Assessing the Impact of Light Rail Transit on Land Values and Tax Revenues

Carol Atkinson-Palombo

Center for Transportation and Livable Systems (CTLS)
270 Middle Turnpike Unit 5202
Storrs, CT 06269-5202

U.S. Dept. of Transportation
1200 New Jersey Ave, SE
Washington DC 20590
202-366-4000

Light Rail Transit (LRT) has become an increasingly popular centerpiece of smart growth policies in the United States, yet considerable debate exists about its cost-effectiveness. The overarching question that we propose to address in this research is “What are the early returns to proposed and under-construction Light-Rail Transit (LRT) investment and how do these vary across neighborhoods?” It may take 10 or 20 years to generate substantial revitalization of a given neighborhood, but early increases in property value and tax collections may stimulate neighborhood change, and defray capital costs. A related question is “What is the best methodology to address the responses to LRT at the neighborhood level, while simultaneously measuring changes in property values at a regional scale rather than focusing solely on station area impacts?” We propose also to separate agglomeration effects (such as neighborhood commercial and residential development) from those associated with change in mobility (such as better access to downtown). Accordingly, a team of researchers with expertise in smart growth, spatial and temporal accessibility at the metropolitan scale derived from GIS-T models, and a proxy for neighborhood change derived from data on new construction. Modeling techniques capable of disentangling neighborhood and regional effects include Geographically Weighted Regression (GWR) and a local Spatial and Temporal Autoregressive (STAR) model. Comparison of out-of-sample predictive ability and spatial autocorrelation in the models’ residuals will inform on performance, and provide a more nuanced understanding of the dynamics set into motion by decisions to build LRT, quantify the spatial distribution and extent of early returns on investment at the metropolitan scale, and provide more effective decision-making on this smart growth tool. The housing market and credit crisis that unfolded shortly after the study began required research efforts to be redirected towards understanding the causes and both local and systemic impact on house of foreclosures and how and why they vary across neighborhoods. Further pressure from a spike in gasoline prices in mid-2008 created increased debate in both academic and policy circles about the sustainability of the traditional suburb, with some even going so far as to proclaim that “the suburbs are dead”. Accordingly, This paper uses numerous statistical tests including Principal Components Analysis, Cluster Analysis, Binary Logistic Regressions, and a derived spatial contagion variable to understand the Spatio-temporal pattern of foreclosures in Maricopa County, Arizona, from January 2007 through February 2009. Datasets include individual housing units that became Real Estate Owned (REO), alongside zip-code level data on credit scores and mortgage loan characteristics as controls. The analyses show how the foreclosure crisis began in neighborhoods dominated by low-income minority households and spread to other locations across the metropolitan area, but avoided neighborhoods dominated by households comprised of very high-income white people with high levels of home equity. Embedding these impacts within a broader geo-political economy perspective reveals that although some households may wish to isolate themselves from low-income neighborhoods, spatial contagion is not the only form of contagion that occurs. Nevertheless, policy-makers should consider incorporating spatial contagion in housing market risk models.
ACKNOWLEDGEMENTS

The author would like to thank the following graduate students who have worked on this project: Jason Patalinghug, George Bentley, Brandon Cramer, Walter D’Lima. The author would also like to thank Kelvin Wong for Hong Kong University, a Fulbright scholar, for contributing 70% of the cost of the LPS data, and John Clapp for his insightful comments that helped the PAR versus WAR paper get accepted and published in Urban Studies.

The author would also like to thank CLTS and the DOT for extending the contract dates to accommodate medical leave following a serious automobile accident in Summer 2009.

DISCLAIMER

The contents of this report reflect the views of the authors, who are solely responsible for the facts and the accuracy of the material and information presented herein. This document is disseminated under the sponsorship of the U.S. Department of Transportation University Transportation Centers Program and the Center for Transportation and Livable Systems in the interest of information exchange. The U.S. Government and the University of Connecticut assume no liability for the contents or use thereof. The contents do not necessarily reflect the official views of the U.S. Government or the University of Connecticut. This report does not constitute a standard, specification, or regulation.
# TABLE OF CONTENTS

**EXCUTIVE SUMMARY** ............................................................................................................. 1

**1.0 INTRODUCTION** .................................................................................................................. 3

1.1 BACKGROUND .......................................................................................................................... 3

1.2 PROBLEM STATEMENT ............................................................................................................. 4

1.3 ORIGINAL OBJECTIVES OF THE PROJECT .......................................................................... 5

1.4 REVISED OBJECTIVES OF THE PROJECT .......................................................................... 5

1.5 STRUCTURE OF THIS REPORT .............................................................................................. 6

**2.0 STUDY: CAPITALIZATION EFFECTS OF LRT IN PARK AND RIDE VERSUS WALK AND RIDE STATION AREAS** ........................................................................................................ 7

2.1 LITERATURE REVIEW ............................................................................................................. 7

2.2 METHODOLOGY ...................................................................................................................... 9

2.2.1 Light Rail in metropolitan Phoenix, Arizona ........................................................................ 9

2.2.2 Data ................................................................................................................................... 11

2.2.3 Analysis of Land-use mix ................................................................................................ 12

2.2.4 Analysis of Overlay Zoning ............................................................................................. 12

2.2.5 Analysis of PAR and neighborhood type ...................................................................... 12

2.2.6 Transaction Data .............................................................................................................. 12

2.2.7 Hedonic models ................................................................................................................ 12

2.2.8 Hypotheses ....................................................................................................................... 13

2.3 RESULTS ................................................................................................................................ 14

