THREE ESSAYS ON HOUSING MARKET AND SPATIAL DISAMENITIES

by

Lin Cui

B.S. in Economics, Tsinghua University, Beijing, 2006

Submitted to the Graduate Faculty of

Arts and Sciences in partial fulfillment

of the requirements for the degree of

PhD in Economics

University of Pittsburgh

2011

UNIVERSITY OF PITTSBURGH

ARTS AND SCIENCES

This dissertation was presented

by

Lin Cui

It was defended on

July 18, 2011

and approved by

Dr. Randall Walsh, University of Pittsburgh, Economics

Dr. Mark Hoekstra, University of Pittsburgh, Economics

Dr. Werner Troesken, University of Pittsburgh, Economics

Dr. Dennis Epple, Carnegie Mellon University, Economics

Dissertation Advisor: Dr. Randall Walsh, University of Pittsburgh, Economics

THREE ESSAYS ON HOUSING MARKET AND SPATIAL DISAMENITIES

Lin Cui, PhD

University of Pittsburgh, 2011

Spatial disamenities can affect neighborhood quality in many ways, and carefully quantifying such effects is essential for policy making. The first chapter of my dissertation focuses on the impact of foreclosures and vacancies on crime. To overcome confounding factors, a difference-in-difference research design is applied to a unique data set containing geocoded foreclosure and crime data from Pittsburgh, Pennsylvania. Results indicate that while foreclosure alone has no effect on crime, violent crime increases by more than 15% once the foreclosed home becomes vacant. The second chapter examines the spillover effects of foreclosures and vacancies on the quantities and prices of properties sold in neighboring areas using the same foreclosure data and similar econometric design. Estimation results show that both foreclosure and vacancy reduce the neighboring houses' probabilities of sale. Also, there is little impact on houses with lower quantity index, and the effects disappear when foreclosed house is reoccupied. This paper is the first study to document the quantity shifts of homes sold at the time of nearby foreclosure in different sections of the housing market as a result of changes in both the demand side and the supply side. The last chapter examines the impact of new shale gas drilling technologies in the Marcellus region on rural residential property values using data from three counties with most drilling activities in Pennsylvania. The results suggest that property values are negatively correlated with the presence of nearby gas wells, though the effects are not statistically significant. Due to mineral right transfer issues, the estimates in this study are the result of two competing effects. The estimated coefficients may be lower-bounds of the actual impact.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

I especially thank my main advisor, Randall Walsh, for his invaluable guidance and support. I also thank Sabina Deitrick, Bob Gradeck and Chris Briem at Pittsburgh Neighborhood and Community Information System, who provided the housing, foreclosure and crime data and, just as importantly, shared their expertise in working with those data. I am grateful for the comments and suggestions provided by Dennis Epple, Mark Hoekstra and Werner Troesken.

1.0 FORECLOSURE, VACANCY AND CRIME

1.1 INTRODUCTION

There are many social problems arising from foreclosure. On individual levels, families undergoing foreclosure can lose accumulated home equity and access to future stable housing; on social levels, foreclosure can have implications for surrounding neighborhoods and larger communities. One potential impact of increased foreclosures in a community is crime.

Sociologists have long theorized a link between neighborhood characteristics and the geographical distribution of crime (the social disorganization theory). For example, Shaw et al. (1929) suggested that high crime rates occur in areas that are characterized by physical deterioration. On a more conceptual level, Faris (1948) stated that crime rates are reflections of the degree of disorganization of the control mechanisms in a society. A modern version is Wilson and Kelling's (1982) broken windows theory, which posits that neighborhood-level disorder is a precursor to serious crime. Skogan (1990) further categorized disorder as social or physical. Social disorder refers to delinquent behavior, such as public drinking, and physical disorder refers to visual signs of negligence, such as abandoned buildings. The idea is that disorder reduces a community's willingness to maintain social control and provide better opportunities for crime.

Skogan (1990) views foreclosed[1] and vacant buildings are a type of neighborhood physical disorder. The problem starts when a homeowner facing foreclosure takes less care of the house. While the property can still be occupied, it may already show visible signs of disrepair. This may signal to potential criminals a lower level of surveillance in the nearby area and thus increase their incentive to commit crimes. Later, if the property becomes vacant, the lack of surveillance is more apparent and those neglected and abandoned buildings can offer criminals places to gather and conduct their activities. The above discussion suggests that both foreclosure and vacancy are positively associated with crime rates, with vacancy possibly having a stronger impact.

While the relationship between foreclosures and crime has received widespread attention in the news media, to date there has been little careful empirical work on this subject. The two exceptions are Immergluck and Smith (2006), and Goodstein and Lee (2010). Using Chicago area foreclosure and crime data for the year 2001, Immergluck and Smith find that a 1 percent increase in the foreclosure rate leads to a 2.3 percent increase in violent crime rate in one tract. Goodstein and Lee (2010) look at foreclosure and crime in a panel data setting and conclude that foreclosure increases burglary and some other property crime at county level. On the vacancy and crime front, there is a single paper by Spelman (1993). Using field data on building conditions within one neighborhood, he finds that blocks with vacant properties have higher crime rates compared to blocks with fully-occupied buildings.

[1] Because there are multiple stages in the foreclosure process, there is no consensus on the exact definition of foreclosure in the literature. For the rest of this paper, I use the terms "foreclosure" and "foreclosure filing" interchangeably to refer to an earlier stage in the foreclosure process, when a lis pendens has been filed. I am aware of the fact that some of the papers I cite here may define foreclosure as a later stage in which a property is sold at sheriff sale.

The results of Spelman and Immergluck and Smith are suggestive, but these studies rely on cross-sectional data and thus face a potentially serious omitted variable problem. The locations of foreclosures/abandoned buildings are likely to be correlated with unobserved neighborhood factors that determine crime rates. Further, increases in the number of foreclosed/vacant properties may correspond to other changes that cause higher levels of crime, or could simply be the result of crime[2]. In addition, neither of the studies estimates the impact of different stages of foreclosure (foreclosure vs. vacancy) on crime. Spelman (1993) focuses only on vacancy (completely independent of the foreclosure process), and Immergluck and Smith (2006) conclude that a higher foreclosure rate leads to more crime but actually only consider foreclosure driven vacancies[3]. Goodstein and Lee (2010) avoid the potential caveats of cross-sectional studies by looking at foreclosure and crime in multiple years. However, given that crime is predominantly a local issue, it is likely to correlate with neighborhood specific trends, which cannot be captured in their county-level analysis.

With information on both foreclosure filing and foreclosure-led vacancy, this study is the first to separate the impact of foreclosure from the impact of vacancy during the foreclosure process[4]. The mechanisms through which the two impact neighborhood crime rates are similar, as discussed above, but the effects can be different in scale. In addition, this study provides new evidence on the impact of length of vacancy on crime, concluding that longer terms of vacancy

---

[2] Although the author is not aware of any study that examines the impact of crime on foreclosure/vacancy rate, various studies have proposed a reciprocal connection between crime and neighborhood characteristics (Faris, 1948, Skogan, 1986 and Bursik, 1988), and have empirically confirmed the link (Examples are Liska and Bellair, 1995 and Hipp, 2010).

[3] They define foreclosure as the eviction of borrowers after sheriff sale; hence in their study a foreclosed property is essentially a vacant property.

[4] There is one study that separately examines the effect of foreclosure and non-foreclosure vacancy on nearby property values (Mikelbank, 2008). However, his analysis is plagued by the fact that the vacancy data is rather inaccurate and incomplete.

have stronger effect on crime. Though seemingly obvious, results in this analysis has important policy relevance as the communities and local government can always reduce the damages of foreclosure by recycling the distressed properties more quickly, even when it is difficult to prevent foreclosure altogether.

More importantly, this study is also the first to exploit both intertemporal and cross-sectional variance in foreclosure and foreclosure-led vacancy, and their effects on crime. In this way, unlike the existing literature, I am able to control for the possibility of confounding unobservables. Further, data on the exact locations of foreclosure-filings, vacancies and crime incidents allow me to exploit variations of crime within small, relatively homogenous areas surrounding the foreclosed and/or vacant properties. Knowledge of the exact timing of foreclosure-filing and vacancy allows me to confirm the absence of substantive preexisting differences in crime rate of treatment and control areas. As demonstrated later, my analysis strongly shows the importance of the robust identification strategy used here.

Taking a difference in difference approach, I find that violent crime rates are more than 15% higher in areas within 250 feet of foreclosed and vacant properties compared to areas only slightly farther away. Results indicate that foreclosure alone has no effect on crime. Effects on property crime are found to be similar, but are less precisely estimated.

I proceed as follows, Section 1.2 of the paper presents background information on the current foreclosure crisis and the foreclosure process in my study area Pittsburgh, Pennsylvania. In Section 1.3, I describe the data used in this study. In Section 1.4, I describe the empirical methodology and present graphical evidence on the impact of foreclosure and vacancy. Empirical results are presented in Section 1.5 and I conclude in Section 1.6.

1.2 BACKGROUND

Since 2006, the United States has experienced a significant increase in home mortgage foreclosures. As of the third quarter of 2009, residential mortgage delinquency rates, as reported by the Mortgage Bankers Association (MBA), stood at a seasonally adjusted 9.64 percent – the highest level ever reported since the MBA began tracking foreclosures in 1972. Though the foreclosure crisis was mainly attributed to predatory lending practices and subprime mortgages, the MBA reports that the fastest-rising segment of foreclosures in recent months has been traditional prime mortgages, reflecting problems within macroeconomic factors, such as the unemployment rate. Given the current economic situation, the crisis will likely continue to present a significant policy challenge for local governments.

The proposed research considers the impacts of foreclosure and vacancy in the context of Pittsburgh, Pennsylvania. In Pennsylvania all foreclosures are carried out through the court system and a lender must follow a state-level judicial process in order to foreclose on a property. The process begins when the borrower fails to make payments for at least 60 days. At that time, the lender can initiate the foreclosure process by sending a Notice of Intent to Foreclose. If the borrower pays all dues and fees within 30 days, the default is “cured.” However, if the borrower is either unable or unwilling to resolve the debt, the entire balance of the mortgage becomes due immediately. The lender can then file a suit to obtain a court order to foreclose on the property (foreclosure). Sometimes the borrower resolves with the lender or successfully sells the property to another permanent owner before the sheriff sales date and the property does not become vacant. Otherwise the lender can choose to sell the property at sheriff sale, setting the opening bid for at least the outstanding loan amount. Typically under this process the property is not sold and will return to the lender (Had there been any potential buyer in the property at a price equal to the outstanding loan amount, the owner could have sold it earlier), most of the time a bank or a mortgage company. The borrower is evicted after the sheriff sale. The foreclosed property is then classified as a real estate owned (REO) property and stays vacant until it is sold to a new permanent owner.

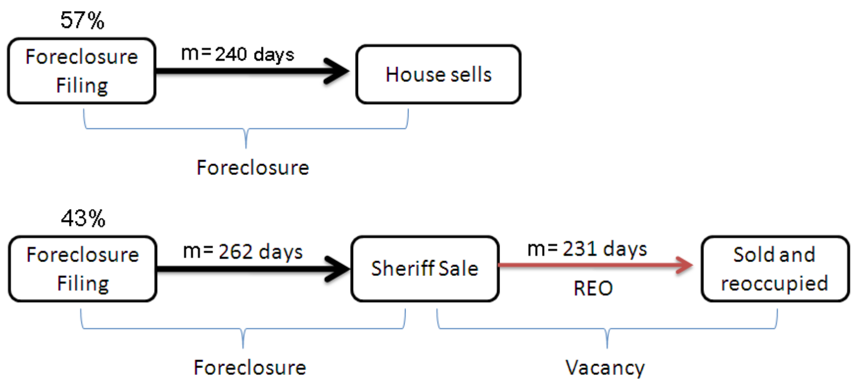

Figure 1.1 - Foreclosure Process

As described above, a typical foreclosure case consists of multiple stages: foreclosure filing, sheriff sale, and sale to a new permanent owner (REO sale). Note that some foreclosure filings are finally resolved and do not reach the point of sheriff sale. Figure 1.1 provides an illustration of the two possible outcomes following foreclosure filing. In this study, 57% of the foreclosure filings result in property sale to another permanent owner before sheriff sale (outcome 1) while 43% experience a period of vacancy until they are finally resold (outcome 2).

To differentiate foreclosure per se from foreclosure-led vacancy, I define up to four stages of the foreclosure process: a pre-foreclosure stage that takes places before the date of foreclosure filing, a foreclosure stage between the date of filing and the date of sheriff sale (or the date of property sale for outcome 1), a vacancy stage, and a reoccupation stage that takes place after the

REO sale date. Due to the judicial nature of foreclosure in Pittsburgh, the whole process typically takes one to two years to complete. As shown in Figure 1.1, the median length of foreclosure stage is 240 days for those without sheriff auction, and 262 days for those experiencing vacancy. The median length of vacancy is 231 days.

Although foreclosure activity reached record high in the third quarter of 2009, Pennsylvania has not been hit as hard as many other regions. The foreclosure rate was 2.58 foreclosures per 1000 households for Pennsylvania during the third quarter of 2009, a 15.48% increase from the third quarter of 2008, while the national average is 7.35 foreclosures per 1000 households, a 22.50% increase from the third quarter of 2008.

1.3 DATA

Synthesizing information from multiple sources, my data are distinguished, in part, by an unusually fine level of geographic precision. The data enable me to exploit the exact timing and locations of foreclosure, vacancy and crime incidents. Here, I describe the data sources, the hierarchical coding of the crime records, and the method to identify periods of vacancy.

1.3.1 Data Sources

Foreclosure filing: The foreclosure filing data are obtained from City of Pittsburgh court records. This file contains information on every foreclosure filing in the city from 2006 to November 2009, such as the date of filing, the parcel ID of the property receiving foreclosure filing, borrower and lender names, and the current stage of filing. Settled and discontinued cases are deleted from the sample. Note that only residential[5] properties are included in this analysis.

Housing transaction: The property transaction data come from Allegheny County Recorder of Deeds Office, which contains sale date, price, parcel ID and buyer and seller names for every property transaction since 1986. Foreclosure filings are linked to all subsequent property transactions by parcel ID to determine the periods of vacancy and reoccupation. Details about the linking process are provided in Section 1.3.3.

Housing and neighborhood characteristics: The housing characteristics data are obtained from Allegheny County's Office of Property Assessments. Most of the information is taken from the last county-wide reassessment in 2002. These assessment data contain housing conditions such as square feet, number of bedroom and year structure was built for every property in the city. Data on neighborhood characteristics come from two sources. Selected block-level demographic characteristics (such as race and age) are taken from the 2000 Census of Population and Housing. Information only available at more aggregate levels (such as education and income) is not included due to the lack of geographic precision. Pre-existing (2005) crime counts come from the crime data described below.

Crime: Crime data is obtained from the Police Department of the City of Pittsburgh. This data includes type of crime and the exact time and street address of each reported crime incident from 2005 to 2009. All the records are geocoded to identify nearby crimes for each foreclosed property.

[5] A property is defined as residential if its structure type falls into one of the following categories: single family, two to four family, row house and townhouse.

1.3.2 Hierarchical Coding of Crime Records

As a participant in the Uniform Crime Reporting (UCR) Program, Pittsburgh's police department follows UCR's guideline of classifying and reporting offenses. All offenses are first classified into 26 categories in a particular order, with homicide being the highest in the hierarchy. In case of a multiple-offense situation, the police department will record only one offense that is the highest on the hierarchy list and not the other offense(s) involved. For example, one crime incident described as both robbery and homicide will appear as homicide but not robbery.

Table 1.1 - Categories of Crime Incidents

Crime	Code	Pct	Crime	Code	Pct
Violent		8.86%	Other		55.90%
Murder-Manslaughter	01	0.13%	Stolen property	13	0.65%
Forcible rape	02	0.25%	Vandalism	14	14.64%
Robbery	03	3.91%	Weapon violations	15	1.16%
Aggravated assault	04	4.57%	Prostitution	16	0.66%
Property		35.24%	Sex offenses	17	0.82%
Burglary	05	8.56%	Drug violations	18	6.91%
Larceny – Theft	06	21.98%	Gambling	19	0.02%
Motor vehicle theft	07	4.51%	Family violence	20	0.16%
Arson	08	0.19%	Drunken driving	21	2.46%
Other		55.90%	Liquor law violations	22	0.09%
Forgery	09	1.31%	Public drunkenness	23	0.49%
Simple Assault	10	15.78%	Disorderly conduct	24	2.96%
Fraud	11	2.36%	Vagrancy	25	0%
Embezzlement	12	0.12%	Other	26	5.33%

Percentages are calculated from all crimes in the City of Pittsburgh in 2006.

Table 1.1 provides a description of the 26 crime categories and percentages of each type of offense recorded in the data. Note that violent crimes are coded in highest hierarchical order, followed by property crimes. As a result the coding rule will not change the total number of violent crimes but all other crimes will be under-reported. The degree of under-reporting increases while moving down the hierarchy list.

Due to this coding rule, I focus my analysis on violent and property crimes, as they have higher priorities to be coded, and thus provide a more accurate measure of the actual reported number of crime incidents.

1.3.3 Identifying Periods of Vacancy

A key challenge of this analysis is to carefully identify the exact period of vacancy for each foreclosed property. In most cases, a foreclosed property is seized by the lender at the sheriff sale, and it becomes vacant immediately thereafter. The property then stays vacant until it is resold to a new permanent owner (REO sale). As a result, the REO status can typically define most foreclosed properties' vacancy periods.

The REO status is identified by two dates: the sheriff sale date and a subsequent REO sale date. Linking foreclosure filings to home sale data enables me to track the complete transaction history of each foreclosed property. In addition, all sheriff sales in the deed recorder are categorized as "sheriff deed" rather than "deed", which serves as a clear identifier. Therefore, a sheriff sale date is assigned to a foreclosure if on that date the property is recorded on a sheriff deed with the seller as the borrower who fails to sell the house before the sheriff sale. An REO sale date is thus defined as a subsequent transaction date when the REO property, under the name of a bank or mortgage company, is finally sold to a new permanent owner.

