coefficient in each sales category, $\sum G$ is sum of each category's Gini coefficient.

⁹

Group	Farm per Group	Income per Group	Accumulated Income	Gini
1	A1	E1	K1=E1	G1=(2*K1-E1)*A1
2	A2	E2	K2=E2+K1	G2=(2*K2-E2)*A2
3	A3	E3	K3=E3+K2	G3=(2*K3-E3)*A3
4	A4	E4	K4=E4+K3	G4=(2*K4-E4)*A4
Total	$\sum A$	$\sum E$		$\sum G$
Inequality Measure				Gini=1- $\sum G / \sum A / \sum E$

4. Model

In this study, the determinants of income can be analyzed by making sales value and total profit of agricultural products sold a function of irrigation and a set of other county level agricultural characteristics. The basic model is

$$\ln y_{it} = \alpha + \beta X_{it} + \gamma irrigation_{it} + a_i + u_{it}, \quad t = 1,2 \quad (4.2)$$

where $\ln y_{it}$ is total market value of crops sold per acre, total market value of agricultural products sold per acre, or total profit of agricultural products sold per acre in county i in year t expressed as the natural logarithmic form. There is a reason to express dependent variables as the natural logarithmic form. The dependent variables are county-level sales or profit value per acre, the estimates of sales values vary from county to county since they have different type of operations with different sizes. The descriptive statistics show that sales values have large variations because the standard deviations are larger than the means of sales values. Thus, the transformed form of y_{it} corrects for the heteroskedasticity resulting from differences in operations. The vector X_{it} contains farmer age, commodity credit loan per acre, and precipitation. This study includes both growing season accumulative precipitation from April to October and precipitation standard deviation derived from growing season precipitation. When the rain level is not consistent, farmers will more likely to adopt irrigation to help increase agriculture productivity. It is thus hypothesized that larger precipitation standard deviation will increase the level of irrigation adaption. Variable *Irrigation* is the interest of this study, it is measured as percentage of irrigated land in each county. The term a_i captures all unobserved, time-invariant factors that affect y_{it} ; and the error term u_i is the idiosyncratic error or time-varying error, it represents unobserved factors that change over time and affect y_{it} .

There are three problems with this basic model. First, because this dataset concludes agricultural data for census year 1997 and 2002, the poolability test is used to examine if data are poolable so that individual time periods have the same constant slopes of regressors. The large F statistics rejects the null hypothesis of poolability¹⁰, so the panel data are not poolable with respect to time. Thus, a year dummy is added in the regression, which is called least squares dummy variable (LSDV) regression. Second, even if it is assumed that the idiosyncratic error u_{it} is uncorrelated with x_{it} and the variable of interest $irrigation_{it}$, the estimations by OLS is biased and inconsistent if a_i is correlated with x_{it} or $irrigation_{it}$. The bias is often called heterogeneity bias, it is really just bias caused from omitting a time-constant variable. In order to correct for it, a vector of soil quality variables are added in the regression, which is time-invariant in the dataset. Third, there are no state fixed effects to account for all unobserved differences across states, such as state agricultural programs. The improved model is

$$\ln y_{it} = \alpha + \delta d2 + \theta state + \beta X_{it} + \gamma irrigation_{it} + \lambda a_i + u_{it}, \quad t = 1, 2 \quad (4.3)$$

where $d2$ is a dummy variable that equals zero when $t=1$ and one when $t=2$; $state$ is a vector of state dummies; a_i represents time-constant variable soil quality, which includes measures of sand content, susceptibility to floods, soil erosion (K-Factor), slope length, permeability, wetland, moisture capacity, and salinity.

Irrigation is suspected endogenous in the model since there are obstacles for poor farmers to adopt irrigation technology due to its high cost. In other words, the farmers' wealth may affect their decisions of adopting irrigation technology. Thus, the crop sales value or the total agricultural products sales value could affect farmers' irrigation level. Two-stage least squares (2SLS) regression is appropriate to address the endogeneity problem. An instrumental variable

