2014

An Empirical Analysis of the Role of Residential Real Estate Investment in the Economic Development of the Northeast Region of the United States

Praveena Rajam Jayaraman

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>

Recommended Citation

Jayaraman, Praveena Rajam, "An Empirical Analysis of the Role of Residential Real Estate Investment in the Economic Development of the Northeast Region of the United States" (2014). *Graduate Theses, Dissertations, and Problem Reports*. 5882.

<https://researchrepository.wvu.edu/etd/5882>

This Dissertation is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Dissertation in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Dissertation has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.

**An Empirical Analysis of the Role of Residential Real Estate Investment in the Economic
Development of the Northeast Region of the United States**

Praveena Rajam Jayaraman

Dissertation Submitted to the
Davis College of Agriculture, Natural Resources and Design
at West Virginia University
in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy
in
Natural Resource Economics

Tesfa Gebremedhin, Ph.D., Chair
Peter Schaeffer, Ph.D.
Tim Phipps, Ph.D.
Donald Lacombe, Ph.D.
Santiago Pinto, Ph.D.
Wesley Burnett, Ph.D.

Agricultural and Resource Economics Program
Division of Resource Management
West Virginia University

Morgantown, West Virginia
2014

Key Words: Residential Real Estate, Economic Development, Simultaneous Equations, Spatial
Panel Methods

Copyright 2014 Praveena Rajam Jayaraman

Abstract

An Empirical Analysis of the Role of Residential Real Estate Investment in the Economic Development of the Northeast Region of the United States

Praveena Rajam Jayaraman

Residential real estate investment has been recognized as an agent of economic development since the 1970s because residential real estate investment is a major economic activity with large multiplier effects. Residential real estate improvement is also linked to many external social and economic benefits. Previous studies have examined the role of residential real estate in economic development through approaches such as the effects of employment and income, household saving, labor productivity, health productivity and growth from real estate investment, as well as home ownership effects. However, recent discussions about the relationships between residential real estate investment and economic development include whether a change in residential real estate investment affects economic development of one region or also affects other neighboring regions. The main objective of this research is to estimate the impacts of residential real estate investment on the economic development of the Northeast region of the United States using a simultaneous equations 3SLS regression, and a spatial Durbin model with a spatial panel data set.

This research analyzes the relationship between residential real estate investment and economic development represented by changes in population, employment, median income, and median housing value. The interdependency among the explanatory variables is estimated by employing a system of simultaneous equations. County-level data from the U.S. Census Bureau for the period of 1980 to 2010 are used in this non-spatial model. Empirical results are expected to show whether changes in residential real estate investment can be used as a leading indicator to forecast changes in population, employment, or income in the county.

Spatial dependence is an important factor in regional economic development analysis, especially in terms of population, employment, and median income. Location is an inherent part of residential real estate, even at the county level, real estate exhibits spatial dependence. Counties that are neighbors are more alike than counties that are spatially far apart. Because of the potential for spatial dependence, this study also uses a spatial panel method to analyze the relationship between residential real estate investment and economic development represented by changes in population, employment, and median income by including spatial dependency.

This research contributes to the housing literature by investigating the simultaneous interaction of population, employment, income, and residential real estate investment decisions. This research also extends existing studies by utilizing a panel spatial Durbin model to investigate the spatial autocorrelation associated with real estate investment decisions. Spatial autocorrelation is shown to exist, and future regional development policies must account for the development policies of neighboring locations. To stimulate regional economic development, policy makers may need to have accumulated information available to ascertain whether they should pursue policies to influence the location and utility decisions of firms or people.

Acknowledgments

I would like to acknowledge and express my heartfelt gratitude and thanks to all those people whose help, support, and guidance motivated me towards the successful completion of my dissertation. I would like to thank my advisor, Dr. Tesfa Gebremedhin, for his guidance, support, constructive comments and suggestions, being available to meet with me at all times, motivation towards the successful completion of my dissertation, and to my academic life at West Virginia University. He has been a great source of confidence for me, and I am extremely grateful and fortunate to have him as my major advisor. It was a wonderful time and a great pleasure working with him.

I thank every committee member whose suggestions, comments and corrections helped me towards the successful completion of my dissertation. I would like to extend my heartfelt thanks and gratitude to Dr. Peter Schaeffer, Dr. Timothy Phipps, Dr. Donald Lacombe, Dr. Santiago Pinto and Dr. Wesley Burnett for giving constructive comments, encouragement, and guidance.

I am also thankful to Mrs. Lisa Lewis, Mrs. Melanie Jimmy, and Mrs. Ellen Hartley-Smith for their administrative assistance and extending help whenever I needed and for their care, encouragement, and support.

I would also like to acknowledge and extend my deepest gratitude to Dr. Mark Sperow and Dr. Jerald Fletcher for providing graduate assistantships through the projects from the National Energy Technology Laboratory (NETL), Department of Energy projects for the first four years of my academic career which helped me with the successful completion of my course work. I would like to thank Dr. Jay Malarcher and Dr. Elizabeth Oppe for giving me a graduate assistantship for two years through Summit Hall and Arnold Hall, WVU. I also extend my heartfelt gratitude and thanks to Dr. Bernadette Jungblut for giving me a graduate assistantship towards the end of my graduate program. Her constant help, motivation, and funding helped me to successfully complete my dissertation.

I would like to extend my heartfelt thanks to my friend Marc Mulkey and his mother Mrs. Daneva Mulkey for editing my dissertation and helping me improve my writing.

Special thanks to my family, my cute niece and nephews Surekha, Srijan, Monish, Tanush and Husna for making my life beautiful. I would also like to thank my colleagues, faculty, staff, and friends with a special mention of Sujitha, Ricky, Yeshwanth, and Lord Ganesha for all their support and encouragement.

Dedication

I dedicate this work to my grandparents Mr. Adinarayana Reddy and Mrs. Rajamma Adinarayana Reddy; Mr. Ramalingam Reddy and Mrs. Renukambal Ramalingam Reddy, and my great grandmother Mrs. Ramalakshammam; my parents Mr. L.A. Jayarama Reddy and Mrs. Kalavathy Jayaraman; my uncles and aunts Mr. Radhakrishnan Reddy and Mrs. Anusuya Radhakrishnan, Mr. D. Narendran and Mrs. Sarojini Narendran, Mr. Mohana Sundaram and Mrs. Sumathi Monahasundaram; my brothers, sisters, and sisters-in-law Mr. Dinakar Babu, Mrs. Shanthi Dinakar Babu, Mr. Prem Kumar, Mrs. Dakshayani Prem Kumar, Mr. B. Narendran, Mrs. Premalatha Narendran, Mr. Sanjay Rajendran, Mrs. Saranya Narendran, and Mr. Lokesh Narendran for their immense love, support, and trust in me.

Table of Contents

Abstract	i
Acknowledgments.....	ii
Dedication.....	iii
Table of Contents	iv
List of Tables	vi
List of Figures.....	vii
Chapter 1.0 Introduction.....	1
1.1 Introduction and Problem Statement.....	1
1.2 Objectives.....	5
1.3 Study Area	5
1.4 Organization.....	13
Chapter 2.0 Literature Review	14
Introduction	14
2.1 Defining Housing and Residential Real Estate Investment	14
2.2 Relationship between Residential Real Estate Investment and Economic Development	17
2.3 Residential Real Estate Investment and Economic Development Models	24
2.3.1 Simultaneous Equations Models.....	24
2.3.2 Spatial Econometric Models	26
2.4 Summary of Past Models and Shortcoming of Past Studies	28
Chapter 3.0 Theoretical Model	30
Introduction	30
3.1 Residential Real Estate Location Decisions.....	30
3.2 Specification of Variables.....	34
3.3 Data Types and Sources	36
Chapter 4.0 Methodology	37
Introduction	37
4.1 Simultaneous Equations 3SLS Model.....	37
4.2 Spatial Model	41

4.2.1 Moran's I	41
4.2.2 Spatial Panel Model	43
Chapter 5.0 Empirical Analysis and Results	48
Introduction	48
5.1 Simultaneous Equations Model Results.....	48
5.1.1 Population Equation	49
5.1.2 Employment Equation	51
5.1.3 Income Equation	53
5.1.4 Residential Real Estate Equation	55
5.2 Spatial Panel Results	58
5.2.1 Spatial Durbin Model Results.....	58
Chapter 6.0 Summary and Conclusions	66
6.1 3SLS Simultaneous Equations Model Conclusions.....	67
6.2 Spatial Panel Model Conclusions	68
6.3 Policy Implications	69
6.4 Limitations and Future Work	70
References	73
Appendix	78

List of Tables

<i>Table 1: Definition of Variables</i>	36
<i>Table 2: Three Stage Least Square Results of Population Equation</i>	51
<i>Table 3: Three Stage Least Square Results of Employment Equation</i>	53
<i>Table 4: Three Stage Least Square Results of Income Equation</i>	55
<i>Table 5: Three Stage Least Square Results of Residential Real Estate Investment Equation</i>	57
<i>Table 6: Spatial Durbin Model Spatial Panel Results</i>	65

List of Figures

<i>Figure 1: Study Area</i>	6
<i>Figure 2: Population in the Study Area (Data Source: U.S. Census, 2013)</i>	9
<i>Figure 3: Employment in the Study Area (Data Source: U.S. Census, 2013)</i>	9
<i>Figure 4: Urban Population in the Study Area (Data Source: U.S. Census, 2013)</i>	10
<i>Figure 5: Rural Population in the Study Area (Data Source: U.S. Census, 2013)</i>	10
<i>Figure 6: Building Permits in the Study Area (Data Source: U.S. Census, 2013)</i>	11
<i>Figure 8: Occupied Houses in the Study Area (Data Source: U.S. Census, 2013)</i>	12
<i>Figure 9. Median Housing Value per County 1980 (Data Source: U.S. Census, 1980)</i>	79
<i>Figure 10. Median Housing Value per County 1990 (Data Source: U.S. Census, 1990)</i>	80
<i>Figure 11. Median Housing Value per County 2000 (Data Source: U.S. Census, 2000)</i>	81
<i>Figure 12. Median Housing Value per County 2010 (Data Source: U.S. Census, 2010)</i>	82
<i>Figure 13. Population per County 1980 (Data Source: U.S. Census, 1980)</i>	83
<i>Figure 14. Population per County 1990 (Data Source: U.S. Census, 1990)</i>	84
<i>Figure 15. Population per County 2000 (Data Source: U.S. Census, 2000)</i>	85
<i>Figure 16. Population per County 2010 (Data Source: U.S. Census, 2010)</i>	86
<i>Figure 17. Employment per County 1980 (Data Source: U.S. Census, 1980)</i>	87
<i>Figure 18. Employment per County 1990 (Data Source: U.S. Census, 1990)</i>	88
<i>Figure 19. Employment per County 2000 (Data Source: U.S. Census, 2000)</i>	89
<i>Figure 20. Employment per County 2010 (Data Source: U.S. Census, 2010)</i>	90
<i>Figure 21. Median Income per County 1980 (Data Source: U.S. Census, 1980)</i>	91
<i>Figure 22. Median Income per County 1990 (Data Source: U.S. Census, 1990)</i>	92
<i>Figure 23. Median Income per County 2000 (Data Source: U.S. Census, 2000)</i>	93
<i>Figure 24. Median Income per County 2010 (Data Source: U.S. Census, 2010)</i>	94
<i>Figure 25. Median Housing Value per county Moran's I test result.</i>	95
<i>Figure 26. Population per county Moran's I test result.</i>	96
<i>Figure 27. Employment per county Moran's I test result.</i>	97
<i>Figure 28. Median Income per county Moran's I test result.</i>	98

Chapter 1.0 Introduction

1.1 Introduction and Problem Statement

Residential real estate investment has been recognized as an important agent of economic development since the 1970s because residential real estate investment is a major economic activity with large multiplier effects. Residential real estate improvement is also linked to many external social and economic benefits. Previous studies have examined the role of residential real estate in economic development through various approaches, including the effects of employment and income (Leung, 2004), household saving (Turner and Luea, 2009), labor productivity (Ofori and Han, 2003), health productivity and growth from real estate investment (Arku and Harris, 2005), as well as home ownership effects (Carruthers and Mulligan, 2005); (Carruthers and Mulligan, 2008)).

In the United States, real estate is an important investment for individual investors. In the Census Bureau's Survey of Income and Program Participation (SIPP) (Census, 2000), residential real estate was the largest class of assets held by individuals, amounting to 78.7% of total household asset value. Of this, homes represented 67.2%, rental properties 4.9%, and other real estate such as vacation homes and land holdings 6.6% of total investment portfolios. By contrast, the value of commercial real estate has decreased by 40% since 2007 in the United States (Census, 2000).

Wealth is a major factor in household financial security in the U.S. Wealth increases occur in times of economic prosperity when households tend to grow their assets and decreases during economic hardships as households draw upon these accumulated financial resources. The

contribution of wealth as a major financial security in the households in the U.S. is evidenced by the pattern observed for median household net worth during the period 2000-2011. Median household net worth showed a 30% increase from 2000-2005 and then showed a decrease of 35% by 2011, consequentially showing an overall decrease of 16% between 2000 and 2011 (Gottschalck, Vornovytskyy and Smith, 2013). The Northeast region showed a significant increase and decrease of median household net worth during 2000-2011, with a 7% decrease between 2000 and 2005 and a 26% decrease between 2005 and 2011 (Gottschalck, Vornovytskyy and Smith, 2013). While there was no statistically significant change exhibited by median net worth excluding home equity during 2000-2005, it showed a significant decrease during 2005-2011 (Gottschalck, Vornovytskyy and Smith, 2013). Therefore, the changes in overall median net worth observed over the past decade have been driven primarily by one of the major components of wealth, that is the household assets held by the American people (Gottschalck, Vornovytskyy and Smith, 2013).

According to Deloitte LLP (Deloitte, 2009), loss of jobs and reductions in consumer spending negatively affected all types of real estate investment, in general, and office and retail properties in particular. Rental rates and real estate prices decreased due to high vacancy rates of properties. However, in 2010 a potential recovery in economic growth of the country was leading to increases in property values again.

Some important issues associated with real estate; population, income, cost, quality, and affordability of real estate, all influence residential real estate prices. According to the U.S. Census Bureau (Census, 2011), the population of the Northeast region is approximately 73 million, equal to 23.4% of the U.S. population. The population in urban areas in the region increased by 18%

from 1980 to 2000. At the same time the rural population decreased by 18%. One possible reason for the simultaneous increase of the urban and decrease of the rural population could be due to the migration of rural population to urban areas for employment.

After the national recession in the early 1980s, the Northeast region recovered rapidly according to the FDIC (2010). In this region, commercial and residential real estate markets grew quickly due to strong regional employment and economic growth during 1982-1988. However in the late 1980s, economic growth in the region declined due to a decrease in employment and slow personal income growth, and overbuilt real estate markets intensified the effects. Residential real estate costs and quality of life issues continued to be a problem for low-income populations especially in rural areas. More than 42% of unassisted low-income renters had severe residential problems paying for housing in the region during the same year. In spite of the problem of acquiring affordable quality housing and available credit in rural areas, ownership of real estate is one of the best methods of asset accumulation for low-income rural households.