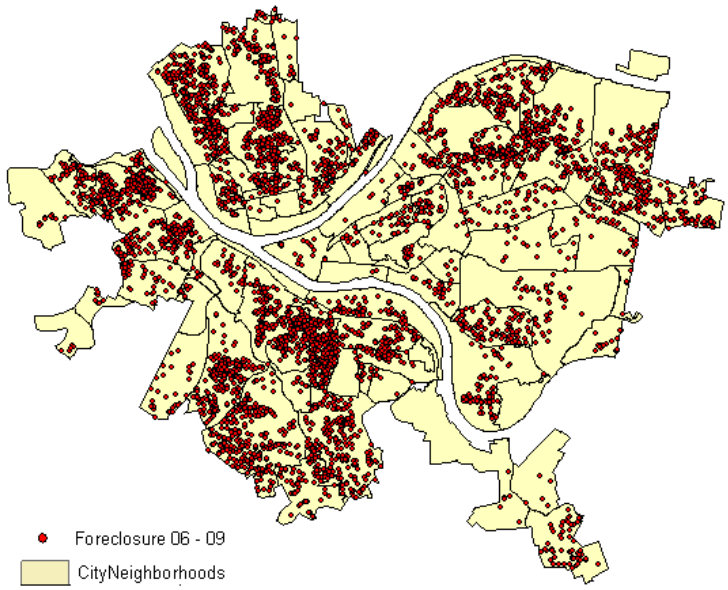

Figure 1.2 - Foreclosures in Pittsburgh, 2006-2009

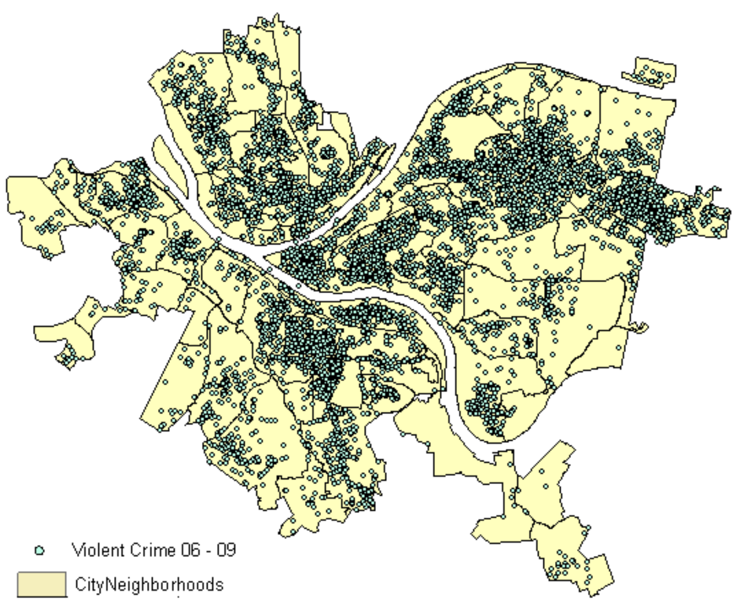

Figure 1.3 - Violent Crime in Pittsburgh, 2006-2009

Occasionally, the REO status does not coincide with the period of vacancy, for instance, when the borrower abandons the property before the sheriff sale date. To address this issue I use data on gas shut off dates to assist in the identification of vacant foreclosed properties. This dataset is obtained from the 3 major gas companies providing service to virtually every property in the City of Pittsburgh. This data contains a list of addresses with no gas usage on a specific day every December from 2006 to 2009.

Combining information on REO and gas shut off status, I begin by defining the starting point of vacancy as the date a foreclosed property is sold to a bank/mortgage company (sheriff sale), and the ending point of vacancy as the date of the next transaction, when the property is resold to a new permanent owner (REO sale). If the property has an REO period and the gas shutoff month occurred prior to the sheriff sale date, I will assign the December 31st of the gas shut off year as the starting date of vacancy. Among the 3,282 properties foreclosed between 2006 and 2009 in Pittsburgh, 1,403 experienced vacancy. Among those, the vacancy periods of 1,213 (86%) properties are solely defined by REO status.

Nevertheless, it is likely that the method described above can only generate close approximations of the actual vacancy periods. As a result, some of the foreclosed properties may be classified as vacant when they are in fact occupied, or vice versa. This will lead to a downward (attenuation) bias in my estimates of the differences in crime rates between vacancy and non-vacancy periods.

1.4 EMPIRICAL STRATEGY

Crime incidents are not randomly scattered across a metropolitan area. Potential criminals look for locations with high opportunity of finding a target and low risk of being caught. Foreclosed houses typically wear visible signs of disrepair, signaling to potential offenders that the neighborhood lacks surveillance and social control. Foreclosed and vacant houses not only show signs of disrepair but can also serve as a private place for potential criminals to plan crimes or engage in other activities that would attract too much attention if done in public view. Thus, it is likely that areas immediately surrounding such houses will see a higher rate of crime incidents than areas farther away. It is also expected that foreclosure alone has a smaller impact on crime than vacancy does.

For several reasons, violent crimes are likely to be the best measures of actual incidents not only because they have the highest priority to be coded in the data, but also because they are most likely to be reported.

As is documented[6] in both the academic literature and the popular press, foreclosures tend to cluster in lower-income neighborhoods with higher portion of minority residents and subprime mortgages. The correlation between foreclosure locations and both observable and unobservable neighborhood characteristics makes it difficult to identify the effect of foreclosure and vacancy on crime rates by simply comparing areas with and without foreclosed houses. If the locations of foreclosures correlate with some unobservable neighborhood characteristics that affect crime rates, cross-sectional analysis will yield biased estimates. Further, cross-sectional analysis cannot rule out the possibility that the observed foreclosures are the result of crime rather than the cause.

[6] Examples are Gerardi, Shapiro,and Willen (2007), and Immergluck and Smith (2004).

While previous studies compare data aggregated up to county, census tract or block level, my data provides the specific locations of crime incidents and foreclosed properties, the dates of foreclosure filings, as well as the start and the end of vacancy period. The specific location data allow me to compare crime rates within small proximate areas in which neighborhood characteristics are more homogenous than in aggregate comparisons of the existing literature.

As discussed in Section 1.1, the possible mechanisms of foreclosure and crime indicate that the effect on crime is likely to be a continuous decreasing function of the distance to the foreclosure site. As a result, it is difficult to define pure treatment and control areas, as shown in the first graph of Figure 1.4. Instead, the defined control area may be partly treated, and the defined treatment area may be partly untreated, as shown in the second and third graphs of Figure 1.4. In both cases, any differences in crime counts between the treatment and control areas will be underestimated. Similar to potential measurement error in the timing of vacancy, this would lead to attenuation bias in my estimates of the effect of foreclosure and vacancy on crime.

I proceed by defining the treatment area as a circle with a foreclosure site as the center point, and control as area that lies directly adjacent to the treatment circle. Because the number of crimes in a given location is a function of both crime density and the location's area, control locations are defined so as to be identical in area to treatment locations, as illustrated in Figure 1.5.

To minimize the possibility of clustering yet to obtain reasonable counts of crime incidents, I define my treatment area as all locations within 250 feet of a foreclosed property. Control area is defined to be a set of immediately proximate locations adjoining this 250 feet circle. I use the region between 250 and $250*\sqrt{2}$ ($\cong$353.6) feet of the foreclosed and/or vacant properties to make the treatment and control areas equal in size. Figure 1.5 provides an illustration of the treatment and control areas surrounding a foreclosed property.

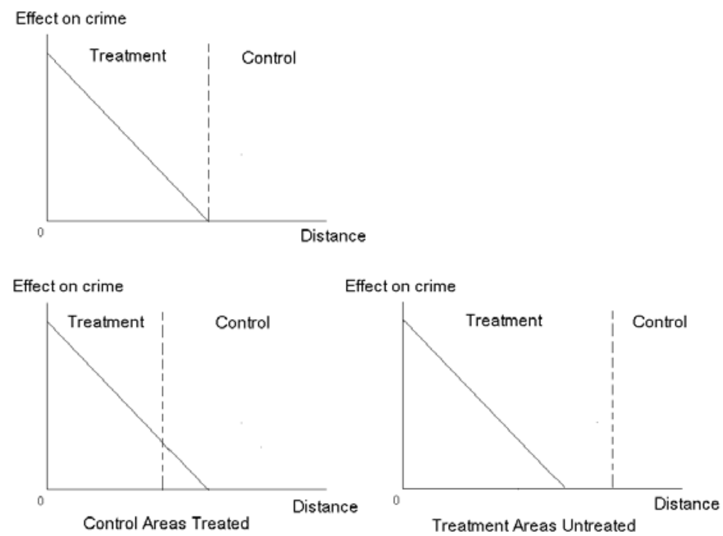

Figure 1.4 - Relationship between Distance to Foreclosure and Crime

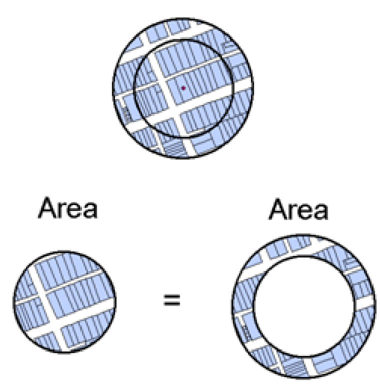

Figure 1.5 - Treatment and Control Areas Surrounding a Foreclosed Property

Because my estimation strategy hinges on the relative similarity of treatment and control areas, I compare a list of housing and neighborhood characteristics in the two groups. For housing characteristics, I first identify all houses located in a treatment area, and then calculate the average housing characteristics for each treatment ring (for instance, the average square feet and number of bedrooms). I repeat the process for houses located in control areas. For neighborhood characteristics, average demographic[7] and crime measures are calculated for each treatment/control area.

As a diagnostic test, I estimate the differences in housing and neighborhood characteristics across the treatment and control groups. I run a series of regressions with each housing or neighborhood characteristics as the dependent variable and a treatment indicator as the only independent variable. If the coefficient on the treatment indicator is insignificant across all housing and neighborhood characteristics, it will be suggestive that treatment and control groups are readily comparable (at least in terms of observables).

Table 1.2 compares characteristics of houses inside the treatment and control rings, as well as demographics in Census 2000 and crime rates in 2005 for these areas. To better characterize pre-existing housing conditions, I also include average last sale price and average days since last sale for transactions taking place on or before December 31, 2005. Love and affection sales are excluded from the sample. Note that the total number of houses inside treatment or control areas is 36,410 and 38,985, respectively. Although no statistical estimation of the difference can be provided, the two numbers are fairly close. Given that the treatment and control areas have

[7] Demographic measures are constructed by mapping Census 2000 block boundaries and treatment/control areas, calculating the percent area of each block inside a treatment/control and the demographic counts in each intersection area, and aggregating the block-level counts to the areas of control and treatment, assuming everything is of equal geographic distribution within a block. I only include block-level demographics because the average size of a block is small enough to create a reasonable measure for my treatment/control areas.

similar number of houses, it is plausible to assume that there are also similar numbers of people living in the two areas[8]. As shown in column (1) and (2), all housing and neighborhood measures look very similar in both groups, suggesting that the small treatment and control areas are relatively homogenous.

The third column of Table 1.2 reports point estimates and standard errors of coefficients on the indicator of being in the treatment group. While all other coefficients are statistically insignificant, the regression result indicates that current homeowners in the control areas stay in their houses for longer time. However, considering the difference in days since last sale is merely 55 compared with a baseline level of 5,583, there is little evidence of noticeable differences in observed housing and neighborhood characteristics between the two groups. Overall, the results in all three columns suggest the similarity of the treatment and control areas.

Given that treatment and control are of the same area and there is little difference in population in the two groups, it is feasible to directly compare crime counts in those two areas. To build the panel data set that facilitates this study's difference-in-differences approach, it is necessary to aggregate crime totals for fixed time periods for each location. To guarantee reasonable counts of each crime type, I aggregate by quarter. For each foreclosed property, quarterly crime counts in the treatment and control areas for different quarters are labeled according to the specific timing of foreclosure filing and vacancy of the reference foreclosed property[9]. As a result my unit of observation is the number of violent or property crimes in a treatment or control area in one calendar quarter.

---

[8] There is no precise measure of population in 2000 Census because population is aggregated to block level.

[9] For example, crimes surrounding property i in the fourth quarter of 2006 are labeled as "a quarter before foreclosure" if property i receives a foreclosure filing in the first quarter of 2007.

Table 1.2 - Characteristics of All Properties in Treatment and Control Areas

	Around Properties with Foreclosure Filings		
	Within 250 feet	Within 250-353 feet	Differences
Housing Characteristics			
Square Footage	1,582	1,597	-14.45
	(385.2)	(371.8)	(9.499)
Assessment Value	49,839	50,173	351.29
	(90,055)	(66,585)	(830.66)
Lot Area (in square feet)	3,712	3,740	-53.44
	(1,977)	(4,488)	(50.30)
Year Since Built	88.07	88.05	0.063
	(16.69)	(15.85)	(0.402)
Bedroom	2.960	2.975	0.0019
	(0.503)	(0.494)	(0.0118)
Bathroom	1.358	1.369	-0.0079
	(0.307)	(0.308)	(0.0074)
Last Sale Price (till 2005)	44,896	47,707	-2,811
	(71,496)	(84,007)	(1,958)
Days Since Last Sale (till 2005)	5,583	5,638	-55.54**
	(1,202)	(975.1)	(27.46)
Average story height	1.897	1.893	0.003
	(0.246)	(0.236)	(0.006)
Garage	13.08%	13.99%	
Full Basement	90.6%	93.17%	
Sample size	36,410	38,985	
Neighborhood Characteristics			
#Violent crime in 2005	0.694	0.650	0.044
	(1.332)	(1.338)	(0.033)
#Property crime in 2005	2.561	2.520	0.041
	(2.929)	(3.385)	(0.078)
% Black	27.15	27.16	-1.28e-5
	(31.64)	(31.94)	(0.008)
% Hispanic	0.93	0.98	-0.0005
	(1.03)	(1.45)	(0.0003)
% Male aged 15-24	6.48	6.53	-0.0005
	(2.51)	(2.88)	(0.0007)
Sample size	3,282	3,282	

Standard errors are reported in parentheses. Housing characteristics are in 2002. Information on demographics is taken from Census 2000. All neighborhood characteristics are measured at block level. Estimates in the third column are ω, the coefficient on D_i^{250} in $X_i = \alpha + \omega D_i^{250} + \varepsilon_i$.

As noted above, a foreclosed property can experience as many as four stages: pre-foreclosure, foreclosure, vacancy and reoccupation. To capture most of the variations in each stage, I define pre-foreclosure period as up to 6 quarters before the calendar quarter a property gets a foreclosure filing, foreclosure period as the foreclosure filing quarter plus up to 8 quarters thereafter but before the property is sold (prior to sheriff sale) or becomes vacant, vacancy period as the vacancy start quarter, the vacancy end quarter plus 8 quarters in which the property stays vacant, and reoccupation period as up to 4 quarters after the quarter the property gets re-occupied.

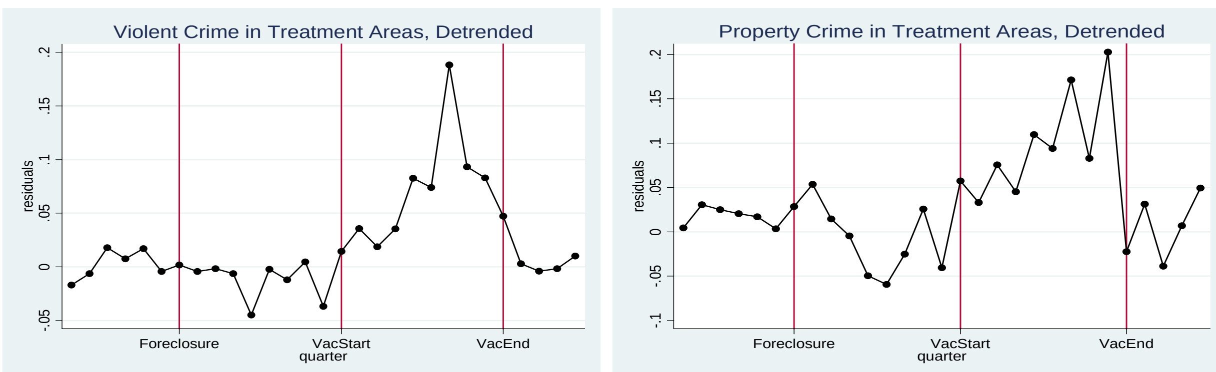

Figure 1.6 - Crime Trend by Quarter

I proceed by providing graphical evidence. If the proximity to foreclosed and vacant properties has a positive impact on crime, we should see numbers of crime in areas near foreclosure sites rise subsequent to foreclosure filing and the start of vacancy. Moreover, we should observe a larger impact of vacancy. Finally, the impact on treatment areas is expected to be stronger than on control areas as a result of geographic proximity, which indicates that the pattern should be seen absence of control.

Figure 1.6 shows violent and property crime trends for treatment areas after controlling for quarter fixed effects. In the first graph, the number of violent crimes in foreclosure quarters remains largely the same as in pre-foreclosure quarters. Violent crime rate increases during the quarters of vacancy and largely drops to pre-foreclosure level after the foreclosed properties are reoccupied. These trends suggest that vacancy increases violent crime but foreclosure alone does not have a strong effect. The fact that crime rate drops after re-occupancy strengthens this interpretation. In the second graph, we see that the pattern is similar for property crime rates, though with more noise.

While Figure 1.6 is certainly suggestive, to control for the possibility that *changes* in both crime and treatment status are being driven by some unobserved confounding factor, I proceed by incorporating into my analysis control areas that lie directly adjacent to the treatment areas.

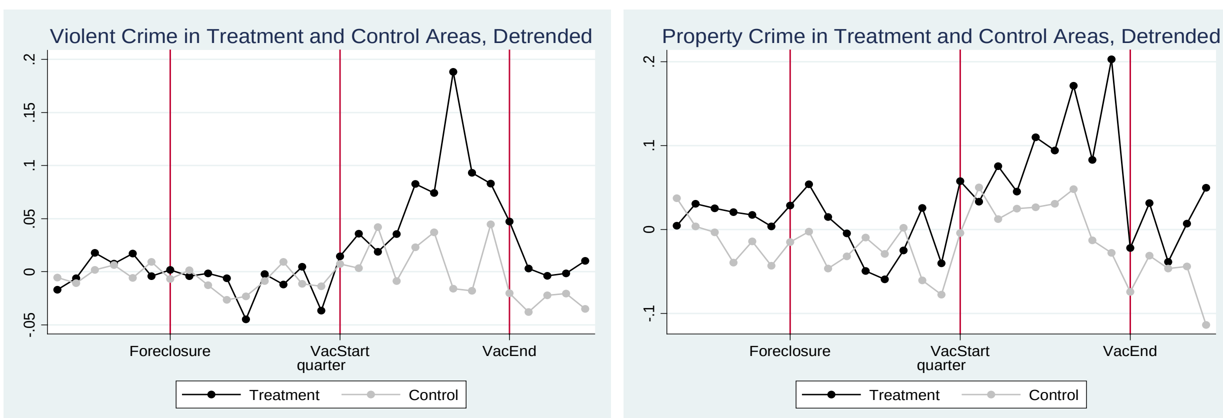

Figure 1.7 - Crime Trend by Quarter by Treatment and Control Groups

Figure 1.7 provides a visual treatment of the complete identification strategy, showing the violent and property crime trends in both treatment and control areas after controlling for quarterly fixed effects. Consistent with the notion of a continuous and decreasing spatial effect of vacancy on crime, there is visual evidence that the control areas experience increases in crime rates in vacancy quarters, but at a much smaller scale than the treatment areas. Again, there is no evidence that foreclosure per se leads to increased crime.

1.5 ESTIMATION FRAMEWORK

Before estimating empirical models inspired by the graphical approach discussed above, I first estimate the impact of foreclosure and vacancy on crime using a cross-sectional model in the spirit of the previous literature. A set of validity checks are performed and causal interpretation of results are evaluated. Given the similarity of the treatment and control areas defined in the previous section, a difference-in-differences model is then introduced and estimated.

The cross-sectional specification takes the following form:

$$\#(Crime_i) = \alpha + \beta_0 X_i + \beta_1 Fore_i + \beta_2 Vac_i + \varepsilon_i \tag{1.1}$$

The unit of observation is a Census 2000 block group. $\#(Crime_i)$ is the number of violent crimes in block group i. X_i is a vector of neighborhood characteristics in Census 2000 that might be expected to affect crime rates. $Fore_i$ and Vac_i are block-group level foreclosure rate and foreclosure-led vacancy rate. To be more consistent with Immergluck and Smith (2006), those rates are measured by the number of foreclosures/vacancies divided by the number of owner-occupied housing units in the same block group. As the neighborhood characteristics are measured around 2000, I denote $Fore_i$ as foreclosure rate and Vac_i as foreclosure-led vacancy rate in 2006.