¹⁰ $F = \frac{(e'e - \sum e'_t e_t)/(T-1)K}{\sum e'_t e_t/T(n-K)} = F[(T-1)K, T(n-K)]$, where $e'_t e_t$ is SSE of the OLS regression at time t

for irrigation is needed to conduct the first stage regression. Caswell and Zilberman (1986) find that farmers in locations with relatively low land quality and expensive water (i.e. deep wells or groundwaters) are more likely to adopt drip and sprinkler irrigation systems. The same findings are also found in Shreastha and Gopalakrishnan's (1993) work. Another recent study conducted by Molnar and Sydnor (2010) shows that a major reason that farmers are reluctant to adopt irrigation technology in some Southeast U.S. counties is due to lack of groundwater. In this study, I use the number of wells in each county as a proxy to measure the groundwater availability, which works as an instrumental variable for irrigation adoption. The first stage model is

$$Irrigation_{it} = \alpha' + \delta'd2 + \theta'state + \beta'X_{it} + \lambda'a_i + \eta wells_{it} + u_{it}, t = 1,2 \quad (4.4)$$

where X_{it} includes all the independent variables in the second stage; $wells$ is number of wells in each county.

Income inequality is analyzed as a function of a set of irrigation and other farm characteristics. The model is expressed as

$$y_{it} = \alpha + \delta d2_t + \beta X_{it} + \gamma irrigation_{it} + \varepsilon_{it}, t = 1,2 \quad (4.5)$$

The dependent variable is Gini coefficient calculated from Census of Agricultural total agricultural products sold values categories, independent variables include average farm size, average commodity credit loans and irrigation. I use average acreage of irrigated land, number of irrigated farm, and percentage of irrigated land to represent irrigation respectively. According to the treadmill theory, it is hypothesized that an increase in the acreage or number of irrigated farm will lead to an increase in the agricultural income inequality at first, but continually increasing irrigated land or irrigated farms will drive down marginal profit of agricultural products and thus decrease income inequality. So it is hypothesized that the relationship between irrigation and

income inequality is nonlinear. Thus, the square forms of irrigation variables are added to examine this hypothesis.

5. Results

Table 4.1 reports the descriptive statistics for study variables. For the nine states, the average total market value of crops sold is \$354 per acre, ranging from \$5 per acre to \$23,172 per acre. The average total market value of agricultural products sold is \$466 per acre, ranging from \$20 per acre to \$24,836 per acre. The average profit of agricultural products sold is \$85 per acre, ranging from -\$674 per acre to \$4,542 per acre. The operational product sales values appear unequal to some extent. The average irrigated land is 6%, ranging from 0.1% to 78%. The number of wells in each county ranges from 0 to 4247.

Table 4.1 Descriptive Statistics for Study Variables, 9 Southeast States, 1997 and 2002 (N=568)

Variable	Mean	Std. Dev.	Minimum	Maximum
Income variables				
Total market value of crops sold (\$/acre)	353.94	940.56	5.01	23172.41
Total market value of agricultural products sold (\$/acre)	466.24	835.78	19.62	24836.12
Total profit of agricultural products sold (\$/acre)	85.27	199.86	-673.97	4542.88
Gini coefficient	0.81	0.12	0	0.98
Irrigation				
Percentage of irrigated land	6.07	12.54	0.01	78.08
Land quality				
Fraction sand	0.25	0.37	0	1
Fraction flood-prone	0.18	0.24	0	1
K Factor	0.26	0.11	0.03	0.49
Slope length	148.33	80.46	21.26	785.71
Permeability	4.96	4.35	0.21	20
Wetlands	0.18	0.16	0.004	0.81
Moisture capacity	0.13	0.05	0.04	0.30
salinity	0.003	0.02	0	0.25
Precipitation				
April average precipitation	4.05	1.01	1.79	6.32
May average precipitation	4.58	0.86	2.86	6.98
June average precipitation	4.71	0.92	3.18	8.77
July average precipitation	5.03	1.07	2.78	8.41
August average precipitation	4.54	1.34	2.51	9.4
September average precipitation	4.11	0.92	2.72	7.77
October average precipitation	3.54	0.61	2.26	5.69
Standard deviation	0.93	0.47	0.20	3.22
Cumulative precipitation	30.57	3.65	24.48	42.15
Other variables				
Number of wells	170.39	336.60	0	4247
Average age of principle operator	55.82	2.02	44.80	62.20
Average commodity credit loans (\$/acre)	5.29	9.74	0	79.63