2.3.1 Cluster Analysis ................................................................................................................ 14

2.3.2 Neighborhood type and overlay zoning ...................................................................... 14

2.3.3 Neighborhood type and PAR/WAR ............................................................................. 15

2.3.4 Capitalization effects ....................................................................................................... 15

2.3.5 Spatial Autocorrelation .................................................................................................... 20

2.4 EVALUATION AND CONCLUSIONS .................................................................................... 20

2.5 NOTES ..................................................................................................................................... 22

**3.0 THE GEOGRAPHY OF FORECLOSURE** ........................................................................... 25

3.1 INTRODUCTION ...................................................................................................................... 25

3.2 CAUSES OF FORECLOSURE .................................................................................................. 25

3.2.1 Background ....................................................................................................................... 26

3.2.2 Explanatory Variables of Foreclosure .......................................................................... 26

3.2.3 Geographic approaches to understanding foreclosure ................................................. 27

3.2.4 Economic Approaches to the Foreclosure Crisis .......................................................... 28

3.2.5 Contagion over space and time ..................................................................................... 29

3.2.6 The Effect on Foreclosures of the 2008 Surge in Gasoline Prices ............................... 29

3.2.7 Evaluation of the literature ............................................................................................. 29

3.3 EMPIRICAL ANALYSIS ON FORECLOSURES .................................................................... 30

3.3.1 Study Area .......................................................................................................................... 30

3.3.2 Conceptual approach ....................................................................................................... 30

3.3.3 Data .................................................................................................................................. 31

3.3.4 Preliminary Analysis of Raw Data .................................................................................. 32

3.3.5 Inferential Statistics ......................................................................................................... 36
3.3.6 Predicting foreclosure ............................................................... 36
3.3.7 Cluster analysis of neighborhood types ............................................................... 37
3.3.8 ANOVA tests of foreclosure rates ............................................................... 37
3.3.9 Modeling foreclosure by neighborhood type with and without spatial contagion .... 37
3.4 RESULTS ......................................................................................... 38
3.4.1 Predicting foreclosure ........................................................................ 38
3.4.2 Cluster analysis of neighborhood types ....................................................... 40
3.4.3 ANOVA tests of foreclosure rates ................................................................. 42
3.4.4 Modeling foreclosure by neighborhood type without/with spatial contagion .... 44
3.5 EVALUATION AND DISCUSSION .................................................... 47

4.0 EMPIRICAL ANALYSIS OF ADDITIONAL IMPACT ON FORECLOSURES OF HIGHER GASOLINE PRICES ................................................................. 49
4.1 INTRODUCTION ......................................................................................... 49
4.2 LOGISTIC REGRESSION ANALYSIS ................................................................. 52
4.3 HYPOTHESIS ............................................................................................... 53
4.4 RESULTS .................................................................................................... 53
4.5 DISCUSSION AND CONCLUSIONS ............................................................. 54

5.0 LIGHT RAIL TRANSIT’S AFFECT ON LAND VALUE—HOW MUCH COMES FROM NEW CONSTRUCTION? ................................................................. 55
5.1 INTRODUCTION ........................................................................................... 55
5.2 ABSTRACT .................................................................................................... 55

6.0 ADDITIONAL WORK IN PROGRESS ..................................................... 57
6.1 METHODOLOGY AND GIS ........................................................................ 57
6.1.1 Abstract: Network Interpolation of Population Values ................................... 57
6.2 IMPACT OF FORECLOSURE ON SURROUNDING HOUSE PRICES. DOES THE EFFECT DIFFER IN DIFFERENT TYPES OF NEIGHBORHOODS? .................. 57
6.3 COMPARISON OF THE PERFORMANCE OF HOUSE VALUES IN PROPERTIES IN LRT STATION AREAS IN AN UP-MARKET AND IN A DOWN-MARKET AND ITS IMPLICATIONS FOR THE CONSIDERATION OF RESILIENT CITIES ............... 57
6.4 EXTENSION OF ANALYSIS TO CALIFORNIA ........................................... 58

7.0 EDUCATION-RELATED INITIATIVES ....................................................... 59
7.1 TEACHING ................................................................................................. 59
7.2 COLLOQUIUM ............................................................................................ 59

8.0 FIELD STUDY, DECEMBER 2009 ............................................................. 61
8.1 REAL ESTATE MARKET DYNAMICS IN LRT CORRIDOR .................................. 61
8.2 LAND USE DYNAMICS ............................................................................. 61
8.3 MULTI-MODAL SYSTEM DEVELOPMENT .................................................. 61
8.4 SOCIAL AND DEMOGRAPHIC CHANGE ....................................................... 61

APPENDICES

APPENDIX A: Selected Photographs of Field Study, LRT system, Arizona, December 2009
LIST OF TABLES

Table 2.1: Neighborhood Types ................................................................................................... 14
Table 2.2: Percentage of housing units subject to overlay zoning by Neighborhood Type ....... 15
Table 2.3: Park-and-Ride (PAR) by Neighborhood Type ............................................................ 15
Table 2.4: Descriptive Statistics of housing transactions "Before" overlay zoning was announced
Table 2.5: Key hedonic model coefficients for single-family houses .......................................... 18
Table 2.6: Key hedonic model coefficients for condos ................................................................ 19
Table 3.1: Spatial Datasets ............................................................................................................ 31
Table 3.2: Results of Principal Components Analysis: Weights of Individual Variables and how
they load onto various components....................................................................................... 39
Table 3.3: Results of Binary Logistic Regression Model Predicting Foreclosure as a function of
individual components ........................................................................................................ 40
Table 3.4: Mean Values for each cluster group and entire dataset ............................................... 41
Table 3.5: Model Estimates and odds ratios for binary logistic regression of foreclosure rate (by
quarter) as a function of neighborhood type without and with spatial contagion variable... 45
Table 3.6: Model fits for binary logistic regression model without/with spatial contagion variable
............................................................................................................................................... 46
Table 4.1: Descriptive statistics of neighborhood variables ......................................................... 50