Carlino and Mills' (1987) simultaneous equations model estimated employment and population changes in U.S. counties and explained the migration patterns in U.S. areas with high family incomes that generate relatively higher demand for goods and services leading to higher levels of service and commercial employment but lower levels of manufacturing employment. Lower levels of manufacturing employment were likely influenced by relatively higher land prices in areas with high family income and potentially more expensive residential real estate (Carlino and Mills, 1987).

This research will further attempt to delve into the question of whether “people follow jobs” or “jobs follow people” (Carlino and Mills, 1987). Policy makers need information to stimulate regional economic growth, especially if they should pursue policies to influence the location and utility decisions of firms or of people. In conjunction with information about population and income, this research plans to identify a relationship between residential real estate as a measure of the value people place on living in a location and employment as a measure of how firms value a location.

Regional geographic variation within the U.S. influences population location decisions (Carlino and Mills, 1987). Counties in the Sunbelt region were more attractive than counties in colder regions of the U.S. Interregional differences within the U.S. were important, while intraregional differences were less important. Accordingly, there were large differences between counties located in different regions of the country, but only small differences between counties in the same region. Counties in the Northeast were statistically different from counties in the South and other regions, but counties within the Northeast region were relatively similar to each other. If counties outside the Northeast were compared to counties in the Northeast, bias reflecting amenities like climate would affect the results. This study will not directly consider the effects of amenities on housing location decisions. In order to limit the introduction of bias reflecting amenities into this study, only counties in HUD Regions I, II, and III will be utilized, as they have been shown to be statistically similar to each other.

State policy makers and local leaders need to have a full grasp of the relationship between investments in residential real estate and economic development. It will be hard to design and

follow proper state or local policy for economic development without a better understanding of what types of economic sectors and development programs are most appropriate in attracting businesses, alleviating poverty, and influencing economic development. Also, own county policy development requires knowledge about the impact of neighboring counties' influence on economic development.

1.2 Objectives

The overall objective of this study is to provide policy makers with information on the relationship between residential real estate investment and economic growth in the Northeast of the U.S. region. The specific objectives are to:

1. Identify the spatial distribution of residential real estate investment showing development patterns across the study's time period.
2. Identify and estimate the impacts of residential real estate investment on the economic development of the Northeast region.
3. Based on the research findings, draw policy implications for the economic development of the Northeast region of the U.S.

1.3 Study Area

This study focuses on the counties of the Northeast region and the District of Columbia of the United States for the census years between 1980 and 2010; the region is shown in Figure 1. County level data was collected for endogenous and exogenous variables from the U.S. Census Bureau, Censtats Databases and the U.S. Department of Agriculture, Economic Research Service. The study area coincides with the definition of the U.S. Department of Housing and Urban Development (HUD) Regions I, II and III.

The study area is chosen because this region is economically diverse and well developed. It is the nation's most densely populated region. Additionally, this region also shows cultural

diversity and it is also one of the most urbanized regions of the United States with 59.6% of its 434 counties considered metro areas (USDA-ERS, 2013).

The study area comprises 434 counties of HUD regions I, II, and III, which include Connecticut, Delaware, Maryland, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Virginia, District of Columbia., and West Virginia.

Figure 1: Study Area

According to the U.S. Census Bureau and Bureau of Labor Statistics, the Northeast region has diverse urban characteristics and displays spatial variations in economic growth. It also has resources and economic opportunities to enhance regional economic development. In order to develop further, policy makers need to introduce appropriate policies to improve the Northeastern environment for population growth, employment growth, income growth, and residential real estate investment.

Economic conditions in all three HUD regions have made significant progress during 2013. The New England Region I has seen a significant economic improvement led by job gains in Massachusetts, which accounted for 48% of the total nonfarm jobs and 59% of growth in the region (HUD, 2013). During 2013, every sector in New England Region I, except the manufacturing and financial sectors, gained jobs. Except for Vermont, every state in the region showed an increase in sales of single-family homes and condominiums. Also, the median sales prices and the multifamily building activity in the region showed a significant increase from the year 2012, which made this region achieve the highest growth rate compared to other regions (HUD, 2013)

The U.S. Department of Housing and Urban Development (2013) states that home sales and median home sale prices increased through most of the New England Region I. Home sales in the New England Region I increased by 4.2% during August 2013 when compared to August 2012 (HUD, 2013). Median single-family home sales prices also increased in all the states except Vermont, where they decreased by 2.3% (HUD, 2013).

Massachusetts and New Hampshire showed the highest growth in permits for single-family housing for the period 2012 to 2013 in the region with 32% and 26% increases respectively. Rhode Island showed an increase of 10%, Connecticut 7%, and Maine 4%. During the same period Vermont saw a decrease of 6% in single-family home permits (HUD, 2013).

HUD Region II (New York/New Jersey) has been showing continuous improvement in its economic conditions. The economy of the New York/New Jersey region expanded in 2013 with the region's nonfarm payrolls showing a 1.5% job growth; the education and health sectors increased by 2.1%, and the professional and business service sector grew by 2.5% (HUD, 2013).

According to a Housing and Urban Development (HUD, 2013) report, the housing market in the New York/New Jersey region improved with an increase in home sales and prices. A total of 4,600 single-family homes permits were issued, a 30% increase over the same period in 2012. New Jersey showed a 47% increase of 47% of building permits for single-family homes. In total, the New York/ New Jersey region accounted for about 80% of the increase in homebuilding activity in the study region (Census, 2013).

Economic conditions in HUD Region III, the Mid-Atlantic region, have improved throughout the region since 2010, with an increase in job growth, home sales volume and sales prices (HUD, 2013). Maryland had the highest growth in nonfarm payrolls (1.8%), accounting for one third of the job gains in the Mid-Atlantic region. The professional and business services sector and the education and health services sector showed 3.7% and 3.2% job growth. Virginia showed a 1% nonfarm payroll growth, while Pennsylvania, Delaware, West Virginia and District of Columbia showed significant increase in job growth (Henderson, 2012).

According to a U.S. Census Bureau (2013) report, home sales prices increased throughout the Mid-Atlantic region with increases ranging from 2.4% in Virginia to 8.6% in the District of Columbia. Single-family construction activity measured by housing permits issued increased throughout the Mid-Atlantic (by 19%) region compared to 2012. Two-thirds of these permits were issued in Virginia and Pennsylvania. Maryland and Delaware had a 15% and 25% increase, respectively, while the District of Columbia added 50% more permits and West Virginia 20% (Census, 2013).

The graphs following below compare the Northeast region with the entire U.S. for the four decennial census years, using population growth, job growth, urban and rural split, single-family housing permits, and number of vacant and occupied houses (single-family).

Figure 2: Population in the Study Area (Data Source: U.S. Census, 2013)

Note: Percentages are absolute, and represent the total contribution of the Northeast Region towards the U.S. total population.

Figure 3: Employment in the Study Area (Data Source: U.S. Census, 2013)

Note: Percentages are absolute, and represent the total contribution of the Northeast Region towards the U.S. total employment.

Figures 2 and 3 show the comparison between the Northeast region and the entire U.S. with respect to population and employment. From both figures it is clear that the Northeast region accounts

for about one-fourth of total U.S. population and also one-fourth of the total job growth in the U.S. There was significant job growth between 1980-2000 and though there was an increase in the number of jobs added during 2000-2010, the entire country showed very little employment growth due to the loss of 3.2 million jobs during the recession that started in December 2007 (Henderson, 2012).

Figure 4: Urban Population in the Study Area (Data Source: U.S. Census, 2013)

Note: Percentages are absolute, and represent the total contribution of the Northeast Region towards the U.S. total urban population.

Figure 5: Rural Population in the Study Area (Data Source: U.S. Census, 2013)

Note: Percentages are absolute, and represent the total contribution of the Northeast Region towards the U.S. total rural population.

Figures 4 and 5 show the comparison between urban and rural population in the Northeast region and the U.S. The urban population of the Northeast region contributes to about 25% of the whole U.S. population and there is a constant increase in the urban population from 1980-2010. This is because of the faster growth of metro areas during 1980-2000. Though the contribution of suburb, exurb and outer suburb growth during 2000 slowed urban population growth, data show an increase in urban population as a whole. This is because the classification is based on the U.S. Census Bureau's Urban and Rural classification for the census years 1980, 1990, 2000 and 2010 which includes all these suburban, exurban and outer suburban areas (Frey, 2012). On the other hand the population in the Northeast region showed a decline in the rural population growth rate from 3.2% during 1980-1990 to 0.4% during 2000-2010.

Figure 6: Building Permits in the Study Area (Data Source: U.S. Census, 2013)

Note: Percentages are absolute, and represent the total contribution of the Northeast Region towards the U.S. total building permits.

Figure 6 shows the total number of single-family housing permits for the Northeast region and the entire U.S., respectively. Permits issued in the Northeast region reflect the business cycle of the entire country during these four periods. The total number of permits increased from 1980 to 2000 and decreased in 2010, but permits issued in the Northeast during 2010 constitute a

much larger (26.97%) share than the U.S. total in previous years because of the relative strong recovery of the housing market in this region (HUD, 2013).

Figure 7: Vacant Houses in the Study Area (Data Source: U.S. Census, 2013)

Note: Percentages are absolute, and represent the total contribution of the Northeast Region towards the U.S. total vacant houses.

Figure 7: Occupied Houses in the Study Area (Data Source: U.S. Census, 2013)

Note: Percentages are absolute, and represent the total contribution of the Northeast Region towards the U.S. total occupied houses.

Figure 7 shows the vacant housing stock in the Northeast region and the U.S. as a whole for the decennial census years. Vacant single family housing increased in all years in the study

period except 2000, when it experienced a slight decrease. The proportion of vacant houses in the Northeast to the rest of the U.S. falls throughout the time period.

Figure 8 shows the occupied single family housing units in the Northeast and in the U.S. as a whole for the decennial census years. The number of occupied single family housing units increased in all years in the study period. As with vacant housing, occupied housing in the Northeast, though increasing in numbers, constitutes a declining share of the occupied housing stock for the entire U.S.

The change in spatial patterns over time in median housing values, population, employment, and median income at the county level is shown in figures 9-24 in the Appendix. The spatial pattern of changes is interesting and shows that certain areas within the study region tended to concentrate economic development gains, while other areas tended to experience less economic development, and by some measures even economic decline.

1.4 Organization

This dissertation has eight chapters. Chapter 1 is an introduction that relates the problem statement, study area, and objectives. Chapter 2 presents the review of previous research concerning residential real estate and economic development, simultaneous equations methods, and spatial panel methods. Chapter 3 provides the theoretical foundation for this dissertation, as well as the specification of variables and data used. Chapter 4 discusses the method of analysis consisting the simultaneous equations model and the spatial panel models. The empirical results and their analysis and interpretation constitute Chapter 5, and Chapter 6 concludes this

dissertation with a summary of findings, conclusions, a discussion of policy implications, and areas for future research. Chapter 7 provides references. Chapter 8 is an appendix.

Chapter 2.0 Literature Review

Introduction

This chapter provides a review of the most relevant studies on the relationship between residential real estate investment and economic development. The chapter is organized into four sections. The first section provides a review about the definition of housing and residential real estate investment. The second section provides a review about residential real estate investment and economic development. The third section reviews estimation models and their use in providing insight into the relationship between residential real estate investment and economic development. The last section summarizes past models, their shortcomings, and discusses the contribution of this research.

2.1 Defining Housing and Residential Real Estate Investment

Investing in residential real estate is both a consumption and a major investment decision (Plaut, 1987). Henderson and Ioannides (1979) designed a model of tenure choice where housing serves two purposes, a consumption good and an investment holding in a portfolio. A consumer simultaneously chooses the optimal level of housing consumption and optimal portfolio holdings. When consumption demand is less than investment demand, the consumer owner-occupies that portion of investment equaling consumption demand and rents out the remainder. If

consumption demand is greater than investment demand, the consumer cannot own part of his consumption and must rent (Henderson and Ioannides, 1979).

For middle-income households in particular, home ownership is a very one-sided investment portfolio, with substantial financial risks (Forrest and Murie, 1989). Housing remains the largest asset investment of most American families. Some studies have shown that owning a house often put households in a favorable position relative to those who remained tenants for life (Badcock, 1989, Hamnett, 1991, Kendig, 1984). Government subsidies encourage home ownership, but not everyone is in a financial position to take advantage.

Engelhardt (1995) examines how changes in housing values cause homeowners to alter their savings and consumption behavior. If residential real estate is an investment, then an increase in the value of that investment should cause owners to alter their consumption and saving behavior, likewise with a decrease. Using Panel Study of Income Dynamics (PSID) data from 1984 and 1989, Engelhardt (1995) showed that owners who experienced gains in the value of their investment in residential real estate did not tend to save more over time, while owners whose residential real estate investments lost value tended to save more income in an attempt to compensate. This behavior demonstrates that owners realize residential real estate investment is a primary savings vehicle for a large section of the U.S. population, and that when this investment decreases in value owners tend to increase savings.

An alternative view in economics is that housing prices are driven primarily by construction costs. This view was laid out by Grebler, Blank and Winnick (1956). This model considers that people did not view housing as a speculative asset: almost all of the value of

houses has been value of structure, a manufactured product. From this alternative model, buying a product and holding it for resale in order to make money makes no sense whether the product is a house or tables and chairs (Schiller, 2007).

Gyourko et al. (2006) argue that great cities will indefinitely outperform the economy in general. They found that some “superstar cities” have shown long-term, that is 50-year, appreciation above national averages (Gyourko, Mayer and Sinai, 2006). They used Census decadal owners’ evaluations of the value of their homes, but found only relatively small excess returns to homes in those cities. They reported much smaller differences across cities than people expect. Their paper found that Los Angeles grew at 2.46% a year from 1950-2000, but this is far below the kind of expectations we have seen recently (Gyourko, Mayer and Sinai, 2006). Moreover, in the decennial Census data there is no correction for quality change, and yet homes have been getting larger in the superstar cities, so the actual appreciation of existing homes was likely less (Schiller, 2007).

DiPasquale, Forslid, and Glaeser (2000) have found that homeowners tend to be more involved in local government, are more informed about their political leaders, and join more organizations than renters. This view has led to widespread political support for policies that encourage homeownership over much of the world, including the mortgage interest deduction in the U.S.

Alternately, there are many sensible reasons for people to rent rather than own. Some people who cannot currently bear the responsibilities of household management, who are likely to move soon, or who have other plans for their time, should rent rather than own. Renting rather than owning encourages a better diversification of investments; many homeowners have very

undiversified investment portfolios, and these investments are often highly leveraged (Schiller, 2007). Creating too much attention to housing as investment may encourage speculative thinking, and therefore, excessive volatility in the market for homes. Encouraging people into risky investments in housing may have bad outcomes (Schiller, 2007).

The physical nature of land and houses as forms of capital requires a different treatment than financial or human capital. This leads Mayer and Somerville (2000) to examine the effect of housing construction in the general economy of a region and how new housing construction often leads to both recessions and recoveries. Land and housing is a physical capital, and is not as mobile as other forms of capital. When sold, the capital still occupies the same location. Also, housing capital physically depreciates and can be removed from the market.

2.2 Relationship between Residential Real Estate Investment and Economic Development

According to Shaffer et al., (2006) economic growth and economic development are frequently used interchangeably. Though the concepts are related, they are different. Economic growth is the growth in employment, income, and resources of production. Economic development goes beyond economic growth to incorporate the institutional and structural change in the capacity to act, innovate, and move forward in all aspects of life. Therefore “economic growth can occur without development, and development can occur without growth” (Shaffer, Deller and Marcouiller, 2006).