Table 4.2 reports the first stage OLS regression for irrigation adoption. Besides all the exogenous variables in the second stage, the instrumental variable number of wells is added in the equation. The relationship between the number of wells and irrigation is strongly positive. 1% increase in number of wells will increase irrigated land by 2.2% The Hausman test is used to test endogeneity in all the models, large chi-square statistics 8.25 (P=0.004) and 8.17 (P=0.0043) confirm that there are endogeneity problems in the regression for total market value of crops sold. However, there is no robust evidence that the irrigation is endogenous to total market value of agricultural products sold and total profit of agricultural products sold. Thus, IV-2SLS

regression is preferred to the OLS regression for dependent variable total market value of crops sold.

Table 4.2 First Stage OLS Regression Results for Irrigation

	Dependent Variable	
	Percentage of irrigated land	
	Model 1	Model 2
<i>Instrumental variable</i>		
ln(Number of wells)	2.20*** (0.41)	2.27*** (0.41)
<i>Land quality</i>		
Fraction sand	0.04 (0.03)	0.01 (0.03)
Fraction flood-prone	-0.05* (0.03)	-0.05 (0.28)
K Factor	0.23 (0.14)	0.19 (0.14)
Slope length	-0.002 (0.01)	0.002 (0.01)
Permeability	-0.66* (0.39)	0.09 (0.37)
Wetlands	-0.07* (0.04)	-0.003 (0.04)
Moisture capacity	-0.04 (0.38)	0.12 (0.38)
Salinity	0.76** (0.35)	0.83** (0.36)
<i>Precipitation</i>		
Standard deviation of precipitation	7.85*** (3.07)	
Cumulative precipitation		-0.50** (0.22)
<i>Other variables</i>		
ln(Average commodity credit loans) (\$/acre)	1.74*** (0.32)	1.90*** (0.32)
Age	-34.53*** (7.40)	-34.12*** (7.43)
Age ²	0.30*** (0.07)	0.29*** (0.07)
Year2002	2.53*** (0.88)	2.16** (0.88)
State fixed effect	Yes	Yes
Adj R ²	0.67	0.67

Note: ***Statistically significant at 1%; **Statistically significant at 5%; *Statistically significant at 10%.
The standard errors are in parentheses.

Table 4.3 reports the OLS regression and Table 4.4 reports the IV-2SLS regression for the effect of irrigation on total market value of crops sold, total market value of agricultural products sold and total profit of agricultural products sold respectively. Most of the coefficients are statistically significant and have the expected sign. Most importantly, irrigation is positively related to total market of crops sold and total market value of agricultural products sold. The coefficients estimated in IV-2SLS are about two and half times the magnitude as those estimated in the OLS regression. Holding other variables constant, an increase of 1% of irrigated land leads to total market value of crop sold increasing by 5% per acre, and total market value of agricultural product sold increasing by 0.5% per acre. On average, in 2002, the total market value of crops sold decreases by 36% per acre compared to 1997.

A surprising result is that there are no statistically significant relationships between the average commodity credit loan and the market value of crops sold estimated by all the models in 2SLS regression. In addition, there are no statistically significant causality between commodity credit loan and market value of agricultural products sold as well as total profit of agricultural products sold in the OLS regression.