LIST OF FIGURES

Figure 2.1: Map of Light Rail Transit system in metropolitan Phoenix, Arizona .................... 11
Figure 3.1: Study Area .................................................................................................................. 32
Figure 3.2: Spatial Distribution of Loan-to-Value Ratios ............................................................ 33
Figure 3.3: : Spatial Distribution of Credit Scores ................................................................. 34
Figure 3.4: Spatial Distribution of Loan Types ............................................................................ 34
Figure 3.5: Foreclosure rates from Q1 2007 through Q1 2009................................................. 35
Figure 3.6: Neighborhood types based on social, economic, credit, and loan variables ......... 43
Figure 3.7: : Foreclosure rates by quarter for neighborhood type ........................................... 44
Figure 4.1: Conceptual Model to Understand the Impact of Oil Prices on Foreclosures ......... 49
Figure 4.2: Map of metropolitan Phoenix. Each grey square represents a 1 mile x 1 mile grid that
contains 10 or more housing units. Areas in black represent the neighborhoods included in
the study. ............................................................................................................................... 50
Figure 4.3: Scaled trend of prices of Crude oil, Gasoline, foreclosure rate, and Dow Jones
Industrial Average during the study period 2007-Q1 – 2009-Q1 ......................................... 53
EXECUTIVE SUMMARY

Light Rail Transit (LRT) systems have been built in an increasingly number of cities in the United States over the past twenty years. Those involved in building LRT more recently have the benefit of learning from experiences of early adopters. One of the most recent systems in metropolitan Phoenix, Arizona, is interesting from a number of perspectives. First, the City of Phoenix adopted overlay zoning ordinances ahead of system construction to reduce uncertainty associated with the redevelopment of station areas. Second, with the starter system comprising only twenty miles of in-lane tracks connecting the downtowns of Tempe and Phoenix, within a vast expanse of low density metropolitan development, the system is an important test case about the effectiveness of LRT in already sprawling suburban developments. Third, construction of the system spanned one of the most pronounced boom and bust cycles in the real estate market that the region has ever experienced. The metropolitan area has been hit extremely hard by the housing market crash, credit crunch, and foreclosure crisis that have dominated the economic landscape since 2008. However, residential foreclosure rates in the LRT corridor were surprisingly low compared to outlying neighborhoods, particularly those in the recently-developed western municipalities on the suburban fringe.

This study—originally envisaged as a rather straightforward analysis of the impact of LRT and zoning changes on local and regional property values—morphed into a much more complex investigation of property values. Market developments necessitated that the investigators step back from the rather narrow purview of LRT-related literature and examine the macro, regional, and local causes and consequences of foreclosure in the housing markets. This in-depth study of a rapidly-evolving topic has produced some interesting insights that were well received at an international conference in Sweden of housing market specialists. One of the most prominent findings of the analysis into the geography of foreclosures was that foreclosure rates were highest at the suburban fringe where low-income minority residents had taken out subprime mortgages with very little down payment. The concept of spatial contagion explained how foreclosures spread from one neighborhood to the next across the metropolitan area. This analysis at the metropolitan-wide scale used some novel approaches that have not been used in this context before—a combination of fishnetting, daisimetric mapping, cluster analysis and binary logistic regression modeling. The results set the stage for a study that examines whether or not the impact of foreclosures on property prices differs by neighborhood type. This analysis is still underway.

A second study that has flowed directly from the foreclosure analysis is an examination as to whether or not the increase in gasoline prices that took place in Summer 2008 exacerbated the foreclosure crisis. This analysis has proven to be quite tricky to implement because of a lack of reliable data about household-level VMT and its responsiveness to higher gasoline prices. Despite these challenges, the analysis is worth pursuing because more anecdotal evidence has suggested that this has been the case. Others have jumped directly to the conclusion that the combination of mortgage market dynamics and higher gasoline prices in a post peak oil environment signals the death of the suburbs. Alarmist statements such as these require rigorous
analysis to understand in a more balanced and rational manner, precisely how risks may play out and what public policy-makers may be able to do in the future to mitigate those risks.

Once we are able to disentangle the effects that higher gasoline prices and spatial contagion may have had on property prices and therefore foreclosure, we will be in a position to move ahead and examine more precisely why properties in the LRT corridor appear to have fared better than properties in outlying locations in the recent housing market boom. The results of this final component of the study will inform the very current debate about location efficiency, and the more general concept of resilient cities.