An often discussed idea in the housing literature is the assertion that a long-run equilibrium relationship exists between house prices and fundamentals, such as income, population and costs associated with home ownership. The validity of this assumption has

important implications for how residential real estate dynamics are modeled. If the assumption is accurate, so that residential real estate prices are cointegrated with fundamentals, then the error-correction specifications of Abraham and Hendershott (1996), Malpezzi (1999) and Capozza et al. (2002) are appropriate. Gallin (2006) used standard tests to show that there is little evidence for cointegration of house prices and various fundamentals at the national level. In addition, Gallin (2006) showed that bootstrapped versions of more powerful panel-data tests, applied to a panel of 95 U.S. metropolitan areas over 23 years, also do not find evidence for cointegration. This shows that the level of residential real estate investment does not appear to have a stable long run equilibrium relationship with the level of fundamentals such as income or population.

Whether house prices are cointegrated with economic fundamentals such as income or employment was investigated by Zhou (2010). The existence of cointegration would imply that a reduction in the price of housing was caused by market forces. More specifically, cointegration means that there is a long run equilibrium relationship between residential real estate investment and economic fundamentals (Zhou, 2010). Conversely, failure to find cointegration suggests that a reduction in the price of housing would be random. Zhou (2010) selected ten cities (Boston, Chicago, Cleveland, Dallas, Los Angeles, New York, Philadelphia, Richmond, Seattle and St. Louis), and tested first for linear cointegration, then for nonlinear cointegration. Cleveland showed evidence of linear cointegration. Six of the nine remaining cities (Chicago, Dallas, Philadelphia, Richmond, Seattle and St. Louis) and the country as a whole showed evidence of nonlinear cointegration leaving three cities (Boston, Los Angeles and New York) with no evidence of linear or nonlinear cointegration (Zhou, 2010). These results show that the fundamentals of

residential real estate investment are diverse at the city level, reflecting the local nature of the housing market. Local conditions including employment and income affect residential real estate investment decisions.

Using U.S. post-war data for empirical analysis, Wen (2001) investigated the relationship between economic growth and fixed capital formation. The results showed that differentiating residential investment from business investment is important in analyzing the relationship between capital formation and economic growth. Most household savings are in the form of real estate, and economic booms often follow real estate booms and economic recessions follow real estate declines. He also showed that capital formation in the household sector unambiguously and unilaterally affects the growth of GDP, which in turn affects capital formation in the business sector.

Housing demand and population growth are also interconnected (Frame, 2008). Economic theory suggests that an increase in population should increase demand for housing, which can be measured as increases in the value of residential real estate. Although strong evidence of endogeneity problems hampered his research, Frame (2008) showed that changes in median aggregate income have a negative relationship with returns to housing and local population growth. This result is important because it illustrates the interaction between population and residential real estate investment and shows that the relationship is more complicated than traditional microeconomic theory suggests.

Using an error correction model, Johnes and Hyclak (1999) investigated potentially significant interactions between housing and labor markets to explain the average wage, unemployment rate, labor force, and average house prices in urban areas, using quarterly data

from four cities, (Fort Lauderdale, FL, Hartford, CT, Houston, TX, and Milwaukee, WI) for the 1980s. Unemployment and labor force changes affect house prices which have a significant effect on the size of the labor force. Labor markets are important both to potential jobseekers and employers. A deeper labor pool increases the probability that employers find qualified workers and increases the probability that job seekers become employed. The attractiveness of the labor market can be expected to be capitalized into housing prices.

Saks (2008) argues that the large variation in house prices across the U.S. is due to the differences in the supply of housing. The elasticity of housing supply impacts local labor markets due to the influence of house prices on migration. Because of this influence on migration, local employment depends critically on the construction industry to accommodate increases in housing demand. The elasticity of housing supply affects the increase in labor demand on employment, wages and house prices. These predictions are assessed empirically by tracing out the effect of an increase in labor demand on metropolitan area housing and labor markets (Saks, 2008). A three variable Vector Auto-Regression (VAR) is used to estimate the shocks on the dynamics of housing prices, employment and wages. The results showed that locations with higher housing regulations led to lower residential construction and larger increase in housing prices in response to labor demand (Saks, 2008). Also, in the long run an increase in labor demand results in lower employment in metropolitan areas, resulting in lower elasticity of housing supply. These results showed that interactions between local labor markets and the housing supply play an important role in the employment development of a particular region.

Another approach to explain the changes in migration, housing market and labor market in response to employment shocks was examined by Zabel (2012). He explains that the

households' decision to move not only depends on job prospects but also on the relative cost of housing. Household mobility in turn is significantly affected by the housing market as the ability to move depends on the relative cost of housing across metropolitan statistical areas. The housing market affects households through two channels. First, homeowners are less likely to move than renters due to their higher mobility cost as selling a home and buying a new one takes time and financial resources. Secondly, the relative cost of housing across markets and the likelihood that a household would move from one MSA to another depends on the relative cost of residing in each city.

Quigley (1999) evaluates the relationship between U.S. economic cycles and housing prices. He indicates that few studies have explored the causal role between outcomes in the real estate market and the overall economy. Quigley blames this lack of research on the lack of an empirical model. Without an empirical model to test the relationship between real estate outcomes and economic development, any theoretical relationship remains unverified. This research will attempt to fill in that particular gap, utilizing an empirical model to test the relationship that Quigley identified.

Home ownership may be associated with higher rates of unemployment (Oswald, 1996). The theory advocates that home ownership reduces labor mobility as the costs associated with the sale of physical housing increases costs related to relocation in search of new jobs. This creates labor market inefficiency and higher unemployment rates. Regional unemployment may also be affected by homeownership (Coulson, 2009, Munch, Rosholm and Svarer, 2006).

A decline in housing values may also attribute to reduced geographic mobility (Chan, 2001, Engelhardt, 2003, Ferreira, Gyourko and Tracy, 2010). Valleta (2013) investigates if this

reduced mobility contributed to higher unemployment. Using a maximum likelihood estimation technique, he finds that homeowners in areas with large price declines had longer periods of unemployment than those in areas with more limited price declines, but this experience is also applied to renters who lived in the same areas (Valleta, 2013). This result is supported by other research that also discounts spatial house lock caused by declines in housing values (Donovan and Schnure, 2011, Farber, 2012, Modestino and Dennett, 2013, Schmitt and Warner, 2011).

Battacharya and Kim (2010) use panel data for 20 MSA's in the U.S. to show that employment, real construction cost and the real user cost of housing have a significant impact on real housing prices. Over the 2003–2005 period when subprime lending soared, all 20 MSAs experienced a rapid appreciation in real price (Battacharya and Kim, 2010). They employed a panel fixed effects model and found that there was no long run equilibrium relationship between housing prices and economic fundamentals such as employment, or mortgage rates. Local policy makers should beware of a sudden surge in housing prices that is not sustained by economic fundamentals, especially since the magnitude and effect of such shocks can vary considerably across regions (Battacharya and Kim, 2010).

The economic stability of a household is important for residential real estate investment, as are household income level and income expectations (Goodman, 1988, Haurin and Gill, 1987, Rosenthal, 1989). Homeownership is positively related to income, and the greater the present proportion of income versus future income expectations, the more people move from renting to owning a house (Henderson and Ioannides, 1987). Clark et al. (1994) showed that many first-time buyers have recently experienced a positive income change. Permanency of income is also an important aspect of the decision to buy a house and housing economists have tried to estimate

this part of the household income in models of housing and tenure consumption (Dieleman and Everaers, 1994).

Changes in population and in tastes and preferences also affect residential real estate investment decisions (Green, 1997). Beyond simple population increases or decreases affecting the demand for housing, the population composition is also important. Green (1997) used 1980 and 1990 census data to estimate changes in housing affordability, household type, and tenure choices by household type. The actual results were compared with estimates to show that while housing did become less affordable during his study's time period, changes in household type were also responsible for changes in consumers' residential real estate investment decisions. During the study period, the number of households headed by single and divorced persons increased. These household types were less likely to invest in housing, partly because of affordability and partly because of labor mobility concerns (Green, 1997).

Expectations and uncertainty about income potential can also affect residential real estate investment choices and housing tenure. Robst et al. (Robst, Deitz and McGoldbrick, 1999) used data from the University of Michigan's Panel Study of Income Dynamics to estimate the effects of income uncertainty and its effect on housing decisions. The results showed that homeowners are less uncertain about future earnings, and that renters and owners should therefore be treated as separate groups (Robst, Deitz and McGoldbrick, 1999). Whites are significantly more likely to own than rent, and men are more likely to own than rent (Robst, Deitz and McGoldbrick, 1999). Income uncertainty reduced residential real estate investments for all groups. The greater the income uncertainty, the greater is the reduction in investment.

2.3 Residential Real Estate Investment and Economic Development Models

2.3.1 Simultaneous Equations Models

Carlino and Mills' (1987) simultaneous two-equations model estimated employment and population changes in U.S. counties and explained the migration patterns in the U.S. Additionally, a simultaneous equations approach is used to investigate the impact of population, employment and median income on median housing value. The authors found that areas with high family incomes have relatively higher demand for goods and services, leading to higher levels of service and commercial employment, but lower levels of manufacturing employment. Lower levels of manufacturing employment were likely influenced by relatively higher land prices in areas with high family income and potentially more expensive residential real estate (Carlino and Mills, 1987).

Using a three-equation model, Deller et al. (2001) expanded the work of Carlino and Mills (1987) by adding non-market amenities like natural resources to the analysis. They found that rural areas in the U.S. can capitalize on resource endowments. By effectively managing natural resources, rural areas can promote economic development. Recreational use of natural resource amenities also provided economic benefits (Deller, Tsai, Marcouiller and English, 2001).

Deller and Lledo (2007) estimate population, employment, and income in their simultaneous equations model in attempt to better analyze amenities' role in economic development for rural regions in the U.S. The authors' goals included: expanding the Carlino-Mills model to capture greater dimensions of growth than what is commonly found in the amenity-growth literature, a statistical approach to modeling amenities lends a greater degree of objectivity to what is commonly found in the amenity-growth literature, and investigating

whether the U.S. is sufficiently heterogeneous that impact of amenities or other policy variables may be significantly different depending on location (Deller and Lledo, 2007).

The empirical model of Deller et al. (2001) has been extended to estimate simultaneous relationships of economic development with entrepreneurship (Bashir, 2011, Mojica, 2009), amenities (Kahsai, 2009), environmental regulation (Nondo, 2009), and has been used to model small business growth, migration behavior, local public services and median household income (Gebremeriam, 2006).

A system of simultaneous equations with random effects was used in Robbins et al., (2000) to analyze the relationship between the proportion of small businesses and four determinants of economic growth: productivity, gross state product (GSP), unemployment, and wage inflation at the state level in the U.S. using panel data from 48 states from 1986 to 1995. The study showed that very small businesses provided economic benefits at a macro level. They concluded that as the number of small businesses (20 employees or less) increased, the level of productivity and GSP growth increased at a state level. At the same wages, inflation and unemployment rates were reduced. Therefore, macroeconomic policies were more beneficial to the states that were rich in small businesses.

Previous studies have used regional simultaneous equation models to address the relationship between amenities and economic growth. Rudzitis (1999) examined growth in and around counties with federally designated wilderness and found that employment did not explain migration, while migration did explain employment. Vias (1999) also looked at 254 non-metropolitan counties in the Rocky Mountain West for three time periods, the 1970s, 1980s and

1980-1995 and found that population was the driving force for employment growth, but found an inverse relationship between employment and population. When employment declined, population increased. These studies demonstrated the validity of the simultaneous equations approach, and the relationships between employment and population.

2.3.2 Spatial Econometric Models

The effect of spatial autocorrelation and spatial heterogeneity in housing markets in Dijon, France was investigated by Baumont (2007). Using a hedonic price function to account for spatial effects, neighborhood attributes, and accessibility, a spatial error model showed that lower income areas within Dijon affected surrounding areas. Housing prices are autocorrelated due to their nature as a durable good in a fixed location. In cities and suburbs, houses within a neighborhood are often built at the same time and with similar structural features (Baumont, 2007). Neighborhoods draw from similar labor markets and amenities like schools and parks as well.

Employment, income, net migration, and government expenditures in Appalachia have been shown to be spatially correlated (Gebremariam, et al., 2007). Utilizing panel data and a spatial approach to generalized Three-Stage Least Squares (3SLS), the authors showed the existence of feedback simultaneities between employment, income, migration, and government expenditures. Additionally, growth rates in one county were affected by the growth rates of the neighboring counties (Gebremariam, et al., 2007) .

Teen employment and wages are investigated using a spatial panel approach by Kalenkoski and Lacombe (2011). Since minimum wage laws vary among the states in the U.S.,

there exist neighboring counties with different minimum wage rates. The authors examined the economic effects of this disparity on teens who earn these wages by looking at their incomes and employment. Since employment is also correlated across states, a spatial panel approach is the correct tool for this analysis. Decreases in teen employment caused by increases in the minimum wage are greater when accounting for spatial dependence than in previous studies (Kalenkoski and Lacombe, 2011)

Mayer and Somerville (2000) discussed the spatial role of housing in a metropolitan area. Their results indicated that as housing prices rise, the boundary of a city may increase and houses at the fringe should have the same value as houses that were at the fringe before the expansion of the city boundary. Thus, the general value of housing itself has not changed, but instead the boundary of the city has expanded.

By including a Bayesian component, Deller, Lledo, and Marcouiller, (2008), created a method for choosing variables for their study on the effects of amenities on regional economic growth. The traditional empirical growth literature has been criticized as being ad hoc in the selection of right-hand-side control variables (Deller, Lledo and Marcouiller, 2008). After the Bayesian step, the authors then used spatial error, spatial autoregressive, and spatial Durbin models for estimation.

Jeanty, Partridge, and Irwin (2010) used Michigan Census tract-level data, to estimate a spatial simultaneous equations model that jointly models population change and housing values. The model explicitly considered the spatial interactions between housing price and population change within and across neighborhoods, while also controlling for spurious correlations (Jeanty,

Partridge and Irwin, 2010). The model is estimated using a generalized spatial two-stage (GS2SLS) procedure based on the work of Kelejian and Prucha (2004). Their results demonstrated the significance of substantive spatial interactions with neighboring census tracts. Average tract-level housing values are positively associated with average housing values of neighboring tracts. Population gains in neighboring tracts have a positive influence on population gains in the original tract. In addition, the coefficient estimates provided evidence of feedback simultaneity; neighborhoods are likely to experience an increase in their housing values if they gain population and they are more likely to lose population if they experience an increase in housing values. In decomposing the interactions of population change and housing values into direct (the “own” effect), indirect (spatial spillover effect) and total impacts, the authors found that housing value in a census tract is affected by a change in population growth in both the own and neighboring tracts, while population growth in a tract is only affected by changes in housing values in the same tract. Strong evidence of spatial spillovers in both population and housing values indicate that research models should account for spatial interaction of these variables.

2.4 Summary of Past Models and Shortcoming of Past Studies

Previous studies have contributed to our understanding of the concept of housing, homeowner’s behavior, and the reason for housing as a major household investment. Many studies analyzed the simultaneous relationship between incomes, population, and employment and made an assertion on the existence of long-run equilibrium relationship between housing prices, income, and population.