Table 1.3 presents the results of the cross-sectional analysis associated with Equation 1.1. In columns (1) and (2), where all the variables of interest are measured in 2006, foreclosure rate is found to have a positive and statistically significant impact on violent crime rate. However, the perverse sign on vacancy rate, a supposedly more intensive treatment, raises concerns about confounding unobservables. The concern is further strengthened by the fact that the results on vacancy rate are sensitive to demographic controls.

Table 1.3 - Cross-Sectional Results

# Violent Crime in 2006	(1)	(2)	(3)	(4)
% properties with foreclosure filings in 2006	1.985***	1.083***		
	(0.331)	(0.292)		
% foreclosed and vacant properties in 2006	-2.795**	-0.355		
	(1.235)	(1.026)		
% properties with foreclosure filings in 2009			1.598***	0.772**
			(0.376)	(0.322)
% foreclosed and vacant properties in 2009			-2.622	-0.340
			(1.978)	(1.620)
Include demographic controls	No	Yes	No	Yes
Observations	343	343	343	343

Standard errors are reported in parentheses. Demographic characteristics includes population, median family income in 1999, percent unemployed, percent female headed households, percent on public assistance, percent renters, percent black and percent Hispanic.

To check the validity of the exogeneity assumptions implicit in the cross-sectional approach, I also estimate the model using foreclosure and vacancy rates in 2009 to predict violent crimes in 2006. The idea is that if increase crime is the result rather than the cause of foreclosure/vacancy, crimes committed beforehand should be uncorrelated with foreclosure/vacancy.

The results in column (3) and (4) show that foreclosure rate in 2009 also has a positive and statistically significant impact on violent crime rate in 2006, though the magnitude is smaller. In addition, the results on 2009 vacancy rate do not differ from those in 2006. Given that only 1.7% of 2006 foreclosures are still unresolved in 2009, it is unlikely that results in column (3) and (4) are driven by persistency of individual foreclosures. Therefore, the estimation result of the effect of 2009 foreclosure and vacancy on 2006 violent crime clearly indicates that this simple cross-sectional analysis of aggregate crime and foreclosure levels suffer from omitted variable bias and is inappropriate for identifying the causal link between foreclosure (vacancy) and crime. This finding motivates the use of my difference-in-differences specification which uses spatially proximate controls and compares crime levels both pre- and post-treatment to greatly reduce the potential for confounding unobservables.

Equation 1.2 presents the empirical specification. The study area is reduced to areas within 353 feet of the foreclosed and/or vacant properties. This specification incorporates 3 indicators for the three different stages of foreclosure and interactions of the treatment indicator (within 250 feet of the foreclosure sites) with each of these 3 indicators. Thus, the counterfactual change in crime rates in areas immediately surrounding the foreclosure sites is estimated using crimes in areas just slightly farther away in the same periods. In some specifications, I also include neighborhood-quarter fixed effects (controlling for presence in one of 86 neighborhoods) as well as housing characteristics for the treatment and control areas:

$$\#(Crime_{it}) = \alpha + \beta X_i + \omega_0 D_i^{250} + \omega_1 Fore_{it} + \omega_2 Vac_{it} + \omega_3 ReOcc_{it} + \pi_1 \cdot (D_i^{250} * Fore_{it}) + \pi_2 \cdot (D_i^{250} * Vac_{it}) + \pi_3 \cdot (D_i^{250} * ReOcc_{it}) + \varepsilon_{it} \qquad (1.2)$$

$\#(Crime_{it})$ is the number of violent or property crimes in control or treatment areas surrounding foreclosed property i in quarter t. $Fore_{it}$ equals 1 if in quarter t property i received a foreclosure filing in a previous quarter but is still occupied; Vac_{it} equals 1 if in quarter t property i is vacant; $ReOcc_{it}$ equals 1 if property i is reoccupied in quarter t. Parameters of interest are π_1, π_2 and

π_3.The estimated impact of foreclosure is given by the term π_1, while the impact of foreclosure-led vacancy and the impact of reoccupation are given by π_2 and π_3.

As discussed in Section 1.4, there are up to four stages in a foreclosure process. I use the same definitions of pre-foreclosure, foreclosure, vacancy and reoccupation in estimating Equation 1.2. Below I also explore the impact of different lengths of vacancy on crime rates as a robustness check.

Table 1.4 presents the coefficients of interest from the estimation for Equation 1.2. The dependent variable in columns (1) and (2) is violent crime counts. Housing controls[10] include all average housing characteristics reported in Table 1.2. In the first row, the coefficient on being treated and foreclosed is small and insignificant, indicating that foreclosure alone does not impact crime rates in treatment relative to control areas. However, we find that during the time a foreclosed property stays vacant, the neighboring areas have more violent crimes than areas slightly further away from the vacant property at the same time. The coefficient is 5 percent significant after controlling for average housing characteristics at treatment/control ring level. Given that the average number of violent crimes within 250 feet of a foreclosed property each quarter is 0.2, a coefficient of 0.032 in column (2) translates into an increase of more than 15%. Comparing column (1) and (2), we can see that adding housing controls do not change any results, which confirms the similarity of treatment and control areas.

Table 1.4 also presents other coefficients from Equation 1.2: the effect of being treated and being in foreclosure, vacancy or reoccupation period. The coefficient on being treated is essentially 0 for violent crime, which confirms the similarity between treatment and control. The

[10] Including both housing and Census 2000 demographic controls in equation (2) do not change the results. Nevertheless, I do not report the results with Census controls in Table 1.4 because those demographics are taken at block level and may not precisely measure the actual counts in my localized study areas.

effect of being in foreclosure period is insignificant after controlling for housing characteristics, and the effect of being in vacancy period is close to 0.

Table 1.4 - Difference-in-Differences Results

	# Violent Crime		#Property Crime	
	(1)	(2)	(3)	(4)
Within 250 feet * foreclosure	-0.00224	-0.00269	0.00203	0.00180
	(0.00789)	(0.00791)	(0.0142)	(0.0143)
Within 250 feet * vacancy	0.0338***	0.0324**	0.0245	0.0214
	(0.0130)	(0.0129)	(0.0222)	(0.0223)
Within 250 feet * reoccupied	0.0269*	0.0255*	0.0383	0.0365
	(0.0143)	(0.0143)	(0.0264)	(0.0264)
Within 250 feet	0.00373	0.00493	0.0285**	0.0221*
	(0.00623)	(0.00633)	(0.0123)	(0.0125)
Foreclosure	-0.0114*	-0.00915	-0.0123	-0.00659
	(0.00685)	(0.00680)	(0.0130)	(0.0125)
Vacancy	0.00182	0.000732	-0.0103	-0.0152
	(0.0109)	(0.0107)	(0.0201)	(0.0190)
Reoccupied	-0.0239*	-0.0240*	-0.0261	-0.0265
	(0.0130)	(0.0126)	(0.0232)	(0.0219)
Housing characteristics		Yes		Yes
Year Neighborhood interaction	Yes	Yes	Yes	Yes
SD clustered by	Property Ring	Property Ring	Property Ring	Property Ring
Sample size	85,024	84,907	85,024	84,907

Standard errors are reported in parentheses.

Note that the coefficient of being in reoccupied quarters is negative and statistically significant at 10 percent level, and the coefficient of being treated and in reoccupied quarters is positive and significant at 10 percent level. These results indicate that while both treatment and control areas experience decline in violent crime rates after the foreclosed house is reoccupied, crime rate in treated areas in period of reoccupation is higher than in control areas. One possible explanation is that, similar to the effect of vacancy, the effect of removing vacancy on crime is likely to be a continuous decreasing function of the distance to the foreclosure site, but the geographic scale of that impact may be different. In other words, if vacancy only affects areas up to 250 feet away, reoccupation may impact areas farther away. Therefore, it is possible that both the treatment and control are in effect treated in terms of reoccupation, with the areas within 250 feet of foreclosure more intensely treated. Note that these results are consistent with the graphical evidence. However, given the results on reoccupation are only significant at 10 percent level, the null hypothesis that reoccupation completely offsets the increase from vacancy cannot be rejected.

Column (3) and (4) presents estimates on property crimes. Vacancy still has some impact but the coefficients are no long significant. Possibly this is due to the fact that those coded as property crimes are only the non-violent ones.

Table 1.5 - Difference-in-Differences Results by Length of Vacancy

# Violent Crime	(1) Foreclosed Properties Ever Vacant
Within 250 feet * foreclosure	-0.00679
	(0.0119)
Within 250 feet * vacant 0-3 months	0.0173
	(0.0160)
Within 250 feet * vacant 3-6 months	0.0198
	(0.0162)
Within 250 feet * vacant 6-12 months	0.0394**
	(0.0192)
Within 250 feet * vacant12-18 months	0.0762***
	(0.0290)
Within 250 feet * vacant > 18 months	0.0499***
	(0.0182)
Observations	42,668
#Foreclosed homes	1,403

Standard errors clustered by property rings are reported in parentheses. Housing controls and year*neighborhood interactions are included.

As a final empirical exercise, I explore the impact of different lengths of vacancy on violent crime rates. The motivation here is twofold: first, to rule out the possibility that my main results are driven by compositional effects; second, to better understand the role of vacancy length in violent crime rate changes. Because only half of the foreclosed homes in my sample fell into vacancy and these homes stayed vacant for different amounts of time, compositional effects could potentially create spurious correlation between vacancy and crime. For example, if the pre-treatment levels of crime are higher for houses that experienced longer time of vacancy, the coefficient on vacancy will be positive even if crime rates remain unchanged during vacancy period. In other words, the estimates may simply be picking up pre-existing differences in crime rates of areas experiencing different levels of treatment.

Table 1.6 - Difference-in-Differences Results 2 by Length of Vacancy

# Violent Crime	(1) Vacant for 0-3 months	(2) Vacant for 3-6 months	(3) Vacant for 6-12 months	(4) Vacant for 12-18 months	(5) Vacant for >18 months
Within 250 feet * foreclosure	0.0357	-0.0423*	0.00487	-0.00124	-0.0220
	(0.0291)	(0.0237)	(0.0257)	(0.0314)	(0.0339)
Within 250 feet * vacant first 3 months	0.0352	0.0171	0.0138	0.0316	0.0133
	(0.0404)	(0.0325)	(0.0313)	(0.0503)	(0.0413)
Within 250 feet * vacant for 3-6 months		0.0327	0.0598	-0.0108	-0.0153
		(0.0307)	(0.0367)	(0.0351)	(0.0365)
Within 250 feet * vacant for 6-12 months			0.0342	0.0307	0.0068
			(0.0268)	(0.0374)	(0.0299)
Within 250 feet * vacant for12-18 months				0.0618	0.0801**
				(0.0495)	(0.0367)
Within 250 feet * vacant > 18 months					0.0660**
					(0.0358)
Observations	7,560	8,754	11,106	6,706	8,542
#Foreclosed homes	278	310	364	209	242

Standard errors clustered by property rings are reported in parentheses. Housing controls and year*neighborhood interactions are included.

A direct remedy is to separately estimate the effect of different stages of vacancy on subsamples of foreclosed homes that stayed vacant for various amount of time. If the impacts of different lengths of vacancy remain positive and consistent across all subsamples, I can safely rule out the existence of compositional effects.

Below I present the estimation results of Equation 1.2 on a group of subsamples, adding interaction terms of being in treatment areas and in different stages of vacancy. I first restrict the sample to all homes that experience vacancy after foreclosure and thus reducing the sample size by half. The results are shown in Table 1.5. The coefficients are all of the expected sign and are larger in magnitude as the length of vacancy increases, indicating that longer-term vacancy has a stronger effect on violent crime rates. Moreover, the estimates on more-than-6-months vacancies are statistically significant. Note that the coefficient on foreclosure is close to 0 and insignificant, ruling out the possibility that foreclosure alone impacts crime rate in this subsample of vacant properties and that the results on vacancy simply reflect the differences in pre-vacancy crime levels in treatment and control areas.

Table 1.6 reports estimates of Equation 1.2 with subsamples of foreclosed homes experiencing different lengths of vacancy. While almost all coefficients on vacancy are positive, most of them lose significance, possibly due to reduction in sample sizes. For areas surrounding houses vacant for more than 18 months, the coefficients on having been vacant for longer terms are larger in magnitude and statistically significant, confirming the results in Table 1.5. Finally, almost all coefficients on foreclosure are small and insignificant. Overall, these results are consistent with estimates in Table 1.5 and confirm the absence of compositional effect in the main estimation equation presented earlier.

Given there is little evidence of compositional effect, I again estimate Equation 1.2 using the full sample to better understand the role of vacancy length in violent crime rate changes. Results are reported in Table 1.7. The estimates are similar to those in Table 1.5, but obtain more statistical significance with a larger sample. Together with Table 1.5, the results indicate that presence of houses vacant for longer than 6 months clearly increases violent crime rates, and the impact increases with duration of vacancy.

Table 1.7 - Difference-in-Differences Results 3 by Length of Vacancy

# Violent Crime	(1) All Foreclosed Properties
Within 250 feet * foreclosure	-0.0006
	(0.0071)
Within 250 feet * vacant 0-3 months	0.0208
	(0.0150)
Within 250 feet * vacant 3-6 months	0.0263*
	(0.0157)
Within 250 feet * vacant 6-12 months	0.0441**
	(0.0186)
Within 250 feet * vacant12-18 months	0.0792***
	(0.0283)
Within 250 feet * vacant > 18 months	0.0579***
	(0.0175)
Observations	84,907
#Foreclosed homes	3,282

Standard errors clustered by property rings are reported in parentheses. Housing controls and year*neighborhood interactions are included.

1.6 CONCLUSION

In this paper I use a difference-in-differences research design with carefully constrained control locations to measure the impact of residential foreclosures and vacancies on violent and property crime. Using detailed data on addresses and dates of foreclosures and crime, I estimate that, on average, violent crimes within 250 feet of foreclosed homes increases by more than 15% once the foreclosed home becomes vacant, compared to crimes in areas between 250 and 353 feet away. Foreclosure alone has no effect on crime while effects on property crime are similar but are less precisely estimated. Because I exploit the exact timing and location of foreclosure, vacancy and crime by comparing crime rates in geographically small and homogenous areas at different stages of foreclosure, these results provide a significant improvement upon the existing literature that attempts to identify the impact of foreclosure and vacancy on crime with cross-sectional design or analysis at aggregate levels that are subject to considerable concerns regarding omitted variable bias.

In addition, this paper provides the first evidence on the impact of vacancy length on crime and concludes that longer terms of vacancy have a stronger effect on violent crime compared to shorter-terms of vacancy. While the majority of current federal and state level foreclosure programs are focusing on loan modification, my results strongly indicate that policies aiming at post-foreclosure vacancy reduction will most effectively alleviate the external cost of foreclosure.

2.0 THE EXTERNAL EFFECT OF FORECLOSURE ON HOUSING MARKET TRANSACTIONS

2.1 INTRODUCTION

The United States has experienced a tremendous wave of mortgage foreclosures in recent years. Foreclosure not only affects individuals directly involved, but can also have implications for surrounding neighborhoods and larger communities, such as lowering nearby property values, reducing the local property tax base, increasing blight and crime, disrupting social ties, etc. This paper focuses on foreclosure's external impact on neighborhood property values. It is the first study to document the quantity shifts of homes sold at the time of nearby foreclosure[11] in different sections of the housing market as a result of changes in both the demand side and the supply side.

A foreclosure may affect both the demand for and the supply of nearby properties. There are several possible mechanisms from the demand side. First, owners with delinquent mortgages usually have limited financial means to maintain and/or upgrade their houses, even before they enter the foreclosure process. This in turn leads to physical blight. Sometimes the delinquent

[11] Because there are multiple stages in the foreclosure process, there is no consensus on the exact definition of foreclosure in the literature. For the rest of this paper, I use the term "foreclosure" to refer to an earlier stage in the foreclosure process, when a lis pendens has been filed, and the term "vacancy" refers to a later stage of foreclosure when the property is unoccupied (it is either taken back by the lender (REO) or is abandoned). In this paper, "foreclosure" sometimes also refers to the whole process.

owners are evicted, and their properties will be vacant for some time, which attracts vandalism and crime (for example, Immergluck and Smith, 2006; Cui, 2011). It further exacerbates the blight and makes the neighborhood less desirable for potential homebuyers. Second, it is documented that foreclosed properties are generally sold at deep discounts (Campbell, Giglio, and Pathak, 2009; Clauretie and Daneshvary, 2009; Sumell, 2009; Pennington-Cross, 2006). A homebuyer may use the sale prices of nearby comparable properties as a valuation benchmark and lowers her reservation price for a non-foreclosed property. Lastly, a high concentration of foreclosures could increase the local supply of housing and decrease the demand for non-foreclosed properties.

From the supply side, a homeowner may choose not to sell her house at the time of nearby foreclosure because she will suffer a price discount. Such price effects are documented in various studies on the relationship between foreclosure and neighborhood housing prices[12]. Also, the homeowner may be more eager to sell the house in order to live away from foreclosures. Finally, homeowners may be forced to lower their reservation prices due to changed expectations. Overall, a foreclosure can in theory affect both the price and the probability of sale of a nearby property, but the prediction on the supply side is more ambiguous.

Taking a difference in differences approach, I compare areas immediately surrounding foreclosure sites with areas slightly further away with similar composition of houses and sold properties before foreclosure. Estimation results show that the degree of equilibrium quantity and price shifts due to foreclosure depends on the type of nearby properties. In particular, both quantity and price drop for houses with a higher quantity[13] index (Type H) at the time of nearby

---

[12] These studies are discussed in the literature review section

[13] Here "quantity" may refer to both the quantity and the quality of housing services a house provides. Similar use of this word can be found in housing economics literature

foreclosure, while for houses with a lower quantity index (Type L) only see a price drop. Quantile regression results confirm the differential effects of foreclosure on both types of houses. In addition, vacancy has an overall larger effect on sale volumes than foreclosure does, and such effects mostly disappear after the vacant house is re-occupied by a new permanent owner.

The main contribution of this paper is to document the external effect of foreclosure on different sections of the housing market from both demand and supply perspectives. The results show the changes in both equilibrium quantity and equilibrium price. In contrast, previous studies only focus on the price effect, without considering the potential quantity shifts caused by supply and demand changes.

But why there is little impact on Type L homes? One explanation is that owners of Type H are usually households of higher income and higher wealth level, and thus facing lower liquidity and credit constraint. They are more unlikely to face forced sale[14]. If we assume higher income households live in nicer homes, it is very likely that they can afford to delay the home sales while it is much more difficult for lower income families to do so.

Figure 2.1 presents simplified graphical explanations on the supply and demand shifts for Type L and Type H homes, respectively[15]. While a nearby foreclosure tends to decrease demand for a non-foreclosed property, homeowners are unwilling to adjust their reservation prices and sell the properties at discounts. As a result, the housing supply curve may not shift after

[4] In fact, US Census Bureau shows that in 2004, households within 80-100 income percentiles have a net wealth of $294,087, which is 2.51 times the net wealth of households within 60-80 income percentiles, 4.63 times of households within 40-60 income percentiles, and 8.1 times of those earning 20-40 percent in the income distribution.