One material point that I would like to make regarding this study relates to the availability of data relating to foreclosures. Variables such as credit score, mortgage type, loan-to-value ratio, and age of mortgage are critical controls that are needed to understand the causes of foreclosures. These data are collected by Lender Processing Services (LPS). We were able to purchase some data cubes in which relevant data were aggregated to the zip code spatial scale for $10,000. We believe that the results of the study would have been much more reliable had we been able to have these data at a finer spatial scale. Such data are available but require a subscription costing $200,000 per annum. Some academic researchers have gotten around this limitation by co-authoring publications with researchers from the Federal Reserve Board who have access to the raw data. The price of the data and the fact that it is available to so few people may limit the number and type of questions that can be addressed.
1.0 INTRODUCTION

1.1 BACKGROUND

Light rail transit (LRT) has become an increasingly popular strategy to create more sustainable cities by offering an alternative to automobile-dependent suburban living typified by the single-family house on its individual lot (Duaney, Plater-Zyberk, and Speck 2001; Cervero and Landis 1995; Nozzi 2005; Boone 2004; Guy and Marvin 1999; Banister, Watson, and Wood 1997). Billions of dollars of public funds are invested in LRT systems in anticipation that transportation, along with supportive policies such as zoning changes, will provide a catalyst for development on the spine of transit, known as transit-oriented development (TOD). TOD is fundamentally about creating high density, pedestrian-friendly, self-contained neighborhoods (Bernick and Cervero 1997; Salvensen 1996; TCRP 1997; Cervero and Duncan 2002). Existing literature on this topic has emphasized remaking places into locations that attract jobs, housing, and amenities so that capital and entrepreneurs generate a self-reinforcing cycle between land-use and ridership, and produce urban revitalization (Farris 2001; Suchman and Sowell 1997; Danielsen, Lang, and Fulton 1999; Lang, Hughes, and Danielsen 1997; TCRP 1997, 2002, 2004). Case studies of individual station areas or specific projects have revealed that not all station areas have the capacity to provide amenity-rich urban environments (Bernstein 2004; Leach 2004; Ohland 2004; Feigon, Hoyt, and Ohland 2004; Breznau 2004; Ohland 2004). Discourse has therefore shifted to the idea that TOD will unfold differently across station areas within a given system. Scholars have drawn upon the distinction between “place” and “node” articulated by Bertolini and Split (1998) as the starting point for conveying that some station areas will be “destinations” while others will function as “transportation nodes” to feed passengers onto the transit system (e.g. Bernstein 2004). This “geographic turn” in the conceptualization of TOD is consistent with basic theories of location that suggest that differences in pre-existing characteristics of place, the spatial distribution of policy intervention, forces of agglomeration, and transportation-related characteristics will mediate development (Fujita 1989; Fujita and Thisse 1996; Parr 2002). Having said that, it is currently unclear how and why the effects of LRT differ across neighborhoods, and what the relative impact is of forces of agglomeration and changes in accessibility due to LRT.

Most commonly, quantitative studies focus on measuring the impact of transportation systems on the price of specific types of land-use. The rationale for doing so is that changes in land value precede changes in land use. One criticism of such models is that coefficients are held constant over space and time. Scholars who have investigated spatial autocorrelation in the residuals of hedonic models—the extent to which errors vary systematically over space—have argued that standard hedonic regression models are at best limited, and at worst, scientifically mis-specified (Anselin et al. 1996; Anselin and Getis 1992). In recent years, aided by technological development, researchers have made great strides in a number of separate research domains that relate to quantifying the effect of LRT on land markets. These include accounting for spatial and temporal factors in real estate pricing models (see Case et al. 2004 and the references therein),
software that more easily quantifies spatial autocorrelation in data (Fotheringham, Brunsdon, and Charlton 2002), and more sophisticated techniques to represent distance in GIS models (e.g. Miller and Wentz 2003; Thill 2000; Horner and Murray 2004).

Only two studies have provided information about the effect of LRT on land prices by different neighborhood type. Gatzlaff and Smith (1993) divided their Miami study area in two: the northern area in economic decline and the southern area which was more stable. They found that the strongest price increases were experienced in the better neighborhoods. In his study of Atlanta, Nelson (1992) found the opposite effect: that transit accessibility increased home prices in lower-income census tracts, but decreased values in upper income tracts. While these two empirical studies suggest that the impacts of LRT differ by neighborhood, the empirical results conflicted. Neighborhood and metropolitan area effects can be measured with spatial dependency; observations are not independent of one another but instead neighboring observations have an influence on each other. In fact, scholars have taken the issue one step further, using the scope of spatial dependency in data as a means to identify neighborhoods (Clapp and Wang 2006).

**1.2 PROBLEM STATEMENT**

Numerous techniques have been offered to correct for spatial dependency in datasets, based mainly on econometric techniques (see Can and Megbolugbe 1997; Case et al. 2004 for a comprehensive review and more-recent update). Among these techniques, the relatively new and rapidly-evolving technique of Geographically-Weighted Regression allows regression coefficients to vary over space (Fotheringham, Brunsdon, and Charlton 2002). Allowing flexibility in a model’s coefficients reduces spatial autocorrelation, improving the predictive accuracy of models. Pace (1997) and Pace et al (1998) developed semi-parametric spatial and temporal auto-regressive (STAR) models. STAR models build on spatial autoregressive models (explanatory variables include neighboring prices) and combines them with a temporal autoregressive process (explanatory variables include prices lagged in time). The spatial lags structure indicates the relative premium or discount at any given location as compared to the overall market changes over time revealed by nearby property transactions. This implies that one can construct a temporal index (or a map of locational premia/discounts) for a particular location (or for a point in time). So the estimated model presents the evolution of a price surface. This research will be the first time that GWR and STAR have been used to look at the impacts of LRT in the United States. The STAR models will quantify how proximity to LRT been capitalized into the price of single-family homes for discrete time periods, while the GWR will show how price impacts vary across neighborhoods. Taken together, the model results will inform on the impact that LRT has on metropolitan-wide property taxes.