In reality endogenous variables like employment, income and population will affect housing prices while simultaneously affecting each other. This can be shown by a system of simultaneous equations which incorporates all these important variables as endogenous variables along with housing prices.

Previous studies showed a theoretical relationship between increases in population, employment, income, and residential real estate investment, but the simultaneous effects of all the four endogenous variables (housing prices, income, employment and population) has been ignored by previous studies. This research will explore the simultaneous relationship between these variables, and their interaction in the process of economic development using 3SLS estimation of a simultaneous system of equations.

Another important factor to be considered in studies is the spillover effect or neighbors' interaction. Because homeowner's are mobile, they may choose to own a house in one county and commute to a neighboring county for work. Homeowner preferences can put large spatial distances between work and employment locations. This effect will be neglected if spatial effects are not taken into consideration.

Also, previous studies have ignored the indirect effects of the housing investments from the surrounding and thus making it difficult to assess the overall effects of housing investments on the regional economic development. OLS analyses ignore spatial autocorrelation and leads to inefficient standard errors and estimators. Policies based on OLS data may be misleading, with undesired and inefficient effects. This study also uses a spatial panel technique to extend past studies further by considering spatial spill-over effects from the neighboring counties. This will

capture not only the direct and indirect effects, but also the total effects which helps in the estimation of the impact of the surrounding counties on the regional economic development.

Chapter 3.0 Theoretical Model

Introduction

Housing is a major investment for many households in the U.S (Brueckner and Pereira, 1997). However, unlike other financial investments, housing also provides significant consumption benefits. Residential real estate investment is driven by dual consumption and investment motives (Henderson and Ioannides, 1983). This research assumes that the consumption of housing is primarily a function of the location of the residential real estate investment, and that consumers choose to invest based in that location due to the bundle of investment and consumption goods that housing at that location provides.

This chapter is organized into three sections. The first section provides a theoretical model of residential real estate investment based on the choice of location. The second section specifies the variables used in this research. The third section lists the types and sources of the data.

3.1 Residential Real Estate Location Decisions

Location decisions of firms and households serves as one of the driving forces for regional development. The behavior of firms and households impacts the regional labor market by demanding and supplying labor. The housing market and housing value in turn are affected by the behavior of firms and households and also influenced by the regional labor market's

behavior. This section presents a model to analyze the interaction between the location decisions of firms and their effect of spatial variation in the economic development. The model was developed by Roback (1982) and is discussed with some minor changes. It is a general equilibrium model which allows for wages and land rents to interact in the location decision of households and firms. Firms and individuals simultaneously choose where to locate based upon the costs and benefits of the qualities of each location (Roback, 1982).

The basic assumptions for a simplified model include: both capital and labor are completely mobile across cities; the quantity of land is fixed for cities, but allowed to be change between commercial and residential land uses; workers ignore leisure, have homogeneous preferences, and supply one unit of labor independent of the wage rate. If every location has a vector of characteristics s and each worker can only produce and consume a good x , then each worker will attempt to maximize their utility with respect to their consumption of x , and the amount of residential land consumed l^c with respect to their budget constraint. That is:

$$(1) \max U(x, l^c; s) \text{ subject to } w + I = x + rl^c,$$

where wage and rental payments are w and r , respectively, and non-labor income is I (I does not depend on location), and s is a level of amenities at each location.

The indirect utility function V associated with equation (1) is:

$$(2) V(w, r; s) = k,$$

where k is a constant utility level, exogenous to individual locations. The equilibrium condition implies that utility at all locations is equalized in an interregional equilibrium at some constant k . If this doesn't happen, some workers would have an incentive to move to maximize their utility.

Wages and rents across locations must be equal, or else workers will move to where they can have a higher utility. Households choose locations by considering the trade-offs between wages (w), the residential land price (r), and amenities provided at each location (s).

Firms face a similar location choice. Firms use land, l^p , and the number of workers, n , to produce good x . Firms face constant returns to scale, and choose locations to minimize the cost of producing good x . Firms choose production locations to minimize total cost of producing the composite good $X = f(n, l^p, s)$ by considering the tradeoff between input prices, given the quantity of s (s can be productive or unproductive in the production process) in the location. The production function is assumed to exhibit constant returns to scale and diminishing marginal returns to each factor. So, at any given location, a firm chooses the best combination of labor (n) and land used in production (l^p) to minimize the total production cost (r and w are the price of land and wage rate respectively):

$$(3) \min wn + rl^p \quad \text{subject to } X = f(n, l^p, s).$$

Costless mobility and open entry assures that firms everywhere produce at a minimum cost and that marginal cost equals unit price. The optimal solution of equation 3 yields the demand functions for land in production and labor:

$$(4a) \ l^p = l^p(w, r; s),$$

$$(4b) \ n^d = n^d(w, r; s).$$

By assumption, firms reach their equilibrium when average cost equals price, assumed to be 1:

$$(5) \ C(w, r, s) = 1.$$

If average cost exceeds price, firms can choose to move to other cities where profits are higher. Conversely, if average costs is below price, the city may attract firms from places with lower profits. In order to fully describe the problem, the wages and rents must be at equilibrium within both labor and real estate markets.

Regions working towards attracting new firms and households should insure that the production cost per unit is below unity ($C(w, r; s) \leq 1$) and the indirect utility should be above the region's eventually desired equilibrium ($V(w, r; s, I) \geq k$). This will give the region a comparative advantage in cost and quality of life to attract new firms and households until it is eventually equalized. At any location, an increase in land price must be offset by an increase in wages to maintain consumer indifference. Increases in wages must be offset by reductions in rent to maintain average cost equal to unity. The equilibrium conditions for firms and households together determine the equilibrium levels of wage and land price. Roback (1982) uses a simplified model to show that wages, the number of employees per city, land rents, and amenities all combine in equilibrium to describe cities. This research will attempt to utilize this theoretical relationship among these variables.

Key assumptions of the Roback (1982) model are relaxed when dealing with empirical econometric models. Roback assumes that every worker has a job and housing, but both unemployment and homelessness exist within the study area. Since census data are used for this study, both unemployed persons and homeless persons will be included in the population variable. Additionally, Roback assumes that workers have homogenous preferences, no leisure time, and supply the same amount of labor regardless of the wage rate. These assumptions are also relaxed for the empirical estimations presented in chapter 5.

3.2 Specification of Variables

The empirical models are used to analyze the effect of residential real estate investment in regional economic development using changes in population, employment, and median income. The model will be explained as a system of equations with endogenous variables as a function of residential real estate, human capital, economic, and demographic variables. All endogenous and exogenous variables are defined in Table 1. The natural logarithms of all variables that are not equal to zero are used for estimation. Due to the nature of the variables collected there are no negative values.

The median housing value (HVM) in each county, measured in dollars, is the residential real estate dependent variable. Explanatory variables for residential real estate investment include the number of banks divided by the population in each county (BNK), the number of new building permits divided by the population (PER), the number of vacant housing units divided by the population (VAC), and the number of occupied housing units divided by the population (OCC). Banks and permits relate the availability of financing and new housing constructions, while vacancy and households relate the supply of vacant housing and the potential demand for housing respectively.

The population dependent variable is population per county (POP). This variable is directly related to HVM, as areas with an increase in population during the study period should have an increase in the demand for residential real estate. Government expenditures per county (GOV) and the number of occupied houses in a county (OCC) are used to explain population. Government spending and occupied houses are scaled by county population.

The employment dependent variable is employment per county (EMP). This variable is directly related to HVM, as areas with an increase in employment during the study period should have an increase in the demand for residential real estate. The explanatory variable for employment is number of businesses per county (BUS), as areas with more businesses, may have increased demand for housing for workers, increasing residential housing values. The business variable is scaled by population per county.

The dependent variable change in median income per county is INC. INC is directly related to HVM, as an increase in the median income per county means an increase in the demand for residential real estate, leading to a higher value for residential real estate. Explanatory variables for income include families living in poverty per county (POV) and people 25 years or older with a college degree per county (EDU). Poverty is a reflection of the standard of living in the surrounding area. Education is an instrument for the ability of workers to earn a good wage. Poverty and education are scaled by population per county.

Explanatory variables used as instruments for every equation include the population above 65 (P65), and population that is under 18 (P18). These variables relate segments of the population that are generally not in the labor force, and affect the amount of housing that must be purchased by households. Additionally, the P65 group generally has a lower income than the working age population (population between ages 18 and 65), also affecting residential real estate. A dummy variable denoting metropolitan counties is also included in every equation. The metropolitan designation is taken from and follows the USDA Economic Research Service definition of metropolitan counties (USDA-ERS, 2013). Some counties in the 1980 and 1990 time periods are not designated metropolitan counties, but become metropolitan in later time periods

due to population growth. The urban rural continuum codes are aggregated from nine categories into a binary dummy variable. Originally all nine categories were considered, but they resulting matrix was nonsingular. Schaeffer, Kahsai, and Jackson (2013) provides insight into the idea that rural and urban are not a strict dichotomy, but instead that issues which affect one invariably affect the other.

Table 1: Definition of Variables

Variable	Definition
Endogenous variables	Per county
POP	Population
EMP	Employment
INC	Median income
HVM	Median housing value
Exogenous variables	
P18	Number of people below the age of 18
P65	Number of people above the age of 65
MET	Dummy variable denoting metropolitan county
GOV	Federal government expenditure / population
BUS	Number of businesses / population
POV	Families below the poverty line / population
EDU	People 25 years+ with college degree / population
BNK	Number of banks / population
PER	Number of new housing permits / population
VAC	Number of vacant houses / population
OCC	Number of occupied houses / population

The natural logarithms of all variables are used for estimation except for GOV, EDU, PER, and MET as these variables contain zero as the values for some counties.

3.3 Data Types and Sources

The empirical model analyzes the effects of residential real estate investment on regional economic development using changes in population, employment, median income and median housing value. This study focuses on 434 counties of the Northeast region of the U.S. This study

uses a panel data set for the period 1980-2010. County-level data for endogenous and exogenous variables were collected from the U.S. Census Bureau, and the Economic Research Service of the U.S. Department of Agriculture.

Chapter 4.0 Methodology

Introduction

Chapter 3 examined the theoretical link between residential real estate investment and economic development. Chapter 4 develops the empirical models. The chapter is organized into two sections. Section 4.1 explores the non-spatial simultaneous equations 3SLS model. Section 4.2 justifies, describes and presents the spatial Durbin model.

4.1 Simultaneous Equations 3SLS Model

As indicated earlier, the focus of this study is to analyze the relationship between residential real estate investment and economic development represented by changes in population, employment, median income and median housing value. The empirical model for the study is derived from the two-equation non-spatial simultaneous equation model of Carlino and Mills (1987) who formulated their model by modifying Steinnes' model (1982). Deller et al. (2001) extended Carlino and Mills' framework to three simultaneous equations, which allowed them to incorporate interdependencies among income, population and employment change. Deller et al. (2001) has been extended to estimate simultaneous relationships of economic development with entrepreneurship (Bashir, 2011, Mojica, 2009), amenities (Kahsai, 2009), environmental regulation (Nondo, 2009), and modeling small business growth, migration behavior, local public

services and median household income (Gebremariam, 2006). This study will further extend the model of Deller et al. (2001) by estimating the simultaneous relationship between residential real estate investment and economic development.

The estimation methods are described and explained in Greene (2002) and Wooldridge (2010). The system of simultaneous equations is a complete system of equations since the number of equations is equal to the number of endogenous variables. The reduced form implies that the model can be solved equation by equation given there are no restrictions on parameter space and that orthogonality holds for the error terms. However, to guarantee that the system of equations has unique solutions, the conditions for identification must be satisfied. These include the rank and order conditions. To satisfy the rank condition, the number of exogenous variables excluded from an equation should be equal to or greater than the number of endogenous variables included in the equation. This is a necessary condition to ensure that the system has at least one solution. The rank condition requires that there is at least one non-zero determinant in the array of coefficients of the excluded variables which appears in the other equations. The rank condition ensures that there is exactly one solution for the structural parameters. In the model, there are more than one excluded variable in each equation, hence, both the order and rank conditions hold.

Ordinary least square (OLS) gives a biased and inconsistent estimate of the structural model if independent variables include endogenous variables. The simultaneity bias comes from the correlation between the right-hand side endogenous variable with the error terms. The models presented below imply simultaneity or reverse causation between dependent variables.

Therefore, the estimation is done using three stage least squares (3SLS) regression, which is used frequently to deal with endogenous variables. 3SLS first calculates the two stage least squares (2SLS) estimates of the identified equations. Then it uses the 2SLS estimates to estimate the structural equations' errors, and uses these errors to estimate the contemporaneous variance-covariance matrix of the errors of the structural equations. Finally, the generalized least squares method is applied to the large equation that represents all the identified equations of the system. 3SLS estimators are consistent and are generally asymptotically more efficient than 2SLS estimators (Kennedy, 2008). The method uses instrumental variables that are uncorrelated with the error terms to compute estimated values of the problematic predictors in the first stage and then uses those computed values to estimate a linear regression model of the dependent variable in the second stage. Since the computed values are based on variables that are uncorrelated with the errors, the result of the two-stage estimation is optimal.

The general form of a four simultaneous equations model defining the interaction between population (P), employment (E), median income (Y), and residential real estate (R) (measured as median housing value) is specified as:

$$(6a) \ P^* = f(E^*, Y^*, R^* / X^P),$$

$$(6b) \ E^* = f(P^*, Y^*, R^* / X^E),$$

$$(6c) \ Y^* = f(P^*, E^*, R^* / X^Y),$$

$$(6d) \ R^* = f(P^*, E^*, Y^* / X^R),$$

where P^*, E^*, Y^* , and R^* represent equilibrium levels of population, employment, median income, and median housing value, respectively, in the i th county; X^P, X^E, X^Y , and X^R are a

set of exogenous variables that have either direct or indirect effects on population, employment, median income, and median housing value.

Equations (6a) to (6d) represent population, employment, median income, median housing value, and exogenous variables in X that determine the equilibriums of population, employment, income, and residential real estate. The general equilibrium conditions specified in equations (6a) to (6d) expressed as a linear relationship can be explained as:

$$(7a) \quad P^* = \alpha_{0P} + \beta_{1P}E^* + \beta_{2P}Y^* + \beta_{3P}R^* + \sum \delta_{1P}X^P + \mu_P,$$

$$(7b) \quad E^* = \alpha_{0E} + \beta_{1E}P^* + \beta_{2E}Y^* + \beta_{3E}R^* + \sum \delta_{2E}X^E + \mu_E,$$

$$(7c) \quad Y^* = \alpha_{0Y} + \beta_{1Y}P^* + \beta_{2Y}E^* + \beta_{3Y}R^* + \sum \delta_{3Y}X^Y + \mu_Y,$$

$$(7d) \quad R^* = \alpha_{0R} + \beta_{1R}P^* + \beta_{2R}E^* + \beta_{3R}Y^* + \sum \delta_{4R}X^R + \mu_R.$$

The endogenous variables P, E, Y , and R indicate a county's population, employment, median income, and median housing value. Error terms are shown by u_P, u_E, u_Y , and u_R ; and the exogenous variable vector is represented by X . A panel data set is used with four time periods, 1980, 1990, 2000, and 2010. These years are chosen because of the availability of data due to the census being conducted in those years. Deller and Lledo (2007) and Deller et al. (2001) identified that the speed-of-adjustment coefficients are embedded in the coefficients α , β , and δ . This framework permits one to estimate structural relationships while simultaneously isolating the effects of real estate investment measured in terms of median housing value on regional economic development. Thus, the estimation of equations (7a) to (7d) is from short-run adjustments of population, employment, median income, and median housing value to long-run equilibriums (P^*, E^*, Y^* , and R^*).