[15] While Graph (a) shows an upward-sloping housing supply curve for Type L houses, the other possible scenario is that supply of Type L houses is nearly perfectly inelastic. And as a result supply shift in Type L houses is not necessary to reach a new equilibrium similar to what depicted in (a). Suppose the supply curve is perfectly inelastic, and demand curve shifts inward due to foreclosure, the new equilibrium will still be on the same supply curve. Because housing supply fixed, there will be no quantity shift and a price drop, which resembles the changes in equilibrium price and quantity in (a).

foreclosure happens. However, owners of Type L homes tend to face more liquidity and credit constraints and cannot delay the sales. Those owners are forced to lower their reservation prices and the housing supply curve shifts rightward, while owners of Type H homes can wait till the nearby foreclosure is resolved. Therefore, for Type L houses, there are both demand and supply shifts, as shown in Part (a) of Figure I; while for Type H houses, only demand curve shifts, as shown in Part (b). In equilibrium, both types of houses suffer a price reduction, and quantity decreases for Type H homes. But the overall effect on quantity may be ambiguous for Type L homes.

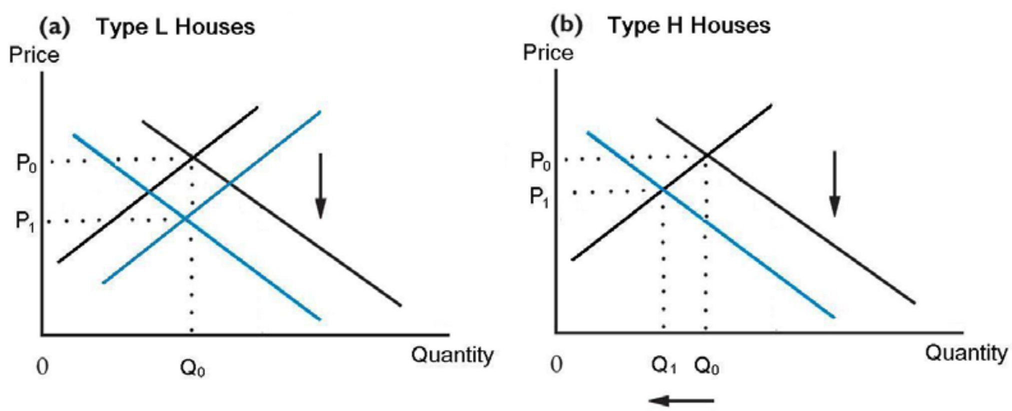

Figure 2.1 - Impact of Foreclosure on Sale Prices and Sale Quantities of Nearby Homes[16]

I proceed as follows, Section 2.2 provides a literature review, Section 2.3 of the paper presents background information on the current foreclosure crisis and the foreclosure process in the study area Pittsburgh, Pennsylvania. In Section 2.4, I describe the data used in this study. In

---

[16] The black lines represent demand and supply curves before the neighborhood foreclosure and the blue lines represent the curves after foreclosure.

Section 2.5, I describe the empirical methodology and present the estimation equations. Results are presented in Section 2.6. I discuss the possible mechanisms of housing supply shifts in Section 2.7 and conclude in Section 2.8.

2.2 LITERATURE REVIEW

The relatively rich literature on foreclosure and nearby property values all confirmed foreclosure can depress sale prices of neighboring houses, though the magnitudes of their estimates vary. Many studies found only modest impact of foreclosure on nearby housing prices. Among those, Immergluck and Smith (2006) relate sales price data for single-family properties in Chicago sold in 1999 to foreclosures in the two prior years and find that a foreclosure within 600 feet of a single-family home could lower its sale price by 0.9 -1.1%. Using data from St. Louis County, Missouri, between 1998 and 2007, Rogers and Winter (2009) find very similar results. Leonard and Murdoch (2009) study sales of single-family homes in Dallas County, Texas, during 2006, and find that properties in some stage of foreclosure depress sales prices and that each foreclosure within 250 feet appears to have an effect of about –0.5 percent on sales prices. Campbell, Giglio, and Pathak (2009) study twenty years of single-family property sales from Massachusetts and estimate that foreclosures within 0.05 mile lower sales prices by about 1%. As an alternative to a hedonic regression, Harding, Rosenblatt, and Yao (2009) use a repeat sales approach in seven metropolitan areas and find that property sales located within 300 feet of a foreclosed property experience about a 1 percent discount per foreclosure. The authors also find that the peak discount occurs at the time of sheriff sale.

Other studies find the external costs of foreclosure to be much higher. Shlay and Whitman (2006) estimate that the presence of a foreclosure-led vacant property in Philadelphia depresses the prices of properties located within 150 feet by $7,627. Lin, Rosenblatt, and Yao (2009) explore property sales data for the Chicago MSA for 2003 and 2006 and delineate foreclosure spillover effects along both physical and temporal dimensions. The authors find spillover effects is 8.7% for properties located within 300 feet of a foreclosure in 2006. Schuetz, Been, and Ellen (2008) study residential property sales and foreclosure filings in New York City between 2000 and 2005 and find that foreclosure filings can depress sale prices of houses within 250 feet by as much as 3.7%.

The different results in the literature may reflect the fact that some of the studies use data from pre-crisis years and set larger treatment areas (for example, treated area is within 600 feet from foreclosure instead of no more than 300 feet from foreclosure). Also, the magnitude of impact seems to depend on the choice of city. At the same time, those studies do reach a few common conclusions. First, all the empirical evidence confirms the negative impact of foreclosure on nearby property sale prices. Second, such discounts likely dissipate quickly the farther away the foreclosed property is from the sale in space. Third, among those papers taking account of different stages of foreclosure, the price effect peaks around the time the foreclosed property becomes vacant.

While the existing studies estimate the impact of foreclosure on prices for surrounding houses that sell, they do not consider potential changes in quantities of sold houses caused by both supply and demand shifts. Furthermore, their estimates on prices may be biased. In the following paragraphs, I explain why the estimation on quantities is potentially more credible.

In the ideal case, researchers observe fair market value of every house at all times, and they can simply compare the property values before and after a nearby foreclosure, even for a house that has never been sold. In reality, we only observe fair market value of a house when it is sold, and homeowners can freely choose when to put their properties on the market, with only a subsample of those houses sold successfully. Because houses sold at the time of nearby foreclosure are selected samples of all houses in that area, it may bias the estimation on price effect. For example, homeowners may choose to sell only when their losses are limited, and those suffering larger dips in home values may not wish to sell at all. In this case estimates on the price effect will only be a lower bound of the actual impact. On the other hand, it is possible that families who sell their homes after their neighbor's foreclosure are more eager to leave the affected area than those who choose to stay. If these families set lower reservation prices, it will bias the price estimates upward. Therefore, the direction of bias on the price estimates is uncertain.

In contrast to limited information on fair market values, we can observe the timing and locations of sales for every house at any moment. As a result, the estimation of quantity shifts should not suffer any selection bias discussed above. Therefore, while it is difficult to credibly estimate the price effect, the direction and magnitude of quantity shift may provide valuable information in understanding the external costs of foreclosure on nearby properties.

2.3 BACKGROUND

Since 2006, the United States has experienced a significant increase in home mortgage foreclosures. As of 2010, foreclosure filing rate nationwide reached 2.23 percent, the highest level ever recorded, up from 0.58 percent in 2006[17]. Though the foreclosure crisis was initially attributed to predatory lending practices and subprime mortgages, the Mortgage Bankers Association reports that the fastest-rising segment of foreclosures in recent months has been traditional prime mortgages, reflecting problems within macroeconomic factors, such as the unemployment rate. Given the current economic situation, the crisis will likely continue to present a significant policy challenge for local governments.

The proposed research considers the impacts of foreclosure and vacancy in the context of Pittsburgh, Pennsylvania. The housing market in Pittsburgh exhibit several distinctive features. First, the housing stock is among the oldest in the country with more than 60 percent of homes built before 1940, and over 25% are owned by homeowners aged 65 and older, according to the 2000 Census. As a result, many houses are poorly maintained, and once they become vacant, the condition of the houses can deteriorate quickly, further depresses the home values in the neighborhood. Second, housing transactions are relatively infrequent. According to the Pittsburgh assessment record, the average length of residence is 16.7 years. Third, housing price is very stable during the current crisis, and Pittsburgh has a much lower foreclosure rate than the national average.

Two types of foreclosure are widely used in the United States: judicial and power of sale. The former involves the supervision of a court and generally takes longer time to process. In Pennsylvania, all foreclosures are judicial and a lender must follow a state-level judicial procedure in order to foreclose on a property. A lender can file a suit when the borrower is at least 90 days behind the mortgage payment. At this time, the two parties may reach an agreement on loan modification or the borrower may successfully sell the house and pay off the mortgage.

[17] Data source: RealtyTrac.com

If none of the options works, the lender has the right to sell the property at an open auction and it goes to the highest bidder. Typically under this process the property is not sold and will return to the lender (Had there been any potential buyer in the property at a price equal to the outstanding loan amount, the owner could have sold it earlier). The borrower is evicted after the sheriff sale. The foreclosed property is then classified as a real estate owned (REO) property and stays vacant until it is sold to a new permanent owner.

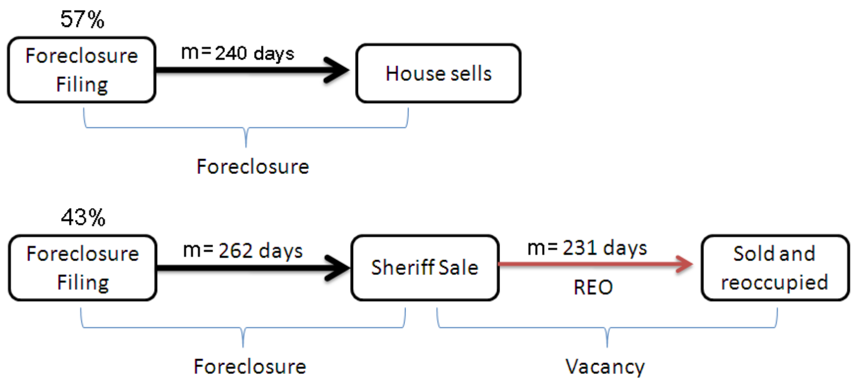

Figure 2.2 - Foreclosure Process

As described above, a typical foreclosure case consists of multiple stages: foreclosure filing, sheriff sale, and sale to a new permanent owner (REO sale). Note that some foreclosure filings are finally resolved and do not reach the point of sheriff sale. Figure 2.2 provides an illustration of the two possible outcomes following foreclosure filing. In this study, 57% of the foreclosure filings result in property sale to another permanent owner before sheriff sale (outcome 1) while 43% experience a period of vacancy until they are finally resold (outcome 2).

To differentiate foreclosure per se from foreclosure-led vacancy, I define up to four stages of the foreclosure process: a pre-foreclosure stage that takes places before the date of foreclosure filing, a foreclosure stage between the date of filing and the date of sheriff sale (or the date of property sale for outcome 1), a vacancy stage, and a reoccupation stage that takes place after the REO sale date. Due to the judicial nature of foreclosure in Pittsburgh, the whole process typically takes one to two years to complete. As shown in Figure 2.2, the median length of foreclosure stage is 240 days for those without sheriff auction, and 262 days for those experiencing vacancy. The median length of vacancy is 231 days.

2.4 DATA DESCRIPTION

The analysis is based upon three sets of data regarding the locations of foreclosure filings, the locations and characteristics of properties and neighborhoods in the city of Pittsburgh, and property sales.

The foreclosure filing data are obtained from the court records. This dataset contains information on every foreclosure filing in the city from 2006 to November 2009, such as the date of filing, the parcel ID of the property receiving foreclosure filing, borrower and lender names, and the current stage of filing. Settled and discontinued cases are deleted from the sample. Note that only residential[18] properties are included in this analysis. From January 2006 to November 2009, there were 3,282 foreclosure filings on residential properties in the city of Pittsburgh. In those cases, an average borrower owed $81,271.3.

[18] A property is defined as residential if its structure type falls into one of the following categories: single family, two to four family, row house and townhouse.

Table 2.1 - Characteristics of Properties in City of Pittsburgh

	All properties	Sold
Housing Characteristics		
Assessed value	71,787	75,028
	(77,502)	(82,405)
Square footage	1,691	1,691
	(785.4)	(755.7)
Lot area (in square feet)	4,416	4,096
	(7,258)	(6,154)
Building age	86.54	87.09
	(25.91)	(26.56)
Bedroom	3.049	3.037
	(1.063)	(1.048)
Bathroom	1.460	1.470
	(0.687)	(0.696)
Average story height	1.942	1.968
	(0.443)	(0.424)
Days since last sale	6,112	
	(6,295)	
Sales price		98,742
		(141,750)
Garage	23.7%	22.2%
Full basement	93.0%	93.1%
Sample size	94,430	25,577
Neighborhood Characteristics		
#Violent crime/block in 2005	0.443	
	(1.185)	
#Property crime/block in 2005	1.452	
	(2.548)	
% Owner occupied	67.09	
	(22.95)	
% Black	25.51	
	(32.12)	
% Hispanic	0.920	
	(2.467)	
% Female householder	42.43	
	(29.10)	
Sample size	7,122	

The second source of data comes from the Allegheny County's Office of Property Assessments. Information on housing characteristics is taken from the last county-wide reassessment in 2002. These assessment data provides comprehensive physical characteristics for each residential property, such as square feet, number of bedroom and year structure was built. Data on neighborhood characteristics come from two sources. Selected block-level demographic characteristics (such as race and age) are taken from the 2000 Census of Population and Housing. Information only available at more aggregate levels (such as education and income) is not included due to the lack of geographic precision. Pre-existing (2005) crime counts come from the Police Department of the City of Pittsburgh. Properties with foreclosure filings are matched with assessment records by their parcel IDs.

I merge the matched foreclosure-assessment data with property sales from January 2005 to October 2010, provided by the Allegheny County Recorder of Deeds Office. The data contains sale date, price, parcel ID and buyer and seller names for every property transaction during that time period. Any transaction between family members and those with sale prices below the 1st or above the 99th percentile of the distribution of raw prices are dropped from the sample. The resulting dataset has a total of 25,577 sales in all 89 neighborhoods in Pittsburgh.

Table 2.1 provides summary statistics of parcels in Pittsburgh, including housing characteristics of various parcels and block-level neighborhood information for Pittsburgh area. The first column provides information on all properties in the city and the second column shows properties that were sold between January 2005 and October 2010. Consistent with the description of Pittsburgh housing stock in Section 2.2, houses are generally old and the time between two transactions for a house is quite long. The average building age is 86.54 years and days since last sale is 6,112, both measures well above the U.S. averages of about 30 years and less than 4000 days[19]. The average sale price in 2005 is $98,742, which is significantly higher than the 2002 assessed value of 75,028 for those properties sold in 2005. If we believe the 2002 assessed value reflect fair market price in that year, it seems that Pittsburgh's housing market experienced less of a boom than national averages during the pre-crisis years. In general, houses sold in 2005 are similar to other houses in terms of characteristics list in Table 2.1. The average property in Pittsburgh is two-story house, has an assessed value of $71,787 on a 0.1 acre lot, and has 1,691 square feet, 3 bedrooms, one and half baths, full basement but no garage.

The next step is to determine the vacancy period for each property with a foreclosure filing. I briefly describe the procedure, which is identical to the method used in Cui (2011). As a first step, foreclosure filings are linked to all subsequent property transactions by parcel ID. Because most foreclosure-led vacancies coincide with being held as REOs, I use REO status as a vacancy indicator. It is identified by two dates: the sheriff sale date and a subsequent date when the property is sold by the lender to a new permanent owner (REO sale). For any foreclosure experiencing these two sales, I define the vacancy period as days between the two sales. Occasionally, I change the start of vacancy to an earlier date if gas shut-off record shows the property has no gas usage some time before the sheriff sale date. Among the 3,282 properties foreclosed between 2006 and 2009 in Pittsburgh, 1,403 experienced vacancy. Among those, 86% of the vacancies are solely determined by REO status, while only 14% are jointly determined by REO and gas shut-off date. Though the procedure may not generate completely accurate dates of vacancy, and may cause measurement error problem, it will understate the difference in outcome variables between vacancy and non-vacancy periods, and will lead to a downward (attenuation) bias, making my estimates lower bounds of the true effects.

---

[19] Data sources: 2000 Census

2.5 EMPIRICAL METHODOLOGY AND ESTIMATION FRAMEWORK

As discussed in the first section, foreclosure can affect the demand for nearby houses in many ways. These impacts can be roughly categorized into two groups: the effect of neighborhood disamenity and the effect of added housing supply. The first is a typical scenario in hedonic analysis of housing markets.

There are difficulties in identifying the hedonic price function for local (dis)amenities). The most well-known one is that variation in the amenity may be correlated with unobservable factors. The second one is that prices are only observed when transactions take place.

In the setting of this paper, the locations of foreclosures/abandoned buildings are likely to be correlated with unobserved neighborhood factors that affect local housing prices. For example, there is evidence that foreclosures tend to cluster in lower-income neighborhoods, and those neighborhoods can be less attractive to home buyers for reasons researchers cannot observe. Therefore, cross-sectional analysis that simply compares housing prices in areas with and without foreclosure will yield biased estimates.

The second point is often overlooked by empirical work on local amenities and property values. A typical practice is to assume that housing supply is fixed in the short run and to estimate only the short-run impact on housing prices. However, without additional evidence from the supply side, the assumption does not guarantee that the estimates in those studies are unbiased or consistent. In fact, researchers and housing specialists have long observed a positive relationship between second-hand housing price and sale volume (Stein, 1995, Genesove and Mayer, 2001). Even if the assumption of fixed housing supply is true, it is possible that people have heterogeneous preferences over foreclosure and the observed prices do not reflect the average willingness to buy or sell in the neighborhood.

The challenges faced in the hedonic analysis of foreclosure and housing market, coupled with the fact that foreclosures add to local housing supply, indicate that the information needed to credibly estimate the price effect may be unattainable. Nevertheless, this paper will still attempt to document the equilibrium price changes, while focusing on the quantity shifts at different sections of the housing market.

I adopt a difference-in-differences framework to control for the possible correlation between foreclosure locations and time-invariant unobserved neighborhood characteristics. With information on the specific locations of all houses and the dates of foreclosure, vacancy and housing transactions, I am able to compare houses within small proximate areas in which neighborhood characteristics are generally homogenous, and to confirm the absence of substantive preexisting differences in housing compositions of treatment and control areas.

As indicated in the literature, the spillover effect of foreclosures is a decreasing function of distance to the foreclosure site, and such impacts are often extremely local. As a result, the treatment group in this study is set as areas within 250 feet from foreclosures and the control areas are between 250 and 750 feet away. The choices of treatment and control areas are consistent with many studies on foreclosure and housing prices. In addition, the treatment and control units are relatively small areas, alleviating concerns that common neighborhood factors may impact both foreclosure and housing transactions. Finally, keeping the treatment area small will help to reduce the possibility of clustering, which can create double-treated areas and bias the estimation results upward. Nevertheless, it is possible that control areas may coincide with other treatment areas due to clustering or the way treatment and control are defined. Also, the treatment may not be fully treated[20]. But even if the control and treatment areas are not correctly defined, any differences in housing composition between the two areas will be underestimated, and it will lead to attenuation bias in the estimates.

