Spatial Analysis of the Financial Crisis: Modeling Financial Clusters Across the New York Metropolitan Area

Silvia Maria Lorenzo
CUNY Hunter College

Recommended Citation
http://academicworks.cuny.edu/hc_sas_etds/11

How does access to this work benefit you? Let us know!
Follow this and additional works at: http://academicworks.cuny.edu/hc_sas_etds
Part of the Geographic Information Sciences Commons
Spatial Analysis of the Financial Crisis:
Modeling Financial Clusters Across the New York Metropolitan Area

by

Silvia Maria Lorenzo

Submitted in partial fulfillment
of the requirements for the degree of
Master of Arts, Geography, Hunter College
The City University of New York

2015

Thesis Sponsor:

December 31, 2015
Date

Dr. Hongmian Gong
First Reader

December 31, 2015
Date

Dr. Carson J.Q. Farmer
Second Reader
Dedication

Para Mami. Gracias por todo.
Acknowledgements

I would like to thank my thesis advisor and graduate advisor, Dr. Hongmian Gong, and my thesis co-advisor, Dr. Carson Farmer, for their guidance, encouragement and support. Thank you to all the faculty and staff of the Geography Department at Hunter College of the City University of New York; and thank you to the Society of Woman Geographers; Saul and Miriam Cohen; Graduate Student Association; Graduate Geography Association; Hunter GIS Club; and OpenStreetMap for your support. I’m especially grateful for all of the constructive feedback received while presenting early chapters of this thesis at the Association of American Geographers’ Annual Meeting, the Graduate Student Association’s Graduate Student Symposium, the Society of Woman Geographers’ members meeting, and at the Fourth Global Conference on Economic Geography. I would also like to extend an extra special thank you to my family for their love and support.
# Table of Contents

Abstract ................................................................................................................................. i

List of Tables .......................................................................................................................... ii

List of Figures ........................................................................................................................ iv

Chapter 1. Introduction ........................................................................................................... 1

Chapter 2. Literature Review ................................................................................................. 3

Chapter 3. Data and Methods ............................................................................................... 16

3.1 Spatial Weights Matrix ................................................................................................. 18

3.2 Spatial Autocorrelation ................................................................................................. 19

3.3 Cluster Analysis ............................................................................................................. 20

3.4 Spatial Regression Analysis ............................................................................................ 22

Chapter 4. Results .................................................................................................................. 25

4.1 Spatial Autocorrelation Results ..................................................................................... 25

4.2 Cluster Analysis Results ................................................................................................. 26

4.3 Spatial Regression Analysis Results ............................................................................... 45

Chapter 5. Conclusion ............................................................................................................. 67

5.1 Future Directions ............................................................................................................ 70

Appendix A. Preliminary Cluster Analysis I .......................................................................... 72

Appendix B. Preliminary Cluster Analysis II ......................................................................... 88

Appendix C. Preliminary Geographically Weighted Regression ......................................... 91

References .............................................................................................................................. 108
Abstract

Clusters of credit and securities intermediaries across the New York metropolitan area are identified during the 2000s using spatial statistics. Geographically weighted regression is then used to model the processes underlying the changes of credit and securities clusters identified with specific attention to taxes, median household income, transportation and employment.
List of Tables

Table 1. State Business Tax Climate Index listing New Jersey with the worst business tax rate ranking of the U.S. .................................................................................................................................7

Table 2. State and Local Sales Tax Rates listing New York with the highest combined rate of the Tri-State region........................................................................................................................................8

Table 3. Global Moran’s I statistic for cluster analysis in 2000, 2005 and 2010..................................................25

Table 4. List of ZIP codes in each cluster of extreme values identified in 2000..................................................30

Table 5. List of ZIP codes in each cluster of extreme values identified in 2005..................................................31

Table 6. List of ZIP codes in each cluster of extreme values identified in 2010..................................................32

Table 7. Summary of Geographically Weighted Regression model performance for credit and securities establishments by employment size in 2000 and 2010........................................48

Table 8. List of variables used for the 2000 GWR regression model analysis and the significance of each variable according to the Exploratory Regression.................................................53

Table 9. List of variables used for the 2010 GWR regression model analysis and the significance of each variable according to the Exploratory Regression.................................................49

Table 10. Global Moran’s I statistic for the residuals for the 2000 and 2010 GWR models...........49

Table 11. List of ZIP codes constituting each cluster of extreme values identified for credit and securities establishments with 250 establishments or more for each ZIP code in 2000.............................................................................................................................................77

Table 12. List of ZIP codes constituting each cluster of extreme values identified for credit and securities establishments with 250 establishments or more for each ZIP code in 2005.............................................................................................................................................78

Table 13. List of ZIP codes constituting each cluster of extreme values identified for credit and securities establishments with 250 establishments or more for each ZIP code in 2010.............................................................................................................................................79
Table 14. Summary of Geographically Weighted Regression model performance for total credit and securities establishments in 2000 and 2010.................................................................93

Table 15. List of variables used for the regression model analyses total credit and securities establishments in 2000 and the significance of each variable according to the Exploratory Regression.................................................................................................................................94

Table 16. List of variables used for the regression model analyses total credit and securities establishments in 2010 and the significance of each variable according to the Exploratory Regression.................................................................................................................................94
List of Figures

Figure 1. State tax rate for the highest income bracket of individuals........................................9

Figure 2. State tax rate for the highest income bracket for corporations..................................9

Figure 3. Map depicts total paid employees in the NAICS 522 and NAICS 523 financial sectors
throughout the 2000s. The MSA for counties in New York, New Jersey and Connecticut are shown.................................................................12

Figure 4. Map depicts annual payroll in the NAICS 522 and NAICS 523 financial sectors
throughout the 2000s for the study area................................................................................13

Figure 5. Map depicts total establishments in the NAICS 522 and NAICS 523 financial sectors
throughout the 2000s. The Metropolitan Statistical Area (MSA) for counties in New York,
New Jersey and Connecticut are shown.................................................................................14

Figure 6. New York Metropolitan Area used as the study area..................................................18

Figure 7. Map of Gi* statistic identifying clusters of extreme values for 2000..........................26

Figure 8. Map of Gi* statistic identifying clusters of extreme values for 2005..........................27

Figure 9. Map of Gi* statistic identifying clusters of extreme values for 2010..........................28

Figure 10. Map of Manhattan/New York County study area shows a cluster of extreme values
for credit and securities establishments in 2000 in the traditional finance district center
of Wall Street ZIP code 10005, while also showing a significant cluster in and near
Midtown................................................................................................................................34

Figure 11. Map of Manhattan/New York County study area in 2005 shows the cluster
of credit and securities establishments of the traditional finance district center
of Wall Street ZIP code 10005 has shrunk in size in Lower Manhattan..............................35

Figure 12. Map of Manhattan/New York County study area shows a cluster
of credit and securities establishments in 2010 in the traditional finance district center
of Wall Street ZIP code 10005............................................................................................36

Figure 13. Map of Hudson County study area shows a cluster of credit and securities
establishments in 2000 for Jersey City ZIP code 07310 and 07302.....................................37
Figure 14. Map of Hudson County study area shows a reduction hot spots of credit and securities establishments in 2005.  

Figure 15. Map of Hudson County study area shows a cluster of credit and securities establishments in 2010 is maintained in Jersey City ZIP code 07310 and also includes Jersey City ZIP code 07302.  

Figure 16. Map of Fairfield County study area in 2000 without a significant cluster of credit and securities establishments.  

Figure 17. Map of Fairfield County shows a cluster of credit and securities establishments in 2005 for Stamford ZIP codes 06901 and 06902.  

Figure 18. Map shows a reduction in hot spots of credit and securities establishments in 2010 for Stamford ZIP code 06901 and Greenwich ZIP code 06830.  

Figure 19. Map shows hot spots of credit and securities establishments in 2000 for Melville, Suffolk County ZIP code 11747.  

Figure 20. Map shows hot spots of credit and securities establishments in 2005 for Melville, Suffolk County ZIP code 11747.  

Figure 21. Map shows hot spots of credit and securities establishments in 2010 for Melville, Suffolk County.  

Figure 22. Map of standardized residuals resulting from the Geographically Weighted Regression analysis for the credit and securities industry in 2000.  

Figure 23. Map of local squared residuals resulting from the Geographically Weighted Regression analysis for the credit and securities industry in 2000.  

Figure 24. Map of standardized residuals resulting from the Geographically Weighted Regression analysis for the credit and securities industry in 2010.  

Figure 25. Map of local squared residuals resulting from the Geographically Weighted Regression analysis for the credit and securities industry in 2010.  

Figure 26. Map of spatial distribution of the median household income coefficient from the Geographically Weighted Regression analysis for 2000.  

Figure 27. Map of spatial distribution of coefficient on employment in finance, insurance and real estate industries from the Geographically Weighted Regression analysis for 2000.
Figure 28. Map of spatial distribution of commuting to working via public transportation coefficient from the Geographically Weighted Regression analysis for 2000........................................59

Figure 29. Map of spatial distribution of the median real estate taxes coefficient from the Geographically Weighted Regression analysis for 2000.................................................................60

Figure 30. Map of spatial distribution of coefficient on employment in information and communication technology industries from the Geographically Weighted Regression analysis for 2000........................................................................................................61

Figure 31. Map of spatial distribution of median household income coefficient from the Geographically Weighted Regression analysis for 2010.................................................................62

Figure 32. Map of spatial distribution of coefficient on employment in finance, insurance and real estate industries from the Geographically Weighted Regression analysis for 2010........................................................................................................63

Figure 33. Map of spatial distribution of commuting to working via public transportation coefficient from the Geographically Weighted Regression analysis for 2010..............................................64

Figure 34. Map of spatial distribution of median real estate taxes coefficient from the Geographically Weighted Regression analysis for 2010........................................................................65

Figure 35. Map of spatial distribution of coefficient on employment in information and communication technology industries from the Geographically Weighted Regression analysis for 2010........................................................................................................66

Figure 36. Map of Gi* statistic identifying clusters of extreme values for credit and securities establishments with 250 establishments or more in 2000.........................................................73

Figure 37. Map of Gi* statistic identifying clusters of extreme values for credit and securities establishments with 250 establishments or more in 2005.........................................................74

Figure 38. Map of Gi* statistic identifying clusters of extreme values for credit and securities establishments with 250 establishments or more in 2010.........................................................75

Figure 39. Map of Manhattan/New York County study area shows a cluster of extreme values for credit and securities establishments with 250 establishments or more for each ZIP code in 2000........................................................................................................79

Figure 40. Map of Manhattan/New York County study area shows the cluster for credit and securities establishments with 250 establishments or more for each ZIP code in 2005 ........................................................................................................80
Figure 41. Map of Manhattan/New York County study area shows the cluster for credit and securities establishments with 250 establishments or more for each ZIP code in 2010. 

Figure 42. Map of Fairfield County study area without a significant cluster of credit and securities establishments with 250 establishments or more for each ZIP code in 2000. 

Figure 43. Map of Fairfield County shows a cluster of credit and securities establishments with 250 establishments or more for each ZIP code in 2005 for Stamford ZIP codes. 

Figure 44. Map shows a reduction in the quantity of credit and securities establishments with 250 establishments or more for each ZIP code in 2010 for Stamford ZIP codes. 

Figure 45. Map of Hudson County study area shows a cluster of credit and securities establishments with 250 establishments or more for each ZIP code in 2000 for Jersey City and Union City ZIP codes. 

Figure 46. Map of Hudson County study area shows a reduction in the quantity of credit and securities establishments with 250 establishments or more for each ZIP code in 2005. 

Figure 47. Map of Hudson County study area shows a cluster of credit and securities establishments with 250 establishments or more for each ZIP code in 2010. 

Figure 48. Map of Gi* statistic identifying clusters of extreme values for total credit and securities establishments for each ZIP code in 2000. 

Figure 49. Map of Gi* statistic identifying clusters of extreme values for total credit and securities establishments for each ZIP code in 2005. 

Figure 50. Map of Gi* statistic identifying clusters of extreme values for total credit and securities establishments for each ZIP code in 2010. 