4.2 Spatial Model

The model described in the previous section ignores the role of space. However, regions are made of multiple counties that influence one another's activities. These interdependent economic activities are likely to show spatial correlation. Consequently, specifying and estimating models without accounting for the spatial correlation results in biased estimators. Additionally, the spatial distribution of the endogenous variables described in the previous section may not be dependent on county level political jurisdiction, the unit for which data are reported. A possible mismatch between the geographical units and the spatial distribution of the variables being studied may result in misspecification problems and misleading policy implications (Anselin, 1988, Anselin and Bera, 1998, Doreian, 1980, 1981, LeSage, 1997). Furthermore, most socioeconomic and demographic variables also tend to have spatial dependence (Doreian, 1980, 1981). For this reason it is necessary to investigate spatial dependence in the data.

4.2.1 Moran's I

To investigate spatial dependence in the data, a global Moran's I test is performed. Moran's I is a measure of spatial autocorrelation. The spatial autocorrelation of the endogenous variables (population, employment, median income, and median housing value) for the Northeast region of the United States can be measured by Moran's I index. It is calculated by using the formula:

$$(8) I = \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j (X_i - \bar{X})(X_j - \bar{X})}{\sum_i (X_i - \bar{X})^2}$$

Where N is the number of spatial units indexed by i and j , X is the variable of interest (population, employment, median income, median housing value), and w_{ij} is an element in the i^{th} row and j^{th} column of the spatial weight matrix where w_{ij} represents the geographic relationship between different counties in the Northeast region of the U.S. The spatial weight matrix used for this analysis is a row-normalized (LeSage and Pace, 2009) 5 nearest neighbor weight matrix based on the physical centroids for the 434 counties in the study area.

Moran's I is a statistical test that can determine if a spatial pattern is the result of clustering, or a random distribution through space. A random distribution will show no spatial pattern, and spatial statistical models may not be required for estimation of parameters of interest.

Moran's I index is defined over the interval $[-1, 1]$ and the significance of Moran's I is tested using Z-statistics. A positive Moran's I value indicates positive spatial autocorrelation or spatial dependence, and a value of one indicates perfect correlation. A negative Moran's I value implies negative spatial correlation or no spatial dependency and the value of negative one indicates perfect dispersion. The value of zero indicates random spatial pattern.

ArcGIS 10.1 was used to run a Moran's I test for the four primary variables of interest. The results of the Moran's I tests are shown in Figures 25 through 28 in the appendix. A clustered spatial pattern shows that data is correlated with location.

The spatial dispersion of median housing value, population, employment, and median income per county are shown to have a positive Moran's I value and are significant at a 1% level. The results shows that the endogenous variables of interest (population, employment, median

income, and median housing value) are not randomly distributed throughout the study area, but are instead clustered and display spatial autocorrelation. Despite the findings that there is a spatial autocorrelation of the four primary variables in the chosen study area, the Moran's I test is only effective when the spatial patterns are consistent across the counties in Northeast region of the U.S. This is due to the reason that the Moran's I assesses the overall spatial pattern and trend and any inconsistent variation across the counties could offset one another and the Moran's I test may reveal non-spatial autocorrelation.

This statistically significant, spatial autocorrelation among the county level data for the four primary variables of population, employment, median income, and median housing value imply that non-spatial regression results may be biased by not including location information in the regression analysis. These results show that the four primary variables of interest are spatially correlated, and suggest that a spatial econometric model may be used in their estimation.

4.2.2 Spatial Panel Model

The focus of this study is to analyze the relationship between residential real estate investment and economic development represented by changes in population, employment, and median income. In cases of simultaneous equations, spatial dependencies appear due to two reasons (Kelejian and Prucha, 1999, 2004). The first reason is that error terms are not only assumed to be spatially correlated but also correlated across equations. The second reason is that the value of the endogenous variables in a given equation is assumed to depend upon a weighted sum of those endogenous variables over neighboring regions.

Spatial dependence is an important factor in regional economic development analysis, especially in terms of population, employment, and per capita income (LeSage and Fischer, 2009).

Spatial panel models allow explanatory variables in neighboring locations in previous or contemporaneous time periods to affect dependent variables (Debarys, Ertur and LeSage, 2012). This is easily described as economic actors in one county are likely to be aware of potential job opportunities, incomes, and living conditions in neighboring counties, and this awareness likely decreases with distance from their current location. A spatial weight matrix will capture this knowledge, and a spatial panel model will account for the ability of economic actors to move from one county to another within the time frame of the panel data set.

The study is derived from a spatial autoregressive model with autoregressive disturbances for the simultaneous equations of Kelejian and Prucha (2004). They built this model by extending Cliff-Ord's single equation model (Cliff and Ord, 1973, 1981). Kelejian and Prucha (2004) extended the model into a simultaneous system to estimate a simultaneous system of cross-sectional equations with spatial dependencies.

Anselin (1988) argued that in the presence of spillover effects, estimation of the econometric model will be biased or inefficient if spatial dependencies are ignored in the model. Anselin (1988) also showed that OLS estimation results are inconsistent.

This means that the non-spatial simultaneous equations should be estimated by incorporating spatial dependency. One approach, which incorporates spatial dependencies, is the spatial Durbin model (SDM). The SDM model nests both the spatial error model (SEM) and spatial autoregressive model (SAR), two other commonly used spatial econometric models. LeSage and Pace (2009) explained that SDM incorporates not only spatial lag of dependent variables but also independent variables. LeSage and Pace (2009) also shows that the SDM model is a superior model choice to either the SEM or SAR models, because in cases of misspecification

about the true data generating process that is being modeled, the SDM model will produce unbiased (but perhaps inefficient) coefficient estimates. Both the SAR and SEM models produce biased coefficient estimates, or the SDM model produces coefficient estimates that are at least as good as SEM or SAR in cases of misspecification (LeSage and Pace, 2009).

Omitted variable bias is a common occurrence in regression models. LeSage and Pace (2009) shows that when an omitted variable in the regressors or disturbances is also spatially dependent, the effect of the bias in the OLS regression is amplified. Omitted variable bias in the presence of spatial dependence may produce greater bias than omitted variable bias in OLS alone. LeSage and Fischer (2009) indicated that SDM also deals with omitted variable bias.

For these reasons this research utilizes a panel spatial Durbin model for estimation. A derivation of the model follows. A simple pooled linear regression model with spatial and temporal specific effects is considered, but without spatial interaction effects,

$$(9) \quad y_{it} = x_{it}\beta + \mu_i + \eta_t + \varepsilon_{it},$$

where i is an index for the cross-sectional dimension (spatial units), with $i=1, \dots, N$, and t is an index for the time dimension (time periods), with $t=1, \dots, T$. y_{it} is an observation on the dependent variable at i and t , x_{it} an $(1, K)$ row vector of observations on the independent variables, and β a matching $(K, 1)$ vector of fixed but unknown parameters. ε_{it} is an independently and identically distributed error term for i and t with zero mean and variance σ^2 , while μ_i denotes a spatial specific effect and η_t denotes a time specific effect. The standard reasoning behind spatial specific effect is that they control for all space-specific time-invariant variables whose omission could bias the estimates in a typical cross sectional study (Elhorst, 2010). The standard reasoning behind a

temporal specific effect is that they control for all time-specific space-invariant variables whose omission could bias the estimates in a typical cross sectional study (Elhorst, 2010).

When specifying interaction between spatial units, the model may contain a spatially lagged dependent variable or a spatial autoregressive process in the error term, known as the spatial lag and the spatial error model, respectively. The spatial lag model posits that the dependent variable depends on the dependent variable observed in neighboring units and on a set of observed local characteristics

$$(10) \quad y_{it} = \delta \sum_{j=1}^N w_{ij} y_{jt} + x_{it} \beta + \mu_i + \eta_t + \varepsilon_{it},$$

where δ is called the spatial autoregressive coefficient and w_{ij} is an element of a spatial weights matrix W describing the spatial arrangement of the units in the sample (Elhorst, 2010). It is assumed that W is a pre-specified non-negative matrix of order N .

The spatial error model, on the other hand, posits that the dependent variable depends on a set of observed local characteristics and that the error terms are correlated across space

$$(11) \quad y_{it} = x_{it} \beta + \mu_i + \eta_t + \phi_{it},$$

$$(12) \quad \phi_{it} = \rho \sum_{j=1}^N w_{ij} \phi_{jt} + \varepsilon_{it},$$

where ϕ_{it} reflects the spatially autocorrelated error term and ρ is called the spatial autocorrelation coefficient.

In both the spatial lag and the spatial error model, stationarity requires that $1/\omega_{min} < \delta < 1/\omega_{max}$ and $1/\omega_{min} < \rho < 1/\omega_{max}$, where ω_{min} and ω_{max} denote the smallest (i.e., most negative)

and largest characteristic roots of the matrix W . While it is often suggested in the literature to constrain δ or ρ to the interval $(-1, +1)$, this may be unnecessarily restrictive (Elhorst, 2010).

An unconstrained spatial Durbin model with spatial and temporal fixed effects looks like

$$(13) \quad y_{it} = \delta \sum_{j=1}^N w_{ij} y_{jt} + x_{it} \beta + \sum_{j=1}^N w_{ij} x_{ijt} \gamma + \mu_i + \eta_t + \varepsilon_{it},$$

where i is an index for the cross-sectional dimension (spatial units), with $i=1, \dots, N$, and t is an index for the time dimension (time periods), with $t=1, \dots, T$. y_{it} is an observation on the dependent variable at i and t , x_{it} an $(1, K)$ row vector of observations on the independent variables, and β a matching $(K, 1)$ vector of fixed but unknown parameters. The variable ε_{it} is an independently and identically distributed error term for i and t with zero mean and variance σ^2 , while μ_i denotes a spatial specific effect and η_t denotes a time specific effect. The variable γ , just as β , is a $(K, 1)$ vector of fixed but unknown parameters.

The term $\sum_{j=1}^N w_{ij} y_{jt}$ denotes the interaction effect of the dependent variable y_{it} with the dependent variables y_{jt} in the neighboring counties, where w_{ij} is the i, j -th element of a $(N \times N)$ spatial weigh matrix W . The term $\sum_{j=1}^N w_{ij} x_{ijt} \gamma$ denotes the interaction effect of the dependent variable y_{it} with the weighted average effects of the neighboring counties on the independent variable x_{ijt} .

The hypothesis $H_0: \gamma=0$ can be tested to investigate whether this model can be simplified to the spatial lag model and the hypothesis $H_0: \gamma+\delta\beta=0$ whether it can be simplified to the spatial

error model (Elhorst, 2010). This research will utilize the spatial Durbin model, as LeSage and Pace (2009) show that in cases of misdiagnosis of the true data generating process, SDM will produce unbiased coefficient estimates. Additionally, the SDM is better at providing unbiased coefficient estimates in the presence of omitted variables (LeSage and Fischer, 2009).

Chapter 5.0 Empirical Analysis and Results

Introduction

This chapter concentrates on estimation of the empirical models for determining the relationship between residential real estate investment and regional economic development. Regional economic development is indicated by growth in population density, employment, and median income per county. The empirical models are estimated using the 3SLS method and the spatial Durbin model with a panel data set. This chapter consists of two sections. Section 5.1 presents the results of the 3SLS simultaneous equations model. Section 5.2 presents the results of the spatial Durbin model.

5.1 Simultaneous Equations Model Results

This section concentrates on estimation of the empirical models for determining the relationship between residential real estate investment and regional economic development. The simultaneous equation empirical model is estimated using the three stage least squares (3SLS) method and the statistical package Stata version 12. 3SLS is used to overcome the problem of correlation of the error terms of each equation, especially the population and employment equations. 3SLS also accounts for all restrictions on parameters in the system of simultaneous

equations. The endogeneity of the four variables, population, employment, median income, and median housing value is tested with a Hausman test and found to be endogenous. A strongly balanced panel data set with 1736 observations for all the counties in the Northeast region is used to estimate a fixed effects model using 3SLS via Stata.

There are problems attempting to split the panel into rural and urban counties for separate estimation. Due to the transition of some counties from rural to urban throughout the time period, the urban and rural panels become unbalanced. Fixed effects estimation eliminates these counties from the regressions.

5.1.1 Population Equation

The results of the population growth equation for the Northeast region using 3SLS are presented in Table 2. The population growth equation is estimated against endogenous variables of employment (EMP), median household income (INC), and residential real estate investment as measured by median housing values (HVM); and control variables are included to measure economic effects. The overall fit (R^2) of the empirical population equation is 92%. This is likely due to correlation between the population, population under the age of 18, population over the age of 65, and employment variables.

The empirical results show that population is positively and significantly related to employment at the 1% level which explains that an increase in the number of jobs also increases population. An increase in jobs in a region or county brings in more people and their families for employment or in search of employment. This in turn increases the population. Thus, the increase in population requires more government personnel like police, fire fighters, and teachers, creating an overall net gain of jobs in a county.

A significant and negative relationship (at the 1% level) between population and median household income indicates that population decreases as median household income increases. This implies when income increases people tend to move away from densely populated regions and towards less densely populated regions. This could be evidence of suburbanization, but may not make sense at the county level. Alternately this may be evidence of a general demographic transition, where rising incomes are responsible for falling fertility rates and population increasing at a decreasing rate in developed countries (Sinding, 2009).

The relationship between median housing value and population is positive and significant at the 1% level. The results show that as the residential real estate investment measured by median housing value increases, population will also increase. The increase in population may be attributed to migration from other counties or increases in the household population. Also, the increase in housing value is a potential measure of improved amenities, like school districts and attractive neighborhoods. This shows the preferences of people who choose to live in areas with more or better amenities even if it costs more money.

P18 and P65 are both positively related to population and are significant at the 1% level. This shows that when both P18 and P65 increase population in a county also increases. This is obvious because both P18 and P65 are part of the population.

Government spending programs include amenities that improve living conditions within a county, but are not statistically significant. This might be due to the reason that the variable used here is federal expenditures and may not capture the effects of improvements in amenities due to local government spending. Occupied houses within a county is positive and significant at the 1% level, showing that an increase in number of occupied houses in a county increases the

population. This result is as expected as we expect the population of a county to grow when a new household moves into a county.

The dummy variable denoting metropolitan counties is not statistically significant. The metropolitan counties variable is expected to be significant as metropolitan counties include both suburbs and exurbs and captures the migration of people between cities, suburbs and exurbs. The variables occupied houses and metropolitan counties are both correlated with population, these variables were included because a Wald test showed that they were jointly significant.