Urban decision-makers would benefit from a more detailed understanding of how and why public policies have different impacts across space and time. Understanding how different neighborhoods are affected during different phases of planning and construction are an important dimension of the debate on value recovery—that is the recouping of public expenditures through taxation—especially because some policy-makers have suggested the use of “zones” of taxation.

---

1 The nearest neighbor component of STAR models allows the spatial (temporal) extent of the lag to increase in more sparsely populated areas (time periods), and contract in areas (time periods) with higher density.
1.3 ORIGINAL OBJECTIVES OF THE PROJECT

The overarching objective of the project was to improve our ability to measure the economic impacts of LRT to inform the debate about public investments in transit systems. The relationship between land use and transportation in contemporary urban environments is complex and understanding of this important topic would benefit from an interdisciplinary perspective. Existing research into the effects of LRT on the land markets is limited in that researchers have focused primarily on station-area rather than metropolitan-wide benefits, used static descriptions of neighborhoods obtained from Census Data, and simplistic measures of accessibility based on Euclidian distance. Furthermore, while considerable progress has been made in crafting models that simultaneously measure effects at the neighborhood scale and reduce spatial autocorrelation in model residuals, such models have yet to be applied to the topic of LRT’s impacts on the land markets.

Accordingly, we proposed to build a series of models that build bring together cutting-edge techniques in spatial econometrics, GIS, and GIS-T. Models would assess the effect of the LRT system in Phoenix, Arizona, from 1995-2008, using land use and transaction data at the individual parcel level. Innovative variables would include metrics comprised of the total tax value of new construction that will serve as a proxy for neighborhood change, and measures of network distance derived using GIS-T that take into account the changing surface of metropolitan accessibility created by ongoing freeway construction. We planned to create GWR and local STAR models, and compare the results against a baseline model to compare model performance. The end result was to be a methodology to not only quantify changes in property values at a regional scale rather than focusing solely on station area impacts, but one that also shows how those effects differ across the metropolitan region. An additional feature of the methodology was its ability to separate mobility-related benefits from those associated with agglomeration effects.

The plan was to translate the model results into monetary terms to provide a dollar value of the early returns to public investment in LRT for three distinct phases: pre-announcement, post-announcement/pre-construction, and construction. Our main policy-related objective was to inform the public debate about the cost-effectiveness of this particular alternative to ongoing automobile dependency.

1.4 REVISED OBJECTIVES OF THE PROJECT

The original project proposal was submitted in March 2008 and the official start date of the project was 22nd August 2008. In September 2008, shortly after the study began, Lehman Brothers declared bankruptcy as a result of financial losses associated with sub-prime mortgage lending. This event marked the beginning of the housing market crisis, the near-collapse of the US banking sector, and the ensuing downturn in the global economy. The housing market in Phoenix, Arizona, where the case study for this project is located, was hit particularly hard by these events. A wave of foreclosures spread across the housing market in metropolitan Phoenix. A preliminary investigation of foreclosure data obtained from the Maricopa County Assessor’s Office revealed two interesting phenomena. First was that foreclosures were not evenly spread across the metropolitan area. Some places (notably retirement communities in up-market neighborhoods) had no foreclosures; whereas other places had a very high concentration or
clustering of foreclosure activity. Second was that foreclosure rates in the Light Rail Transit corridor (defined as the number of foreclosures divided by the number of housing units) was much lower than in outlying areas. This latter pattern was surprising because prior research had shown that property values around station areas had risen much higher than comparable properties outside in anticipation of LRT operation. Another important factor that may have further strained household budgets was a spike in oil and gasoline prices that took place in Summer 2008.

Accordingly, the objectives of the project were revised as follows. First, to conduct a thorough review of the literature to obtain a full understanding of the causes of the housing market crisis, foreclosure activity, and the effect that foreclosure activity has on the value of surrounding properties. Second, use empirical data for metropolitan Phoenix, Arizona, (and potentially selected cities in California—particularly the very hard hit Inland Empire) to quantify the above mentioned effects. A third objective is to incorporate gasoline prices to understand the extent to which the increases in gasoline prices may have exacerbated housing market foreclosures. Fourth, to use the insights gained from the above-mentioned research in a model to investigate how housing in the LRT corridor performed relative to locations outside the corridor, using foreclosures and VMT as statistical controls.

1.5 STRUCTURE OF THIS REPORT

This report consists of five sections. Section two contains a study that was published in Urban Studies. Section three contains a journal article currently in review with Urban Geography. Section 4 contains excerpts from a journal article currently undergoing revision with the Journal of Transportation and Land Use. Section five contains the abstract that summarizes a paper entitled “Light Rail Transit’s Affect on Land Value—How much comes from new construction?” that will be submitted to Environment and Planning A. Section six summarizes additional work in progress.
2.0 STUDY: CAPITALIZATION EFFECTS OF LRT IN PARK AND RIDE VERSUS WALK AND RIDE STATION AREAS

2.1 LITERATURE REVIEW

Light Rail Transit (LRT) is an increasingly popular technology used to produce urban environments that are ostensibly more sustainable. Traditional theories of land economics suggest that improving the surface of accessibility by adding LRT to urban environments will increase land values and promote development of more compact environments clustered onto the spine of transit (Fujita, 1989). Empirical research of numerous systems built in the 1980s and 1990s revealed that overlaying rail systems onto streetscapes configured for the automobile had little impact on land use patterns (Landis and Zhang, 1995; Landis et al., 1995). Research into barriers to land use change in contemporary cities (TCRP, 1997) led to the formulation of a set of best practices to encourage land use change (Belzer et al., 2004; TCRP, 2002, 2004). Current thinking is that urban transportation and land use planning need to be integrated and include tools to ensure that density, design, and the mix of development are conducive to transit—a concept known as Transit-Oriented Development (TOD) (Calthorpe, 2004; Cervero and Kockelman, 1997; Down, 1992, 1997; Duaney et al., 2001).