Because my estimation strategy hinges on the relative similarity of treatment and control areas, I compare a list of housing and neighborhood characteristics in the two groups for all properties and those sold in 2005, before the time of foreclosures in my sample. All the results are shown in Table 2.2 and Table 2.3. For housing characteristics, I calculate the average housing characteristics for each treatment ring. I repeat the process for houses located in control areas. For neighborhood characteristics, average demographic measures are calculated for each treatment/control area. Differences in those characteristics across the two groups are also estimated. I run a series of regressions with each housing or neighborhood characteristics as the dependent variable and a treatment indicator as the only independent variable. If the coefficient on the treatment indicator is insignificant across all housing and neighborhood characteristics, it will be suggestive that treatment and control groups are readily comparable (at least in terms of observables). In Table 2.2 and 2.3, I also compare the list housing characteristics in and outside treated areas, using the procedure describe above. For the rest of the city, housing characteristics are obtained for each property and averages are calculated.

The first two columns in Table 2.2 compare the characteristics of all residential properties in and outside the treatment areas, while the first and fourth columns compare the properties in treatment and control areas. Differences between column (1) and (2) are reported in column (3), while differences between column (1) and (4) are reported in column (5). We can see that houses

[20] In any given year in the sample, there are less than 3% of foreclosures that are within 500 feet from another one. Excluding properties close to those foreclosures do not alter the results.

Table 2.2 - Characteristics of All Properties

	Within 250 feet	Rest of City	Within 250-750 feet	Differences 1	Differences 2
Housing Characteristics					
Assessment Value	49,839	96,621	50,185	-40,525***	-356.64
	(90,055)	(107,115)	(65,754)	(906.94)	(826.70)
Square footage	1,582	1,801	1,588	-205.8***	-5.046
	(385.2)	(933.1)	(624.3)	(5.614)	(3.845)
Lot area	3,712	5,394	3,763	-1,596***	-50.76
	(1,977)	(10,292)	(4,257)	(56.65)	(27.09)
Building age	88.07	83.16	88.24	5.520***	-0.114
	(16.69)	(28.49)	(23.10)	(0.179)	(0.184)
Bedroom	2.960	3.136	2.961	-0.142***	-0.002
	(0.503)	(1.119)	(0.965)	(0.007)	(0.005)
Bathroom	1.358	1.587	1.352	-0.207***	-0.006
	(0.307)	(0.788)	(0.569)	(0.005)	(0.005)
Average story height	1.897	1.909	1.902	-0.054***	-0.005*
	(0.246)	(0.475)	(0.400)	(0.003)	(0.002)
Garage	13.08%	31.54%	14.10%		
Full basement	90.62%	92.45%	90.88%		
Sample size	36,410	58,020	74,125		
Neighborhood Characteristics					
% Owner occupied	66.31		66.32		-0.002
	(23.00)		(22.94)		(0.002)
% Black	25.98		25.44		0.044
	(32.20)		(32.10)		(0.032)
% Hispanic	0.945		0.946		0.001
	(2.377)		(2.481)		(0.002)
% Female householder	47.54		47.54		0
	(23.18)		(23.25)		(0.002)
Sample size	3,282		3,282		

Standard errors are reported in parentheses. Colum 2 includes all housing characteristics of properties located outside the treatment areas. Housing characteristics are in 2002. Information on demographics is taken from Census 2000. All neighborhood characteristics are measured at block level. Estimates in the third column are ω, the coefficient on D_i^{250} in $X_i = \alpha + \omega D_i^{250} + \varepsilon_i$.

Table 2.3 - Characteristics of Properties Sold in 2005

	Within 250 feet	Rest of City	Within 250-750 feet	Differences 1	Differences 2
Housing Characteristics					
Assessment Value	53,377	112,285	54,376	-55,602***	997.4
	(43,942)	(124,384)	(53,336)	(3,089)	(740.7)
Square footage	1,579	1,797	1,580	-169.9***	-1.249
	(597.1)	(923.1)	(622.1)	(24.66)	(5.729)
Lot area	3,805	4,731	3,909	-1,205***	-103.9
	(2,365)	(4,945)	(2,669)	(127.2)	(87.71)
Building age	87.99	80.74	87.20	9.700***	0.782***
	(23.29)	(28.50)	(22.93)	(0.805)	(0.291)
Bedroom	2.919	3.122	2.933	-0.128***	-0.015
	(0.998)	(1.135)	(1.025)	(0.032)	(0.012)
Bathroom	1.312	1.615	1.323	-0.217***	-0.011
	(0.589)	(0.816)	(0.609)	(0.022)	(0.008)
Average story height	1.956	1.930	1.936	0.063***	0.020
	(0.392)	(0.446)	(0.394)	(0.013)	(0.015)
Garage	19.50%	31.92%	19.74%		
Full basement	93.53%	91.69%	93.80%		
Sample size	2,339	2,636	4,688		

Standard errors are reported in parentheses. Colum 2 includes all housing characteristics of properties located outside the treatment areas. Housing characteristics are in 2002. Estimates in the third column are ω, the coefficient on D_i^{250} in $X_i = \alpha + \omega D_i^{250} + \varepsilon_i$.

in treated areas are generally older and smaller, with much less assessed value than houses located in other areas of the city, which is demonstrated in column (3), as all differences are 1% significant. In contrast, for all houses in treatment and control, they are very similar in housing characteristics, and they are located in similar neighborhoods in terms of racial compositions and homeownership rates.

Table 2.3 summarizes the characteristics of residential properties sold in 2005. Similar to Table 2.2, the first 2 columns show that houses sold in 2005 are very different in and outside treatment rings, while there is little difference between the houses in treatment and control areas. The only exception is building age. Column (5) shows that houses sold in treatment areas in 2005 are on average 0.8 year older than those sold in control areas in the same year, and the difference is statistically significant. However, considering the difference is merely 0.8 compared to an average building age of more than 80 years, there is little evidence of noticeable differences. Overall, the results in Table 2.2 and Table 2.3 suggest the similarity of the treatment and control areas for all houses and those sold in 2005, and they demonstrate the importance of localized control and treatment areas.

The next step is to define the different treatment stages. To better understand the effect of multiple stages of foreclosure, a foreclosure filing period is separated from a vacancy period, as discussed in Section 2.2. It is likely that foreclosure alone has a smaller impact on the neighboring houses than vacancy does, as the physical condition is generally worse when a house is abandoned after it is foreclosed, making the neighborhood less attractive to buyers.

As this paper focuses on the impact of foreclosure on housing compositions of properties sold in the nearby areas, the outcome variables will be the quantities and characteristics of houses sold at the time of nearby foreclosures. One problem is that the housing characteristics are determined by many factors, such as the characteristics of the house itself and its location, and it will be troublesome to estimate the changes in all those measures separately. As a result, I compose a quantity index for each house. The index is defined as a predicted 2005 sale price. Here is the equation used to calculate the index:

$$\log\left(price_{is}^{2005}\right) = \alpha_s + \beta X_{is} + \varepsilon_{is} \tag{2.1}$$

The sample consists of houses sold in 2005, prior to foreclosures. $\log\left(price_{is}^{2005}\right)$ is the logarithm of sale price for house i in neighborhood s in 2005. α_s is a neighborhood fixed effect, and X_{is} is an extensive list of housing and neighborhood characteristics[21] for house i. After estimating Equation 2.1 the coefficients are used to construct the quantity measure Q for all properties in the city:

$$Q_{is} = \widehat{\log(price_{is}^{2005})} = \widehat{\alpha_s} + \hat{\beta} X_{is} \tag{2.2}$$

Equation 2.2 includes all houses in the city. Q is essentially a linear combination of all the housing and neighborhood characteristics for house i in neighborhood s. Therefore it can serve as a composite measure for housing quantity, and it is positively linked to 2005 sale price.

The benefit of constructing Q is twofold. First, it serves as a dependent variable in estimating the impact of foreclosure on quantities and characteristics of houses sold in the neighborhood. Second, as discussed above, foreclosure may have differential impacts on households of different characteristics. Without data on individual households, we can only infer such information from

---

[21] Housing characteristics includes square feet, lot size, numbers of total room, bedroom and bathroom, number of fireplace, building height, year since built, garage, 20 indicators for house style, 8 for exterior material, 4 for roof, 4 for basement, 18 for heating method, and 25 for building condition. Neighborhood characteristics include block-level violent and property crime rates, homeownership rate, percent Black and Hispanic, and percent female householder. The demographic information is from 2000 Census. I exclude the characteristics measured at block-group level due to lack of geographic precision.

Q. As a result, three sets of difference-in-differences equations will be estimated. The outcome variable of the first one will be quantity of houses sold (when the unit of observation is a house, it is equivalent to probability of sale) and in the second equation the dependent variable will be quantity measure Q. Because of the potential heterogeneity in households, the same equations will be estimated on subsamples of houses with different levels of Q[22]. Finally, the price effect will be estimated.

Equation 2.3 presents the empirical specification on foreclosure and probability of sale for nearby houses. The study area is reduced to all residential properties located within 750 feet of the foreclosed and/or vacant houses. The sample includes all properties located inside the control and treatment areas. The unit of observation is a property. This specification incorporates 3 indicators for the three different stages of foreclosure and interactions of the treatment indicator (within 250 feet of the foreclosure sites) with each of these 3 indicators. The outcome variable is an indicator of sale in quarter t. Thus, the counterfactual change in probability of sale of houses in areas immediately surrounding the foreclosure sites is estimated using probability of sale of houses in areas just slightly farther away in the same periods.

$$\begin{aligned} Sale_{ijt} &= \alpha_{st} + \beta X_i + \omega_0 D_{ij}^{250} + \omega_1 Fore_{jt} + \omega_2 Vac_{jt} + \omega_3 ReOcc_{jt} \\ &\quad + \pi_1 \cdot (D_{ij}^{250} * Fore_{jt}) + \pi_2 \cdot (D_{ij}^{250} * Vac_{jt}) + \pi_3 \cdot (D_{ij}^{250} * ReOcc_{jt}) + +\varepsilon_{ijt} \end{aligned} \quad (2.3)$$

$Sale_{ijt}$ equals 1 if house i in the area of foreclosed property j in neighborhood s is sold in quarter[23] t. α_{st} is a neighborhood-quarter fixed effect. In all, there are 89 neighborhoods in the study area. X_i is the observable housing and neighborhood characteristics for house i. $Fore_{jt}$

[22] Type H houses are those with higher Q, while Type L houses are those with lower Q

[23] Here a quarter is the same as a calendar quarter

equals 1 if in quarter t property j received a foreclosure filing in a previous quarter but is still occupied; Vac_{jt} equals 1 if in quarter t property j is vacant; $ReOcc_{jt}$ equals 1 if property j is reoccupied in quarter t. Parameters of interest are π_1, π_2 and π_3. The estimated impact of foreclosure on probability of sale in treated areas is given by the term π_1, while the impact of foreclosure-led vacancy and the impact of reoccupation are given by π_2 and π_3. In order to examine foreclosure's differential impact on different types of houses, Equation 2.3 is estimated on the subsample of houses with every 10th quantile of Q.

Equation 2.4 presents the difference-in-differences equation on the characteristics of houses sold close to a foreclosure. The sample includes all property sales within 750 feet of foreclosures. The unit of observation is a sale. The outcome variable is the quantity index of sale i which takes place in quarter t.

$$Q_{ijt} = \alpha_t + \omega_0 D_{ij}^{250} + \omega_1 Fore_{jt} + \omega_2 Vac_{jt} + \omega_3 ReOcc_{jt} + \pi_1 \cdot (D_{ij}^{250} * Fore_{jt}) + \pi_2 \cdot (D_{ij}^{250} * Vac_{jt}) + \pi_3 \cdot (D_{ij}^{250} * ReOcc_{jt}) + \varepsilon_{ijt} \quad (2.4)$$

Q_{ijt} is the quantity index of sale i in quarter t in the vicinity of foreclosure j. α_t is a quarter fixed effect. Similar to Equation 2.3, $Fore_{jt}$ equals 1 if in quarter t property j received a foreclosure filing in a previous quarter but is still occupied; Vac_{jt} equals 1 if in quarter t property j is vacant; $ReOcc_{jt}$ equals 1 if property j is reoccupied in quarter t. Parameters of interest are π_1, π_2 and π_3. Note that Equation 2.4 does not include any housing and neighborhood controls, because the dependent variable Q is already a linear combination of those controls. As we are interested in learning the impact of foreclosure on the distribution of Q, a quantile regression is estimated on the different quantiles of Q.

Finally, Equation 2.5 presents the difference-in-differences specification on the sale prices of the non-foreclosed properties. The sample includes all property sales within 750 feet of foreclosures. The unit of observation is a sale. The outcome variable is the logarithm of price of sale i in quarter t in the vicinity of foreclosure j.

$$\log(P_{ijt}) = \alpha_{st} + \beta X_i + \omega_0 D_{ij}^{250} + \omega_1 Fore_{jt} + \omega_2 Vac_{jt} + \omega_3 ReOcc_{jt}$$
$$+ \pi_1 \cdot (D_{ij}^{250} * Fore_{jt}) + \pi_2 \cdot (D_{ij}^{250} * Vac_{jt}) + \pi_3 \cdot (D_{ij}^{250} * ReOcc_{jt}) + \varepsilon_{ijt} \qquad (2.5)$$

Similar to Equation 2.5, α_{st} is a neighborhood-quarter fixed effect. X_i is the observable housing and neighborhood characteristics for house i. Parameters of interest are π_1, π_2 and π_3. In order to examine foreclosure's differential impact on different types of houses, Equation 2.5 is estimated on the subsample of houses with different levels of Q.

2.6 RESULTS

Table 2.4 presents the coefficients from the estimation for Equation 2.3 on the subsample of houses with every 10^{th} quantile of Q. Housing and neighborhood controls include all the characteristics reported in Table 2.2. The first four rows report coefficients β, ω_0, ω_1 and ω_2, respectively. For most of the subsamples, those coefficients are essentially 0, except the indicator of vacancy for a few subsamples with higher Q. Those results suggest that treatment and control groups are statistically similar and the average effects of being in the treated periods across the two groups are close to 0. As for houses with higher Q, it may be the case that vacancy reduces the probability of sale for houses in both treatment and control areas, which suggest that control area may be treated too.

The coefficients of interest for Equation 2.3 are reported in the last three rows of Table 2.4. The results clearly show the more significant impacts of foreclosure and vacancy on houses with higher Q. The first four columns show that for houses of Q lower than 4^{th} decile, neither foreclosure nor vacancy has statistically significant impact, and the magnitudes of those estimates are relatively small, though the signs are all negative. The last six columns tell an entirely different story: for houses with Q higher than the 4^{th} decile, both foreclosure and vacancy have negative and significant impact on their probabilities of sale. There are larger impacts of vacancy than foreclosure for each subsample, and the effects are increasing in magnitudes as Q increases. In a word, at the time of nearby foreclosure (and vacancy), houses with higher Q are less likely to sell while houses with lower Q are not affected, and the effect of vacancy dominates. The results on treated*reoccupied indicate that when the neighborhood foreclosure is cured, those negative impacts on sale volume of nicer houses disappear, suggesting that the housing market bounces back when the neighborhood disamenity is removed.

Across all 10 subsamples, the average probability of sale for a house in a specific quarter is roughly 0.13. Therefore, the treated*vacancy coefficient (-0.00305) on houses with Q of the highest 10^{th} percentile translates into a 23% decrease in the probability of sale for those houses at the presence of nearby foreclosed and vacant property. Likewise, the highest treated*foreclosure coefficient (-0.00168) translates into a 13% decrease in the probability of sale. Both numbers indicate significant impacts of foreclosure and vacancy on the nicer homes in the neighborhood.

Table 2.5 presents the quantile regression results for Equation 2.4 on various quantiles of Q. As mentioned above, the regression does not include any housing and neighborhood controls, as they are incorporated into Q. The results suggest that the impacts of foreclosure and vacancy on the housing quantities of sold properties in the neighborhood grow stronger when moving to the higher quantiles, while they have no effect at the lower end of the housing quantity distribution. In particular, there is little impact at the 10^{th} and 25^{th} quantile, but the coefficient on treated*vacancy is statistically significant at the median, 75^{th} and 90^{th} quantile, while the coefficient on treat*foreclosure is significant at the 75^{th} and 90^{th} quantile. As expected, the impact of vacancy is more significant than the impact of foreclosure, and the effects disappear when the neighboring foreclosure is resolved. The OLS results are similar to the median regression; however, the impact of foreclosure here is not statistically significant due to higher standard error.

When comparing the estimation results from Equation 2.3 and 2.4, we can see that one reinforces the other: results from Equation 2.3 suggest that houses with higher quantity measure are less likely to sell due to neighborhood foreclosure and vacancy, and Equation 2.4 indicates that the higher quantiles of housing distribution decrease because there are fewer sales for houses with higher quantity measure, while the lower quantiles are not affected as houses with lower quantity measure sell as usual.

Because results in Table 2.4 and 2.5 show little correlation between foreclosure-related events and composition of sales of houses with lower quantiles of Q in the neighborhood, it reduces the concern that any attempt to estimate the price effect of foreclosure on those houses may be biased due to the simultaneous changes in quantity and timing of sales. As the price effect estimation provides more direct information on the external cost of foreclosure, it may be worthwhile to examine the effect of foreclosure on sale prices of houses with lower quantiles of Q. In fact, the price estimates on all houses will be informative.

Table 2.6 presents the OLS regression results on sale prices on the subsample of houses with different quartiles of Q. From the results we can see that for different levels of Q, foreclosure-led

Table 2.4 - Difference-in-Differences Results on Probability of Sale

Probability of sale (0 or 1)	(1) Q=1st decile	(2) Q=1-2 decile	(3) Q=2-3 decile	(4) Q=3-4 decile	(5) Q=4-5 decile
Treated	-6.44e-05	-9.36e-05	-0.000427	-0.000432	-0.000182
	(0.000392)	(0.000403)	(0.000405)	(0.000421)	(0.000413)
Foreclosure	0.000302	-0.000277	-0.000102	-0.000377	-0.000409*
	(0.000213)	(0.000233)	(0.000241)	(0.000248)	(0.000244)
Vacancy	6.58e-06	0.000223	-0.000469	-0.000730**	-0.000247
	(0.000271)	(0.000284)	(0.000301)	(0.000307)	(0.000317)
Reoccupied	-0.000547	-0.000436	-0.000239	-0.000598	0.000522
	(0.000375)	(0.000383)	(0.000375)	(0.000397)	(0.000394)
Treated*foreclosure	-0.000256	-0.000332	-0.000361	-0.000839	-0.00128*
	(0.000595)	(0.000618)	(0.000650)	(0.000655)	(0.000636)
Treated*vacancy	-0.000455	-0.000730	-0.00112	-0.00161	-0.00171*
	(0.000868)	(0.000868)	(0.000905)	(0.001000)	(0.000986)
Treated*reoccupied	-0.000104	-0.000840	0.00149	-0.000615	-0.000112
	(0.00110)	(0.00101)	(0.00106)	(0.00103)	(0.00104)
Observations	1,261,515	1,270,754	1,265,921	1,263,045	1,258,770

Table 2.4 (continued)

Probability of sale (0 or 1)	(1) Q=5-6 decile	(2) Q=6-7 decile	(3) Q=7-8 decile	(4) Q=8-9 decile	(5) Q=9-10 decile
Treated	-0.000438	-0.000460	0.000315	-0.000499	0.000335
	(0.000395)	(0.000394)	(0.000408)	(0.000393)	(0.000415)
Foreclosure	0.000459*	-7.52e-06	0.000363	0.000198	-0.000292
	(0.000236)	(0.000233)	(0.000227)	(0.000221)	(0.000221)
Vacancy	-0.000391	-0.000641*	-0.000648**	-0.000585*	-7.67e-05
	(0.000318)	(0.000321)	(0.000319)	(0.000319)	(0.000357)
Reoccupied	0.000703*	-0.000561	-0.000178	0.000146	-0.000179
	(0.000397)	(0.000376)	(0.000364)	(0.000371)	(0.000398)
Treated*foreclosure	-0.00118*	-0.000905	-0.00136**	-0.00168***	-0.00156**
	(0.000619)	(0.000623)	(0.000633)	(0.000612)	(0.000627)
Treated*vacancy	-0.00192*	-0.00238**	-0.00231**	-0.00268**	-0.00305**
	(0.000993)	(0.00101)	(0.00115)	(0.00118)	(0.00129)
Treated*reoccupied	-0.000283	0.000510	0.000295	-0.000583	-0.000790
	(0.00104)	(0.00104)	(0.00109)	(0.00102)	(0.00116)
Observations	1,261,064	1,261,186	1,257,654	1,257,428	1,244,748

Standard errors clustered at foreclosure ring level are reported in parentheses. All regressions control for housing and neighborhood characteristics X_i and quarter*neighborhood fixed effects α_{st}.