Table 2: Three Stage Least Square Results of Population Equation

Variable	Coefficient	z statistic	p value
Ln EMP	0.5118157*	23.22	0.000
Ln INC	-0.211198*	-11.97	0.000
Ln HVM	0.2045593*	16.36	0.000
Ln P18	0.2552898*	18.54	0.000
Ln P65	0.1080035*	17.67	0.000
GOV	0.0000941	0.98	0.328
Ln OCC	0.0263798*	3.55	0.000
MET	0.0058699	1.5	0.133
CONST	2.175779*	38.94	0.000
N	1736		
R ²	0.9289		

*, **, and *** denote significance at 1%, 5%, and 10% respectively

5.1.2 Employment Equation

The employment growth equation is estimated against endogenous variables of population (POP), median income (INC), and residential real estate investment as measured by median housing values (HVM); and control variables are included to measure economic effects. The results of the employment growth equation for the Northeast region using 3SLS are presented in Table 3. The overall fit (R^2) of the empirical employment equation is 79%.

The empirical results show that employment is positively and significantly related to population at the 1% level. This explains that an increase in the number of people also increases employment which agrees with the theory “jobs follow people” which is also supported by the results of Steinnes and Fischer (1982). This makes sense because an increase in population increases the demand for goods and services. More people require at least basic services, and jobs providing basic services increase the employment in a county and a region.

A significant (at the 1% level) and positive relationship between employment and median income indicates that employment increases as median income increases. As incomes increase, consumers spend more on goods and services. This increased spending leads to employment. This results supports Deller’s (2001) extension of Carlino and Mill’s (1987) model where per capita income is hypothesized to positively affect employment.

The relationship between median housing value and employment is negative and significant at the 1% level. Counties with higher median housing values have low employment. This may be caused by housing markets that are highly segmented and the high value housing might be totally separated from the other lower value housing markets within a county. Alternately, this result may be an illustration of residential land uses outbidding commercial land uses for locations. Areas with significant residential amenities may not be conducive to commercial use, and vice versa.

The number of businesses in a county is negatively and significantly related to employment at the 5% level. This is an unexpected result since all businesses should have at least one, and probably more than one, employee. Again, this may be attributed to the reason that businesses might just congregate on large cities in a county. So the overall effect of these

businesses is not properly captured when the employment for the whole county is considered. Metropolitan counties are not significant. Metropolitan counties have relatively higher populations and businesses, both variables should affect employment.

Table 3: Three Stage Least Square Results of Employment Equation

Variable	Coefficient	z statistic	p value
Ln POP	1.113938*	76.66	0.000
Ln INC	0.5278751*	24.2	0.000
Ln HVM	-0.433919*	-23.75	0.000
Ln BUS	-0.00768***	-1.85	0.064
MET	0.005063	0.58	0.562
CONST	-2.595718*	-17.53	0.000
N	1736		
R ²	0.7853		

*, **, and *** denote significance at 1%, 5%, and 10% respectively

5.1.3 Income Equation

The median income equation is estimated against the endogenous variables population (POP), employment (EMP), and residential real estate investment as measured by median house values (HVM) and control variables are included to measure economic effects. The overall fit (R^2) of the empirical median income equation is 93%. The results of the median income equation for the Northeast region using 3SLS are presented in Table 4.

Population is not significantly related to median income. As with the income result from the population equation, a large increase in the population of any county is likely to decrease the median income. Firstly, more workers earn low wages than high wages, so any increase of population without specifically targeting high wage earners should have the effect of decreasing the median income in a county. Secondly, the easiest source of population growth is through having children. Children are counted as people in the population variable, but they generally

earn no wage at all. This will also result in decreasing the median income of a county.

Employment is positively and significantly related to median income at the 1 % level. This shows that counties with higher levels of employment have higher median incomes. This makes sense, as wage income is the primary source of income for most people in the U.S. and in the Northeast region.

The relationship between residential real estate investment and median household income is positive and significant at the 1% level. As housing values increase, the median income of the residents of those houses should also increase in order to maintain equilibrium rents. If incomes do not rise, owners will potentially sell the homes to capture the increase in value and move to a new location with a housing consumption investment bundle more closely aligned with their original preferences.

The poverty rate is not statistically significant. Poverty is generally defined by having a low income, and so this result is surprising. Negative externalities from living in areas with high rates of poverty can negatively impact consumer's levels of residential real estate investment. Additionally depreciation of the physical capital of housing stock is associated with lower rents. This allows families in poverty to have a place to live and not be forced into homelessness, but does discourage reinvestment in the depreciating housing assets.

The education rate is significantly (at the 5% level) and positively related to median income. This result is as expected. A higher proportion of educated people in a county increase median income in that county. Population under 18 and population over 65 are negatively and significantly related to median income at the 1% level. The signs for these two coefficients were expected to be negative, as these segments of the population are generally not in the workforce

and so do not earn a large income. Metropolitan counties are not statistically significant. Counties with more residents may have a higher median income, potentially to offset higher costs of living.

Table 4: Three Stage Least Square Results of Income Equation

Variable	Coefficient	z statistic	p value
Ln POP	-0.2453314	-1.35	0.177
Ln EMP	1.389655*	11.54	0.000
Ln HVM	0.6690106*	32.74	0.000
Ln POV	-0.0206902	-0.78	0.433
EDU	0.054521**	2.01	0.045
Ln P18	-0.792454*	-6.07	0.000
Ln P65	-0.279165*	-4.64	0.000
MET	-0.028838	-1.36	0.175
CONST	1.42174***	1.75	0.079
N	1736		
R ²	0.9267		

*, **, and *** denote significance at 1%, 5%, and 10% respectively

5.1.4 Residential Real Estate Equation

The residential real estate equation is estimated against the endogenous variables population (POP), employment (EMP), and median income (INC) and control variables are included to measure economic effects. The overall fit (R^2) of the empirical Residential Real Estate equation is 92%. The results of the residential real estate equation for the Northeast region using 3SLS are presented in Table 5.

Population is positively and significantly related to residential real estate investment at the 1% level. This shows that a large population positively impacts residential real estate prices. People need housing, and the demand for housing spurs investment in the residential real estate sector. Increased demand for housing will increase the levels of residential real estate investment in the county.

Employment per county has a significant (at the 1% level) and negative relationship with

residential real estate investment as measured by median housing value. This may be a reflection of the effect of commuting behavior, where consumers live in one county, but drive to work in another (Jeanty, Partridge and Irwin, 2010). Also, considering this research uses county level data, this negative result may be applicable only when people choose to live near the county boundaries, like when large cities are located close to state borders. Alternately, this result may arise because of the prevalence of rental housing units in areas of high employment. The significant rental markets of the large cities within the study area may be responsible for the negative relationship.

A significant (at the 1% level) and positive relationship between residential real estate and median income indicates that residential real estate investment is positively affected by areas with high median incomes. As incomes increase, consumers invest some of their income in housing. This can take the form of purchases of higher quality housing, or renovations and additions to existing housing that lead to higher median values.

The proportion of banks to population per county is negatively and significantly related (at the 1% level) to housing value. This variable serves as a measure of the availability of financing in each county. Most home owners in the U.S. finance their home purchases or renovations through the banking sector. The ability to acquire a local source of lending for residential real estate investment should lower the transactional costs of that investment.

Number of new building permits (scaled by population) issued in each county is positively related to median housing value and is significant at the 5% level. Theoretically, an increase in building permits signals a future increase in the supply of housing available. This should lead to the price of the current supply of housing decreasing, as future demand will be met with new

construction and not just bidding up the price of current housing.

The number of vacant houses per county (scaled by population) is positively and significantly related to median housing value at the 1% level. The number of vacancies is an indication of the demand for housing, with a low number of vacancies demonstrating a high demand for housing in an area, so an increase in median housing value makes sense.

The number of occupied housing units per county (scaled by population) is negative and significant at the 1% level. This is also a measure of the demand for housing, as each household theoretically needs housing in which to live. The number of occupied households is correlated with population, with areas of high population also having a high number of households so people tend to move away from high populated areas to suburbs or exurbs which have better amenities and thus higher median housing value. Metropolitan counties are not statistically significant. Metropolitan counties have more people by definition, and the demand for housing should lead to an increase in residential real estate investment.

Table 5: Three Stage Least Square Results of Residential Real Estate Investment Equation

Variable	Coefficient	z statistic	p value
Ln POP	2.325384*	27.76	0.000
Ln EMP	-2.07463*	-24.83	0.000
Ln INC	1.21961*	86.25	0.000
Ln BNK	-50.03298*	-3.22	0.001
PER	0.355919**	2.02	0.044
Ln VAC	0.0492293*	3.58	0.000
Ln OCC	-0.181652*	-2.62	0.009
MET	0.0123771	0.71	0.479
CONST	-5.647077*	-19.41	0.000
N	1736		
R ²	0.9166		

*, **, and *** denote significance at 1%, 5%, and 10% respectively

5.2 Spatial Panel Results

The empirical results presented in the previous section were established upon an assumption of no spatial dependence among the endogenous variables. A residential real estate investment choice is inherently spatial in nature. Every house exists in a location, and the combined bundle of investment and consumption goods that arise from owning the house is also spatial. The empirical results discussed above are biased and inconsistent without considering spatial dependence. This section provides estimation results of the spatial model developed in Chapter 4. The spatial Durbin model (SDM) analyzes that spillover effects by including spatial lags for dependent and independent variables and is used as a method of spatial analysis. This model captures the direct and indirect effects of residential real estate investment on neighboring counties in regional economic development.

Interpretation of the coefficients in spatial model is important. Kirby and Lesage (2009) stated that in spatial models changes in the independent variables x_i are represented by a direct effect on the county's median residential real estate investment as measured by the median housing value and a spatial spillover (indirect) effect on neighboring county's median residential real estate investment.

5.2.1 Spatial Durbin Model Results

The results of the SDM model including both county and year fixed effects are given in Table 6. The spatial weight matrix used for this analysis is a row-normalized 5 nearest neighbor weight matrix ($434 * 434$) based on the physical centroids for the 434 counties in the study area. Splitting the sample into rural and urban counties creates problems with the dimensions of the

weight matrix. Estimation of only rural counties does not consider the spatial spillovers of nearby urban counties. Additionally, the weight matrix is no longer a matrix of the five nearest neighbors. Attempts to use the original weight matrix will result in incompatible matrix algebra. Urban only estimates experience the same problems.

The overall fit can be measured by the R^2 of 0.99 which is initially viewed as very high. Elhorst (2010) recommended that squared correlation coefficient be used as an alternative measure of goodness-of-fit because of the nature of fixed effects. The squared correlation coefficient for this model is 0.69. The difference between the R^2 and the squared correlation coefficient indicates how much of the variation in the dependent variable is explained by the fixed effects. The fixed effects portion of this model explains approximately 30% of the variation in the dependent variable.

The results from this SDM model indicate that there is a significant level of spatial autocorrelation in the dependent variable, with the δ parameter equal to 0.76 and significant at the 1% level. This result confirms the usage of the SDM model as appropriate to deal with the spatial autocorrelation. This result is also used to calculate the proper marginal effects. LeSage and Pace (2009) show that the marginal effect of a change in an explanatory variable depends on both the direct and indirect effects which depend on the value of δ .

The primary focus of this research is to determine the effects of changes in county level median income, population density, and employment on residential real estate investment as measured by median housing value. The theory states that increases in these three explanatory variables should increase median housing values, as they should cause increases in the demand for housing.

The direct effect of a change in a county's income measures how a change in a particular county's income affects median housing value in that same county. From Table 6, the direct effect of a change in a county's median income on the median housing value is 0.67 and is significant at the 1% level. The indirect effect is 1.35, and the total effect is 2.02 and both are significant at the 1% level. . This means that an increase in the median income in a county will increase the median housing values within the same county, and will increase the median housing values in neighboring counties. This follows from economic theory that an increase in income leads to an increase in the consumption of all goods including an increase in the consumption of housing, and some of this increase can occur in neighboring counties through leakages via migration of people and firms across the time periods in the study.

The total effect shows that changes in the median income affects median housing value, considering own-county and neighboring county spillover effects. The total effect estimate shows that, as the median income increases by 10%, median housing value increases by 20.15%. An increase in income leads to an increase in residential real estate investment. As consumers spend more income on housing, the demand for housing increases

The direct effect of a change in a county's population measures how a change in population affects median housing value in that same county. From Table 6, the direct effect of a change in a county's population on the median housing value is -0.10 and is significant at the 5% level. The indirect effect estimate is 0.75, and the total effect is 0.65 with both significant at the 5% level. As a county increases its population, median housing value in neighboring counties (as defined by the 5 nearest neighbors weight matrix) also increases while the median housing value in the same county decreases. The total effect estimate shows that, as the population

increases by 10%, median housing value increases by 6.5%. An increase in population leads to an overall increase in residential real estate investment, as demand for housing increases.

The direct effect of a change in a county's employment measures how a change in employment affects median housing value in that same county. From Table 6, the direct effect of a change in a county's employment on the median housing value is 0.18 and is significant at the 1% level. This means that as a county increases its own employment by 10%, median housing value in that same county increases by 1.8%. This might be because an increase in employment in a county will lead to more people earning a wage, a portion of which can be invested in residential real estate. Since most of the people in U.S. consider housing a good long term investment, people with jobs and incomes tend to invest in housing and thus there is increase in median housing value.

The indirect effect estimate is -0.47, and is significant at the 1% level. As a county increases its employment, median housing value in neighboring counties (as defined by the 5 nearest neighbors' weight matrix) decreases. This might be because counties are a large geographic area and for someone employed in the interior of a county, investing in housing in the neighboring county might be a bad idea considering the commuting distance. The total effect of a change in employment is -0.31 and is not statistically significant. The total effect estimate shows that employment as measured by the number of jobs in each county is not a significant factor in determining the level of residential real estate investment in counties in the Northeast region of the U.S.

Of the three primary explanatory variables, the median income of a county had the greatest effect on residential real estate investment in own and neighboring counties. The next

most influential variable is population per county. While employment is not statistically significant in the total effect this lends support to the “jobs follow people” hypothesis. Regional economic development policies should favor enticing population growth through migration. Additionally, policies to improve the income of residents in each county will have a beneficial impact on residential real estate investment in own and surrounding counties.

The direct effect of the ratio of businesses per population in each county is 0.04 and is significant at the 1% level and shows that increase in business in a county increases the employment of the own county. The indirect effect is 0.08 and is not significant. The total effect is 0.13 and is not significant. The number of firms located in each county affects own county residential real estate investment, but does not affect the neighboring counties. Policies to encourage firms to locate within a particular county benefit only the county itself and not the surrounding neighbors with regard to residential real estate investment. This lends some support for policies to bring firms to a county as means of economic development, but if the firms are leaving a neighboring county there is no overall gain to the region as a whole, just a movement of resources from one county within the study area to another county within the same study area.

The direct effect of the number of families in poverty divided by the population per county is 0.03 and is significant at the 10% level. The indirect effect is 0.09 and the total effect is 0.13 and both are not significant. Poverty increases residential real estate investment in own counties, but does not influence surrounding counties. It may be an issue of data aggregation, since both rich counties and poor counties in the study area can have similar poverty rates, but different

median housing values. The aggregation of poverty into counties instead of cities may be biasing the results.

The effect of the number of building permits issued (scaled by population) per county is 0.32 and significant at the 1% level. The indirect effect is -2.00 and is significant at the 10% level. The total effect is -1.68 and is not significant. Building permits in own county have a positive effect on housing prices, while permits in neighboring counties have a larger, negative effect. An increase in the supply of housing in own county may signify that own county is an attractive location, and thus a small positive effect on price makes sense. An increase in neighboring county permits increases the future supply of housing in neighboring locales. This may signify that neighboring counties are attractive locations and competition for location may reduce own county housing prices. Also increased building permits in neighboring counties may signal an increase in the future supply of housing, leading to a decrease in current housing prices.