Case studies of individual station areas or projects have shown that not all locations within a given system have the capacity to generate pedestrian- and transit-friendly environments (Breznau, 2004; Feigon et al., 2004; Leach, 2004; Ohland, 2004a, 2004b). It is becoming increasingly understood that various stations serve different but complementary functions within a given system (Dittmar and Poticha, 2004). Some stations, for example those in downtowns with a high concentration of employment and entertainment, are effectively destinations. At the other end of the node-destination spectrum described by Bertolini and Split (1998) some stations need to host the parking lots that are essential to expanding the commuter-shed or coverage of the system and ensure that LRT integrates with other modes of transportation within the region. Variations in station function were captured by Dittmar and Poticha (2004) in a typology that defined six common classes of TOD and distinguished their roles within regional systems: urban downtown, urban neighborhood, suburban town center, suburban neighborhood, neighborhood transit zone, and commuter town.

Numerous benefits of TOD are cited in the literature, but few agreed-upon ways to measure impacts exist aside from studies that focus on land prices, ridership, and reductions in VMT (Renne and Wells, 2005). Even for a seemingly straightforward study of capitalization benefits, disentangling the effects that may result from LRT-related changes in accessibility from those generated by supplementary policies such as changes in zoning to increase density, and shape the design and composition of urban development, is challenging. Statistically, it is difficult to separate out the importance of being close to transit stops from public policy incentives, like zoning bonuses, in explaining land value increases. In many instances, they are likely to be codependent, with zoning incentives necessary if proximity to transit is to yield dividends, and
proximity to transit necessary if density bonuses and other zoning “perks” are to pay off (TCRP, 2004).

Despite these challenges, many studies have been undertaken to investigate the relationship between LRT and the land markets, theoretically the precursor to land use change (see Hess and Lombardi (2004) and Duncan (2008) and the references therein for extensive reviews)\textsuperscript{ii}. One limitation of the majority of existing studies is that they have quantified the effect of LRT on single-family houses because the data are more abundant and more readily available (see Cervero and Duncan, 2001 for further details). A recent contribution to the literature pertinent to this study is a comparative study of capitalization benefits of LRT in San Diego on single-family houses and condominiums (Duncan, 2008). He found that condominiums—which he argued probably better reflect the density preferences of the TOD market than do single-family houses—experienced higher capitalization benefits than single-family houses (17\% versus 6\%).

Duncan’s (2008) study is important because preferences about housing density are fundamental to LRT/TOD and new urbanism more generally, where the emphasis is on providing an alternative to suburban living (Dittmar and Poticha, 2004; Duaney et al., 2001). Scholars of housing policy and planning believe that a niche market exists for downtown housing consisting of people with preferences for proximity not only to centrally-located jobs but also nightlife, urban amenities, and vibrant urban environments (Danielsen et al., 1999; Farris, 2001; Lang and Danielsen, 2005; Lang et al., 1997; Sohmer and Lang, 2001). Evidence is also mounting that people’s preferences are shifting towards smaller lot/house sizes and shorter commutes in the face of uncertainty about global oil prices in a post-peak oil era and increasing recognition of the need to reduce greenhouse gas emissions (Ewing et al., 2008). Researchers have suggested that demand for high-density, mixed-use housing far outstrips supply (see Ewing at al., 2008, and the references therein), while others have shown that people live at lower densities than they would like (Levine and Frank, 2007; Levin and Inam, 2004). Heterogeneous preferences have been incorporated into both simulation models (examples include Brown and Robinson, 2006; Fernandez et al., 2005), and hedonic models (Bourassa et al., 2003; Fik et al., 2003; Goodman and Thibodeau, 1998; Michaels and Smith, 1990).

A recent study by Kahn (2007), in which he investigated the impact of 14 rail systems built in the US from 1970 to 2000 on neighborhood demographics, highlighted two additional points of relevance to this paper. First, is a distinction between stations with/without a parking lot—which he defined as park-and-ride (PAR) and walk-and-ride (WAR) communities. The rationale for quantifying these effects separately was that WAR communities were more likely to offer a new urbanist lifestyle, while PAR communities may experience a lower quality of life due to increased noise, traffic, and congestion. Kahn found that WAR generated a premium of 3\% in house prices while PAR generated no change. In Boston, where WAR and PAR effects showed the largest divergence, WAR showed a 7\% increase in house prices, while PAR experienced a 5\% decline in house prices. The second point in Kahn’s paper that is relevant to this study is his discussion and treatment of endogeneity issues in the modeling process. Research on infill projects and on NIMBY-ism in general suggests that pre-existing land use—intertwined with pre-existing zoning—and community capital will influence what type of TOD evolves at each station area because transportation agencies tend to follow the path of least resistance (Altshuler
and Luberoff, 2003; Dear, 1992; Farris, 2001; Lake, 1993). This suggests the need for greater consideration of how transportation, land use and land prices interrelate.