Figure 51. Map of standardized residuals resulting from the Geographically Weighted Regression analysis for total credit and securities establishments in 2000. 

Figure 52. Map of local squared residuals resulting from the Geographically Weighted Regression analysis for total credit and securities establishments in 2000. 

Figure 53. Map of standardized residuals resulting from the Geographically Weighted Regression analysis for total credit and securities establishments in 2010.
Figure 54. Map of local squared residuals resulting from the Geographically Weighted Regression analysis for total credit and securities establishments in 2010.................................97

Figure 55. Map of spatial distribution of commuting to working via public transportation coefficient from the Geographically Weighted Regression analysis for 2000.................................99

Figure 56. Map of spatial distribution of coefficient on employment in finance, insurance and real estate industries from the Geographically Weighted Regression analysis for 2000 ..................................................................................................................100

Figure 57. Map of spatial distribution of the median household income coefficient from the Geographically Weighted Regression analysis for 2000..............................................101

Figure 58. Map of spatial distribution of the median real estate taxes coefficient from the Geographically Weighted Regression analysis for 2000..............................................102

Figure 59. Map of spatial distribution of commuting to working via public transportation coefficient from the Geographically Weighted Regression analysis for 2010.................................103

Figure 60. Map of spatial distribution of median household income coefficient from the Geographically Weighted Regression analysis for 2010..............................................104

Figure 61. Map of spatial distribution of coefficient on employment in finance, insurance and real estate industries from the Geographically Weighted Regression analysis for 2010 ..................................................................................................................105

Figure 62. Map of spatial distribution of median real estate taxes coefficient from the Geographically Weighted Regression analysis for 2010..............................................106

Figure 63. Map of spatial distribution of coefficient on employment in information industry from the Geographically Weighted Regression analysis for 2010.................107
1. Introduction

Considered the worst financial crisis in history, the global financial crisis of 2007-2008 is associated with unprecedented economic losses in terms of both size and scale. With the influx of research assessing the economic impact of the financial crisis, a spatial analysis based on empirical data is needed to better understand the underlying processes of the financial geography as well as the spatial significance of the financial crisis in New York, a key global financial center also considered the origin of the crisis. The existence and location of financial clusters specializing in credit and securities throughout the New York metropolitan area are identified for the time period before, during the height of the financial bubble, and after the crisis across ZIP code, county and state boundaries using various spatial statistical methods.

Building on these results, this thesis identifies factors influencing the formation of clusters of financial activity specific to credit and securities intermediaries across 538 ZIP codes comprising the New York metropolitan area using a Geographically Weighted Regression for 2000 and 2010. Processes underlying the formation of financial geographies in the New York metropolitan area are examined with specific attention to tax regimes, employment, technology, transportation as well as household income. This analysis thus provides empirical research and unique methodologies useful for financial risk management and public policy initiatives aimed at addressing economic sustainability within and across state boundaries, while also developing the groundwork for further research on a spatial analysis of the global financial crisis.
Thesis Objectives

The goal of this thesis is to examine the role of the global financial crisis on regional financial geographies across the New York metro area by using spatial statistics to identify clusters of credit and securities intermediaries and to model the processes underlying the changes of credit and securities clusters at the scale of the ZIP code across the Tri-State region. To address this goal, the research accomplishes the following objectives:

• Chapter 2 reviews literature on financial crises from economic and economic geography disciplines. Grounded in this research, clusters of the financial services industry specific to credit and securities intermediaries are defined for this paper; and factors contributing to the formation of credit and securities clusters are identified.

• Chapter 3 develops methods to (1) examine the effects and significance of space on the location of credit and securities intermediaries, (2) empirically identify hot spots of clusters of the credit and securities industry throughout the New York metro area, and (3) to model the processes underlying the clusters using Geographically Weighted Regression models over time.

• Chapter 4 discusses policy implications for each variable included as part of the regression models at a regional and local scale over time.
2. Literature Review

As an introduction to a special issue of the Journal of Economic Geography focusing on geographies of finance, Engelen and Faulconbridge (2009) assert the need for “geographers to shift their focus of research to finance”, the crux of the present day economy. Published towards the end of the global financial crisis, this issue signals the influx of research across disciplines using longitudinal case studies to examine the processes working synergistically to spur on the global financial crisis (Aalbers 2009; Harvey 2011; Lee, Clark, et al 2009; Sassen 2011; Leichenko, O’Brien and Solecki 2010; Pani and Holman 2014). The literature states the dominance of international financial centers in large urban cities signals the importance of finance and, specifically, credit and securities markets in the present day economy. As reinforced in both economic and economic geography literature, New York is considered the global financial center of the U.S. and the origin of the global financial crisis (Bernanke 2013; Aalbers 2009; Martin 2010; Lee, Clark, et al 2009; Stiglitz 2010; Sassen 2011; Harvey 2011). Within this global financial center, the growth of financial intermediaries specializing in credit and securities signal the onset of financial crises (Bernanke 1983; Gertler and Kyotaki 2010; Sassen 2001; Stiglitz 2010). For this reason, this thesis uses empirical data on the financial services industry specific to credit and securities intermediaries in the New York metropolitan area for 2000 and 2010 marking the beginning and end of the 2000s as well as the time period before and after the height of the global financial crisis in 2008 for the basis of the spatial regression analysis.
The global financial crisis demonstrated the importance of geography across disciplines (Ross 2014; Krugman 2011; Stiglitz 2010; Porter 1998; Harvey 2011) and marked a new turning point in the use of geography to examine regional economies. Economic literature up until recently failed to account for geographic factors which were traditionally considered exogenous to economic models (Krugman 1998; Ross 2014; Porter 1998). As a conceptual framework to address local studies of industries and regional economies, research on clusters of economic activity across political boundaries gained considerable attention as part of economics and economic geography research (Storper and Walker 1989; Storper and Scott 1995; Porter 1998; Martin and Sunley 2003). Cluster analysis forms an integral part of the larger conceptual framework of agglomeration economies used to account for the role of geography on concentrations of economic activity (Porter 1996; Rosenthal and Strange 2004). Economic literature defines clusters as “geographic concentrations” of interconnected companies and institutions in a particular field (Porter 1998: 4; OECD 2010; World Bank 2009). As part of this research, clusters are primarily defined by the interconnections of business transactions and relationships across industries and institutions, and thus are not confined to political boundaries. Clusters are considered key to agglomeration economies and economic growth; and thus cluster research has been used as a policy tool for regional economic growth. Indeed, cluster research and the emphasis on geography has increased its impact on public policy as seen with the use of cluster analysis by the OECD (2001, 2007, 2010) and the World Bank (2009). Still, there is lack of empirical support for the identification and formation of clusters (Martin and Sunley 2003; Krugman 1998), and there is a need for an empirical spatial analysis as part of economic geography research (Martin 2003; Monteiro 2011; Peet and Thrift 1989; Storper and Walker
1989; Krugman 1998; Krugman 2011; Wu, Ji and Su 2012; Tonts and Taylor 2013). Without the use of an empirical spatial analysis, the spatial effects of the financial crisis remain largely unexamined, thus limiting the value of economic geography research for public policy initiatives. Grounded in economic geography research, this thesis will identify the concentration of credit and securities intermediaries which played the largest role in the credit crunch of the financial crisis as well as the factors that influenced the location of clusters of credit and securities markets using spatial statistics and exploratory spatial data analysis in the framework popularized by Anselin (1989) while also making use of Geographically Weighted Regression models. The spatial analysis methodologies and findings discussed in this paper may prove beneficial to regional planning policies as well as financial risk management.

This thesis will examine the influence of real estate taxes, household income, transportation networks, employment and technological advances on the clustering of the credit and securities industry in 2000 and 2010 across the New York metropolitan area at the scale of the ZIP code. These factors are considered to influence the formation and location of financial services clusters (Porter 1998) and also serve as sources of agglomeration economies by matching labor markets and providing knowledge spillovers (Rosenthal and Strange 2004; Duranton and Puga 2004).

Advances in technology are considered an important factor influencing the location of financial services establishments (Longcore and Rees 1996; Gong and Keenan 2012) and are also considered a factor contributing to the overall growth of the finance industry via the internationalization of securities trading in the 1980s (Sassen 2001; Harvey 2011; Longcore and
Rees 1996). The dispersion of the traditional financial district was examined in the 1980s (Longcore and Rees 1996), where, faced with spatial constraints of limited plot sizes and limited technological capacity, financial establishments sought newer, larger buildings outside of Wall Street (Longcore and Rees 1996). This observed move coincided with the growth of the securities and credit industry as well as advances in technology (Longcore and Rees 1996, Gong and Keenan 2012). Office spaces in Midtown Manhattan in New York City allowed for the building infrastructure capable of supporting the technology and computing needs of the growing financial institutions (Longcore and Rees 1996).

Transit is considered an important factor of agglomeration economies across industries and urban environments (Venables 2007; Tabuchi 1998; Japzon and Gong 1995; Gong and Keenan 2012). Midtown Manhattan serves as the prominent transportation hub in New York City with Grand Central Terminal, the largest and busiest railroad transit facility in the world, and Penn Station, one of the largest in North America, providing direct connection from Midtown Manhattan to Stamford, Connecticut and Jersey City, New Jersey as well as connections to corporate and academic institutional investors in the Tri-State region of New York, New Jersey and Connecticut (Elphinstone 2007). Transportation networks connecting Fairfield County, Connecticut to NYC allows workers to maintain close proximity to the financial sector of Wall Street with the combined benefits of a comparably lower real-estate value, lower property taxes, high income levels, as well as lower crime rates to allow for a higher quality of living (Elphinstone 2007; CNN 2005; CNN 2010).
Tax-regimes, or the taxes of each level of government, are an important consideration for the location of financial establishments and also serve as an important factor determining the residential preference for employees and investors (Stowell 2012: 256; Fung and Hseih 1999: 318). At the state level, according to the State Business Tax Climate index, the tax regime in Connecticut is considered more favorable to businesses than the tax regime in New York and New Jersey, which were ranked the lowest of the fifty states in the country (Tax Foundation 2015). Similarly, at the county level, Nassau County and Westchester County in New York State have the highest taxes and the least establishments of financial firms, while Fairfield County, Connecticut has the lowest taxes and the most financial firm establishments (City of Stamford 2012; NYS ORPTS 2013a; NYS ORPTS 2013b; NYS ORPTS 2013e; NYS ORPTS 2013f; Pruner 2011).

Tax exemptions and deductions are a way for the government to provide incentives for businesses (Kocieniewski 2011a, Kocieniewski 2011b; Williams 1988).

Table 1. State Business Tax Climate Index listing New Jersey with the worst business tax rate ranking of the U.S. (Tax Foundation 2015)

<table>
<thead>
<tr>
<th>State</th>
<th>2011 Rank</th>
<th>2012 Rank</th>
<th>2015 Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Connecticut</td>
<td>40</td>
<td>40</td>
<td>42</td>
</tr>
<tr>
<td>New York</td>
<td>48</td>
<td>49</td>
<td>49</td>
</tr>
<tr>
<td>New Jersey</td>
<td>50</td>
<td>50</td>
<td>50</td>
</tr>
</tbody>
</table>
Table 2. State and Local Sales Tax Rates listing New York with the highest combined rate of the Tri-State region (Tax Foundation 2015)

<table>
<thead>
<tr>
<th>State</th>
<th>State Sales Tax Rate</th>
<th>Rank</th>
<th>Local Sales Tax Rate</th>
<th>Combined Rate</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Connecticut</td>
<td>6.35%</td>
<td>11</td>
<td>None</td>
<td>6.35%</td>
<td>31</td>
</tr>
<tr>
<td>New Jersey</td>
<td>7.00%</td>
<td>2</td>
<td>-0.03%</td>
<td>6.97%</td>
<td>22</td>
</tr>
<tr>
<td>New York</td>
<td>4.00%</td>
<td>38</td>
<td>4.48%</td>
<td>8.48%</td>
<td>7</td>
</tr>
</tbody>
</table>

Since the 1990s, New Jersey, New York and Connecticut have competed for businesses by providing tax breaks and state subsidies (Bagli 2012; Timmons 2001; Slatin 2006). General Electric, the largest corporation in the U.S. and also a large institutional investor headquartered in Fairfield County, Connecticut, has avoided paying federal and state corporate taxes since 2008 (Kocieniewski 2011a; Kocieniewski 2011b; Forbes 2012). Since the late 1990s, the growth of General Electric Company’s financial division, GE Capital, accounted for a significant share of General Electric Company’s profitability (Lohr 2012; General Electric 2012). Most of the earnings in GE Capital result from their investment in hedge funds (New York Times 2012).
Figure 1. State tax rate for the highest income bracket of individuals (Tax Foundation 2012)

Figure 2. State tax rate for the highest income bracket for corporations (Tax Foundation 2012)
Before the global financial crisis, Fairfield County, Connecticut had approximately 9.5 percent of the global hedge fund assets and was considered one of the primary centers for hedge funds worldwide (Hedge Fund Intelligence 2007; Teo 2009). Hedge funds are included as part NAICS 523 securities intermediation of the financial services industry. Of the world’s 351 funds with more than $1 billion in assets during the height of the financial bubble, 143 were based in Upper East Side, NYC and Greenwich, Connecticut (Gross 2007). According to a survey of hedge funds conducted in 2007, Connecticut emerged as the third most popular center for hedge funds globally with about 30 hedge funds based in towns such as Greenwich, Stamford, and Westport – all located in Fairfield County - with a combined asset of almost $170 billion, about 10.5% of total hedge funds in the world (Hedge Fund Intelligence 2007). The growth of securities intermediation in this area is largely attributed to Connecticut’s favorable tax climate (Stowell 2012: 256; Fung and Hseih 1999: 318; Bagli 2012).