Table 4.3 OLS Regression Results for Determinants of Income

	Dependent variable					
	ln(Total market value of crops sold) (\$/acre)		ln(Total market value of agricultural products sold) (\$/acre)		ln(Total profit of agricultural products sold) (\$/acre) ¹¹	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
<i>Irrigation</i>						
Percentage of irrigated land	0.02*** (0.002)	0.02*** (0.002)	0.002 (0.002)	0.005* (0.003)	0.003 (0.005)	0.005 (0.005)
<i>Land quality</i>						
Fraction sand	-0.001 (0.002)	-0.002 (0.002)	0.002 (0.002)	0.002 (0.002)	0.006 (0.004)	0.007* (0.004)
Fraction flood-prone	0.001 (0.001)	0.001 (0.001)	-0.002 (0.002)	-0.001 (0.002)	-0.002 (0.003)	-0.001 (0.003)
K Factor	0.01 (0.01)	0.01 (0.01)	-0.02** (0.01)	-0.02** (0.01)	-0.03 (0.02)	-0.02*** (0.02)
Slope length	0.0003 (0.0004)	0.003 (0.0004)	0.001** (0.0004)	0.001** (0.0005)	0.0003 (0.001)	0.0001 (0.001)
Permeability	0.06*** (0.02)	0.07*** (0.02)	-0.01 (0.02)	0.02 (0.02)	0.01 (0.05)	-0.002 (0.05)
Wetlands	0.01*** (0.002)	0.01*** (0.002)	-0.01*** (0.002)	-0.005** (0.002)	-0.003 (0.005)	-0.005 (0.005)
Moisture capacity	-0.03 (0.02)	-0.03 (0.02)	0.04 (0.02)	0.02 (0.02)	0.06 (0.05)	0.03 (0.05)
salinity	-0.001 (0.02)	-0.007 (0.02)	0.01 (0.02)	-0.01 (0.02)	0.05 (0.05)	0.02 (0.05)
<i>Precipitation</i>						
Standard deviation of precipitation	0.30** (0.14)		0.80*** (0.16)		0.45 (0.33)	
Cumulative precipitation		0.01 (0.01)		0.05*** (0.01)		0.09*** (0.03)
<i>Other variables</i>						
ln(Average commodity credit loans) (\$/acre)	0.05*** (0.02)	0.06*** (0.02)	-0.01 (0.02)	0.005 (0.02)	-0.07 (0.04)	-0.06 (0.04)
Age	0.38 (0.40)	0.41 (0.40)	0.99** (0.42)	1.06** (0.47)	-0.52 (1.11)	-0.61 (1.10)
Age ²	-0.003 (0.004)	-0.004 (0.004)	-0.01** (0.004)	-0.01** (0.004)	0.004 (0.01)	0.004 (0.01)
Year2002	-0.30*** (0.05)	-0.31*** (0.05)	0.001 (0.05)	-0.03 (0.05)	-0.16 (0.11)	-0.18 (0.11)
State fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Adj R ²	0.52	0.51	0.45	0.43	0.22	0.23

Note: ***Statistically significant at 1%; **Statistically significant at 5%; *Statistically significant at 10%.
The standard errors are in parentheses.

¹¹ Total profit of agricultural product sold is calculated by subtracting total farm production expenses from total market value of agricultural product sold.

Table 4.4 Two Stage Least Squares (2SLS) Regression Results for Determinants of Income

	Dependent variable	
	ln(Total market value of crops sold (\$/acre))	
	Model 1	Model 2
<i>Irrigation</i>		
Percentage of irrigated land	0.05*** (0.01)	0.05*** (0.01)
<i>Land quality</i>		
Fraction sand	-0.002 (0.002)	-0.002 (0.002)
Fraction flood-prone	0.003 (0.002)	0.003* (0.002)
K Factor	0.004 (0.01)	0.005 (0.01)
Slope length	0.0003 (0.0004)	0.0002 (0.0004)
Permeability	0.07*** (0.02)	0.07*** (0.02)
Wetlands	0.01*** (0.003)	0.01*** (0.002)
Moisture capacity	-0.03 (0.02)	-0.03 (0.02)
salinity	-0.03 (0.02)	-0.03 (0.02)
<i>Precipitation</i>		
Standard deviation of precipitation	0.08 (0.18)	
Cumulative precipitation		0.02* (0.01)
<i>Other variables</i>		
ln(Average commodity credit loans (\$/acre))	-0.01 (0.03)	-0.005 (0.03)
Age	1.22** (0.57)	1.21** (0.56)
Age ²	-0.01** (0.005)	-0.01** (0.005)
Year2002	-0.36*** (0.06)	-0.36*** (0.06)
State fixed effect	Yes	Yes
Adj R ²	0.42	0.43

Note: ***Statistically significant at 1%; **Statistically significant at 5%; *Statistically significant at 10%.
The standard errors are in parentheses.

Table 4.5 reports the effects of irrigation on income equality. Three variables are chosen to measure irrigation, they are average irrigated land (acres/farm), irrigated farm number, and percentage of irrigated land. The square forms of these variables are added to examine the

agricultural treadmill theory. All three models confirm that income inequality increases with increased irrigation adoption, but when more and more farmers adopt this technology, the marginal benefit becomes smaller and smaller, the income inequality thus decreases. An 10% increase in irrigated number of farms increases Gini coefficient by 0.021, but when the number of irrigated farms exceeds 483¹² on average, the Gini coefficient then start to decrease, which means income inequality begins to drop. Similar results can also be calculated by using percentage of irrigated land, when more than 20% of land is irrigated, income inequality will drop.