Table 2.5 - Quantile Regression Results

$Q=log(\widehat{price}^{2005})$	(1) OLS	(2) 10th	(3) 25th	(4) Median	(5) 75th	(6) 90th
Treated	-0.00159	0.00795	-0.00491	0.000973	-0.00175	-0.00128
	(0.00925)	(0.00735)	(0.00701)	(0.00514)	(0.00553)	(0.00800)
Foreclosure	-0.00370	-8.39e-05	0.00752	0.00587	-0.00240	-0.00334
	(0.0193)	(0.00625)	(0.00594)	(0.00435)	(0.00469)	(0.00679)
Vacancy	-0.0100	0.0142	-0.00950	0.0109	0.0101	-0.00524
	(0.0302)	(0.0102)	(0.00981)	(0.00722)	(0.00781)	(0.0113)
Reoccupied	-0.0037	-0.00652	-0.00544	-0.00439	-0.0024	-0.000241
	(0.0302)	(0.00891)	(0.00858)	(0.00633)	(0.00681)	(0.00984)
Treated*foreclosure	-0.0194	-0.00287	-0.00517	-0.0120	-0.0204**	-0.0295**
	(0.0122)	(0.0128)	(0.0123)	(0.00899)	(0.00967)	(0.0140)
Treated*vacancy	-0.0315*	-0.00312	-0.0179	-0.0387**	-0.0840***	-0.124***
	(0.0156)	(0.0251)	(0.0239)	(0.0175)	(0.0189)	(0.0273)
Treated*reoccupied	-0.0110	-0.00208	-0.0181	-0.0124	-0.0195	-0.0198
	(0.0203)	(0.0189)	(0.0181)	(0.0132)	(0.0142)	(0.0206)
Observations	195,935	195,935	195,935	195,935	195,935	195,935

Standard errors clustered at foreclosure ring level are reported in parentheses. All regressions control for housing and neighborhood characteristics X_i and quarter*neighborhood fixed effects α_{st}.

Table 2.6 - OLS Results on Sale Prices with Subsamples of Quantiles of Q

Log (Sale Price)	(1) 25th	(2) 50th	(3) 75th	(4) 100th
Treated (within 250 feet)	-0.0137	-0.0168	-0.0250**	-0.0195
	(0.0200)	(0.0171)	(0.0120)	(0.0105)
Foreclosure	-0.00667	0.0057	0.00743	-0.00609
	(0.0158)	(0.0159)	(0.0104)	(0.00814)
Vacancy	-0.0441	-0.0236	-0.0339	-0.0204
	(0.0236)	(0.0234)	(0.0173)	(0.0128)
Reoccupied	-0.0236	0.00642	-0.00601	-0.0206
	(0.0278)	(0.0242)	(0.0166)	(0.0148)
Treated*foreclosure	-0.0423	-0.0383	-0.0427*	-0.0248
	(0.0346)	(0.0309)	(0.0248)	(0.0185)
Treated*vacancy	-0.1208**	-0.1020**	-0.0800*	-0.0717**
	(0.0502)	(0.0527)	(0.0424)	(0.0310)
Treated*reoccupied	-0.00784	-0.0517	-0.00941	-0.0223
	(0.0540)	(0.0516)	(0.0342)	(0.0349)
Observations	48,976	48,990	48,984	48,985

Standard errors clustered at foreclosure ring level are reported in parentheses. All regressions control for housinand neighborhood characteristics X_i and quarter*neighborhood fixed effects α_{st}.

vacancies depress the price by roughly 10%, though the magnitudes vary for houses with different Q. Overall, the price results show that foreclosure alone has limited effect on sale prices of nearby homes, and the effect mostly disappear when the foreclosed house is reoccupied. The estimates on price effect are higher than the average estimates in the literature, but closer in magnitudes to the results in Lin, Rosenblatt, and Yao (2009).

While the degrees of price drop appear to vary for houses with different Q, it is helpful to directly compare the estimates across those subsamples. Table 2.7 presents the marginal effect of being in a higher quartile of Q at the time of nearby vacancy on sale prices. The coefficient is obtained from re-estimating Equation 2.5, with the interaction term of vacancy by treatment replaced by an interaction term of Q, vacancy and being treated. This coefficient is reported in the first row of Table 2.7. The estimate is close to 0, indicating that the price effect is rather constant across Q.

Similar to Cui (2011), I explore the impact of different lengths of vacancy on the composition of houses sold at the time of nearby foreclosures, mainly to rule out the possibility that the results are driven by heterogeneity in areas experiencing various lengths of foreclosure-led vacancy. For example, nicer houses in the vicinity of properties vacated for longer terms may be less likely to sell even before foreclosures take place. To rule out the possibility that the estimates on composition effect are the result of pre-existing differences in treated areas, I re-estimate Equation 2.4 using OLS estimation on the sample of property transactions with all levels of Q. However, the re-estimation is on subsamples of foreclosed homes that stayed vacant for various amount of time. The difference-in-differences indicator on being treated and in the period of vacancy, as shown in Equation 2.4, are replaced by a set of dummy variables on being treated and in different periods of vacancy.

Table 2.7 - Comparing Magnitudes of Estimates in the Price Regression

Log (Sale Price)	(1)
Treated * Q * Vacancy	-0.0018
	(0.0079)
Foreclosure	0.0024
	(0.0129)
Vacancy	-0.0164
	(0.0134)
Reoccupied	0.00642
	(0.0152)
Treated*foreclosure	-0.0313
	(0.0221)
Treated*reoccupied	-0.0221
	(0.0254)
Observations	195,935

Standard errors clustered at foreclosure ring level are reported in parentheses. All regressions control for housing and neighborhood characteristics X_i and quarter*neighborhood fixed effects α_{st}.

Table 2.8 - OLS Results on Q by Length of Vacancy

$Q=log(\widehat{price}^{2005})$	(1) Vacant for 0-3 months	(2) Vacant for 3-6 months	(3) Vacant for 6-12 months	(4) Vacant for 12-18 months	(5) Vacant for >18 months
Treated * foreclosure	-0.0182	-0.0191	-0.0114	-0.0262	-0.0228
	(0.0249)	(0.0158)	(0.0121)	(0.0261)	(0.0249)
Treated * vacant first 3 months	-0.0134	-0.0124	-0.0140	-0.0125	-0.0182*
	(0.0165)	(0.0083)	(0.0108)	(0.0146)	(0.0109)
Treated * vacant for 3-6 months		-0.0291*	-0.0163	-0.0178	-0.0254
		(0.0169)	(0.0119)	(0.0175)	(0.0163)
Treated * vacant for 6-12 months			-0.0326*	-0.0258	-0.0332*
			(0.0187)	(0.0197)	(0.0188)
Treated * vacant for12-18 months				-0.0373*	-0.0395**
				(0.0203)	(0.0175)
Treated * vacant > 18 months					-0.0447**
					(0.0214)
Observations	16,213	18,803	21,751	13,518	15,923
#Foreclosed homes	278	310	364	209	242

Standard errors clustered by foreclosure property rings are reported in parentheses. Housing controls and year*neighborhood interactions are included.

Table 2.8 presents the estimation results of Equation 2.4 on subsamples of house sales taken place near foreclosed homes experiencing different lengths of vacancy. While almost all coefficients on vacancy are positive, most of them lose significance, possibly due to reduction in sample sizes. For areas surrounding houses vacant for more than 18 months, the coefficients on having been vacant for longer terms are larger in magnitude and statistically significant, which is consistent with results in Table 2.5. Moreover, foreclosure alone seems to have a negative impact, though none of the coefficients are statistically significant, possibly due to loss of sample size. Given the results on housing compositions, it is unlikely that the estimates on probability of sale and the price effect presented in Table 2.4 and 2.6 are the result of heterogeneity in treatment. Overall, these results are consistent with estimates in Table 2.5 and confirm the absence of compositional effect in the setting of this paper.

2.7 DISCUSSION

The estimation results show reduced sales volume at the upper end of housing market due to foreclosure, which is consistent with the observation of positively linked housing price and sales volume by researchers and housing specialists. While it is important to document such effects, the mechanisms behind the results are worth exploring. The quantity changes indicate that the second-hand housing supply may have shifted or is not perfectly inelastic in the short run.

As mentioned in previous sections, homeowners may be more eager to leave the neighborhood because of foreclosures. In this case, they will lower the reservation prices and supply will shift to the right. But my results show the opposite: some people tend to stay their houses when their neighbor’s house is foreclosed. Therefore, it may be the case that the second-hand housing supply curve is upward-sloping: a decreased demand for houses near foreclosure will lead to less frequent transactions.

Researchers have conjured several hypotheses to explain the positive relationship between housing price and sales. Genesove and Mayer (2001) attribute this phenomenon to seller's loss aversion. Using property listing information for Boston condominiums, they show that owners subject to nominal losses set higher asking prices and their properties stay much longer on the market. They argue that according to prospect theory (Tversky and Kahneman, 1991), sellers are averse to realizing losses in that they set up a reference point equal to the last sale price. Any offer lower than that point is considered a loss and the sellers are unwilling to accept such an offer.

Though I do not observe listing price and time on the market, my results exhibit similar pattern: transactions for higher-end houses are much less frequent when a foreclosure is present in the neighborhood, and the effect disappears when the foreclosure is cured. It may be the case that owners of those houses set higher asking prices and had to wait for longer time for a suitable buyer, if their reference points are higher than the selling prices with foreclosures present. It is also possible that those owners set up a reference point that is different from the last sale price, because in Pittsburgh it takes 16.7 years on average for a homeowner to move to another house. Most likely the nominal sale price is much higher than the buying price. As a result, it is possible that a homeowner's reference point is the fair market value of her house not long before the foreclosure in the neighborhood.

Another explanation is that sellers cannot significantly lower the reservation price because minimum down payment is required for the purchase of a new home (Stein, 1995, Genesove and Mayer, 1997). However, this theory may not be true for Pittsburgh housing market, given that the median time between two sales for a house is 16.7 years. People who stayed in the house for more than 10 years are likely to have paid enough mortgages to gain the amount of cash needed for a new home's down payment after selling their houses. In addition, many of the homeowners are elderly who look for downsizing and will buy smaller houses with less required down payments.

The third theory on housing price and sale volume is that sellers are slow to adjust their expectations. This hypothesis is empirically equivalent to loss aversion story, in that sellers do not change their reference points quickly.

On a separate note, because foreclosure is only temporary, home sellers have more incentive to wait for the foreclosure to disappear before selling their houses.

However, if many foreclosures can be resolved within a couple years, and the neighborhood condition will improve then, why is there price reduction at the first place? Because the long-term attractiveness of the neighborhood is generally not affected by one or two foreclosures, a rational buyer may be willing to purchase a nearby house at pre-foreclosure price. Possible explanations include the uncertainty about the length of vacancy and that some buyers may choose to move again in a few years. Those buyers will place a lower reservation price for a house near foreclosure, thus reducing the average prices of non-foreclosed properties in the neighborhood.

2.8 CONCLUSION

This paper has shown that foreclosures affect quantities and prices of houses that are sold in the neighboring areas, especially for higher-end houses. Data from Pittsburgh during the years of large scale foreclosure crisis show that: 1. both foreclosure and vacancy reduce the neighboring houses' probabilities of sale, and the effect is as much as 23% for vacancy and 13% for foreclosure; 2. there is little impact on houses with lower quantity measure; 3. the impact of vacancy is more significant; 4. the effects disappear when foreclosed house is reoccupied. 5. consistent with literature, foreclosure depresses sale prices of the neighboring houses. These impacts are most likely caused by homeowners' loss aversion. At the same time, I conjecture that owners of Type H houses are less liquidity and credit constrained and are therefore can afford to delay the sale, while owners of Type L houses are forced to sell.

The paper's results have broad implications for the understanding of second-hand housing market and spatial externality. First, the second-hand housing supply may be more elastic than previously assumed. Second, hedonic analysis on housing market should be more cautious when using home sale prices to estimate the impact of local (dis)amenities, as the quantity of sale may be affected too.

Lastly, the conclusions of this paper indicate that the previous literature may have underestimated the external costs of foreclosure by focusing only on the price effect. As shown in this study, home sellers living close to foreclosures will suffer the losses from sale price reductions and any additional costs associated with delaying of the sales. Those who did not end up selling their houses may face the same extra costs of sale delays, because they may have wanted to sell but failed. Such costs include real estate taxes, utility bills, maintenance expenses, and all possible indirect costs (such as loss of investment opportunities). Although estimating the exact costs associated with sale delays are beyond the scope of this paper, such losses can be substantial. Another issue overlooked by the literature is that the effect of foreclosure can be different across various types of housing, as demonstrated in this study. Again, the reason the literature underestimated the true costs of foreclosure is that hedonic analysis may yield biased results because the supply curves have shifted too.

3.0 SHALE GAS DRILLING AND THE RURAL HOUSING MARKETS

3.1 INTRODUCTION

The recent Marcellus Shale drilling boom has brought heated debates over future regulations in several northeastern states. With the introduction of new drilling techniques, drilling and production of shale gas have had unimaginable economic impacts to the Marcellus region. At the same time, the rapid development of shale gas has created concerns on potential health and environmental risks. Despite the numerous media stories, there are very few studies on the impact of shale gas drilling. This paper intends to provide the first piece of evidence on the perceived environmental and health risks of shale gas drilling by estimating its impact on nearby property values.

At the core of shale gas development are two key technologies: horizontal drilling and hydraulic fracturing. Techniques used to hydraulically fracture horizontal wells often involve injecting large volumes of chemically charged water into wells. Such practice has attracted critical interest regarding risks posed to groundwater and surface water. Other potential risks and disamienties include air pollution, gas explosion, road damage and noise. As a result, the values of properties located close to gas wells are expected to drop because of decreases in the perceived neighborhood amenity levels.

Although the area has a tradition of natural gas drilling activities, the Marcellus Shale formation was not considered to be an important gas resource until a 2008 report[24] showed that it contains the most natural gas in all the shale formations in the US. From the econometric standpoint, the sudden discovery of Marcellus Shale's natural gas potential may serve as a great natural experiment for hedonic analysis, because the geological factors determining well locations are mostly independent of pre-existing residential settlement patterns.

The only study to date focusing on the issue of shale gas drilling is Osborn et al. (2011). They find methane contamination of drinking water associated with shale gas extraction. The methods employed in these studies, however, have not been the typical techniques employed by economists.

Though there has been no literature on the external effect of shale gas drilling, many studies have examined the effect of local disamenities on nearby property values. Among them, some have looked at the external impact of oil and gas facilities. A few studies on the impact of sour gas facilities found no impact on prices (Lore and Associates Ltd., 1988; Serecon, 1997). These studies grouped relatively small samples of properties according to their proximity to infrastructure and compared prices across these groupings, or used price regression that included few property variables. One study on the same topic documented significant negative impact on rural property values (Boxall et al., 2005). Other studies focusing on similar local disamenities sometimes find reasonably large effects on land prices. Examples include refineries (Flower and Ragas, 1994), changes in water quality (Leggett and Bockstael, 2000), and electricity transmission lines (Hamilton and Schwann, 1995).

[24] Terry Englander and Gary Lash estimated that the Marcellus might contain more than 500 trillion cubic feet of natural gas.

This paper reports efforts to determine the impact of proximity to natural gas drilling sites on rural residential property values using data from three counties with most shale gas drilling activities in Pennsylvania. In theory, the impact comes from the environmental and health disamenities associated with hydraulic fracturing and horizontal drilling. The estimation results show that at this moment, shale gas drilling has no measurable impact on nearby property sales. Such lack of evidence may reflect two data limitations of this study.

The first issue is sample size. Because Marcellus gas drilling is a new phenomenon and sales are infrequent in the study area, the number of post-drilling transactions of properties located close to gas wells is rather limited, even though the study area includes more than half of the shale gas wells in Pennsylvania. The second issue is that property sales may involve mineral right transfers, which cannot be observed separately from surface right transfers. This leads to a downward bias in estimating the negative impact. A detailed discussion of those two issues can be found in data and empirical strategy sections.

I proceed as follows; Section 3.2 of the paper presents background information on the Marcellus gas boom. In Section 3.3, I describe the data used in this study. In Section 3.4, I describe the empirical methodology and present the estimation equations. Results are presented in Section 3.5. Discussions on estimation results are provided in Section 3.6. I conclude in Section 3.7.

3.2 BACKGROUND

3.2.1 Overview on Marcellus Shale Gas Drilling

Experts have known for years that natural gas deposits existed in deep shale formations, but until recently the vast quantities of natural gas in these formations were not thought to be recoverable. During the past decade, through the use of hydraulic fracturing technologies, combined with sophisticated horizontal drilling, shale gas has become an increasingly important source of natural gas in the United States.

In these shale formations, the challenge is recovering the gas from very tiny pore spaces in a low permeability rock unit. To stimulate the productivity of wells in organic-rich formations, companies drill horizontally through the rock unit and then use hydraulic fracturing to produce artificial permeability. Done together, horizontal drilling and hydraulic fracturing can make a productive well where a vertical well would have produced only a small amount of gas.

The Marcellus formation is a Devonian black shale lying about one mile beneath much of Ohio, West Virginia, Pennsylvania and New York. Because of the proximity to densely populated areas in the northeastern and the central part of the country, gas recovered from the Marcellus Shale has a distinctive transportation advantage. And the presence of an enormous volume of natural gas has a great economic significance. In additional to the positive impact on the stability of natural gas supply of the surrounding region, drilling activities have brought thousands of jobs, significant incomes for residents who lease their land to the drillers and revenues for the state and local governments.

However, there are considerable environmental safety and health concerns surrounding the current implementation of hydraulic fracturing practice. As mentioned in Section 3.1, hydraulic fracturing practice may be associated with contamination of groundwater and surface water, air pollution, gas explosion, road damage and noise.