The direct effect of the ratio of college educated population to total population per county is 0.38 and is significant at the 1% level. The indirect effect is 0.78 and is not significant. The total effect is 1.16 and is significant at the 1% level. A higher proportion of educated individuals in a county increases residential real estate investment, potentially through factors not captured via the income variable.

From Table 6, the direct effect of a change in the number of vacant single family housing units (scaled by population) per county on the median housing value is 0.07 and is significant at the 1% level. The indirect effect estimate is 0.34, and the total effect estimate is 0.41. Both are significant at the 1% level. This means that as the number of vacant houses per county increases by 1%, median housing value in that same county increases by 4.1%. Vacant houses per county

have a positive impact on housing prices. This may be a measure of the demand for housing, as housing must be vacant in the transitional state between first building and purchase or between owners. A high enough price for housing may entice owners to sell and move elsewhere to maximize their housing consumption and investment bundle.

The direct effect of the number of occupied housing units (scaled by population) per county on residential real estate investment is -0.21 and is significant at the 1% level. The indirect effect is 0.11 and the total effect is -0.10; both not significant. Occupied housing units have a negative effect on residential real estate investment in own county, a surprising result. This result is similar to the population result, as occupied housing units are correlated with population. An increase in occupied housing decreases own county median housing prices. This may be another result of data aggregation at the county level.

The direct effect of a metropolitan county on residential real estate investment is 0.017 and is significant at the 10% level. The indirect effect is 0.08 and the total effect is 0.10; both not significant. Metropolitan counties have a positive effect on residential real estate investment on its own county but are not significant with the neighboring counties. Counties with more people experience more residential real estate investment, but being a neighbor to urban counties has no effect.

The differential between the direct effects and the total effects for all of the coefficients again shows that controlling for spatial dependence is important. Any research that neglects spatial dependence potentially drastically underestimates the effects of change in income, population density, and employment on housing values.

For the SDM model, the total effect of each primary explanatory variable shows that the full impact of any development policy will not only be felt in the own county, but will also affect neighboring counties. This research shows that policies to increase own county median income, population, and improve employment opportunities are effective for economic development. Additionally, the population, employment, and income of neighboring counties will partly determine own county development strategy.

Table 6: Spatial Durbin Model Spatial Panel Results

Variable	Coefficient	Direct	Indirect	Total
Ln INC	0.578285*	0.665843*	1.349659*	2.015503*
Ln POP	-0.149744*	-0.100341**	0.749848**	0.649507**
Ln EMP	0.184949*	0.153524*	-0.47094*	-0.317416
Ln BUS	0.036154*	0.041449**	0.085544	0.126993
Ln POV	0.02799***	0.033637***	0.092066	0.125703
PER	0.454632*	0.322963*	-2.004355***	-1.681392
EDU	0.326146*	0.377718*	0.777712	1.15543*
Ln VAC	0.04841*	0.07054*	0.341947*	0.412487*
Ln OCC	-0.221214*	-0.213168*	0.110637	-0.102531
MET	0.011333	0.017315***	0.08466	0.101976
W Ln INC	-0.087366			
W Ln POP	0.306852			
W Ln EMP	-0.261494			
W Ln BUS	-0.005469			
W Ln POV	0.00301			
W EDU	-0.048009			
W PER	-0.852956			
W Ln VAC	0.052345			
W Ln OCC	0.192546			
W MET	0.013307			
δ	0.756984			
σ^2	0.6889			
R^2	0.9919			

*, **, and *** denote significance at 1%, 5%, and 10% respectively

Chapter 6.0 Summary and Conclusions

The main objective of this research is to analyze the role of residential real estate investments in the economic development of the Northeast region of the United States. The general conclusion of the study is that investment in residential real estate shows positive and significant effects on economic development of the Northeast region of the United States during the period of 1980-2010. The results from both simultaneous equations (non-spatial) and spatial panel models provide policy makers with information on how the investments in residential real estate in the Northeast regions affect the economic development of the region. The spatial clusters of the residential real estate investments across the region during different time periods show how investment changes over time and how that variation over the region requires different policies to enhance economic development in different regions within the study area.

Two econometric models were used for estimation. First, a system of four simultaneous equations model using three-stage least squares (3SLS) is used to estimate the results. Second to estimate the spatial dependency and spatial clustering of the economic development a spatial panel model was used with a five nearest neighbor contiguity weight matrix. A panel dataset comprising demographic and economic variables for the period of 1980-2010 is used for both the models.

The general conclusion of the study is that residential real estate investment is positively associated with regional economic development during the 1980-2010 period. Based on the empirical analyses this study provides policy makers with information on the role of residential real estate investment in the economic development of the Northeast region.

6.1 3SLS Simultaneous Equations Model Conclusions

In the simultaneous equations method, a 3SLS regression was used to study the interactions between population, employment, median income and residential real estate investment as measured by median housing values at the county level. A panel data set of decennial census years was collected for the 434 counties in the Northeast U.S. A four equation system of simultaneous equations was solved using 3SLS and fixed effects. The general conclusion of the 3SLS estimation is that residential real estate investment as measured by median housing values does affect economic development in the region, but the results are mixed.

Population growth had a significant and positive effect on employment and median housing value, but a negative effect on median income. This shows that growth in population leads to growth in employment, as population demands goods and services leading to employment; and that population growth leads to increases in the value of residential real estate investments. A growing population demands more or better housing, leading to price increases. Population growth has a negative effect on median incomes. This may be because population growth may occur in demographic ranges that do not earn income (population under the age of 18 that cannot earn a high wage).

Employment growth had a positive and significant effect on population and median income, but a negative and significant effect on median housing value. Employment growth in a region may lead to migration into that region, increasing the population and presumably the incomes of those workers who get new jobs. Lower median housing values in higher employment

counties may be an example of commuting behavior around major metropolitan areas where jobs are concentrated in one location and many workers commute from outside the county.

Incomes negatively and significantly affect population, but positively and significantly affect employment and housing values. Incomes positively affecting employment and housing values is predicted by economic theory, as higher wages can buy more goods and services, and one of those goods is increased consumption of housing. When income increases people tend to shift their household away from densely populated regions and towards less densely populated regions. This is a demonstration of suburbanization, but may not make sense at the county level.

Median housing values are positively and significantly related to population and incomes, but negatively and significantly related to employment. This shows that areas with relatively high investment in residential real estate attract consumers with relatively more money to spend. Employment in these counties can be negatively affected, as lower wage workers get pushed out to the periphery. The mixed results reveal that the effect of residential real estate investment on the region is generally positive, as expressed by increases in population and income, but the negative effect on employment may arise because the proportion of lower paying jobs is larger than higher paying jobs.

6.2 Spatial Panel Model Conclusions

In the spatial Durbin model estimation, median housing price per county as a measure of residential real estate investment for the Northeast region is estimated. The spatial panel method divides the effect of any change in an exogenous variable into the direct effect, which affects own

county, and the indirect effect, which affects neighboring counties. 434 counties in the Northeast region are estimated using the same dataset as used in the previous 3SLS estimation.

The spatial Durbin model shows that changes in own county's median income and employment have larger positive effects on residential real estate investment than the negative effects that changes in population bring. Additionally, increasing the number of businesses within a county or the number of people with a college degree also have significant own county effects on residential real estate investment.

The actions of neighbors can greatly influence own county's economic development as well. The SDM analysis shows that residential real estate investment in own county is positively influenced by the population and the median income levels of neighboring counties, but that neighboring levels of employment have a negative effect. The indirect effect of neighboring counties is larger than the direct effect for own county measures of economic development. Any individual county within the study area must consider the strengths and weaknesses of its own economic position, as well as the strengths and weaknesses of its neighbors when formulating economic policy. The total effect of any policy change may cause benefit to neighboring counties, while the cost is borne by the own county taxpayers.

6.3 Policy Implications

The mixed results from the 3SLS and the results from the SDM model show that residential real estate investment is an important component of regional economic development. Any economic development plan attempting to increase the levels of residential real estate investment should focus on increasing the population, the number of available jobs in a county,

and on income growth, attracting educated residents or educating residents that already live within the location is another good strategy for economic growth supported by these results.

Spatial panel methods show that any attempt to improve the economic development in a county is greatly influenced by the behavior of their neighboring counties. The total effect of any policy change can be decomposed into the direct effect that affects own county, and the indirect effect that affects neighboring counties through spillovers and linkages. A regional economic development may have a better chance of cooperation and success than any local or county level development plan.

The effect of neighboring counties can be statistically measured, and economic development plans that do not consider the actions of neighboring counties will not achieve their desired results. Regional planning methods will allow for counties to pool resources, absorb costs, and reap the benefits of economic development. Regional planning will prevent counties from potentially free riding on the development plans of their neighbors. Taxpayers in one county will not be unknowingly subsidizing development in other neighboring counties. Instead these costs and benefits will be discussed in open forums with public input. Regional economic development plans will help better inform the citizens, and reach better development outcomes than a piecemeal approach.

6.4 Limitations and Future Work

The study has expanded the examination of economic impacts caused by investments in residential real estate by using a simultaneous equations method and spatial panel method. However, the study has certain limitations. The first limitation of the study is the data set. This

study uses a county level panel data set for the census years 1980, 1990, 2000, 2010. Better results can be obtained by using a continuous panel data set for a period of 10-20 years. Unavailability of the county level panel data set for the region was a limitation to the study. Census tract panel data set instead of county level data set also may give more precise and reliable results, as the smaller geographic area may allow for commuting behavior or other small scale geographic factors to be more accurately modeled. This helps in designing specific policies to the targeted areas rather than designing for a larger groups with regards to economic development.

Another limitation is the choice of variables included in the analysis. Additional data and variables may improve the results. For example, zoning data at the county level would enable a better understanding of the residential construction/permit regulations of each county. The impact of these zoning regulations on residential real estate investments would also provide better insight into county level economic development.

Also, zoning information will often be most powerful when it is examined along with data which might influence policy decisions. For example, in urban/big cities zoning is an issue that attracts lobbying. Adding both a zoning and a lobbying data set for the analysis might help to analyze the power effects in politics by studying the policy changes/decision makings.

A spatial panel analysis comparing the three regions (Region I, II, III) with the other regions in the U.S. (IV, V, VI, VII, VIII, IX, X) could be an interesting study as this can compare and differentiate the Northeast region with other regions and also see what policy decisions might favor the Northeast regions when compared to the other regions.

The 3SLS model demonstrates that population, employment, median income, and median housing value are endogenously related to each other. The 3SLS model uses this relationship to estimate a system of simultaneous equations that describes the equilibrium location choice of households and firms. Unfortunately, the 3SLS model does not account for spatial autocorrelation. Geographic information is an integral part of any location choice, so a model that utilizes geographic information may also do a good job explaining location choices.

The spatial Durbin model can utilize geographic information by utilizing a spatial weights matrix that indicates which counties are neighbors and which counties are not neighbors. This information follows Tobler's (1970) first law of geography, "Everything is related to everything else, but near things are more related than distant things." Counties that are neighbors will have a stronger effect on each other than counties that are not neighbors. The spatial Durbin model deals with spatial autocorrelation very well, but it does not include the simultaneous relationship demonstrated from the 3SLS model. The empirical disconnect between the two models used in this dissertation remains a limitation of the study. Further research into this problem will attempt to incorporate the nature of the simultaneous relationship between the four primary endogenous variables of population, employment, median income, and median housing value with the spatial econometric models that account for the spatial autocorrelation of the same variables.

References

- Abraham, J., and P. Hendershott. 1996. "Bubbles in Metropolitan Housing Markets." *Journal of Housing Research* 7:191-207.
- Anselin, L. 1988. *Spatial Econometrics: Methods and Models*. Cambridge: MIT Press.
- Anselin, L., and A.K. Bera 1998. "Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics." In A. Ullah, and D. Giles eds. *Handbook of Applied Economic Statistics*. New York, Marcel Dekke, pp. 237-289.
- Arku, G., and R. Harris. 2005. "Housing as a Tool of Economic Development Since 1929." *International Journal of Urban and Regional Research* 29:895-915.
- Badcock, B.A. 1989. "Homeownership and the Accumulation of Real Wealth." *Environment and Planning D: Society and Space* 7:69-91.
- Bashir, S. 2011. "An Empirical Analysis of the Role of Entrepreneurship in the Economic Development in Northeast Region." Ph.D. Dissertation, West Virginia University.
- Battacharya, R., and S. Kim. 2010. "Economic Fundamentals, Subprime Lending and Housing Prices: Evidence from MSA-level Panel Data." *Housing Studies* 26:897-910.
- Baumont, C. 2007. "Neighborhood Effects, Urban Public Policies and Housing Values A Spatial Econometric Perspective." In *A Spatial Econometric Perspective*. Université de Bourgogne Laboratoire d'Economie et de Gestion (CNRS).
- Brueckner, J., and A. Pereira. 1997. "Housing Wealth and the Economy's Adjustment to Unanticipated Shocks." *Regional Science and Urban Economics* 27:497-513.
- Capozza, D.R., P.H. Hendershott, C. Mack, and C.J. Mayer. 2002. "Determinants of Real House Price Dynamics." Unpublished, National Bureau of Economic Research.
- Carlino, G., and E. Mills. 1987. "The Determinants of County Growth." *Journal of Regional Science* 27:39-54.
- Carruthers, J.I., and G.F. Mulligan (2005) "Environmental Valuation: Connecting Theory, Evidence, and Public Policy." In *Presented at the meetings of the Associated Collegiate Schools of Planning in Ft. Worth, Texas*.
- Carruthers, J.I., and G.F. Mulligan. 2008. "A Locational Analysis of Growth and Change in American Metropolitan Areas*." *Papers in Regional Science* 87:155-171.
- Census. 2013. *Building Permits Survey: United States and Census Region*. Washington D.C.
- . 2011. *Selected Historical Decennial Census, Urban and Rural Definitions and Data*. Washington DC.
- . 2000. *Survey of Income and Program Participation*. Washington DC.
- Chan, S. 2001. "Spatial Lock-in: Do Falling House Prices Constrain Residential Mobility?" *Journal of Urban Economics* 49:567-586.
- Clark, W.A.V., M.C. Deurloo, and F.M. Dieleman. 1994. "Tenure Changes in the Context of Microlevel Family and Macrolevel Economic Shifts." *Urban Studies* 31:135-152.
- Cliff, A., and J. Ord. 1973. *Spatial Autocorrelation*. London: Pion.
- Cliff, A., and J. Ord. 1981. *Spatial Process: Models and Applications*. London: Pion.
- Coulson, N.E.a.F., Lynn M. 2009. "Housing Tenure and Labor Market Impacts: The Search Goes On." *Journal of Urban Economics* 65:252-264.
- Debarsy, N., C. Ertur, and J.P. LeSage. 2012. "Interpreting dynamic space-time panel data models." *Statistical Methodology* 9:158-171.
- Deller, S.C., and V. Lledo. 2007. "Amenities and Rural Appalachia Economic Growth." *Agricultural and Resource Economic Review* 36:107-132.
- Deller, S.C., V. Lledo, and D.W. Marcouiller. 2008. "Modeling Regional Economic Growth with a Focus on Amenities." *Review of Urban & Regional Development Studies* 20:1-21.