Table 4.5 Regression Results for Farm Sale Value Inequality

	Dependent variable: Gini coefficient		
	(1)	(2)	(3)
Constant	0.85*** (0.01)	0.23 (0.16)	0.82*** (0.01)
<i>Irrigation</i>			
ln(Average irrigated land) (acres/farm)	0.016*** (02)		
ln(Average irrigated land) ²	-0.0036*** (0.0004)		
ln(Irrigated farm number)		0.21*** (0.05)	
ln(Irrigated farm number) ²		-0.017*** (0.004)	
Percentage of irrigated land			0.002*** (007)
(Percentage of irrigated land) ²			-0.00005*** (0.00001)
<i>Other variables</i>			
Average farm size (acres/farm)	-0.0002*** (0.00002)	-0.00002 (0.00002)	-0.0002*** (0.00002)
ln(Average commodity credit loans) (\$/farm)	0.001 (0.002)	-0.000002** (8.88*10 ⁻⁷)	-0.006*** (0.002)
Year2002	0.04*** (0.005)	0.035*** (0.006)	0.04*** (0.006)
Adj R ²	0.48	0.42	0.44

Note: ***Statistically significant at 1%; **Statistically significant at 5%; *Statistically significant at 10%.
The standard errors are in parentheses.

¹² The maximum number of irrigated number of farms is calculated as $2 \cdot 0.017 \cdot x = 0$, $x = 6.18$. So when more than $e^{6.18} = 483$ farms adopt the irrigation technology, income inequality begins to drop.

6. Discussion and Conclusion

This paper addresses a major methodological problem that lies at the core of empirical literature on agricultural income, the potential endogeneity of irrigation used as explanatory variable.

Using number of wells as instrumental variable for irrigation adoption, I find that irrigation has a dramatic causal impact on the market value of crops sold. In addition, this study supports the treadmill theory that irrigation increases income inequality at first when a few farmers adopt this technology, but when more and more farms are involved in the system, income inequality decreases.

The implication of this research is potentially important from a public policy perspective. Farm consolidation, characterized by growing farm sizes, decreasing farm numbers, and shrinking agricultural GDP, is a dynamic process over the past five decades in the U.S. agricultural sector. Rich farmers have power to increase the agricultural supply and affect agricultural price by adopting new technologies. Small scale and poor farmers have to leave the scene because marginal profit was decreasing, their resources are absorbed by those who make the windfall profits or “scale enlargement. Thus, maybe instead of providing substantial subsidies for specific crops, the government should provide loans and technique supports to the poor farmers to adopt irrigation to help them increase on-farm income and alleviate income inequality in the long run.

There is one limitation in this study. Agricultural operation profit maybe a more ideal way to measure income inequality. However, due to data limitations, I can only use the total market value of agricultural products sold to calculate the Gini coefficient. Although it may not reflect the true farmers’ income inequality, this is the most innovative and the best method to proxy the income inequality using the Census of Agriculture dataset.

CHAPTER 5

Three chapters in the dissertation cover research topics including house foreclosure effects in housing economics and irrigation adoption effects in development economics. There are some connections among three chapters.

Chapter 1 and Chapter 2 use the same dataset to study the effects of house foreclosures on surrounding property sales values. However, the research methods are different. Chapter 1 uses cross-sectional data and employs spatial models. The GS2SLS regression is more appealing when the residuals are heteroskedastic and when the finite samples do not meet the normality requirement. The foreclosure effects extend up to 1500 feet of a property. The results present a slight larger spillover effects when compared to other studies. The marginal foreclosure impact is -1.57% within 300 feet, - 0.54% between 300 feet and 600 feet, -0.3% between 600 and 1200 feet, and -0.37% between 1200 feet and 1500 feet.

By using repeat sales from 2000 to 2010, Chapter 2 employs quasi-experiment models, difference-in-differences and propensity score matching methods to analyze the impacts of foreclosures on neighborhood one to four unit residential property values in city of Atlanta. Using repeat sales, the difference-in-differences and propensity score matching methods avoid omitted variable bias which is likely a problem in hedonic models using cross-sectional data.