Connecticut is also home to one of the highest per capita income communities in the U.S. and is home to large academic institutional investors and pension funds (Northeast Utilities Companies 2012). Indeed, according to the Connecticut Department of Labor (Dyer 2007), the NAICS 523 securities industry in Connecticut is among the world’s highest paid and fastest growth industry. Conversely, despite the comparably lower housing and rent expenses, Jersey City has not attracted the concentration of high income earners working and living in Jersey City to the extent of Stamford primarily due to taxes. As compared to Connecticut as a whole, New Jersey has a higher state tax rate as shown in Figures 1 and 2. Jersey City is also considered to have a significantly higher crime rate and lower median household income levels than cities in
Connecticut in 2005 and 2010 (CNN 2005; CNN 2010). At the scale of the county, payroll for the credit and securities industry across the New York metropolitan area has increased from 2000 to 2010, as shown in Figure 4, while employment has decreased as shown in Figure 3. Total establishments of the credit and securities industry have decreased in New York County, New York and Hudson County, New Jersey from 2000 to 2010 as shown in Figure 5. In Fairfield County, Connecticut, however, total credit and securities establishments have increased. At the scale of the ZIP code, data on payroll for the credit and securities industry was not available (U.S. Census Bureau 2013b) and thus cannot be used for the spatial analysis at the ZIP code level.
Figure 3. Map depicts total paid employees in the NAICS\textsuperscript{1} 522 and NAICS 523 financial sectors throughout the 2000s. The MSA for counties in New York, New Jersey and Connecticut are shown (U.S. Census Bureau 2013a).

\textsuperscript{1} NAICS stands for North American Industry Classification System. The NAICS subsectors 522 and 523 will be defined in detail as part of Chapter 2 Data & Methods.
Figure 4. Map depicts annual payroll in the NAICS 522 and NAICS 523 financial sectors throughout the 2000s for the MSA study area (U.S. Census Bureau 2013a).
Having identified factors important to the financial services industry in the New York metro area, this thesis will use spatial statistics to identify clusters of credit and securities intermediaries across the New York metro area during the 2000s before employing the use of Geographically Weighted Regression to model the processes underlying the changes of credit and securities clusters. As identified in the literature, processes influencing the location of financial establishments include real estate taxes, household income, transportation, employment and technological advances. These factors are identified in the literature as having an important influence on the formation and location of financial services clusters (Porter 1998).
and also serve as sources of agglomeration economies by matching labor markets and providing knowledge spillovers (Rosenthal and Strange 2004; Duranton and Puga 2004). A spatial regression analysis of these factors is needed to properly identify and explain the formation of clusters. In doing so, policies aim at ensuring the viability of the financial services industries throughout the New York metro region could greatly benefit by a spatial regression analysis to account for spatial effects of the financial services industry. If overlooked, the spatial effects unique to this industry may prove problematic to the application of public policies across regions. As of yet this analysis on the factors influencing the formation and location of clusters of credit and securities intermediaries across the New York metropolitan area at the ZIP code level for 2000 and 2010 marking the beginning and end of the 2000s as well as the time period before and after the height of the global financial crisis in 2008 for the basis of the spatial regression analysis has not been attempted. A spatial regression analysis is necessary in better understanding the financial crisis in the New York metropolitan area as well as in better understanding the global financial crisis.
3. Data & Methods

The study area includes 538 ZIP codes that form 11 counties across the Metropolitan Statistical Area for New York and the MSA for Connecticut as defined by the U.S. Census (U.S. Census Bureau 2013b). To identify the concentration of credit and securities intermediaries which played the largest role in the credit crunch of the financial crisis, this thesis uses ZIP Business Patterns (U.S. Census Bureau 2013b) to obtain the quantity of financial establishments dealing specifically with credit and securities intermediation as defined by the North American Industry Classification System subsectors 522 and 523. NAICS 522 and 523 subsectors are defined by the North American Industry Classification System (U.S. Census Bureau 2015) as follows:

**NAICS 522: Industries in the Credit Intermediation and Related Activities**

- Subsector group establishments that (1) lend funds raised from depositors; (2) lend funds raised from credit market borrowing; or (3) facilitate the lending of funds or issuance of credit by engaging in such activities as mortgage and loan brokerage, clearinghouse and reserve services, and check cashing services.

**NAICS 523: Industries in the Securities, Commodity Contracts, and Other Financial Investments and Related Activities**

- Subsector group establishments that are primarily engaged in one of the following: (1) underwriting securities issues and/or making markets for securities and commodities; (2) acting as agents (i.e., brokers) between buyers and sellers of securities and commodities; (3) providing securities and commodity exchange services; and (4) providing other services, such as managing portfolios of assets; providing investment advice; and trust, fiduciary, and custody services.
To better assess the size of each establishment across the study area, the quantity of credit and securities establishments for each ZIP code area is multiplied by the bottom range of the employment size of each establishment. This dataset, which serves as the dependent variable of the cluster analysis and regression analysis, will be referred to as the credit and securities industry in this paper. To assess the spatial effects of the credit and securities industry throughout the 2000s and identity credit and securities clusters, several statistical methods need to be implemented (Krivoruchko, Gotway, et al 2003; Anselin 1999). This chapter will define the spatial structure of the dataset using a spatial weights matrix. With a spatial weights matrix, the spatial autocorrelation of the credit and securities industry will be examined using the Moran’s I statistic to verify statistically the existence of clustering throughout the entire study area. The location of clusters will then be identified using Getis Ord Gi* statistic. Moran’ I measures the global level of spatial association and the average Gi* derived z-score measures the overall local level of spatial association. Before an assessment of spatial autocorrelation and hot spot analysis, a spatial weights matrix is needed.
3.1 Spatial Weights Matrix

In order to account for the spatial structure of the dataset, a spatial weights matrix is needed. A spatial weights matrix provides a representation of the spatial structure of the study area (Anselin 1999; Cliff and Ord 1981; Getis and Aldstadt 2010; Mitchell 2005). When computing Moran’s I statistic, the spatial weights matrix $W$ defines the spatial relationships among all features of the dataset. The spatial weights matrix created used the Inverse distance conceptualization of spatial relationships and a Euclidean distance method to calculate distances from each ZIP code (i) to neighboring ZIP code in the study area. Using the spatial weights matrix,
each feature \((i)\) within a distance of the \((j)\) ZIP code is given a weight of one, and all other features are given a weight of zero (Mitchell 2005; Zhang and Murayama 2000). Using Inverse Euclidean distance, the weights which represent distance, are inverted so that nearer features have a higher weight than features that are farther away.

According to Haining (2003) and Cliff and Ord (1981) inverse distance is one of the most common distances used for spatial-weighting matrices. Euclidean distance is considered the preferred distance method for this study due to the use of categorical financial data allocated to ZIP code areas for counties in three contiguous states with disparate transportation networks. Additionally, research studies using exploratory spatial data analysis to examine categorical data based in social science research have used the inverse distance conceptualization of spatial relationships and a Euclidean distance method to calculate distances (Wu, Ji and Su 2011).

### 3.2 Spatial Autocorrelation

The most widely used global statistic to assess the spatial autocorrelation and to measure spatial dependence is the Global Moran’s I (Boots 2003; Mitchell 2005; Cliff and Ord 1972; Cliff and Ord 1973; Getis and Ord 1992; Getis 2008; Goodchild 1986). Moran’s I is used here to measure the spatial autocorrelation of the quantity of credit and securities establishments spanning 538 ZIP codes for 2000, 2005 and 2010 time periods. Derived from the statistic developed by Moran (1948, 1950) to assess the correlation coefficient between random variables, Cliff and Ord (1972, 1973) modified the statistic to test for spatial autocorrelation of regression residuals (Jin and Lee 2010). This modified Moran’s statistic is referred to as the Moran’s Index (commonly referred to
as Moran’s I). Applications of the Moran’s I are used throughout the social and environmental sciences. The global Moran’s I statistic as used in this thesis compares the differences between attribute values of each pair of features in the dataset to the mean attribute value for the dataset. The statistic uses the entire dataset to derive a single mean value for the study area to compare with each feature. In doing so, the assumption is that the study area is homogenous (Boots 2003; Unwin and Unwin 1998; Fotheringham and Brunsdon 1999), and thus the processes underlying the data values are stationary throughout the study area (Boots 2003; Unwin 1996). For this reason, Moran’s I is used to assess the overall pattern and trend of the dataset but may over-generalize the spatial processes of the study area. Moran’s I is based on a normal distribution approximation, using a standardized z-value obtained from expressions for the mean and variance of the statistic (Cliff and Ord 1972; Cliff and Ord 1973; Moran 1950). To accurately measure distances, spatial data was projected to New York State Plane.

3.3 Cluster Analysis

To better account for local spatial processes and to identify clusters, the Getis Gi* is used. In contrast to the Moran’s I, the local measure of spatial autocorrelation such as the Getis Gi* used in this paper, examines spatial dependence in subsets defined with respect to each feature (i) as compared to neighboring features in the data site (j), and thus provides a single value for each data site in the study area (Boots 2002; Boots 2003; Getis and Ord 1992; Ord and Getis 1995; Mitchell 2005). By assessing each feature within the context of neighboring features, the statistic helps avoids the tendency of over-generalizing spatial processes and thus provides local measures of spatial autocorrelation. The Gi* statistic is a z-score used to identify spatial clusters.
of extreme values relative to the mean (Ord and Getis 1995; Boots 2002; Tiefelsdorf and Boots 1997). Using z-scores, analysis using Gi* allows for comparison between different time periods. A high value of Gi* is interpreted as clustering of extreme high values, and a low value of Gi* indicates clustering of extreme low values. Values less than 2.0 standard deviations and greater than -2.0 from the mean indicates no significant clustering of extreme values (Tiefelsdorf and Boots 1997).

To study the spatial effects throughout the timeframe of the financial crisis, the same three time periods were examined for the spatial autocorrelation and cluster analyses: 2000, 2005 and 2010. These dates are used to mark the time periods before the crisis, during the height of the crisis, and the aftermath of the crisis. The Gi* statistic was calculated for the ZIP code study area for 2000, 2005 and 2010. The Gi* statistic assigns a measure of the level of spatial association at each ZIP code area, highlighting areas which have a high number of credit and securities establishments. Computing the Gi* statistic throughout the 2000s allows for a comparison of the credit and securities establishments throughout the New York metropolitan area before and after the height of the financial crisis. For this analysis, a cluster of credit and securities establishments is defined as a grouping of spatially adjacent, statistically significant high values. This delineation agrees with both spatial analysis literature and economic geography literature.
3.4 Regression Analysis

Regression analysis allows for a better understanding of the factors influencing the values of a target variable by statistically estimating the relationships among variables and can be used to assist in predicting values at a different time periods or geographic locations. Regression analysis includes many techniques for modeling and analyzing several explanatory variables. If the residuals of a linear regression are found to cluster or have spatial autocorrelation, a Geographically Weighted Regression (GWR) can then be used to account for the spatial structure unique to a dataset such as the credit and securities industry across the New York metro area. GWR works by allowing for the relationship between the dependent variable and the explanatory variables to vary across the study area. For a detailed review of a geographically weighted regression model see Fotheringham, Brunsdon and Charlton (2003) and Mitchell (2005).