- Deller, S.C., T.H. Tsai, D.W. Marcouiller, and D.B.K. English. 2001. "The Role of Amenities and Quality of Life in Rural Economic Growth." *American Journal of Agricultural Economics* 83:352-365.
- Deloitte. 2009. *Perspectives on Real Estate Uncovering Opportunity in a Distressed Market*. Available at http://www.deloitte.com/assets/Dcom-UnitedStates/Local%20Assets/Documents/FSI/us_fsi_RE_PerspectiveRE_March10.pdf. Accessed on December 13, 2013.
- Dieleman, F., and P. Everaers. 1994. "From Renting to Owning: Life Course and Housing Market Circumstances." *Housing Studies* 9:11-26.
- DiPasquale, D., R. Forslid, and E. Glaeser. 2000. "Incentives and Social Capital: Are Homeowners Better Citizens?" *Journal of International Economics* 50:497-517.
- Donovan, C., and C. Schnure. 2011. "Locked in the House: Do Underwater Mortgages Reduce Labor Mobility?" Unpublished, Freddie Mac, National Association of Real Estate Investment Trusts.
- Doreian, P. 1981. "Estimating Linear Models with Spatially Distributed Data." In S. Leinhardt ed. *Sociological Methodology*. San Francisco, The Jossey-Bass Inc., pp. 359-388.
- Doreian, P. 1980. "Linear Models with Spatially Distributed Data: Spatial Disturbances or Spatial Effects." *Sociological Methods & Research* 9:29-60.
- Elhorst, J.P. 2010. "Spatial Panel Models." In M.M. Fischer, and A. Getis eds. *Handbook of Applied Spatial Analysis*. Vienna, Springer.
- Engelhardt, G. 1995. "House Prices and Home Owner Saving Behavior." *Regional Science and Urban Economics* 26.
- Engelhardt, G.V. 2003. "Nominal Loss Aversion, Housing Equity Constraints, and Household Mobility: Evidence from the United States." *Journal of Urban Economics* 53:171-195.
- Farber, H.S. 2012. "Unemployment in the Great Recession: Did the Housing Market Crisis Prevent the Unemployed from Moving to Take Jobs?" *American Economic Review* 102:520-525.
- FDIC. 2010. *2007 Annual Report*.
- Ferreira, F., J. Gyourko, and J. Tracy. 2010. "Housing Busts and Housing Mobility." *Journal of Urban Economics* 68:34-45.
- Forrest, R., and A. Murie. 1989. "Differential Accumulation: Wealth, Inheritance and Housing Policy Considered." *Policy and Politics* 17:25-39.
- Frame, D.E. 2008. "Regional Migration and House Price Appreciation." *International Real Estate Review* 11:96-112.
- Frey, W.H. 2012. *Population Growth in Metro America Since 1980: Putting the Volatile 2000s in Perspective*. Washington D.C.
- Gallin, J. 2006. "The Long-Run Relationship between House Prices and Income: Evidence from Local Housing Markets." *Real Estate Economics* 34:417-438.
- Gebremeriam, G.H. 2006. "Modeling Small Business Growth, Migration Behavior, Local Public Services and Median Household Income in Appalachia: A Spatial Simultaneous Equations Approach." Ph.D. Dissertation, West Virginia University.
- Gebremeriam, G.H., T.G. Gebremedhin, P.V. Schaeffer, T.T. Phipps, and R.W. Jackson. 2007. "A Spatial Panel Simultaneous-Equations Model of Business Growth, Migration Behavior, Local Public Services and Household Income in Appalachia " Paper presented at American Agricultural Economics Association. Portland, OR, July 29 - August 1, 2007.
- Goodman, A.C. 1988. "An Econometric Model of Housing Price, Permanent Income, Tenure Choice and Housing Demand." *Journal of Urban Economics* 23:327-353.
- Gottschalk, A., M. Vornovysky, and A. Smith. 2013. *Household Wealth in U.S.: 2000 to 2011*. Washington D.C.
- Grebler, L., D. Blank, and L. Winnick. 1956. *Capital Formation in Residential Real Estate*. Princeton: National Bureau of Economic Research and Princeton University Press.

- Green, R.K. 1997. "Follow the Leader: How Changes in Residential and Non-Residential Investment Predict Changes in GDP." *Real Estate Economics* 25:253.
- Greene, W.H. 2002. *Econometric Analysis*. 7 ed., ed. Moston, MA: Cambridge University Press.
- Gyourko, J., C. Mayer, and T. Sinai. 2006. "Superstar Cities." Unpublished, National Bureau of Economic Research.
- Hamnett, C. 1991. "Home Ownership, Housing Wealth, and Wealth Distribution in Britain." Paper presented at Conference Housing Policy as a Strategy for Change. Oslo, Norway.
- Haurin, D.R., and H.L. Gill. 1987. "Effects of Income Variability on the Demand for Owner-Occupied Housing." *Journal of Urban Economics* 22:136-150.
- Henderson, J.V., and Y. Ioannides. 1979. "A Model of Housing Tenure Choice." *American Economic Review* 73:98-113.
- Henderson, J.V., and Y.M. Ioannides. 1983. "A Model of Housing Tenure Choices." *American Economic Review* 73:98-113.
- Henderson, J.V., and Y.M. Ioannides. 1987. "Owner Occupancy: Investment vs Consumption Demand." *Journal of Urban Economics* 21:228-241.
- Henderson, R. 2012. *Industry Employment and Output Projections to 2020*. Washington D.C.
- HUD, U.S. "U.S. Housing Market Conditions, Regional." U.S. Department of Housing and Urban Development.
- Jeanty, P.W., M. Partridge, and E. Irwin. 2010. "Estimation of a Spatial Simultaneous Equation Model of Population Migration and Housing Price Dynamics." *Regional Science and Urban Economics* 40:343-352.
- Johnes, G., and T. Hyclak. 1999. "House Prices and Regional Labor Market." *The Annals of Regional Science* 33:33-49.
- Kahsai, M.S. 2009. "An Analysis of Amenity-Led Rural Economic Development in Northeast Region: A Spatial Simultaneous Equations Approach." Ph.D Dissertation, West Virginia University.
- Kalenkoski, C.M., and D.J. Lacombe. 2011. "Minimum Wages and Teen Employment: A Spatial Panel Approach." Unpublished, Regional Research Institute.
- Kelejian, H.H., and I.R. Prucha. 2004. "Estimation of Simultaneous Systems of Spatially Interrelated Cross Sectional Equations." *Journal of Econometrics* 118:27-50.
- Kelejian, H.H., and I.R. Prucha. 1999. "A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model." *International Economic Review* 40:509-533.
- Kendig, H.L. 1984. "Housing Careers, Life Cycle and Residential Mobility: Implications for the Housing market." *Urban Studies* 4:271-283.
- Kennedy, P. 2008. *A Guide to Econometrics*. 6th ed ed. Cambridge, MA: Wiley-Blackwell.
- Kirby, D.K., and J.P. LeSage. 2009. "Changes In Commuting To Work Times Over The 1990 to 2000 Period." *Regional Science and Urban Economics*.
- LeSage, J.P. 1997. "Regression Analysis of Spatial Data." *Journal of Regional Analysis and Policy* 27:83-94.
- LeSage, J.P., and M.M. Fischer. 2009. "The Impact of Knowledge Capital on Regional Total Factor Productivity." Unpublished, Working Paper Series, Social Science Research Network.
- LeSage, J.P., and R.K. Pace. 2009. *Introduction to Spatial Econometrics*. Boca Raton: Chapman & Hall/CRC Taylor & Francis Group, LLC.
- Leung, C. 2004. "Macroeconomics and Housing: A Review of the Literature." *Journal of Housing Economics* 13:249-267.
- Malpezzi, S. 1999. "A Simple Error Correction Model of House Prices." *Journal of Housing Economics* 8:27-62.
- Mayer, C.J., and C.T. Somerville. 2000. "Residential Construction: Using the Urban Growth Model to Estimate Housing Supply." *Journal of Urban Economics* 48:85-109.

- Modestino, A.S., and J. Dennett. 2013. "Are American Homeowners Locked into their Houses? The Impact of Housing Market Conditions on State-to-State Migration." *Regional Science and Urban Economics* 43:322-337.
- Mojica, M. 2009. "Examining the Role of Entrepreneurship in Economic Development Appalachia." Ph.D. Dissertation, West Virginia University.
- Munch, J.R., M. Rosholm, and M. Svarer. 2006. "Are Homeowners Really More Unemployed?" *The Economic Journal* 116:991-1013.
- Nondo, C. 2009. "An Empirical Analysis of the Interactions between Environmental Regulations and Economic Growth." Ph.D. Dissertation, West Virginia University.
- Ofori, G., and S.S. Han. 2003. "Abraham J.H., P., Bubbles in Metropolitan Housing Markets, *Journal of Housing Research* 7(1996), 191-207." *Abraham J.H., P., Bubbles in Metropolitan Housing Markets, Journal of Housing Research* 7(1996), 191-207 27:37-62.
- Oswald, A.J. 1996. "A Conjecture of the Explanation for High Unemployment in the Industrialized Nations: Part 1." Unpublished, University of Warwick, Department of Economics.
- Plaut, S.E. 1987. "The Timing of Housnig Tenure Transition." *Journal of Urban Economics* 21:312-322.
- Quigley, J.M. 1999. "Real Estate Prices and Economic Cycles." *International Real Estate Review* 2:1-20.
- Roback, J. 1982. "Wages, Rents, and the Quality of Life." *Journal of Political Economy* 90:1257-1278.
- Robbins, D.K., L.J. Pantuosco, D.F. Parker, and B.K. Fueller. 2000. "An Empirical Assessment of the Contribution of Small Business Employment to U.S. State Economic Performance." *Small Business Economics* 15:293-302.
- Robst, J., R. Deitz, and K. McGoldbrick. 1999. "Income variability, uncertainty, and housing tenure choice." *Regional Science and Urban Economics* 29:219-229.
- Rosenthal, L. 1989. "Income and Price Elasticities of Demand for Owner-Occupied Housing in the Uk: Evidence from Pooled Cross-Sectional and Time-Series Data." *Applied Economics* 22:761-775.
- Rudzitis, G. 1999. "Amenities Increasingly Draw People to the Rural West." *Rural Development Perspective* 14:9-13.
- Saks, R. 2008. "Job Creation and Housing Construction: Constraints on Metropolitan Area Employment Growth." *Journal of Urban Economics* 64:178-195.
- Schaeffer, P.V., M.S. Kahsai, and R.W. Jackson. 2013. "Beyond the Rural-Urban Dichotomy: Essay in Honor of Professor A. M. Isserman." *International Regional Science Review* 36:81-96.
- Schiller, R. 2007. "Understanding Recent Trends in House Prices and Home Ownership." Unpublished, National Bureau of Economic Research.
- Schmitt, J., and K. Warner. "Deconstructing Structural Unemployment." Center for economic and Policy Research.
- Shaffer, R., S. Deller, and D. Marcouiller. 2006. "Rethinking Community Economic Development." *Economic Development Quarterly* 20:59-74.
- Sinding, S. 2009. "Population, poverty and economic development." *Philosophical Transaction of the Royal Society B* 364:3023-3030.
- Steinnes, D.N. 1982. "Do People Follow Jobs or Do Jobs Follow People? A Causality Issue in Urban Economics." *Urban Studies* 19:187-192.
- Tobler, W. 1970. "A computer movie simulating urban growth in the Detroit region." *Economic Geography* 46:234-240.
- Turner, T.M., and H. Luea. 2009. "Homeownership, Wealth Accumulation and Income Status." *Journal of Housing Economics* 18:104-114.
- USDA-ERS. 2013. *Rural Urban Continuum Codes*. <http://www.ers.usda.gov/data-products/rural-urban-continuum-codes.aspx#.UznOqPldV8E>. Accessed on 3/25/2013.

- Valleta, R. 2013. "House Lock and Structural Unemployment." *Labor Economics* In review.
- Vias, A.C., and G.F. Mulligan. 1999. "Integrating Economic Base Theory with Regional Adjustment Models: The Nonmetropolitan Rocky Mountain West Region." *Growth and Change* 30:507-525.
- Wen, Y. 2001. "Residential Investment and Economic Growth." *Annals of Economics and Finance*:437-444.
- Wooldridge, J.M. 2010. *Introductory Econometrics*. 3 ed., ed. Boston, MA: Cambridge University Press.
- Zabel, J. 2012. "Migration, Housing Market, and Labor Market Responses to Employment Shocks." *Journal of Urban Economics* 72:267-284.
- Zhou, J. 2010. "Testing for Cointegration Between House Prices and Economic Fundamentals." *Real Estate Economics* 38:599-632.

Appendix

Figures 9-12 show the county level change in median housing value from 1980 to 2010. Figures 13-16 show the county level change in population from 1980 to 2010. Figures 17-20 show the county level change in employment from 1980 to 2010. Figures 21-24 show the county level change in median income from 1980 to 2010. All data used from the figures is from decennial census surveys in 1980, 1990, 2000, and 2010.

Figures 9 through 24 all use quintiles to display the spatial pattern of change over time. Figures 9-12 show that the bottom quintile for median housing value tend to be rural counties in Pennsylvania, New York, and West Virginia. Eastern urban counties are predominantly in the highest quintile for median home value. Population, employment, and median income follow similar trends, but metropolitan counties that include cities like Pittsburgh Pennsylvania, and Buffalo New York remain higher quintile counties in lower quintile regions.

Figure 8. Median Housing Value per County 1980 (Data Source: U.S. Census, 1980)

Figure 9. Median Housing Value per County 1990 (Data Source: U.S. Census, 1990)

Figure 10. Median Housing Value per County 2000 (Data Source: U.S. Census, 2000)

Figure 11. Median Housing Value per County 2010 (Data Source: U.S. Census, 2010)

Figure 12. Population per County 1980 (Data Source: U.S. Census, 1980)

Figure 13. Population per County 1990 (Data Source: U.S. Census, 1990)

Figure 14. Population per County 2000 (Data Source: U.S. Census, 2000)

Figure 15. Population per County 2010 (Data Source: U.S. Census, 2010)

Figure 16. Employment per County 1980 (Data Source: U.S. Census, 1980)

Figure 17. Employment per County 1990 (Data Source: U.S. Census, 1990)

Figure 18. Employment per County 2000 (Data Source: U.S. Census, 2000)

Figure 19. Employment per County 2010 (Data Source: U.S. Census, 2010)

Figure 20. Median Income per County 1980 (Data Source: U.S. Census, 1980)

Figure 21. Median Income per County 1990 (Data Source: U.S. Census, 1990)

Figure 22. Median Income per County 2000 (Data Source: U.S. Census, 2000)

Figure 23. Median Income per County 2010 (Data Source: U.S. Census, 2010)

Figure 24. Median Housing Value per county Moran's I test result.

(Data Source: U.S. Census, 1980, 1990, 2000, 2010)

Figure 25. Population per county Moran's I test result.

(Data Source: U.S. Census, 1980, 1990, 2000, 2010)

Figure 26. Employment per county Moran's I test result.

(Data Source: U.S. Census, 1980, 1990, 2000, 2010)

Figure 27. Median Income per county Moran's I test result.

(Data Source: U.S. Census, 1980, 1990, 2000, 2010)