Compared to Harding, Rosenblatt and Yao's (2009) work, this study improves their model from several aspects. First, besides number of nearby foreclosures, this study controls more property characteristics that are expected to change between sales, including whether the house remodeled or not, sales quarter and sales type. Second, instead of arbitrarily picking out

two repeat sales, this study includes every transaction record sold more than once during the study periods, which gives more precise estimates due to the efficiency gain from more data. Third, using transaction buyers and sellers' names, this study distinguishes between REO and REO sales. Because each REO sale resolves a REO, it helps increase surrounding house sales price as a result. The study separates effects of the price trend over time and the contagion effects of foreclosures. The results confirm negative contagion effects of foreclosures for surrounding sales properties. The difference-in-differences model reports that one more foreclosure (including REO and REO sales) reduces surrounding house sales prices by 2.5% within 300 feet, 1.1% within 600 feet, 0.6% within 900 feet, and 0.4% within 1200 feet. However, after separating REO and REO sales, the effect of REO increases dramatically, one more REO reduces surrounding sales price by 3.7%, while one more REO sales increases sales price by 0.6% within 600 feet.

Both Chapter 1 and Chapter 3 address potential endogeneity problems in the regression. The endogeneity problems in Chapter 1 and Chapter 3 are caused by reverse causality. Because neighborhood house values depreciated by foreclosures may lead to more foreclosures, foreclosures may thus be endogenous to the sales price. The contributions of Chapter 1 include creating an innovative way to examine endogeneity through accounting for foreclosure timing and it also addresses the endogeneity of the spatially lagged dependent variable by using GS2SLS procedures.

Chapter 3 deals with endogeneity with 2SLS regression. Because irrigation is a relatively expensive technology for small-scale farmers and poor farmers, it impedes their opportunities to adopt irrigation technology. Thus, irrigation is potentially endogenous to agricultural sales income. The coefficients estimated in IV-2SLS are about two and half times the magnitude as

those estimated in the OLS regression. Holding other variables constant, an increase of 1% of irrigated land leads to total market value of crop sold increasing by 5% per acre, and total market value of agricultural product sold increasing by 0.5% per acre.

The research results in three Chapters provide important public policy implications. Because property taxes fund local public goods, losses in the property taxes revenues would have a multiplier impact in degrading provision of local public goods. The estimated property tax loss for 10,121 one-to-four unit family houses at Atlanta is about \$2.2 million in 2008. If the full spectrum of houses types and foreclosures were considered, reducing foreclosures would result in an even higher social benefit. Policy makers should consider programs to make foreclosures resolve in a timely manner to avoid tax loss. The resolved foreclosures actually could help increase surrounding house sales prices which is proved by the study results in Chapter 2.

The implication of Chapter 3 indicates that the government should provide loans and technique supports to the poor farmers to adopt irrigation to help them increase on-farm income and alleviate income inequality in the long run.

REFERENCES

- Bai, D., 2008. Irrigation, Income Distribution, and Industrialized Agriculture In the Southeast United States, Thesis, Department of Agricultural Economics and Rural Sociology, Auburn University.
- Bajaj, V., 2007. Increasing Rate of Foreclosures Upsets Atlanta, The New York Times, Available at <http://www.nytimes.com/2007/07/09/business/09auctions.html>.
- Baxter, V., Lauria, M., 2000. Residential Mortgage Foreclosure and Neighborhood Change, House Policy Debate 11(3), 675-699.
- Bertrand, M., Duflo, E., Mullainathan, S., 2004. How Much Should We Trust Difference-In-Differences Estimation? The Quarterly Journal of Economics 119(1), 249-275.
- Bucholtz, S.J., 2004. Generalized Moments Estimation for Flexible Spatial Error Models: A Library for Matlab, Available at facstaff.uww.edu/welschd/jplv7/spatial/gmm_models/sem_gmm.doc.
- Bunce, H.L., Gruenstein, D., Herbert, C.E., Scheessele, R.M., 2000. Subprime Foreclosures: The Smoking Gun of Predatory Lending? Available at http://griequity.astraea.net/resources/industryandissues/financeandmicrofinance/predatory_lending/subprimeforeclosures200602.pdf.
- Calem, P.S., Hershaff, J.E., Wachter, S.M., 2004. Neighborhood Patterns of Subprime Lending Evidence from Disparate Cities, Working Paper, Available at http://ssrn.com/abstract_id=583102.
- Campbell, J.Y., Giglio, S., Pathak, P., 2011. Forced Sales and House Prices. The American Economic Review 101(5), 2108-2123.