In this study I take as my starting point the theory that decisions to build LRT and enact overlay zoning to encourage TOD triggers a well-accepted process of location sorting whereby those who value proximity to LRT and the amenities integral to TOD outbid others for station locations. The basic assumption that LRT and the TOD associated with some of the station areas are built to appeal to those seeking higher density living environments raises a question not only about the effectiveness of using data for single-family house prices to estimate the impacts of urban transit and associated policies (as articulated by Duncan, 2008), but also a methodological issue about the selection of comparables for modeling purposes. Existing studies have not adequately quantified how capitalization benefits for one or more categories of land use vary in different types of land use settings across a given system. I argue that more consideration needs to be given to comparability of neighborhoods in the hedonic modeling process. By that, I mean moving beyond incorporating land use into hedonic models in the form of independent variables (Al-Mosaind et al., 1995; Chen et al., 1998; Duncan, 2008; Nelson, 1999), to using land-use mix as a means of partitioning datasets to create sub-sets of observations to compare capitalization impacts in similar geographic settings.

Accordingly, I use data for the LRT system in metropolitan Phoenix, Arizona, that began operating in December 2008, to inform on how decisions to build LRT and add overlay zoning to existing zoning affect land prices in different geographic settings within the transportation corridor. Four research questions are: (1) “What combinations of land-use mix—or neighborhood type—exist within the transit corridor?” (2) “What is the relationship between neighborhood type and whether or a single-family house or condo in an LRT station area has been subject to overlay zoning?” (3) “What is the relationship between neighborhood type and whether or not a station is a PAR or a WAR?” and (4) “How do capitalization benefits differ for single-family houses and condos in different types of land-use settings?”

This study adds to the existing body of literature by using categorical variables based on land-use mix to partition data to estimate how the capitalization effects of LRT and overlay zoning vary in different geographic settings within a given system in its planning and construction phases. This approach ties recent work on TOD typologies to the capitalization literature, takes into account endogeneity issues, and allows the results to be evaluated within the context of density preferences. Section 2.2, entitled Methodology, contains details of the data and analysis; Section 2.3 contains the results; and Section 2.4 a discussion of the theoretical and methodological implications and policy significance of the study.

2.2 METHODOLOGY

2.2.1 Light Rail in metropolitan Phoenix, Arizona

In the past fifty years, metropolitan Phoenix has experienced rapid growth in population, urbanized area, and vehicle miles traveled (Gober, 2005; Gober and Burns, 2002). For much of that period, transportation policy focused almost exclusively on the automobile (Abbitt, 1990; ARPA, 2006). Public transit—in one form or another—was on the public ballot in Maricopa County for over two decades, with LRT feasibility studies dating back to 1994 (ARPA, 2006;
Although the type of transit has varied over the past two decades, the route has consistently linked major employment centers on Central Avenue with downtown Phoenix, Sky Harbor Airport, and downtown Phoenix. Following numerous unsuccessful votes, Prop2000 was passed by City of Phoenix voters in 2000 to fund the majority of the system, finalize all but one of the station areas, and drafted an ordinance to add overlay zoning along the transit route. The system consists of 20.3 miles of at-grade track that joins the downtowns of Phoenix, Tempe, and Mesa, and 27 stations in a variety of neighborhoods, commercial districts, and an industrial zone.

The case study presents a unique opportunity to separate out the effects of LRT-related accessibility and zoning. Planners at the City of Phoenix strategically implemented zoning ahead of system construction to potentially accelerate development in station areas of land uses conducive to transit (City of Phoenix, 2002, 2004; City of Tempe, 2006; MAG, 2003, 2007). In Phoenix, a draft transit-overlay zoning ordinance began to circulate in 2000. Areas subject to zoning were determined through the participatory planning process. In Tempe, formal transit-overlay zoning was not circulated until 2005, but basically grandfathered all land parcels located within redevelopment zones (City of Tempe, 1997). Areas subject to overlay zoning—which tends to have excluded any large clusters of single-family houses in station areas—are shown in Figure 2.1. This strategic boundary decision means that in some locations transit-overlay zoning roughly corresponds to station area while in others, very little of the station area has overlay zoning, allowing the effects of zoning and accessibility to be measured separately. The period 2001-2007 represented a planning and construction phase in which overlay zoning was finalized, federal funding was obtained to subsidize the system, a county-wide vote was passed to fund extensions to this initial starter segment, and construction took place. As such, I characterize 1995-1999 as a period in which station locations were public knowledge but overlay zoning was not known (“Before”), and 2001-2007 as a period in which both station locations and overlay zoning were public knowledge (“After”).
The overlay-zoning ordinance for Phoenix contains lists of land uses explicitly prohibited (such as car washes and drive-in businesses), conditionally allowed (such as sports stadiums), and explicitly allowed (mixed-use developments and multi-family housing). The ordinance defines transit-oriented development as “a type of development that encourages transit usage by increasing the base of riders through complementary land uses such as office, retail, and housing near transit stations” (City of Phoenix, 2002, p. 5).

The objective of the ordinance is to provide a “transit-supportive environment … [that integrates] auto uses with a complementary mix of land uses, where streets have a high level of connectivity and the blocks are small, all within a comfortable walking and bicycling distance from light rail stations” (City of Phoenix, 2002, p. 6). Tempe’s ordinance also emphasizes development that both promotes alternative transportation modes and encourages a mix of uses and balance of densities and intensities (City of Tempe, 2006). Specific details on maximum allowable densities and particular uses have, in both instances, been left vague.

### 2.2.2 Data

The following datasets were assembled in a Geographic Information System:
Land Parcels: Geo-coded parcel data and sales affidavits\textsuperscript{vi} for the most recent transaction on land parcels from 1995 through 2007. Relevant fields were: property use code, construction date, date and price of most recent sale, lot size, living area, pool size. Source: Maricopa County Assessor’s Office.