Hydraulic fracturing involves injecting hundreds of millions of gallons of chemically charged water (frac fluid) at high pressure underground to fracture a formation and release trapped gas. Anywhere from 10 percent to 40 percent of the water sent down the well during the process returns to the surface, carrying drilling chemicals, very high levels of salts and, at times, naturally occurring radioactive material. The wastewater is usually stored at the drilling site before sending to the treatment plants. At this stage, leaking can be a serious issue. In fact, in the past three years, at least 16 wells in Pennsylvania reported spills, leaks or failures of pits where wastewater is stored, according to Pennsylvania Department of Environmental Protection records. Most of the wastewater is eventually discharged into the waterways because it is difficult to treat, and many treatment plants are not equipped to process such huge volume of wastewater. The gas industry estimates the amount of wastewater needing disposal in Pennsylvania will increase from 9 million gallons per day in 2009 to 20 million gallons per day by 2011. The wastewater contains many unknown toxics and possible radioactive materials that may pose serious threats to clean drinking water for Pennsylvania and other states over the Marcellus Shale formation.

In addition to water contamination, there is evidence that natural gas drilling causes air pollution. Toxic gases are released into the atmosphere by the process of drilling and hydraulic fracturing. In many cases, these are leading to increases in health-related problems for people living and working nearby[25]. Furthermore, occasional explosions at drilling sites have caused

[25] Though no empirical study has confirmed the health impact of drilling, there have been many media reports on increased cases of asthma and other breathing disorders from toxic fumes released in the drilling process.

injuries and death and the destruction of homes[26]. Hydraulic fracturing noise, drilling noise and road damages caused by heavy trucks carrying frac fluid all bring extra disamenities to the adjacent areas.

Finally, a majority of wells have at least one violation that directly endanger the environment and/or the safety of communities: the Pennsylvania Land Trust Association recently compiled data on citations given to Marcellus well drillers and found that the most-cited companies had an average of 0.76 violations per well. Because of all the potential environmental and health risks discussed above, New York State had a moratorium on hydraulic fracturing for one year[27].

3.2.2 Marcellus Gas Drilling in Pennsylvania

In Pennsylvania, the Marcellus Shale rests beneath the entire western half of the state and the northeastern corner. The Pennsylvania Department of Environmental Protection reports that the number of drilled wells in the Marcellus Shale has been increasing rapidly. In 2007 only 27 Marcellus Shale wells were drilled in the state, however, in 2010 the number of wells drilled had risen to 1386. Currently, natural gas companies have leased about 7 million acres of public and private property — about one-quarter of the state's entire land mass.

Figure 3.1 presents the distribution of Marcellus wells in Pennsylvania. It shows that counties with most number of wells are Bradford County, Tioga County and Washington County, the three counties included in this study. The map also shows that areas with most frequent Marcellus activities are located in northeastern Pennsylvania.

[26] For example, two Eastern Pennsylvania gas explosions in early 2011 resulted in a total of six deaths and several injuries

[27] The moratorium is lifted on July 1, 2011.

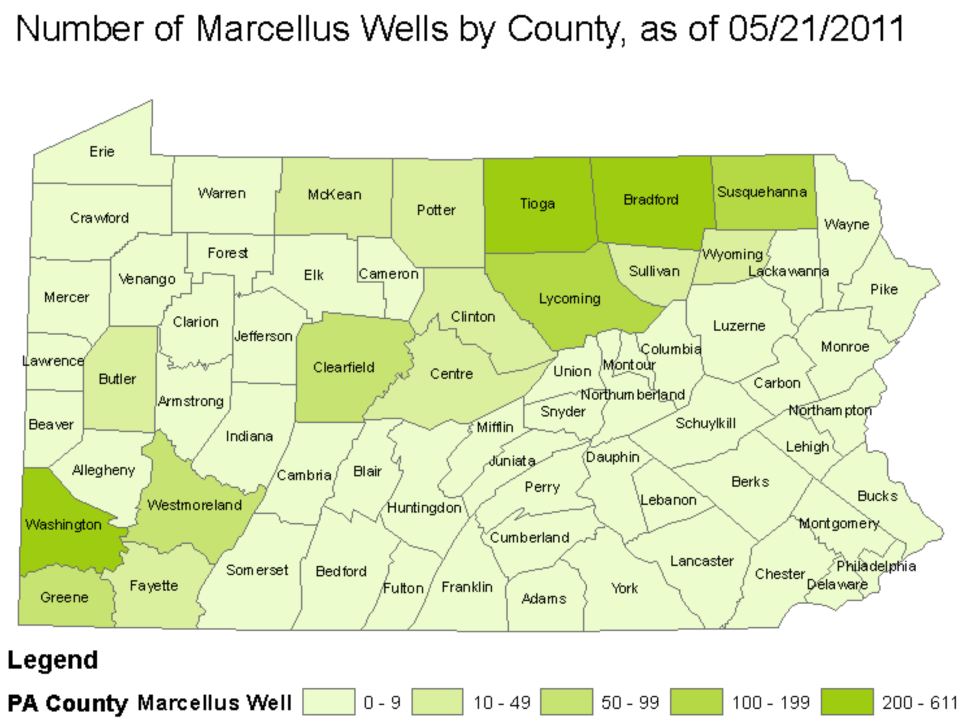

Figure 3.1 - Marcellus Shale Gas Wells in Pennsylvania, by County

3.2.3 Locations of Gas Wells and Leasing from Private Land Owners

Although gas is ubiquitous in the Marcellus region, the highest production potential is where the net thickness of shale is the greatest and where the shale can be drilled at minimum depths. These two factors explain why the northeastern Pennsylvania has seen most activities in gas drilling.

Each Marcellus Shale drill site consists of a 2 to 10 acre drilling pad that typically has 1 to 6 gas wells on it, drilled vertically and then turned horizontally. A single drill site can recover shale gas from underneath an area of 640 acres.

Because mineral rights can be privately owned and separated from surface rights in the United States, many landowners who own the mineral rights to their properties are being approached with offers to lease their rights. The gas companies will need to sign oil and gas leases with every mineral right owner to form the 640-acre drilling unit. The owners are typically compensated with a signing bonus priced by acreage and a minimum 12.5% royalty for all the gas drilled from their lands. However, if the landowners only own the surface rights, they receive no benefit from drilling and they have little control over where and when the gas companies choose to drill[28]. After successfully collecting enough gas leases to drill a unit, the gas companies apply for a drill permit and start building the drilling pad once the permit is granted. It may take several months to complete drilling the wells and have the frac fluid ready to perform hydraulic fracturing. The gas companies are also building thousands of miles of natural gas gathering systems to transport natural gas to major markets, because the total natural gas pipeline capacity currently available is a small fraction of what will be needed.

3.3 DATA

3.3.1 Data Description and Summary Statistics

The data set stems from two sources. Marcellus Shale drilling permit data are obtained from Pennsylvania's Department of Environmental Protection. Combining various data sets available on their website, I construct a file with information on every Marcellus Shale permit in the state.

---

[28] The surface owners will be notified when the gas company obtains a permit and plans to start drilling. They can negotiate compensations for damages occurred to their properties.

These records include the approval date and geo-coded location of those permits, the operator name, and an indicator for horizontal drilling technique. The data set also links each permit to the well drilled subsequently (if any), with information on the start date of drilling, natural gas production in 2009 and 2010, type and amount of waste water produced, and date and description of safety violations for each well. Since horizontal drilling technique has greater environmental and health impacts and is widely adopted by the industry, I remove all non-horizontal permits from the sample. I also exclude a small number of permits approved prior to 2007. In addition, I assign a uniform pair of latitude and longitude to all wells located on the same pad. Between 2007 and May 2011, there were 5,392 unique drilling permits in Pennsylvania, while more than 50% of those (2,758) were located in Bradford County, Tioga County or Washington County.

Second, I also obtain the exact locations (in shapefile format) of all properties from the three counties, with building characteristics attached. The data also includes the most recent transaction date and its sale price for each property. All nonresidential properties are removed from the sample. Also, all transactions between family members are excluded. The property records are merged with drilling permit data by geographic locations using ArcMap. Because all of the data are geo-coded, I am able to cross-check the two datasets and remove any property with Marcellus Shale gas leases. In this sense, the study aims to measure only the external effects associated with shale gas drilling.

Descriptive statistics of the gas permits and wells are reported in Table 3.1. Average characteristics for all counties are reported in column 1 while statistics for each county are reported separately. The first row shows that the average depth of wells drilled is 6,842 feet, which is more than 1 mile deep. For the two counties (Bradford and Tioga) in northeastern part of the state, the wells are noticeably shallower than those in Washington County, which is located in southwestern Pennsylvania. As mentioned in Section 3.2, this geological condition leads to more permits in Bradford County and Tioga County reported in the third row. In addition, Table 3.1 shows that the shallower wells in the two counties have higher production than the deeper wells in Washington. As for violation record, on average, 82.57% of all wells have violated health and safety regulations. Finally, gas drilling in the Marcellus region is still at its early stage, as the median days elapsed since the approval of a gas permit is about one year, and it takes on average 72 days to start the drilling process. Overall, Table 3.1 shows county variations in many key features of gas drilling. However, most of such differences are caused by factors unrelated to the local housing markets.

Table 3.2 provides summary statistics of parcels in the three counties. Overall, Washington County is less rural than the other two counties, with smaller average lot size (3.471 acres) and more frequent housing transactions (thus higher last sale prices). Other housing characteristics are more similar. In general, houses are old and the time between two transactions for a house is quite long. The average building age is 64.28 years and days since last sale is 4,280, which is about 12 years. The fact that housing transactions are rather infrequent (and also that gas drilling started in the recent 4, 5 years) reduces the concern of missing sale information, as only the most recent sale record is included in the data[29].

[29] I randomly selected 20 properties in Tioga County and checked their sales records online from the county assessment database, which contains the 3 most recent sale records, and found no missing sale after 2007.

Table 3.1 - Characteristics of Marcellus Shale Gas Permit and Well by County

	All counties	Bradford	Tioga	Washington
Marcellus Shale permit				
Well depth (feet)	6,842	6,566	6,577	8,047
	(1,708)	(1,553)	(1,613)	(1,757)
Median days since approval	362	357	340	417
Sample size	2,758	1,452	746	560
Marcellus Shale gas well				
Median days between permit approval and start of drilling	72	70	66	85
% producing as of Dec 2010	13.34	10.60	15.42	17.68
	(34.01)	(30.80)	(36.13)	(38.18)
Gas quantity Jul-Dec 2010 (mcf*)	350,276	401,338	346,887	274,783
	(406,471)	(300,601)	(621,078)	(138,092)
Waste water Jul-Dec10 (mcf)	3,695	2,926	5,232	3,864
	(4,308)	(3,430)	(4,783)	(4,992)
#Health & safety violations per well	0.8257	1.096	0.6769	0.5632
(All producing wells)	(1.385)	(1.484)	(1.065)	(1.545)
Sample size	1,266	605	321	340

Standard errors are reported in parentheses. *mcf = 1000 cubic feet

Table 3.2 - Characteristics of Properties by County

Housing Characteristics	All counties	Bradford	Tioga	Washington
Lot area (in acres)	8.004	17.07	16.97	3.471
	(28.70)	(41.73)	(44.35)	(15.83)
Building age	64.28	65.82	62.10	64.34
	(64.79)	(47.91)	(41.80)	(72.63)
Bedroom	2.966	3.138	2.818	2.951
	(0.958)	(1.080)	(1.024)	(0.898)
Bathroom	1.326	1.394	1.231	1.328
	(0.603)	(0.623)	(0.621)	(0.591)
Average story height	1.482	1.483	1.283	1.526
	(0.437)	(0.458)	(0.249)	(0.451)
Square footage	1,671	1,699	NA	1,663
	(744.3)	(703.5)		(755.1)
Assessed land value	13,822	NA	23,933	11,558
	(23,891)		(23,860)	(23,305)
Assessed building value	53,414	NA	62,014	51,489
	(49,843)		(43,991)	(50,862)
Last sales price (valid sale)	102,780	72,476	71,135	118,674
	(136,287)	(97,138)	(96,864)	(150,153)
Days since last sale	4,280	5,554	4,365	3,829
(valid sale)	(3,637)	(5,054)	(3,404)	(2,954)
Sample size	110,144	20,427	16,411	73,306

Standard errors are reported in parentheses

3.3.2 Gas Drilling and Mineral Right Transfer

As mentioned in Section 3.1, it is very difficult to obtain information on mineral right ownership. In fact, even the owners themselves have to resort to title companies sometimes to determine the ownership of mineral rights. As a result, when a sale is observed, it is unclear whether it is only a surface right transfer or a transfer including mineral rights too. Even so, with information on well location and owner name, we are able to identify mineral right ownership in two cases and exclude such properties from the sample.

First, properties with gas permits are excluded. The mineral rights of those properties are leased by the gas drillers. The second column in Table 3.3 presents housing characteristics of residential properties where gas permits are located. In all there are 363 such properties. It is quite obvious that those properties are much different from their neighbors. They are on average 18 times the size of a property without gas permits. The properties with gas permits are in general older and have more rooms and living area. As a result, the assessed land values are much higher. Finally, transactions of the gas permit properties are even more infrequent.

Second, since the transaction records include names of current owners, I am able to identify all sales with buyers from the gas drilling industry. I exclude such sales from the sample because those transactions often involve mineral right transfers that may obscure the effect of gas drilling on surface land values. Housing characteristics of such properties are reported in the last column of Table 3.3. It is evident that those properties are sold recently at much higher prices than their assessed value: the average sale price is almost 1.6 million dollars, and most of them are sold in 2010 or 2011. Other characteristics indicate that they are more similar to properties with existing gas permits than to those without.

Table 3.3 - Characteristics of Properties Associated with Gas Drilling

Housing Characteristics	All Data	With Gas Permits	Sold to Gas Drillers
Lot area (in acres)	8.004	144.21	95.67
	(28.70)	(101.58)	(93.69)
Building age	64.28	107.46	65.00
	(64.79)	(199.74)	(46.10)
Bedroom	2.966	3.592	3.010
	(0.958)	(1.613)	(0.707)
Bathroom	1.326	1.449	1.264
	(0.603)	(0.765)	(0.447)
Average story height	1.482	2.343	1.642
	(0.437)	(0.853)	(0.894)
Square footage	1,671	2,229	1,305
	(744.3)	(1,054)	(565.5)
Assessed land value	13,822	89,849	73,387
	(23,891)	(59,281)	(64,450)
Assessed building value	53,414	66,292	84,620
	(49,843)	(47,485)	(99,685)
Last sales price (valid sale)	102,780	59,326	1,599,750
	(136,287)	(126,588)	(1,825,948)
Days since last sale	4,280	6,408	248
(valid sale)	(3,637)	(6,246)	(141.71)
Sample size	110,144	363	15

Standard errors are reported in parentheses

From Table 3.3, it is obvious that including the properties sold to gas drillers will significantly bias the estimation results upward, because most of such sales take place after drillers obtain the permits. A likely scenario is that drillers buy those lands with mineral rights to reserve for future drilling.

In addition, as discussed in Section 3.2, one drilling unit can cover about 640 acres (1,609 by 1,609 feet), which means that the minimum distance between two drill pads is roughly 3,218 feet (as a result there should be no clustering in treatment areas as long as they are within 1500 feet from the pad). It also implies that any post-drilling property sale taken place within the boundary of a unit has to be a surface-right-only transfer (if they are not sold to gas drillers, as discussed above), because all the mineral rights have to be leased before drilling starts.

3.4 EMPIRICAL STRATEGY AND ESTIMATION FRAMEWORK

As discussed in previous sections, Marcellus gas drilling carries environmental and health risks and will in theory lower the demand for nearby lands. Because of the mechanisms, hedonic analysis can be used to examine the negative effect of drilling on the valuation of nearby properties through reduced sale prices. The following paragraphs provide discussions on how the analysis will be shaped by the specific setting of the problem and the availability of data.

From many perspectives, the locations of gas wells are independent of the choice of housing. First of all, gas companies choose to lease the lands where they can achieve the highest production potentials, and this is mainly determined by the thickness of the Marcellus formation and how deep they are underneath. These geological conditions are most likely unrelated to where people choose to live. Also, for most households, Marcellus gas drilling took place after they settled in the area. In fact, the gas companies showed little interest in drilling until researchers published their new estimates on Marcellus gas potentials in 2008. It was a surprise for both landowners and gas companies. In addition, surface owners and neighbors of landowners who signed the oil and gas leases have no power over where the wells are drilled, though it is true that gas companies will drill their first wells in places where they could gather enough leases to form a unit, and the mineral right owners who signed the leases quickly may be a selected group. Therefore, Marcellus gas drilling serves as a good natural experiment.

On the other hand, it is obvious that gas companies will choose to drill first in areas close to existing pipelines and major roads to reduce transportation costs. Also, Table 3.3 indicates that wells tend to locate in the more rural areas, maybe because more acreage means gas companies could negotiate with fewer people to acquire enough land[30]. But selection on those factors may diminish as more wells are drilled – the gas companies have started building new pipelines and are gradually moving to more urbanized areas as most rural lands are leased.

For all the reasons stated above, Marcellus gas drilling is largely exogenous to housing choices. However, selection on observables such as lot size does exist. As a result, this study will apply both cross-sectional and difference-in-differences design to analyze the effect on property values.

There are two issues related to sale records used in this study. First, because only the most recent sale information is available, it is impossible to adopt repeated sales approach, which is an attractive design as it can control for all the property-specific time-invariant unobserved factors. Second, as discussed in Section 3.1, there are sample size issues in this study. Because

[30] There are also anecdotes that gas companies target at out-of-state mineral right owners because they do not live the area and are possibly not familiar with the bargaining process when leasing.

transactions are infrequent and Marcellus drilling is still a relatively new phenomenon, the number of post-drilling sales is limited, especially in the rural areas adjacent to gas wells.

Table 3.4 - Characteristics of Properties around Gas Permit Locations

	500 feet	1000 feet	1500 feet	2000 feet	2500 feet	3000 feet	6000 feet
Housing Characteristics							
Lot area (in acres)	53.294	41.302	34.557	30.316	27.040	25.637	15.420
	(71.758)	(64.272)	(59.826)	(55.835)	(52.347)	(50.836)	(39.166)
Building age	66.506	66.075	64.612	62.960	61.529	60.981	61.465
	(74.538)	(80.981)	(79.478)	(73.552)	(72.419)	(69.878)	(66.938)
Bedroom	3.238	3.156	3.112	3.070	3.059	3.052	3.002
	(1.236)	(1.160)	(1.119)	(1.080)	(1.074)	(1.064)	(1.012)
Bathroom	1.383	1.356	1.349	1.341	1.349	1.349	1.336
	(0.6942)	(0.6544)	(0.633)	(0.6186)	(0.6252)	(0.6248)	(0.6101)
Average story height	1.446	1.421	1.406	1.399	1.397	1.397	1.434
	(0.4125)	(0.4117)	(0.4108)	(0.4179)	(0.4151)	(0.4150)	(0.4285)
Last sales price (valid sale)	112,636	101,818	92,632	90,618	91,388	92,475	95,681
	(218,919)	(172,734)	(139,358)	(125,803)	(126,155)	(123,178)	(128,113)
Days since last sale	5,197	5,024	4,986	4,966	4,913	4,849	4,638
	(3,932)	(3,706)	(3,704)	(3,667)	(3,652)	(3,625)	(3,595)
Sample size	1,247	2,778	4,779	7,182	9,913	11,095	33,683

Standard errors are reported in parentheses.