Using ZIP Business Patterns (U.S. Census Bureau 2013b), the quantity of credit and securities establishments of each ZIP code across the study area in 2000 and 2010 was used as the dependent variable for the regression model in 2000 and 2010 respectively. For 2000, explanatory variables derived from the 2000 Census Long Form survey included new topics on economic, social and housing characteristics. Beginning in 2010, some of these topics were no longer included as part of the decennial census. For this reason, the explanatory variables used in the 2010 regression model were derived from tables of comparable characteristics of the American Community Survey five-year estimates dataset for 2009 to 2013.
Over 80 other variables were examined as part of a stepwise regression technique used to fit the regression models in 2000 and 2010. For both 2000 and 2010, an exploratory Ordinary Least Squares (OLS) regression was used to test the statistical significance and magnitude of each possible explanatory coefficient variable. Ultimately, the explanatory variables which provided the best model performance were median real estate taxes, median household income, employment in finance, insurance and real estate industries (which includes rental and leasing services), as well as the use of public transportation to commute to work, and employment in the information and communications technology industry. Clustering of the regression residuals was present for the OLS regression model indicating the spatial structure of the data under examination is not properly specified. For a properly specified model, there should be no clustering of standardized residuals. Statistically significant clustering or spatial autocorrelation of the regression model residuals means at least one key explanatory variable is missing from the regression model (Haining 2003; Cliff and Ord 1972). The residuals of the OLS regression were tested for spatial autocorrelation using Global Moran’s I. In the presence of detect spatial autocorrelation, a GWR model is used to account for spatial effects unique to the financial activity under examination. The use of a GWR model to account for spatial effects unique to the financial activity under examination may be useful in the presence of detected spatial autocorrelation.

Having identified spatial autocorrelation in the OLS regression for 2000 and 2010, GWR was then used to model the clusters of financial activity specific to credit and securities in 2000 and 2010. The five variables included in both 2000 and 2010 regressions are real estate taxes, median
household income, employment in the finance, insurance and real estate-related industries, commuting to working using public transportation and employment in the information and communication technology industries. Just as in the OLS regression and the exploratory regression, an additional variable to account for employment in the information and communications technology industry was included in 2010. The regression parameters were the same for both the 2000 and 2010 GWR models: the bandwidth method used cross validation with an adaptive kernel type to allow for the bandwidth distance to change according to the density of the financial establishments across the study area.
4. Results

4.1 Spatial Autocorrelation Results

Results of the Moran's Index, as shown in Table 3, are positive for each time period which means the values in the study area tend to cluster spatially. That is, high values cluster near other high values; low values cluster near other low values (Mitchell 2005). Additionally, given the number of features in the dataset and the variance for the data values overall, the z-score and p-value indicate whether this difference is statistically significant. As shown in the Table 3, the z-score is positive for each time period, showing that there are clusters of high values in the dataset, and the p-value (<0.05) indicates that the clustering is statistically significant, thereby rejecting the null hypothesis of complete spatial randomness for the study area. In other words, the geography and spatial distribution of financial establishments is an important factor to consider.

<table>
<thead>
<tr>
<th></th>
<th>2000</th>
<th>2005</th>
<th>2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moran's Index</td>
<td>0.154157</td>
<td>0.141509</td>
<td>0.137823</td>
</tr>
<tr>
<td>Expected Index</td>
<td>-0.001862</td>
<td>-0.001862</td>
<td>-0.001862</td>
</tr>
<tr>
<td>Variance</td>
<td>0.000062</td>
<td>0.000061</td>
<td>0.000060</td>
</tr>
<tr>
<td>z-score</td>
<td>19.830409</td>
<td>18.312563</td>
<td>17.996002</td>
</tr>
<tr>
<td>p-value</td>
<td>0.000000</td>
<td>0.000000</td>
<td>0.000000</td>
</tr>
</tbody>
</table>
4.2 Cluster Analysis Results

Results of the Getis-Ord Gi* statistic, as shown in Figures 7-9, identified clusters of the credit and securities industry at the scale of the ZIP code area. Figures 7-9 map the spatial distribution of the Gi* derived z-scores for the study area for 2000, 2005 and 2010. The Getis-Ord Gi* hot spots of credit and securities establishments by employment size exist for ZIP codes in Lower Manhattan, Midtown Manhattan and Upper East Side areas of New York County as well as in Jersey City, Hudson County, New Jersey and Melville, Suffolk County, New York for each time period. Though hot spots are evident in Hudson County in 2000, hot spots do not appear in Fairfield County until after 2000.

Figure 7. Map of Gi* statistic identifying clusters of extreme values for 2000.
Figure 8. Map of Gi* statistic identifying clusters of extreme values for 2005.
Figure 9. Map of Gi* statistic identifying clusters of extreme values for 2010.
Table 4. List of ZIP codes constituting each cluster of extreme values identified in 2000.

<table>
<thead>
<tr>
<th></th>
<th>County</th>
<th>Neighborhood</th>
<th>ZIP</th>
<th>Employment Size of Financial Establishments</th>
<th>Gi* z-score</th>
<th>Gi* p-value</th>
<th>Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>New York</td>
<td>Lower Manhattan</td>
<td>10005</td>
<td>20,464</td>
<td>11.01398</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>New York</td>
<td>Midtown</td>
<td>10017</td>
<td>16,882</td>
<td>9.045794</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>New York</td>
<td>Upper East Side</td>
<td>10022</td>
<td>16,300</td>
<td>8.725184</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>New York</td>
<td>Midtown</td>
<td>10019</td>
<td>15,558</td>
<td>8.316601</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>New York</td>
<td>Midtown</td>
<td>10036</td>
<td>9,865</td>
<td>5.188655</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>New York</td>
<td>Lower Manhattan</td>
<td>10004</td>
<td>9,283</td>
<td>4.864072</td>
<td>0.000001</td>
<td>99%</td>
</tr>
<tr>
<td>7</td>
<td>New York</td>
<td>Lower Manhattan</td>
<td>10013</td>
<td>9,011</td>
<td>4.717105</td>
<td>0.000002</td>
<td>99%</td>
</tr>
<tr>
<td>8</td>
<td>New York</td>
<td>Lower Manhattan</td>
<td>10048</td>
<td>8,248</td>
<td>4.302106</td>
<td>0.000017</td>
<td>99%</td>
</tr>
<tr>
<td>9</td>
<td>New York</td>
<td>Midtown</td>
<td>10010</td>
<td>7,238</td>
<td>3.742204</td>
<td>0.000182</td>
<td>99%</td>
</tr>
<tr>
<td>10</td>
<td>New York</td>
<td>Midtown</td>
<td>10020</td>
<td>6,865</td>
<td>3.541721</td>
<td>0.000398</td>
<td>99%</td>
</tr>
<tr>
<td>11</td>
<td>Hudson</td>
<td>Jersey City</td>
<td>07302</td>
<td>6,438</td>
<td>3.299041</td>
<td>0.00097</td>
<td>99%</td>
</tr>
<tr>
<td>12</td>
<td>New York</td>
<td>Lower Manhattan</td>
<td>10006</td>
<td>6,108</td>
<td>3.127772</td>
<td>0.001761</td>
<td>99%</td>
</tr>
<tr>
<td>13</td>
<td>New York</td>
<td>Midtown</td>
<td>10001</td>
<td>5,107</td>
<td>2.571648</td>
<td>0.010122</td>
<td>95%</td>
</tr>
<tr>
<td>14</td>
<td>Hudson</td>
<td>Jersey City</td>
<td>07310</td>
<td>3,968</td>
<td>1.941755</td>
<td>0.052167</td>
<td>90%</td>
</tr>
<tr>
<td>15</td>
<td>New York</td>
<td>Lower Manhattan</td>
<td>10007</td>
<td>3,683</td>
<td>1.792947</td>
<td>0.072981</td>
<td>90%</td>
</tr>
<tr>
<td>16</td>
<td>Suffolk</td>
<td>Melville</td>
<td>11747</td>
<td>3,685</td>
<td>1.780804</td>
<td>0.074944</td>
<td>90%</td>
</tr>
<tr>
<td>17</td>
<td>New York</td>
<td>Midtown</td>
<td>10018</td>
<td>3,461</td>
<td>1.669917</td>
<td>0.094936</td>
<td>90%</td>
</tr>
</tbody>
</table>
Table 5. List of ZIP codes constituting each cluster of extreme values identified in 2005.

<table>
<thead>
<tr>
<th></th>
<th>County</th>
<th>Neighborhood</th>
<th>ZIP</th>
<th>Employment Size of Financial Establishments</th>
<th>Gi* z-score</th>
<th>Gi* p-value</th>
<th>Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>New York</td>
<td>Midtown</td>
<td>10017</td>
<td>18,860</td>
<td>10.624874</td>
<td>0</td>
<td>99%</td>
</tr>
<tr>
<td>2</td>
<td>New York</td>
<td>Upper East Side</td>
<td>10022</td>
<td>18,099</td>
<td>10.185672</td>
<td>0</td>
<td>99%</td>
</tr>
<tr>
<td>3</td>
<td>New York</td>
<td>Midtown</td>
<td>10019</td>
<td>15,850</td>
<td>8.888281</td>
<td>0</td>
<td>99%</td>
</tr>
<tr>
<td>4</td>
<td>New York</td>
<td>Lower Manhattan</td>
<td>10005</td>
<td>15,330</td>
<td>8.58743</td>
<td>0</td>
<td>99%</td>
</tr>
<tr>
<td>5</td>
<td>New York</td>
<td>Lower Manhattan</td>
<td>10013</td>
<td>8,985</td>
<td>4.930065</td>
<td>0.000001</td>
<td>99%</td>
</tr>
<tr>
<td>6</td>
<td>New York</td>
<td>Lower Manhattan</td>
<td>10004</td>
<td>8,757</td>
<td>4.796121</td>
<td>0.000002</td>
<td>99%</td>
</tr>
<tr>
<td>7</td>
<td>New York</td>
<td>Midtown</td>
<td>10020</td>
<td>7,738</td>
<td>4.216852</td>
<td>0.000025</td>
<td>99%</td>
</tr>
<tr>
<td>8</td>
<td>New York</td>
<td>Midtown</td>
<td>10036</td>
<td>7,348</td>
<td>3.989801</td>
<td>0.000066</td>
<td>99%</td>
</tr>
<tr>
<td>9</td>
<td>New York</td>
<td>Midtown</td>
<td>10010</td>
<td>6,341</td>
<td>3.40662</td>
<td>0.000658</td>
<td>99%</td>
</tr>
<tr>
<td>10</td>
<td>New York</td>
<td>Lower Manhattan</td>
<td>10007</td>
<td>6,322</td>
<td>3.397256</td>
<td>0.000681</td>
<td>99%</td>
</tr>
<tr>
<td>11</td>
<td>Fairfield</td>
<td>Stamford</td>
<td>06902</td>
<td>4,417</td>
<td>2.290145</td>
<td>0.022013</td>
<td>95%</td>
</tr>
<tr>
<td>12</td>
<td>New York</td>
<td>Lower Manhattan</td>
<td>10006</td>
<td>4,333</td>
<td>2.252355</td>
<td>0.0243</td>
<td>95%</td>
</tr>
<tr>
<td>13</td>
<td>New York</td>
<td>Midtown</td>
<td>10018</td>
<td>3,961</td>
<td>2.038239</td>
<td>0.041526</td>
<td>95%</td>
</tr>
<tr>
<td>14</td>
<td>Fairfield</td>
<td>Stamford</td>
<td>06901</td>
<td>3,811</td>
<td>1.940915</td>
<td>0.052269</td>
<td>90%</td>
</tr>
<tr>
<td>15</td>
<td>Hudson</td>
<td>Jersey City</td>
<td>07310</td>
<td>3,650</td>
<td>1.852068</td>
<td>0.064016</td>
<td>90%</td>
</tr>
<tr>
<td>16</td>
<td>New York</td>
<td>Midtown</td>
<td>10001</td>
<td>3,339</td>
<td>1.677214</td>
<td>0.093501</td>
<td>90%</td>
</tr>
<tr>
<td>17</td>
<td>Kings</td>
<td>Boerum Hill</td>
<td>11201</td>
<td>3,322</td>
<td>1.663149</td>
<td>0.096283</td>
<td>90%</td>
</tr>
<tr>
<td>18</td>
<td>Suffolk</td>
<td>Melville</td>
<td>11747</td>
<td>3,327</td>
<td>1.661178</td>
<td>0.096678</td>
<td>90%</td>
</tr>
</tbody>
</table>
Table 6. List of ZIP codes constituting each cluster of extreme values identified in 2010.