- Caswell, M., Zilberman, D., 1985. The Choice of Irrigation Technologies in California. *American Journal of Agricultural Economics* 67(2), 224-234.
- Caswell, M., Zilberman, D., 1986. The Effects of Well Depth and Land Quality on the Choice of Irrigation Technology. *American Journal of Agricultural Economics* 68(4), 798-811.
- Chambers, R., 1988. *Managing Canal Irrigation*. Cambridge, UK: Cambridge University Press.
- Chan, S., Been, V., Gedal, M., Haughwout, A., 2010. Mortgage Default Risk: Recent Evidence from New York City, Paper Presented at 36th Eastern Economic Association Annual Conference, Philadelphia, February 26-28.
- Coca-Perraillon, M., 2006. Matching with Propensity Scores to Reduce Bias in Observational Studies. Proceedings of NorthEast SAS Users Group Conference (NESUG), Philadelphia, PA.
- Cochrane, J. D., 1958. The Frequency Distribution of Water Characteristics in the Pacific Ocean. *Deep-Sea Research* 5(2), 111-27.
- Deschênes, O., Greenstone, M., 2007. The Economic Impact of Climate Change: Evidence from Agricultural Output and Random Fluctuations in Weather. *American Economic Review* 97(1), 354-385.
- Foster, J. E., Sen, A., 1997. *On Economic Inequality*, Expanded Edition with A Substantial Annexe. New York, NY: Oxford University Press.
- Gerardi, K., Shapiro, A.H., Willen, P.S., 2007. Subprime Outcomes: Risky Mortgages, Homeownership, Experiences, and Foreclosures, Federal Reserve Bank of Boston Working Paper No. 07-15.

- Gerardi, K., Sherlund, S.M., Lehnert, A., Willen, P., 2008. Making Sense of the Subprime Crisis, in: Elmendorf, D.W., Mankiw, N.G., Summers, L.H. (Eds.), *Bookings Papers on Economic Activity*, Washington, DC: Brooking Institution Press.
- Hanna, B.G., 2007. House Values, Incomes, and Industrial Pollution, *Journal of Environmental Economics and Management* 54, 100-112.
- Harding, J.P., Rosenblatt, E., Yao, V.W., 2009. The Contagion Effect of Foreclosed Properties. *Journal of Urban Economics* 66(3), 164-178.
- Hartarska, V., Gonzalez-Vega, C., 2006. Evidence on the Effect of Credit Counseling on Mortgage Loan Default by Low-Income Households, *Journal of Housing Economics* 15(1), 63-79.
- Heckman, J.J., 1979. Sample Selection Bias as A Specification Error, *Econometrica* 47(1), 153-161.
- Hite, D., Chern, W., Hitzhusen, F., Randall, A., 2001. Property-Value Impacts of an Environmental Disamenity: The Case of Landfills, *Journal of Real Estate Finance and Economics* 22, 185-202.
- Hite, D., 2006. Out of market transactions as neighborhood quality indicators in hedonic house price models, Working Paper, Department of Agricultural Economics and Rural Sociology, Auburn University.
- Huang, Q., Dawe, D., Rozelle, S., Huang, J., Wang, J., 2005. Irrigation, poverty and Inequality in Rural China. *The Australian Journal of Agricultural and Resource Economics* 49, 159-175.
- Hussain, I., 2007. Direct and Indirect Benefits and Potential Disbenefits of Irrigation: Evidence and Lessons. *Irrigation and Drainage* 56, 179-194.