Table 3.4 compare housing characteristics of properties within various distances from the gas permit locations (latitude and longitude). There are three noticeable features. First, properties closer to gas permits has more acreage, reflecting the tendency that gas companies choose to lease in the more rural areas to reduce bargaining costs. Second, because properties closer to gas locations have more lands, they generally sell for higher prices and transactions are less frequent. Third, even though those properties have bigger lots, the buildings on these lands are not much different from buildings on smaller lots in terms of several key housing characteristics, such as number of bedroom and number of bathroom. Overall, though selection on lot size is obvious, other measures relating to housing conditions are more homogenous among properties in various distances from gas permit locations. Due to the possible non-linear effect of lot size on property values in this setting, I include a polynomial[31] of lot size as a set of controls in the estimation equations.

Since there is no consensus on how far the effect of hydraulic fracturing would reach, I experiment with various sizes of impacted areas shown in Table 3.4. Also, because Marcellus gas drilling boom started around 2007, the samples are restricted to arm's length sales[32] in and after 2007. I further define the treatment as the drilling of a well, instead of the permit granting. The argument is that the activity of drilling is visible to nearby residents but the permit approval is not necessarily so[33]. For the same reason, I define the start of treatment period as the date on which the first well on a given pad is drilled.

The cross-sectional specification takes the following form:

$$\log(P_{ijt}) = \alpha + \pi_1 Post\_Drill_i + \beta_0 X_i + \beta_1 \cdot (Mun_j * Q_t) + \varepsilon_{ijt} \tag{3.1}$$

The sample includes every arm's length transaction in and after 2007 of residential property located within certain distance from the drill pad. $\log(P_{ijt})$ is the logarithm of sale price of house i in neighborhood j in quarter t. $Post\_Drill_i$ equals 1 if house i is sold after the date on which drilling of the first well on a given pad starts. X_i is the observable housing characteristics for house i. Mun_j is a municipality fixed effect and Q_t is the year-quarter fixed effect. The parameter of interest is π_1.

---

[31] A square term and a cubic term of lot size are included in some of the specifications

[32] Defined as transactions with sale price greater than $5,000

[33] As discussed in Section 3.2, when a permit is approved, the drillers only need to notify surface owners whose lands will be drilled underneath.

Equation 3.2 presents the difference-in-differences equation with control areas located outside the treatment areas:

$$\text{Log}\left(P_{ijt}\right) = \alpha + \pi_1 Treated_i + \pi_2 Post\_Drill_{it} + \pi_3 \cdot (Treated_i * Post\_Drill_{it})$$
$$+\beta_0 X_i + \beta_1 \cdot (Mun_j * Q_t) + \varepsilon_{it} \qquad (3.2)$$

The sample includes every arm's length transaction in and after 2007 of residential property located in the treatment or control areas. $\log\left(P_{ijt}\right)$ is the logarithm of sale price of house i in neighborhood j in quarter t. $Treated_i$ equals 1 if house i is located inside the treatment ring. Similar to Equation 3.1, $Post\_Drill_i$ equals 1 if house i is sold after the date on which drilling of the first well on a given pad starts. X_i is the observable housing characteristics for house i. Mun_j is a municipality fixed effect and Q_t is the year-quarter fixed effect. The parameter of interest is π_3, which measures the impact of gas drilling on sale price of houses in the treatment areas sold after the start of drilling.

In addition to the price effect, gas drilling may impact probability of sale. Because transaction is rather infrequent, the average probability of sale in a given quarter is close to 0. Therefore, a probit model with difference-in-differences indicators similar to Equation 3.2 is also estimated:

$$\text{Prob}\left(Sale_{ijt} = 1 \mid \boldsymbol{x}\right) = \emptyset(\alpha + \pi_1 Treated_i + \pi_2 Post\_Drill_{it}$$
$$+\pi_3 \cdot (Treated_i * Post\_Drill_{it}) + \beta_0 X_i + \beta_1 \cdot (Mun_j * Q_t)) \qquad (3.3)$$

Note that the sample in Equation 3.3 includes all properties located in the treatment or control areas. $Sale_{ijt}$ equals 1 if there is an arm's length transaction in or after 2007 on property i in municipality j in quarter t.

3.5 RESULTS

Table 3.5 presents the coefficients from the estimation for Equation 3.1. Parameter of interest is reported in the first row, while coefficients on selected housing characteristics are also presented. Each column reports the results of cross-sectional regression on a different sample. For example, in column (1), the sample includes all arms' length sales within 1000 feet of a drill pad since 2007.

The results in the first row of Table 3.5 indicate that drilling does have a negative impact on sale prices of nearby properties, and the effect is gradually decreasing as distance from the drill pad increases. At its peak, drilling depresses sale prices of properties within 1000 feet of a drill pad by 7.8%. At 2500 feet, the impact reduces to 0.48%. However, none of the coefficients are statistically significant, possibly due to rather small sample sizes. Even so, the coefficient at 1500 feet in column (2) has a p-value of 0.163, which is close to being 10% significant. At such distance, the presence of gas wells reduce nearby property sale prices by about 5%.

Table 3.6 presents the difference-in-differences results on sale price. Results on different treatment and control combinations are reported in the 3 columns. For example, the treatment area in column (1) is within 1000 feet of the drill pad while the control area is between 1000 and 3000 feet away; in column (3), the treatment is within 1 mile of the pad and the control is between 1 and 3 miles away.

Across all samples, the indicator of being treated is insignificant, which means the treatment and control areas are relatively homogenous. Similar to the cross-sectional results, the coefficients on difference-in-difference estimators reported in the third row are negative across all treatment and control combinations, though none of them is statistically significant. Note that the magnitudes of the estimator of interest are comparable with the cross-sectional results, which is at least some evidence on the negative impact of gas drilling.

Table 3.7 presents the marginal effects in the probit model (Equation 3.3) on probability of sale, while including the nonlinear terms of lot size. Almost all the marginal effects are close to 0. In other words, gas drilling has little effect on the sale probability of nearby properties.

Taken together, results from Table 3.5, Table 3.6 and Table 3.7 do suggest there might be negative impact of gas drilling on surrounding properties, though none of the results is statistically significant.

Table 3.5 - Cross-Sectional Regression Results on Log Sale Price

Log(Sale Price)	(1) 1000 feet	(2) 1500 feet	(3) 2000 feet	(4) 2500 feet
Post drilling	-0.0708	-0.0481	-0.0251	-0.00484
	(0.0744)	(0.0341)	(0.0319)	(0.0234)
Lot area (in acres)	0.0356**	0.0291***	0.0306***	0.0281***
	(0.0156)	(0.00595)	(0.00470)	(0.00375)
Lot area^2	-0.000303	-0.000205***	-0.000209***	-0.000192***
	(0.000230)	(6.92e-05)	(5.50e-05)	(4.34e-05)
Lot area^3	7.04e-07	3.84e-07***	3.79e-07***	3.65e-07***
	(7.49e-07)	(1.70e-07)	(1.31e-07)	(1.01e-07)
Building age	-0.00149	-0.000564*	-0.000336	-0.000185
	(0.00337)	(0.000311)	(0.000306)	(0.000283)
Bedroom	-0.148	-0.111	-0.0885	-0.0524
	(0.126)	(0.0751)	(0.0536)	(0.0494)
Bathroom	0.557**	0.211*	0.241*	0.231***
	(0.224)	(0.125)	(0.103)	(0.0725)
#Post drilling sales	8	15	23	34
Observations	165	328	530	803
Sales included	>=2007	>=2007	>=2007	>=2007

Standard errors clustered at municipality level are reported in parentheses.

Table 3.6 - Difference-in-Differences Results on Log Sale Price

Log(Sale Price)	(1) 1000 vs. 3000 feet	(2) 2000 vs. 6000 feet	(3) 1 vs. 3 miles
Treated	-0.0558	-0.0501	0.0106
	(0.118)	(0.0475)	(0.0168)
Post drilling	-0.0301	-0.121**	-0.0224
	(0.104)	(0.0507)	(0.0260)
Treated*post drilling	-0.0415	-0.0181	-0.0209
	(0.253)	(0.143)	(0.0556)
Lot Area	0.0259***	0.0314***	0.0292***
	(0.00336)	(0.00276)	(0.00200)
Lot area^2	-0.000150***	-0.000224***	-0.000206***
	(3.57e-05)	(3.19e-05)	(2.24e-05)
Lot area^3	2.39e-07**	4.13e-07***	3.76e-07***
	(9.42e-08)	(7.36e-08)	(4.58e-08)
Building age	-0.000377	-2.06e-05	0.000979*
	(0.000338)	(0.000258)	(0.000499)
Bedroom	-0.0425	-0.00493	0.00243
	(0.0515)	(0.0389)	(0.0227)
Bathroom	0.200***	0.221***	0.219***
	(0.0698)	(0.0352)	(0.0302)
Observations	1,064	4,830	40,744
Sales included	>=2007	>=2007	>=2007

Standard errors clustered at municipality level are reported in parentheses.

Table 3.7 - Probit Regression Results on Probability of Sale

Marginal Effect on Probability of Sale	(1) 1000 vs. 3000 feet	(2) 2000 vs. 6000 feet	(3) 1 vs. 3 miles
Treated	-0.00061	-0.00015	-0.00012
	(0.00316)	(0.0002)	(0.00015)
Post Drilling	-0.00185	-0.00069	-0.00084
	(0.00174)	(0.0004)	(0.00076)
Treated*Post Drilling	-0.00458	0.00078	-0.00243
	(0.00405)	(0.0029)	(0.00184)
Observations	18,862	56,545	186,847

Standard errors are reported in parentheses.

3.6 DISCUSSION

Although the negative relationship between gas drilling and property values seems clear in theory, the estimation results show only weak evidence. There are many possible reasons.

First, because Marcellus gas drilling has only been active on a large scale for about one or two years, there may not be enough observations on either gas well or property transaction, as illustrated in this study. The fact that most wells are in rural areas adds to the issue of limited sample size, since the number of rural properties sold in recent years are much lower than the number of urban housing transactions. For example, there are only 34 post-drilling sales within 2500 feet of a drill pad, out of 803 total sales in the same area since 2007, as shown in Table 3.4. In addition, even if the sample sizes are large enough, it may be too soon to have any significant impact on the housing markets. Because there is no literature on shale gas drilling and housing market, there is no benchmark as of how long it takes to have an impact. In the ideal case, we may be able to see some effect in near future, say five years.

Second, as discussed in Section 3.3, some of the sales may include mineral right transfers, and it will obscure the true effect on surface land values. One possible solution is to obtain the deed records from those counties, which include all gas leases signed by mineral right owners. In that case, we can at least exclude all property sales with those people as sellers, because they are likely selling mineral rights too. At the same time, if the same properties are sold by people with other names, we can be sure that those transactions are surface right only transfers, because we have identified separate mineral right owners.

Table 3.8 - Results Including Sales to Drillers and Those with Gas Wells

Log(Sale Price)	A. Cross-Sectional Results			
	(1) 1000 feet	(2) 1500 feet	(3) 2000 feet	(4) 2500 feet
Post Drilling	0.0108	-0.00998	-9.04e-05	0.0120
	(0.0511)	(0.0260)	(0.0277)	(0.0234)
Observations	168	339	539	815
Sales included	>=2007	>=2007	>=2007	>=2007

	B. Difference-in-Differences Results		
	(1) 1000 vs. 3000 feet	(2) 2000 vs. 6000 feet	(3) 1 vs. 3 miles
Treated	-0.0599	-0.0472	0.0118
	(0.118)	(0.0467)	(0.0169)
Post Drilling	-0.0474	-0.123*	-0.0229
	(0.102)	(0.0501)	(0.0259)
Treated*Post Drilling	0.226	0.0451	-0.0120
	(0.188)	(0.145)	(0.0552)
Observations	1,076	4,844	40,788
Sales included	>=2007	>=2007	>=2007

Standard errors clustered at municipality level are reported in parentheses.

In Section 3.3, two possible scenarios are discussed where the mineral rights are identifiable. To see how including mineral right transfers bias the results downward, I estimate Equation 3.1 and 3.2 with the original samples plus sales to drillers and those with gas wells. Coefficients of interest are reported in row (1) in Panel A and row (3) in Panel B. It is evident that including the possible mineral right sales will reduce the negative impact on property sales. In some settings, the impact turns from negative to positive. Those results indicate that the estimates in this study are lower bound of the actual effect, and if all mineral right transfers can be removed from the sample, the negative effect may be larger and statistically significant.

Lastly, it will be helpful to obtain more sale records on one property. At least it makes a repeat-sale approach possible.

3.7 CONCLUSION

The results of this analysis suggest that the presence of horizontal drilling and hydraulic fracturing have little measurable impacts on the values of neighboring rural residential properties. Given that Marcellus gas drilling is still a new phenomenon, we have confidence that in the longer-term there will be more evidence on such effects.

To our knowledge, this is the first academic study of the implication of the new gas drilling technologies upon property values. And naturally the results should be considered with some caution (and awaits further evidence). Still, the study may be of interest to residents, gas companies and regulators.

BIBLIOGRAPHY

Boxall, P., Chan, W., and McMillan, M., 2005. "The Impact of Oil and Natural Gas Facilities on Rural Residential Property Values: A Spatial Hedonic Analysis", *Resource and Energy Economics*, 27 (2005) 248-269

Bursik, R., 1988. "Social Disorganization and Theories of Crime and Delinquency: Problems and Prospects", *Criminology* 26 (4), 519 – 552

Campbell, J., Giglio, S., and Pathak, P., 2009. "Forced Sales and House Prices", *NBER Working Paper* No. 14866, April

Clauretie, T., and Daneshvary, N., 2009. "Estimating the House Foreclosure Discount Corrected for Spatial Price Interdependence and Endogeneity of Marketing Time". *Real Estate Economics* 37, No. 1:43–67

Cui, L., 2011. "Foreclosure, Vacancy and Crime", *working paper*

Faris, R., 1955. "Social Disorganization", 2nd edition, *New York: The Ronald Press Company*

Flower, P.C., and Ragas,W., 1994. "The Effects of Refineries on Neighborhood Property Values", *Journal of Real Estate Research* 9, 319–338

Genesove, D., and Mayer, C., 1997. "Equity and Time to Sale in the Real Estate Market", *American Economic Review* 87, 255-269

Genesove, D., and Mayer, C., 2001. "Loss Aversion and Seller Behavior: Evidence from the Housing Market", *Quarterly Journal of Economics* 116, 1233-1260

Gerardi, K., Shapiro, A., and Willen, P., 2007. "Subprime Outcomes: Risky Mortgages, Homeownership Experiences, and Foreclosures", *Federal Reserve Bank of Boston Working Paper* 07-15

Goodstein, R. and Lee, Y., 2010. "Do Foreclosures Increase Crime?" *working paper*

Hamilton, S., and Schwann, G., 1995. "Do High Voltage Electric Transmission Lines Affect Property Value?" *Land Economics* 71, 436–444

Harding, J., Rosenblatt, E., and Yao, V., 2009. "The Contagion Effect of Foreclosed Properties", *Journal of Urban Economics* 66, No. 3:164–78

Hipp, J., 2010. "A Dynamic View of Neighborhoods: The Reciprocal Relationship between Crime and Neighborhood Structural Characteristics", *Social Problems*, 57 (2), 205–230

Immergluck, D. and Smith, G., 2004. "Risky Business—an Econometric Analysis of the Relationship between Subprime Lending and Neighborhood Foreclosures", *Chicago, IL: The Woodstock Institute*

Immergluck, D. and Smith, G., 2006. "The Impact of Single-family Mortgage Foreclosures on Neighborhood Crime", *Housing Studies*, 21 (6), 851-866

Immergluck, D., and Smith, G., 2006. "The External Costs of Foreclosure: The Impact of Single-Family Mortgage Foreclosures on Property Values", *Housing Policy Debate* 17, No. 1, 57–79

Kelling, G. and Coles, C., 1996. "Fixing Broken Windows: Restoring Order and Reducing Crime in Our Communities", *the Free Press*

Leggett, C., and Bockstael, N., 2000. "Evidence of the Effects of Water Quality on Residential Land Prices", *Journal of Environmental Economics and Management* 39, 121–144.

Leonard, T., and Murdoch, J., 2009. "The Neighborhood Effects of Foreclosure", *Journal of Geographical Systems* 11, No. 4:317–32

Linden, L., and Rockoff, J., "Estimates of the Impact of Crime Risk on Property Values from Megan's Laws", *American Economic Review*, 2008, 98:3, 1103-1127

Liska, A. and Bellair, P., 1995. "Violent-Crime Rates and Racial Composition: Convergence over Time", *American Journal of Sociology* 101:578–610

Lore, J. and Associates Ltd., 1988. "The Effect of Sour Gas Facilities on Land Values in West Central Alberta", Report Prepared for Shell Canada, 19 pp.

Mikelbank, B., 2008. "Spatial Analysis of the Impact of Vacant, Abandoned and Foreclosed Properties", *working paper*

Osborn, S., Vengosh, A., Warner, N., and Jackson, R., 2011. "Methane Contamination of Drinking Water Accompanying Gas-Well Drilling and Hydraulic Fracturing", *Proceedings of the National Academy of Sciences*, 10.1073/pnas.1100682108

Pennington-Cross, A., 2006. "The Value of Foreclosed Property". *Journal of Real Estate Research* 28, No. 2:193–214

Rogers, W., and Winter, W., 2009. "The Impact of Foreclosures on Neighboring Housing Sales", *Journal of Real Estate Research* 31, No. 4: 455–79

Rossi-Hansberg, E., Pierre-Daniel S, and Owens, R., 2008. "Housing Externalities", *Journal of Political Economy*, 2010, vol. 118, no 3.

Schuetz, J., Been, V., and Ellen, I., 2008. "Neighborhood Effects of Concentrated Mortgage Foreclosures", *Journal of Housing Economics* 17, No. 4:306–19.

Serecon Valuation and Agricultural Consulting Inc. 1997. "Impact of Sour Oil/Gas Facilities on Property Values", Report Prepared for Shell Canada, 14 pp.

Shaw, C. and McKay, H., 1942. "Juvenile Delinquency in Urban Areas", *Chicago: University of Chicago Press*

Shaw, C., Zorbaugh, F., McKay, H. and Contrell L., 1929. "Delinquency Areas", *Chicago: University of Chicago Press*

Sheppard, S., 1999. "Hedonic Analysis of Housing Markets", *Handbook of Regional and Urban Ecoomics*, Chapter 41

Shlay, A., and Whitman, C., 2006. "Research for Democracy: Linking Community Organizing and Research to Leverage Blight Policy", *City & Community* 5 (2): 153-171.

Skogan, W., 1990. "Disorder and Decline: Crime and the Spiral of Decay in American Neighborhoods", *Berkeley: University of California Press*

Spelman, W., 1993. "Abandoned Building: Magnets for Crimes?" *Journal of Criminal Justice*, Vol. 21, 481-495

Stein, J., 1995. "Prices and Trading Volume in the Housing Market: A Model with Down-Payment Constraints," *Quarterly Journal of Economics,* Vol. 110, No. 2, 379–406

Sumell, A., 2009. "The Determinants of Foreclosed Property Values: Evidence from Inner-City Cleveland", *Journal of Housing Research* 18, No. 1:45–61

Tversky, A., and Kahneman, D., "Loss Aversion in Riskless Choice: A Reference Dependent Model," *Quarterly Journal of Economics,* CVI (1991), 1039–1061

Zhengou, L., Rosenblatt, E., and Yao, V., 2009. "Spillover Effects of Foreclosures on Neighborhood Property Values", *Journal of Real Estate Finance and Economics* 38, No. 4, 387–407