<table>
<thead>
<tr>
<th>County</th>
<th>Neighborhood</th>
<th>ZIP</th>
<th>Employment Size of Financial Establishments</th>
<th>Gi* z-score</th>
<th>Gi* p-value</th>
<th>Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 New York</td>
<td>Upper East Side</td>
<td>10022</td>
<td>19247</td>
<td>11.513945</td>
<td>0</td>
<td>99%</td>
</tr>
<tr>
<td>2 New York</td>
<td>Midtown</td>
<td>10017</td>
<td>18526</td>
<td>11.074662</td>
<td>0</td>
<td>99%</td>
</tr>
<tr>
<td>3 New York</td>
<td>Midtown</td>
<td>10019</td>
<td>13192</td>
<td>7.817139</td>
<td>0</td>
<td>99%</td>
</tr>
<tr>
<td>4 New York</td>
<td>Lower Manhattan</td>
<td>10005</td>
<td>11820</td>
<td>6.975885</td>
<td>0</td>
<td>99%</td>
</tr>
<tr>
<td>5 New York</td>
<td>Midtown</td>
<td>10036</td>
<td>9532</td>
<td>5.584168</td>
<td>0</td>
<td>99%</td>
</tr>
<tr>
<td>6 New York</td>
<td>Lower Manhattan</td>
<td>10004</td>
<td>9476</td>
<td>5.543307</td>
<td>0</td>
<td>99%</td>
</tr>
<tr>
<td>7 New York</td>
<td>Midtown</td>
<td>10020</td>
<td>8465</td>
<td>4.935494</td>
<td>0.000001</td>
<td>99%</td>
</tr>
<tr>
<td>8 New York</td>
<td>Midtown</td>
<td>10010</td>
<td>5284</td>
<td>2.988172</td>
<td>0.002807</td>
<td>99%</td>
</tr>
<tr>
<td>9 New York</td>
<td>Lower Manhattan</td>
<td>10013</td>
<td>4921</td>
<td>2.764922</td>
<td>0.005694</td>
<td>99%</td>
</tr>
<tr>
<td>10 Fairfield</td>
<td>Stamford</td>
<td>06901</td>
<td>4883</td>
<td>2.735679</td>
<td>0.006225</td>
<td>99%</td>
</tr>
<tr>
<td>11 Hudson</td>
<td>Jersey City</td>
<td>07310</td>
<td>4740</td>
<td>2.652393</td>
<td>0.007992</td>
<td>99%</td>
</tr>
<tr>
<td>12 Hudson</td>
<td>Jersey City</td>
<td>07302</td>
<td>4720</td>
<td>2.639817</td>
<td>0.008295</td>
<td>99%</td>
</tr>
<tr>
<td>13 New York</td>
<td>Midtown</td>
<td>10018</td>
<td>4385</td>
<td>2.443409</td>
<td>0.014549</td>
<td>95%</td>
</tr>
<tr>
<td>14 Suffolk</td>
<td>Melville</td>
<td>11747</td>
<td>3755</td>
<td>2.046578</td>
<td>0.0407</td>
<td>95%</td>
</tr>
<tr>
<td>15 New York</td>
<td>Lower Manhattan</td>
<td>10007</td>
<td>3701</td>
<td>2.021354</td>
<td>0.043243</td>
<td>95%</td>
</tr>
<tr>
<td>16 New York</td>
<td>Lower Manhattan</td>
<td>10016</td>
<td>3109</td>
<td>1.663478</td>
<td>0.096217</td>
<td>90%</td>
</tr>
<tr>
<td>17 Fairfield</td>
<td>Greenwich</td>
<td>06830</td>
<td>3116</td>
<td>1.656742</td>
<td>0.097572</td>
<td>90%</td>
</tr>
</tbody>
</table>
The spatial distribution of clustering in extreme values is varied throughout the New York metropolitan area for 2000, 2005 and 2010. In 2000, there were a total of 17 ZIP codes identified as hot spots of extreme values of credit and securities establishments across the New York metro area, which increased to 18 ZIP codes in 2005, and which dropped to 17 in 2010. As further detailed in Tables 4-6, 13 ZIP codes were found to remain for the time period before the financial crisis, 2000; the time period during the height of the financial bubble, 2005; and immediately after the financial crisis, 2010. Of these 13 ZIP codes, four are located in Lower Manhattan, six are located in Midtown and one in the Upper East Side of New York County as mapped in Figures 10-12. Lower Manhattan ZIP code 10005, of the Wall Street financial district, had the highest Gi* z-score for the entire New York metro area in 2000. For 2005, Midtown Manhattan had the region’s highest Gi* z-score; and in 2010, the Upper East Side of Manhattan had the region’s highest Gi* z-scores. As shown in Tables 4-6, more hot spots of the credit and securities industry were located in ZIP code areas in Midtown Manhattan than in the financial district of Lower Manhattan for 2000, 2005 and 2010.

Outside of New York County, the hot spot analysis indicates Jersey City, Hudson County and Melville, Suffolk County also have a statistically significant cluster of credit and securities establishments by employment size for 2000, 2005 and 2010. Credit and securities clusters were expected in Hudson County, New Jersey and Fairfield County, Connecticut as observed in the literature. The hot spot identified in ZIP code 11747 of Melville, Suffolk County was not highlighted in the literature and, thus, is an unexpected hot spot of the credit and securities industry.
In Fairfield County, clusters were primarily spotted in Stamford for 2005 and 2010. The literature, however, highlights Greenwich as the primary location for credit and securities clusters. To address this issue, an alternative cluster analysis was done as shown in Appendix A and B. In Appendix B, the cluster analysis used the total quantity of credit and securities establishments. Using the total quantity of credit and securities establishments, hot spots were identified in Greenwich for 2000, 2005 and 2010. While Appendix B focuses on the total quantity of financial establishments, Appendix A focused on areas with a large quantity of financial establishments. To do so, the cluster analysis for Appendix A examined areas with an excess of 250 financial establishments. Similar to the cluster analysis used for this paper (see Figures 7-9), the cluster analysis in Appendix A found hot spots in Stamford in 2005 and 2010, but did not find hot spots in Greenwich for 2000, 2005 and 2010.

Comparing the cluster analysis of Figures 7-9 and the cluster analysis in Appendix A and Appendix B provides a better understanding of the types of clusters that concentrate in Fairfield County, Connecticut. The clusters that appear in Greenwich, as shown in Appendix B, but do not appear in this chapter indicate that the credit and securities establishments in Greenwich are smaller by employment size but larger in quantity of establishments. Conversely, the clusters identified in Stamford for 2000, 2005 and 2010, as shown in Figures 7-9, but only appear in 2010 for the cluster analysis in Appendix B indicate that the credit and securities establishments in Stamford have less establishments but more employees per establishment.
Figure 10. Map of Manhattan/New York County study area shows a cluster of extreme values for credit and securities establishments in 2000 in the traditional finance district center of Wall Street ZIP code 10005, while also showing a significant cluster in and near Midtown.
Figure 11. Map of Manhattan/New York County study area in 2005 shows the cluster of credit and securities establishments of the traditional finance district center of Wall Street ZIP code 10005 has shrunk in size in Lower Manhattan. Hot spots in and near Midtown remain.
Figure 12. Map of Manhattan/New York County study area shows a cluster of credit and securities establishments in 2010 in the traditional finance district center of Wall Street ZIP code 10005, while also showing a significant cluster in and near Midtown.
Figure 13. Map of Hudson County study area shows a cluster of credit and securities establishments in 2000 for Jersey City ZIP code 07310 and 07302.
Figure 14. Map of Hudson County study area shows a reduction in hot spots of credit and securities establishments in 2005. Jersey City ZIP code 07310 remains as a hot spot.
Figure 15. Map of Hudson County study area shows a cluster of credit and securities establishments in 2010 is maintained in Jersey City ZIP code 07310 and also includes Jersey City ZIP code 07302.
Figure 16. Map of Fairfield County study area without a significant cluster of credit and securities establishments in 2000.
Figure 17. Map of Fairfield County shows a cluster of hot spot of credit and securities establishments in 2005 for Stamford ZIP codes 06901 and 06902.
Figure 18. Map shows a reduction in the hot spot of credit and securities establishments in 2010 for Stamford ZIP code 06901 and Greenwich ZIP code 06830.
Figure 19. Map shows the hot spot of credit and securities establishments in 2000 for Melville, Suffolk County ZIP code 11747.
Figure 20. Map shows the hot spot of credit and securities establishments in 2005 for Melville, Suffolk County ZIP code 11747.
Figure 21. Map shows the hot spot of credit and securities establishments in 2010 for Melville, Suffolk County ZIP code 11747.
4.3 Spatial Regression Results

Results of the GWR, as shown in Table 7 and Figures 22-25, indicate the regression model for 2000 is a better fit than the 2010 regression model of the financial services industry specific to credit and securities intermediaries. According to the R-Squared statistic, the regression in 2000 accounts for 56.43 percent of the variation in the credit and securities industry, the dependent variable, with an *adjusted* R-Squared of 48.58 percent. By comparison, the regression in 2010 accounts for 47.92 percent of the variation in the dependent variable with an *adjusted* R-Squared of 39.50 percent. The 2000 model explains approximately 40 percent of the variation of the dependent variable, the credit and securities industry across the New York metropolitan area at the scale of the ZIP code. The 2010 model, by comparison, explains well over 45 percent of the location and formation of the credit and securities industry. As shown in Tables 8 and 9, the factors influencing the location of clusters for both the 2000 and 2010 regression model pertain to median real estate taxes, median household income, transportation, employment in the finance, insurance and real estate-related industries as well as employment in the information industry.

Despite the higher R-Squared and *adjusted* R-squared values for the 2000 regression model, other measures of model performance indicate the 2010 regression model is a better fitting model than 2000 regression model. The sum of squared residuals in 2010 is lower than the sum of squared residuals in 2000; and the regression model for 2010 has a smaller effective number, a smaller sigma, and a smaller AICc as shown in Table 7. The optimal adaptive neighbors is higher in 2010 than in 2000. This means each local computation in the 2010 model was based on more
neighbors than the local computations for the model in 2000. This also means that spatial effects were more pronounced in 2000.

As shown in Figures 22 and 24, the map of standardized residuals highlights the under and over predictions of the model for each year. Maps of standardized residuals for each time period show statistically significant clustering, with the map for the 2010 standardized residuals showing a greater extent of clustering, as confirmed with the Global Moran’s I statistic used on the regression residuals listed in Table 10. For 2010, given the z-score of 18 and a p-value less than 0.05, there is a high probability the clustering is not random. For 2000, given the z-score of -0.13 and a p-value greater than 0.05, indicates there is a low probability that the clustering is attributable to spatial randomness.

The maps of local R-Squared, Figures 23 and 25, show the ZIP code areas where the model is performing best. For both time periods the model performs best within and nearest the five counties of New York City (New York, Queens, Kings, Bronx and Richmond counties) as well as Union City and Jersey City in Hudson County, New Jersey, and Stamford in Fairfield County, Connecticut.