- Imbens, G.W., Wooldridge, J.M., 2007. Difference-In-Differences Estimation. National Bureau of Economics Research Working Paper.
- Immergluck, D., Smith, G., 2005. Measuring the Effect of Subprime Lending on Neighborhood Foreclosures Evidence from Chicago, *Urban Affairs Review* 40(3), 362-389.
- Immergluck, D., Smith, G., 2006. The External Costs of Foreclosures: The Impact of Single-Family Mortgage Foreclosures on Property Values, *House Policy Debate* 17(1), 57-79.
- Immergluck, D., 2009. Neighborhoods in the Wake of the Debacle: Intrametropolitan Patterns of Foreclosed Properties, Working Paper, Available at http://ssrn.com/abstract_id=1533786.
- Kelejian, H.H., Prucha, I.R., 1998. A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances, *Journal of Real Estate Finance and Economics* 17(1), 99-121.
- Kelejian, H.H., Prucha, I.R., 1999. A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model, *International Economic Review* 40(2), 509-533.
- Leonard, T., Murdoch, J.C., 2009. The Neighborhood Effect of Foreclosure, *Journal of Geographical Systems* 11, 317-332.
- Lesage, J.P., 1998. Spatial Econometrics. Available at <http://www.spatial-econometrics.com/html/wbook.pdf>.
- Lichtenberg, E., 1989. Land Quality, Irrigation Development, and Cropping Patterns in the Northern High Plains. *American Journal of Agricultural Economics* 71(1), 187-194.
- Lin, Z., Rosenblatt, E., Yao, V.W., 2009. Spillover Effects of Foreclosures on Neighborhood Property Values, *Journal of Real Estate Financial Economics* 38, 387-407.
- Lipton, M., 2007. Farm Water and Rural Poverty Reduction in Developing Asia. *Irrigation and Drainage* 56, 127-146.

- Miguel, E., 2004. "Economic Shocks and Civil Conflict: An Instrumental Variable Approach".
The Journal of Political Economy 112(4): 725-753.
- Portney, P. R., 1981. Housing Prices, Health Effects, and Valuing Reductions in Risk of Death,
Journal of Environmental Economics and Management 8, 72-78.
- RealtyTrac, 2011. 2011 Year-End Foreclosure Report: Foreclosures on the Retreat. Available at
<http://www.realtytrac.com/content/foreclosure-market-report/2011-year-end-foreclosure-market-report-6984>.
- Rogers, W.H., Winter, W., 2009. The Impact of Foreclosures on Neighboring Housing Sales,
Journal of Real Estate Research 31(4), 455-480.
- Schuetz, J., Been, V., Ellen, I.G., 2008. Neighborhood Effects of Concentrated Mortgage
Foreclosures, *Journal of Housing Economics* 17, 306-319.
- Shlay, A., 2006. Low-income homeownership: American dream or delusion? *Urban Studies*
43(3), 511-531.
- Sheu, M.L., Hu, T.W., Keeler, T.E., Ong, M., Sung, H.Y., 2004. The Effect of A Major Cigarette
Price Change on Smoking Behavior in California: A Zero-Inflated Negative Binomial
Model, *Health Economics* 13, 781-791.
- Shrestha, R.B., Gopalakrishnan, C., 1993. Adoption and Diffusion of Drip Irrigation
Technology: An Econometric Analysis. *Economic Development and Cultural Change*
41(2), 407-418.
- Skaggs, R., 2000. Drip Irrigation in the Dessert: Adoption, Implications, and Obstacles. Paper
Presented to Western Agricultural Economic Association.
- Skogan, W.G., 1990. Disorder and Decline: Crime and the Spiral of Decay in American
Neighborhoods. University of California Press, Berkeley and Los Angeles, CA.

- Smith, L. E., 2004. Assessment of the Contribution of Irrigation to Poverty Reduction and Sustainable Livelihoods. *Water Resources Development* 20(2), 243-257.
- Towe, C., Lawley, C., 2010. The Contagion Effect of Neighboring Foreclosures on Own Foreclosures, Working Paper, Available at http://www.smartgrowth.umd.edu/research/pdf/Towe&Lawley_ForeclosureContagion_120110.pdf.
- Uchoa, E. Atlanta Has One of the Highest Foreclosure Rates in the Nation. Available at <http://www.articlesbase.com/small-business-articles/atlanta-has-one-of-the-highest-foreclosure-rates-in-the-nation-461590.html>.
- Vandell, K.D., 1991. Optimal Comparable Selection and Weighting in Real Property Valuation, *ARERUEA Journal* 19(2), 213-239.