Table 7. Summary of Geographically Weighted Regression model performance for credit and securities establishments by employment size in 2000 and 2010.

<table>
<thead>
<tr>
<th>Measures of Model Performance</th>
<th>2000</th>
<th>2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neighbors</td>
<td>106</td>
<td>123</td>
</tr>
<tr>
<td>Residual Squares</td>
<td>775424487</td>
<td>751664078.7</td>
</tr>
<tr>
<td>Effective Number</td>
<td>83.034019</td>
<td>75.446452</td>
</tr>
<tr>
<td>Sigma</td>
<td>1305.510272</td>
<td>1277.532131</td>
</tr>
<tr>
<td>AICc</td>
<td>9303.192973</td>
<td>9238.631097</td>
</tr>
<tr>
<td>R2</td>
<td>0.564376</td>
<td>0.479247</td>
</tr>
<tr>
<td>R2Adjusted</td>
<td>0.485829</td>
<td>0.39507</td>
</tr>
</tbody>
</table>
Table 8. List of variables used for the 2000 GWR regression model analysis and the significance of each variable according to the Exploratory Regression.

<table>
<thead>
<tr>
<th>Explanatory Variables in 2000</th>
<th>Significance %</th>
<th>Positive %</th>
<th>Negative %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median Household Income</td>
<td>62.50</td>
<td>25.00</td>
<td>75.00</td>
</tr>
<tr>
<td>Employment in the Finance, Insurance or Real Estate Industry</td>
<td>81.25</td>
<td>100.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Commuting to work using public transportation</td>
<td>43.75</td>
<td>43.75</td>
<td>56.25</td>
</tr>
<tr>
<td>Median Real Estate Taxes</td>
<td>56.25</td>
<td>0.00</td>
<td>100.00</td>
</tr>
<tr>
<td>Employment in Information and Communications Tech Industry</td>
<td>100.00</td>
<td>100.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Table 9. List of variables used for the 2010 GWR regression model analysis and the significance of each variable according to the Exploratory Regression.

<table>
<thead>
<tr>
<th>Explanatory Variables in 2010</th>
<th>Significance %</th>
<th>Positive %</th>
<th>Negative %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median Household Income</td>
<td>50.00</td>
<td>50.00</td>
<td>50.00</td>
</tr>
<tr>
<td>Employment in the Finance, Insurance or Real Estate Industry</td>
<td>100.00</td>
<td>100.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Commuting to work using public transportation</td>
<td>62.50</td>
<td>62.50</td>
<td>37.50</td>
</tr>
<tr>
<td>Median Real Estate Taxes</td>
<td>18.75</td>
<td>18.25</td>
<td>81.75</td>
</tr>
<tr>
<td>Employment in Information Industry</td>
<td>100.00</td>
<td>100.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Table 10. Global Moran’s I statistic for the residuals for the 2000 and 2010 GWR models.

<table>
<thead>
<tr>
<th>Measures of Spatial Autocorrelation</th>
<th>2000</th>
<th>2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moran's Index</td>
<td>-0.002929</td>
<td>0.016308</td>
</tr>
<tr>
<td>Expected Index</td>
<td>-0.001862</td>
<td>-0.001862</td>
</tr>
<tr>
<td>Variance</td>
<td>0.000067</td>
<td>0.000064</td>
</tr>
<tr>
<td>z-score</td>
<td>-0.130571</td>
<td>17.996002</td>
</tr>
<tr>
<td>p-value</td>
<td>0.896115</td>
<td>0.000000</td>
</tr>
</tbody>
</table>
Figure 22. Map of standardized residuals resulting from the Geographically Weighted Regression analysis for the credit and securities industry in 2000.
Figure 23. Map of local squared residuals resulting from the Geographically Weighted Regression analysis for the credit and securities industry in 2000.
Figure 24. Map of standardized residuals resulting from the Geographically Weighted Regression analysis for the credit and securities industry in 2010.
Figure 25. Map of local squared residuals resulting from the Geographically Weighted Regression analysis for the credit and securities industry in 2010.
Figures 26-35 map the spatial distribution of each coefficient included in the regression model for 2000 and 2010. The positive or negative relationship of each explanatory variable for each ZIP code area is also provided. The impact of each coefficient on the credit and securities industry, the dependent variable, is depicted differently depending on the strength of the correlation and the positive or negative influence of each coefficient. As opposed to the summary of variable significance derived from the exploratory OLS regression analysis (as listed in Tables 8 and 9), maps of the spatial distribution of coefficient values in Figures 26-35 show variation at the scale of the ZIP code area across the New York metro area for each coefficient for the 2000 and 2010 models. While the summary of variable significance derived from the exploratory OLS regression model is helpful in understanding a broad relationship between dependent and independent variables, the maps of the spatial distribution of GWR model coefficients show the extent of regional variation that exists at the ZIP code level. This spatial distribution is useful for planning studies at a local and regional level by mapping the empirically-based commonalities across county, city and state boundaries given the coefficients identified in the regression model.

The spatial distribution of median real estate taxes in 2000 and 2010 depicts the influence of this coefficient on the credit and securities industry as shown in Figures 29 and 34. As expected, the relationship of the real estate taxes coefficient is stronger for ZIP code areas in Fairfield County, Connecticut than for other ZIP code areas in Hudson County, New Jersey where fewer credit and securities intermediaries are found. This finding agrees with the literature which suggests that
Connecticut is found to have the lowest property taxes of the New York metro area. For 2000 and 2010, there is a negative relationship for ZIP codes in Manhattan/New York County where Wall Street, the traditional financial district, as well as for ZIP codes in Fairfield County and for those in Hudson County, New Jersey. As expected for these areas, as the property taxes increase, the quantity of credit and securities intermediaries decreases.

For 2000 and 2010, the impact of median household income on the quantity of credit and securities establishments was found to vary across the region, as shown in Figures 26 and 31. A negative correlation is found for the median household income for ZIP code areas with the greatest quantity of financial clusters such as Lower and Midtown Manhattan, as well as for areas where concentrations of the credit and securities industry were observed in the literature. A positive correlation was found in Melville, Suffolk County in 2000 only. Melville was identified as a hot spot in the results of the cluster analyses in section 4.2. The spatial distribution for median household income indicates that as median household income increases for ZIP areas identified as hot spots of the credit and securities industry, the employment size of the credit and securities industry of that ZIP code area decreases. This agrees with the data at the scale of the county reviewed for this thesis, as shown in Figures 3 and 4, where payroll for the credit and securities industry has increased from 2000 to 2010 while employment has decreased (U.S. Census Bureau 2013a).

For 2000 and 2010, the impact of using public transportation to commute to work on the credit and securities industry also varies across the region, as shown in Figures 28 and 33. For ZIP codes
in Fairfield County, Connecticut there is a positive correlation between the use of public transportation to commute to work and the employment size of credit and securities establishments of the ZIP code area. Conversely, for ZIP codes in Manhattan, the use of public transportation to commute to work is negatively correlated with the employment size of credit and securities establishments. That is, the lower the use of public transportation to commute to work, the greater the employment size of credit and securities establishments. For Manhattan, this may mean that people residing in the ZIP codes with the highest concentration of credit and securities intermediaries may use a private mode of transportation to commute to work; or, alternatively, this negative correlation may mean that people already live within the ZIP code area where the credit and securities clusters are observed in New York County. For Fairfield County, the positive correlation suggests that as the use of public transportation to commute to work increases, the employment size of credit and securities establishments increases. In addition to this positive correlation, the coefficient value of the use of public transportation to commute to work was also the highest for Fairfield County, Connecticut in 2000; and it was the highest of the clusters identified across the region in 2010. Thus, the spatial distribution for the use of public transportation to commute to work indicates that there is regional variation across the New York metro area for the impact of this coefficient on the clustering of the employment size of credit and securities establishments. For Fairfield County, there is a stronger reliance on the use of public transportation to commute to work for people residing in the ZIP code area where a concentration of credit and securities establishments by employment size was identified. For areas in Manhattan/New York County, Hudson County and Suffolk County, an
alternative mode of transportation is used more than public transportation to commute to work for these areas where clusters of credit and securities industry were identified.

For 2000 and 2010, there is an overall positive correlation between employment in the finance, insurance, real estate, rental and leasing industries and the concentration of the credit and securities industry across the region. As shown in Figures 27 and 32, the positive relationship is stronger in ZIP code areas within or nearest the clusters in New York County, New York and Hudson County, New Jersey. This positive correlation is expected, as the synergistic advantages of agglomeration economies are shown to positively influence the formation and location of credit and securities clusters by providing matching labor markets and knowledge spillovers across industries. The spatial distribution of this coefficient agrees with research on clusters (Porter 1998) and agglomeration economies (Rosenthal and Strange 2004; Duranton and Puga 2004).

A positive correlation also exists for employment in the information industry, which includes information and communications technology, across the region for 2000, as shown in Figure 30. This relationship is expected as literature shows that credit and securities firms rely far more on information and communication technology as part of their daily business operations. The correlation is strongest in New York and New Jersey where the largest firms by employment size are located. The positive relationship between employment in the information industry and the concentration of the credit and securities industry for 2010, as shown in Figure 35, signifies the technological advancements shaping the geography of credit and securities establishments.
Figure 26. Map of spatial distribution of the median household income coefficient from the Geographically Weighted Regression analysis for 2000.
Figure 27. Map of spatial distribution of coefficient on employment in finance, insurance and real estate industries from the Geographically Weighted Regression analysis for 2000.
Figure 28. Map of spatial distribution of commuting to working via public transportation coefficient from the Geographically Weighted Regression analysis for 2000.
Figure 29. Map of spatial distribution of the median real estate taxes coefficient from the Geographically Weighted Regression analysis for 2000.
Figure 30. Map of spatial distribution of coefficient on employment in information and communication technology industries from the Geographically Weighted Regression analysis for 2000.
Figure 31. Map of spatial distribution of median household income coefficient from the Geographically Weighted Regression analysis for 2010.
Figure 32. Map of spatial distribution of coefficient on employment in finance, insurance and real estate industries from the Geographically Weighted Regression analysis for 2010.
Figure 33. Map of spatial distribution of commuting to working via public transportation coefficient from the Geographically Weighted Regression analysis for 2010.
Figure 34. Map of spatial distribution of median real estate taxes coefficient from the Geographically Weighted Regression analysis for 2010.
Figure 35. Map of spatial distribution of coefficient on employment in information and communication technology industries from the Geographically Weighted Regression analysis for 2010.
5. Conclusion

Using spatial statistics to account for the spatial structure of the financial services industry specific to credit and securities intermediaries across the New York metro area during the 2000s, clusters of the credit and securities industry were empirically identified and the processes underlying their formation were modeled. Use of a GWR model that accounts for space by allowing for spatially-varying relationships is beneficial for regional planning policies that span across city and state boundaries of the New York metro area yet addresses policy implications that vary locally across this region. Additionally, the use of a spatial weights matrix using ZIP code parameters as designed specifically for this paper informed the spatial autocorrelation test statistic, cluster analysis and exploratory OLS regression analysis. This spatial weights matrix also serves as a unique method designed specifically for this spatial analysis that is beneficial for future research on credit and securities industry across the New York metro area.

Having confirmed the existence of clusters in the credit and securities industry of the New York metro area using Moran’s I, the cluster analyses empirically identified statistically significant clusters for three areas within New York County and three areas outside New York County. Within New York County, the concentrations of the credit and securities industry were found in the financial district of Lower Manhattan, as well as in Midtown Manhattan and in the Upper East Side for 2000, 2005 and 2010. As expected, the clusters identified within New York County had the highest extreme values of the Tri-State region as indicated by the Gi* z-score. The cluster analyses at the ZIP code level allowed for a more nuanced understanding of these
clusters within New York County. Although the 10005 ZIP code of Wall Street in Lower Manhattan had the highest Gi* z-score of the entire study region for 2000, the 10017 ZIP code of Midtown Manhattan had the highest Gi* z-score in 2005 and the 10022 ZIP code of the Upper East Side had the highest Gi* z-score in 2010. Outside New York County, the cluster analyses also identified clusters in Jersey City, Hudson County, New Jersey; Stamford, Fairfield County, Connecticut; and a new cluster was identified in the 11747 ZIP code area of Melville, Suffolk County, New York. As indicated by the Gi* z-score, the Melville cluster had the lowest extreme values of the entire region for 2000 and 2005 with a p-value greater than 0.05; and, in 2010, the Greenwich ZIP code 06830 had the lowest extreme values of the credit and securities industry with a p-value greater than 0.05.

The regression analyses provide empirical support for processes underlying the formation of credit and securities clusters across the New York metro area in 2000 and 2010 as they pertain to (1) real estate taxes; (2) median household income; (3) employment in finance, insurance, real estate, rental and leasing industries; (4) employment in the information and communication technology industry; and (5) use of public transportation to commute to work. By extension, this thesis also provides empirical support for the role of agglomeration economies on the clustering of the credit and securities industry as examined by the knowledge spillovers and matching labor markets of the finance, insurance, real estate, rental and leasing industries as well as the information industry. These five factors account for approximately 40 percent of the clusters of the credit and securities industry identified in 2000. In 2010, these five factors account for well over 40 percent of the location and concentration of the credit and securities industry.
Policy implications applicable across the region derived from this spatial analysis suggest that competitive real estate taxes along with employment in finance, insurance, real estate, rental and leasing industries as well as employment in the information industry attract and support the viability of credit and securities intermediaries at the scale of the ZIP code across the region. This is evident when comparing the cluster analysis in section 4.3 with the regression analysis in section 4.4 for across the New York metro area. By comparison, factors such as the use of public transportation to commute to work and median household income vary significantly across the region at the ZIP code level. For this reason, policies that take into consideration public transportation and median household income should be focused locally to account for this regional variation. For ZIP code areas in Stamford City in Fairfield County, Connecticut; Jersey City in Hudson County, New Jersey; Lower Manhattan, Midtown Manhattan and the Upper East Side of New York County, New York are where the model performed best and where concentrations of credit and securities establishments are observed as part of the cluster analysis. The cluster analyses and regression analyses included as part of this spatial analysis of the financial crisis agrees with the literature confirming that real estate taxes, employment in finance, insurance and real estate industries as well as employment in the information industry and the use of public transportation are variables which could be used to predict clusters of credit and securities industry for future time periods and at different geographic locations.
Chapter 8.1 Future Directions

This paper serves as the basis for future research beyond the scope of the factors identified, as there still exists spatial autocorrelation of the regression residuals for the regression models. The presence of spatial autocorrelation, as shown in Table 10, indicates that at least one key explanatory variable is missing. Factors and underlying processes unaccounted for as part of the regression models presented in this paper include the 9/11 terrorist attacks on the World Trade Center in Lower Manhattan as well as bank legislation and government subsidies. Quantifying these variables while also ensuring the data is congruous across the study region is needed to better explain the formation and movement of credit and securities clusters across the New York metro area and to better understand the changing financial geographies of this region. Crime rate is another important consideration for future spatial regression models.

For future research, the accuracy of comparing regression models between different time periods will improve with each new release of the American Community Survey (ACS) five-year estimates. All of the variables from the 2000 model used for this thesis were derived from the Census 2000 Long Form survey, which was replaced using ACS data in 2005. The use of data derived from the Census 2000 Long Form survey is the only data publicly available at the scale of the ZIP code for the explanatory variables used in the 2000 regression model. Although the variables used for the 2000 and 2010 regression models derive from tables on comparable economic characteristics, the surveys from which the data are derived are inherently different and caution should be used when comparing the explanatory variables between 2000 and 2010. Instead, this thesis serves to understand the influence of these variables for each respective time
period. In order to ensure an empirically-driven spatial analysis, the variables used in the regression models presented here are quantitative variables, which are congruous across the entire study area spanning three state boundaries. Having identified clusters and factors influencing these clusters, the methods to assess these factors will continue to be examined as a way to enhance the model and expand on the spatial analysis for future research.
Appendix A. Preliminary Cluster Analysis I

This Appendix reports results of an alternative cluster analysis of credit and securities establishments using 250 establishments or more for each ZIP code as the dependent variable to identify clusters of large credit and securities establishments in 2000, 2005 and 2010. For this alternative cluster analysis, large establishments are defined as establishments totaling 250 or more per ZIP code area. The maps indicate the spatial distribution of the Gi* derived z-scores for the study area for each time period. The Getis-Ord Gi* hot spot analyses of credit and securities establishments exist for ZIP codes in Lower, Midtown and Upper East Side area of New York County as well as Jersey City, Hudson County, New Jersey for each time period. Though hot spots are evident in Hudson County in 2000, hot spots do not appear in Fairfield County until after 2000.

In 2000, there were a total of 17 ZIP codes identified as hot spots, which dropped to 14 ZIP codes in 2005 and 2010. As further detailed in Tables 11-13, eight ZIP codes constituted significant clusters throughout the time period before, during and immediately after the financial crisis. Though the Gi* score varied with each time period, these eight ZIP codes were found to remain significant over time. In 2000, these clusters were predominant in the lower Manhattan cluster and also found in and near Midtown Manhattan. Outside of New York, Jersey City remained a hotspot for extreme values of credit and securities establishments for each time period. For 2005 and 2010, dominant clusters were found in midtown Manhattan, which had the highest/largest Gi* score. Lower Manhattan had the highest/largest Gi* score in 2000. As
evidenced by the NAICS 522 (credit intermediation) and NIACS 523 (securities intermediation) financial sectors, credit and securities establishments show a movement from the old financial district center in Wall Street and lower Manhattan to that of Midtown Manhattan, while simultaneously showing an increase in Stamford, Connecticut and Jersey City, New Jersey.

Figure 36. Map of Gi* statistic identifying clusters of extreme values for credit and securities establishments with 250 establishments or more for each ZIP code in 2000.
Figure 37. Map of Gi* statistic identifying clusters of extreme values for credit and securities establishments with 250 establishments or more for each ZIP code in 2005.
Figure 38. Map of Gi* statistic identifying clusters of extreme values for credit and securities establishments with 250 establishments or more for each ZIP code in 2010.
Table 11. List of ZIP codes constituting each cluster of extreme values identified for credit and securities establishments with 250 establishments or more for each ZIP code in 2000.

<table>
<thead>
<tr>
<th>County</th>
<th>City</th>
<th>ZIP</th>
<th>Gi* z-score</th>
<th>Gi* p-value</th>
<th>Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>New York</td>
<td>Lower Manhattan</td>
<td>10005</td>
<td>12.51492</td>
<td>0</td>
<td>99%</td>
</tr>
<tr>
<td>New York</td>
<td>Midtown</td>
<td>10019</td>
<td>9.464488</td>
<td>0</td>
<td>99%</td>
</tr>
<tr>
<td>New York</td>
<td>Midtown</td>
<td>10017</td>
<td>8.449083</td>
<td>0</td>
<td>99%</td>
</tr>
<tr>
<td>New York</td>
<td>Upper East Side</td>
<td>10022</td>
<td>6.416046</td>
<td>0</td>
<td>99%</td>
</tr>
<tr>
<td>New York</td>
<td>Lower Manhattan</td>
<td>10004</td>
<td>5.396426</td>
<td>0</td>
<td>99%</td>
</tr>
<tr>
<td>New York</td>
<td>Midtown</td>
<td>10036</td>
<td>4.884077</td>
<td>0.000012</td>
<td>99%</td>
</tr>
<tr>
<td>New York</td>
<td>Midtown</td>
<td>10010</td>
<td>4.881425</td>
<td>0.000012</td>
<td>99%</td>
</tr>
<tr>
<td>New York</td>
<td>Midtown</td>
<td>10020</td>
<td>3.377549</td>
<td>0.000106</td>
<td>99%</td>
</tr>
<tr>
<td>New York</td>
<td>Lower Manhattan</td>
<td>10006</td>
<td>3.374624</td>
<td>0.000739</td>
<td>99%</td>
</tr>
<tr>
<td>New York</td>
<td>Lower Manhattan</td>
<td>10280</td>
<td>3.371328</td>
<td>0.000748</td>
<td>99%</td>
</tr>
<tr>
<td>Hudson</td>
<td>Jersey City</td>
<td>7302</td>
<td>3.362303</td>
<td>0.000773</td>
<td>99%</td>
</tr>
<tr>
<td>Hudson</td>
<td>Jersey City</td>
<td>7310</td>
<td>2.854363</td>
<td>0.004312</td>
<td>99%</td>
</tr>
<tr>
<td>New York</td>
<td>Lower Manhattan</td>
<td>10048</td>
<td>2.35536</td>
<td>0.018505</td>
<td>90%</td>
</tr>
<tr>
<td>New York</td>
<td>Midtown</td>
<td>10001</td>
<td>2.349403</td>
<td>0.018804</td>
<td>90%</td>
</tr>
<tr>
<td>New York</td>
<td>Lower Manhattan</td>
<td>10007</td>
<td>1.84313</td>
<td>0.065138</td>
<td>90%</td>
</tr>
<tr>
<td>Hudson</td>
<td>Union City</td>
<td>7087</td>
<td>1.836512</td>
<td>0.066282</td>
<td>90%</td>
</tr>
<tr>
<td>Nassau</td>
<td>Hyde Park</td>
<td>11042</td>
<td>1.832445</td>
<td>0.066885</td>
<td>90%</td>
</tr>
</tbody>
</table>
Table 12. List of ZIP codes constituting each cluster of extreme values identified for credit and securities establishments with 250 establishments or more for each ZIP code in 2005.

<table>
<thead>
<tr>
<th>County</th>
<th>City</th>
<th>ZIP</th>
<th>Gi* z-score</th>
<th>Gi* p-value</th>
<th>Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 New York</td>
<td>Midtown</td>
<td>10017</td>
<td>10.83954</td>
<td>0</td>
<td>99%</td>
</tr>
<tr>
<td>2 New York</td>
<td>Midtown</td>
<td>10019</td>
<td>10.28647</td>
<td>0</td>
<td>99%</td>
</tr>
<tr>
<td>3 New York</td>
<td>Upper East Side</td>
<td>10022</td>
<td>7.527585</td>
<td>0</td>
<td>99%</td>
</tr>
<tr>
<td>4 New York</td>
<td>Lower Manhattan</td>
<td>10005</td>
<td>7.52521</td>
<td>0</td>
<td>99%</td>
</tr>
<tr>
<td>5 New York</td>
<td>Midtown</td>
<td>10020</td>
<td>6.425833</td>
<td>0</td>
<td>99%</td>
</tr>
<tr>
<td>6 New York</td>
<td>Midtown</td>
<td>10036</td>
<td>4.767674</td>
<td>0.000002</td>
<td>99%</td>
</tr>
<tr>
<td>7 New York</td>
<td>Lower Manhattan</td>
<td>10004</td>
<td>4.761839</td>
<td>0.000002</td>
<td>99%</td>
</tr>
<tr>
<td>8 New York</td>
<td>Lower Manhattan</td>
<td>10006</td>
<td>4.214801</td>
<td>0.000025</td>
<td>99%</td>
</tr>
<tr>
<td>9 New York</td>
<td>Midtown</td>
<td>10010</td>
<td>4.212731</td>
<td>0.000025</td>
<td>99%</td>
</tr>
<tr>
<td>10 Hudson</td>
<td>Jersey City</td>
<td>7310</td>
<td>3.105257</td>
<td>0.001901</td>
<td>99%</td>
</tr>
<tr>
<td>11 Fairfield</td>
<td>Stamford</td>
<td>6902</td>
<td>3.101451</td>
<td>0.001926</td>
<td>99%</td>
</tr>
<tr>
<td>12 New York</td>
<td>Lower Manhattan</td>
<td>10007</td>
<td>2.557294</td>
<td>0.010549</td>
<td>90%</td>
</tr>
<tr>
<td>13 Kings</td>
<td>Boerum Hill</td>
<td>11201</td>
<td>2.001198</td>
<td>0.045371</td>
<td>90%</td>
</tr>
<tr>
<td>14 Fairfield</td>
<td>Stamford</td>
<td>6901</td>
<td>1.997584</td>
<td>0.045762</td>
<td>90%</td>
</tr>
</tbody>
</table>
Table 13. List of ZIP codes constituting each cluster of extreme values identified for credit and securities establishments with 250 establishments or more for each ZIP code in 2010.

<table>
<thead>
<tr>
<th>County</th>
<th>City</th>
<th>ZIP</th>
<th>Gi* z-score</th>
<th>Gi* p-value</th>
<th>Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>New York</td>
<td>Midtown</td>
<td>10017</td>
<td>10.25193</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>New York</td>
<td>Midtown</td>
<td>10019</td>
<td>10.25182</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>New York</td>
<td>Lower Manhattan</td>
<td>10005</td>
<td>9.149681</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>New York</td>
<td>Midtown</td>
<td>10020</td>
<td>7.508328</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>New York</td>
<td>Midtown</td>
<td>10036</td>
<td>5.858979</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>New York</td>
<td>Upper East Side</td>
<td>10022</td>
<td>5.858478</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>New York</td>
<td>Lower Manhattan</td>
<td>10004</td>
<td>4.75412</td>
<td>0.000002</td>
</tr>
<tr>
<td>8</td>
<td>New York</td>
<td>Midtown</td>
<td>10010</td>
<td>3.657976</td>
<td>0.000254</td>
</tr>
<tr>
<td>9</td>
<td>Hudson</td>
<td>Jersey City</td>
<td>7302</td>
<td>3.655137</td>
<td>0.000257</td>
</tr>
<tr>
<td>10</td>
<td>Hudson</td>
<td>Jersey City</td>
<td>7310</td>
<td>3.106254</td>
<td>0.001895</td>
</tr>
<tr>
<td>11</td>
<td>New York</td>
<td>Lower Manhattan</td>
<td>10007</td>
<td>2.560311</td>
<td>0.010458</td>
</tr>
<tr>
<td>12</td>
<td>New York</td>
<td>Lower Manhattan</td>
<td>10013</td>
<td>2.559133</td>
<td>0.010493</td>
</tr>
<tr>
<td>13</td>
<td>Fairfield</td>
<td>Stamford</td>
<td>6902</td>
<td>2.55315</td>
<td>0.010675</td>
</tr>
<tr>
<td>14</td>
<td>Fairfield</td>
<td>Stamford</td>
<td>6901</td>
<td>2.553147</td>
<td>0.010675</td>
</tr>
</tbody>
</table>
Figure 39. Map of Manhattan/New York County study area shows a cluster of extreme values for credit and securities establishments with 250 establishments or more for each ZIP code in the traditional finance district center of Wall Street ZIP code 10005, while also showing a significant cluster in and near Midtown.
Figure 40. Map of Manhattan/New York County study area in 2005 shows the cluster for credit and securities establishments with 250 establishments or more for each ZIP code of the traditional finance district center of Wall Street ZIP code 10005 has shrunk in size nearest the shore of the southern tip of Manhattan. Hotspots in and near Midtown remain.
Figure 41. Map of Manhattan/New York County study area shows a cluster for credit and securities establishments with 250 establishments or more for each ZIP code of the traditional finance district center of Wall Street ZIP code 10005, while also showing a significant cluster in and near Midtown in 2010.
Figure 42. Map of Fairfield County study area without a significant cluster of credit and securities establishments with 250 establishments or more for each ZIP code in 2000.
Figure 43. Map of Fairfield County shows a cluster of hot spots credit and securities establishments with 250 establishments or more for each ZIP code in 2005 for Stamford ZIP codes 06901 and 06902.
Figure 44. Map shows a reduction of hot spots for the quantity of credit and securities establishments with 250 establishments or more for each ZIP code in 2010 for Stamford ZIP codes 06901 and 06902.
Figure 45. Map of Hudson County study area in 2000 shows a cluster of credit and securities establishments with 250 establishments or more for Jersey City ZIP code 07310 and 07302 as well as a hotspot in Union City ZIP code 07087.
Figure 46. Map of Hudson County study area in 2005 shows a reduction of hot spots for the quantity of credit and securities establishments with 250 establishments or more. Jersey City ZIP code 07310 remains as a hot spot in 2005.
Figure 47. Map of Hudson County study area in 2010 shows a cluster of credit and securities establishments with 250 establishments or more for each ZIP code is maintained in Jersey City ZIP code 07310 and also includes Jersey City ZIP code 07310.
Appendix B. Preliminary Cluster Analysis II

This Appendix presents results of an alternative cluster analysis of credit and securities establishments using total credit establishments for each ZIP code as the dependent variable.

Figure 48. Map of Gi* statistic identifying clusters of extreme values for total credit and securities establishments for each ZIP code in 2000.
Figure 49. Map of Gi* statistic identifying clusters of extreme values for total credit and securities establishments for each ZIP code in 2005.
Figure 50. Map of Gi* statistic identifying clusters of extreme values for total credit and securities establishments for each ZIP code in 2010.
Appendix C. Preliminary GWR

Results of an alternative GWR model for 2000 and 2010 using total credit and securities establishments as the dependent variable. As shown in Table 14, the regression model for 2010 is a better fit than the 2000 regression model of financial clusters. According to the R-Squared statistic, the regression in 2000 accounts for 42.25 percent of the variation in the dependent variable, the quantity of credit and securities intermediaries with an adjusted R-Squared of 31.75 percent. By comparison, the regression in 2010 accounts for 42.57 percent of the variation in the dependent variable with an adjusted R-Squared of 33.28 percent. Both the 2000 and 2010 explain over 30 percent of the dependent variable, which is the quantity of credit and securities intermediaries at each ZIP code across the New York Metropolitan Area. As shown in Tables 15 and 16, the factors influencing the location of clusters for both the 2000 and 2010 regression model pertain to real estate taxes, household income, transportation, employment in the finance-related industries.

As shown in Figures 51 and 53, maps of standardized residuals for each time period show statistically significant clustering, with the map for the 2010 standardized residuals showing a greater extent of clustering, as confirmed with the Global Moran’s I statistic used on the regression residuals. The clustering is apparent in New York County in both models with similar quantities of over and under predictions in 2000 and 2010.
The maps of local R-Squared, Figures 52 and 54, show the model performs best within and near the ZIP codes within and nearest the five counties of New York City (New York, Queens, Kings, Bronx and Richmond counties) as well as Union City and Jersey City in Hudson County, New Jersey, and Stamford in Fairfield County, Connecticut.

Other measures of model performance agree with the R-Squared and adjusted R-Squared to show the 2010 regression model is a slightly better fitting model. The optimal adaptive neighbors is higher in 2010 than in 2000. This means each local computation in the 2010 model was based on more neighbors than the local computations for the model in 2000. Other measures of model performance indicate the 2010 model is a better fit than the 2000 model. The sum of squared residuals in 2010 is lower than the sum of squared residuals in 2000; and the regression model for 2010 has a smaller effective number, a smaller sigma, and a smaller AICc.

Table 14. Summary of Geographically Weighted Regression model performance for total credit and securities establishments in 2000 and 2010

<table>
<thead>
<tr>
<th>Measures of Model Performance</th>
<th>2000</th>
<th>2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neighbors</td>
<td>87</td>
<td>123</td>
</tr>
<tr>
<td>ResidualSquares</td>
<td>2347203</td>
<td>1895553.561</td>
</tr>
<tr>
<td>EffectiveNumber</td>
<td>83.32812</td>
<td>75.446452</td>
</tr>
<tr>
<td>Sigma</td>
<td>72.00846</td>
<td>64.154624</td>
</tr>
<tr>
<td>AICc</td>
<td>6161.732</td>
<td>6031.861763</td>
</tr>
<tr>
<td>R2</td>
<td>0.422532</td>
<td>0.425665</td>
</tr>
<tr>
<td>R2Adjusted</td>
<td>0.317507</td>
<td>0.332826</td>
</tr>
</tbody>
</table>
Table 15. List of variables used for the regression model analyses total credit and securities establishments in 2000 and the significance of each variable according to the Exploratory Regression.

<table>
<thead>
<tr>
<th>Explanatory Variables in 2000</th>
<th>Positive Significance %</th>
<th>Negative Significance %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median Household Income</td>
<td>75.00</td>
<td>25.00</td>
</tr>
<tr>
<td>Employment in the Finance, Insurance or Real Estate Industry</td>
<td>100.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Commuting to work using public transportation</td>
<td>100.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Median Real Estate Taxes</td>
<td>25.00</td>
<td>75.00</td>
</tr>
</tbody>
</table>

Table 16. List of variables used for the regression model analyses total credit and securities establishments in 2010 and the significance of each variable according to the Exploratory Regression.

<table>
<thead>
<tr>
<th>Explanatory Variables in 2010</th>
<th>Positive Significance %</th>
<th>Negative Significance %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median Household Income</td>
<td>50.00</td>
<td>50.00</td>
</tr>
<tr>
<td>Employment in the Finance, Insurance or Real Estate Industry</td>
<td>100.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Commuting to work using public transportation</td>
<td>62.50</td>
<td>37.50</td>
</tr>
<tr>
<td>Median Real Estate Taxes</td>
<td>43.75</td>
<td>56.25</td>
</tr>
<tr>
<td>Employment in Information Industry</td>
<td>100.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>
Figure S1. Map of standardized residuals resulting from the Geographically Weighted Regression analysis for total credit and securities establishments in 2000.
Figure 52. Map of local squared residuals resulting from the Geographically Weighted Regression analysis for total credit and securities establishments in 2000.
Figure 53. Map of standardized residuals resulting from the Geographically Weighted Regression analysis for total credit and securities establishments in 2010.
Figures 54-63 show the spatial distribution of each coefficient included in the regression model for both 2000 and 2010. The spatial distribution of property taxes in 2000 and 2010 agrees with the literature which shows Connecticut has the lowest property taxes of the New York metro area. The relationship is stronger than that compared to other ZIP code areas in Hudson County, New Jersey where fewer credit and securities intermediaries are observed in the literature. For 2000, this relationship is also positive for ZIP codes in lower Manhattan/New York County where Wall Street is located. In 2010, however, there is a negative relationship for ZIP codes in Manhattan/New York County. This suggests the higher the property taxes in this area, the lower the quantity of total credit and securities establishments will be for this area.
For commuting via public transportation, as shown in Figures 51 and 59, a negative relationship suggests the lower the use of public transportation for people residing in a given ZIP code area as a means of commuting to work, the higher the quantity of credit and securities intermediary establishments will be in that ZIP code area. Maps on employment in the finance, insurance and real estate-related industries, Figures 56 and 61, show there is an overall positive correlation between employment in the finance-related industries and the concentration of credit and securities intermediaries for both 2000 and 2010 regression models. As shown in Figures 56 and 61, the positive relationship is stronger in ZIP code areas within or nearest the clusters in New York and New Jersey. A positive correlation also exists for employment in the information industry, a variable included in the 2010 regression model but not used in the 2000 regression model. Employment in the information industry was not used in the 2000 model because it resulted in a poorly fitted regression model.
Figure 55. Map of spatial distribution of commuting to working via public transportation coefficient from the Geographically Weighted Regression analysis for 2000.
Figure 56. Map of spatial distribution of coefficient on employment in finance, insurance and real estate industries from the Geographically Weighted Regression analysis for 2000.
Figure 57. Map of spatial distribution of the median household income coefficient from the Geographically Weighted Regression analysis for 2000.
Figure 58. Map of spatial distribution of the median real estate taxes coefficient from the Geographically Weighted Regression analysis for 2000.
Figure 59. Map of spatial distribution of commuting to working via public transportation coefficient from the Geographically Weighted Regression analysis for 2010.
Figure 60. Map of spatial distribution of median household income coefficient from the Geographically Weighted Regression analysis for 2010.
Figure 61. Map of spatial distribution of coefficient on employment in finance, insurance and real estate industries from the Geographically Weighted Regression analysis for 2010.
Figure 62. Map of spatial distribution of median real estate taxes coefficient from the Geographically Weighted Regression analysis for 2010.
Figure 63. Map of spatial distribution of coefficient on employment in information industry from the Geographically Weighted Regression analysis for 2010
References


Anselin, L. (1989) *What is special about spatial data?: alternative perspectives on spatial data analysis.* Santa Barbara, CA: National Center for Geographic Information and Analysis