Overlay zoning: Transit-related overlay zoning. Source: City of Phoenix, Valley Metro, City of Tempe.


Census Data: Boundaries for Census Tracts as of 2000.

### 2.2.3 Analysis of Land-use mix

Each parcel in the Assessor Office dataset has a land use code, which I reduced to four categories: vacant, residential, TOD-prohibited (if the land use was mentioned specifically in the overlay zoning ordinance as not allowed), and TOD-allowed (if the overlay zoning ordinance would allow that particular use of land). I fish-netted\textsuperscript{vii} the study area into equal-sized cells that approximated the location and size of the most compact Census Tracts, selected those cells that intersected with all land parcels within the transportation corridor (n = 61), calculated the proportion of each type of land use in each cell, and used cluster analysis to produce neighborhood types based on land-use mix.

### 2.2.4 Analysis of Overlay Zoning

An ANOVA test determined whether or not the proportion of single-family houses and condos with overlay zoning was statistically significantly different by neighborhood type.

### 2.2.5 Analysis of PAR and neighborhood type

An ANOVA test determined whether there were statistically significant differences in the proportion of LRT stations with a parking lot by neighborhood type.

### 2.2.6 Transaction Data

I selected all single-family houses that were sold in either 1995-1999 (“Before”) and 2001-2007 (“After”)\textsuperscript{viii}. A second dataset included all condos\textsuperscript{ix} with transactions in 1995-1999 (“Before”) and 2001-2007 (“After”)\textsuperscript{x}.

### 2.2.7 Hedonic models

The generalized form of the hedonic price model for housing is as follows, where $P$ is the selling price, $L$ is a combination of lot-specific characteristics, $N$ is a combination of neighborhood traits, and $T$ is a combination of LRT-related variables.
\[ P = f(L, N, T) \] (2-1)

This may be expanded to:
\[ P = f(L, N, T) = \beta_0 + \sum \beta_i L + \sum \beta_j N + \sum \beta_k T + t + \varepsilon \]

(2-2)

Where: \( \beta_0 \) is the constant term, \( \beta_i \) is the regression coefficient of lot-specific attributes \( L \), \( \beta_j \) is the regression coefficient of neighborhood-specific attributes \( N \), \( \beta_k \) is the regression coefficient of LRT-related attributes \( T \), \( t \) is time, and \( \varepsilon \) is a stochastic or error term. Variables were used to reflect lot-specific characteristics, neighborhood-specific attributes, and LRT-related attributes\( ^{\text{xiii}} \).

Lot-specific variables (L) are: Lot (the size of the land parcel in square feet), Liv (the residential living square footage in the property)\( ^{\text{xii}} \), Pool (a binary variable that reflects whether or not a unit has a private swimming pool), and five binary variables that group the age of housing by era of construction (pre-1920, 1921-1945, 1946-1970, 1971-1990, and post-1990). Neighborhood variables (N) are: DistFree (the straight-line distance to the nearest freeway in feet), DistAmen (the straight-line distance to PVI or peak-value intersection in the nearest downtown, either Tempe or Phoenix); and binary variables for each of the 54 census tracts in the study area. LRT-related (T) variables are BinaryWD (a binary variable that describes whether or not a land parcel is within \( \frac{1}{2} \)-mile walking-distance of a proposed LRT station or not)\( ^{\text{xiii}} \) and BinaryTOD (a binary variable that describes whether or not a land parcel is subject to overlay zoning)\( ^{\text{xiv}} \).

All of the variables in vectors L and N are statistical controls. The variables of interest in this study are in the T vector—BinaryWD and BinaryTOD. In order to control for the effects of time, 49 dummy variables were included in the models for each quarter from Q1 1995 through Q1 2007. I used a semi-log specification because it was robust to various problems (Cropper, Deck and McConnell, 1988).

### 2.2.8 Hypotheses

RQ(1): Location theory and the empirical analysis conducted by Dittmar and Poticha (2004) suggest that various types of neighborhood will exist along the alignment, but the specific categories are an open question.

RQ(2): Theories of NIMBY-ism (Dear, 1992; Pendall, 1999) suggest that some residential communities will have more influence than others in opposing overlay zoning.

RQ(3): Theories of land economics (Fujita, 1989) suggest that locations with a high proportion of land use devoted to amenities will not have a parking lot because other land uses will be more valuable.

RQ(4): The variable BinaryWD should be statistically significant and positive in both the “Before” and “After” periods (because station locations were publically known) for station areas with WAR, but lower for station areas with PAR because of the noise and congestion associated the station-area parking lots reduces the capitalization benefits associated with improved accessibility. Because overlay zoning allows for more dense future development, theories of
profit-maximization suggest that BinaryTOD should be positive and significant in the “After” period, when it was known in the market, and statistically insignificant in the “Before” period (Fujita, 1989). The coefficients for both BinaryWD and BinaryTOD should be greater for condos than for single-family houses because this housing type is a better reflection of self selectors’ density preferences. Similarly, theories of location sorting and associated empirical research (Matthews and Turnbull, 2007; Plaut and Boarnet, 2003; Tu and Eppli, 1999) suggest that the coefficients for BinaryWD should be higher in neighborhoods that offer a new urbanist lifestyle, than those that are more traditionally residential.

The coefficients of the control variables, Lot, LivSqFt, and Pool, should be positive indicating that the greater the values, the higher the sale price. DistAmen, the Euclidian distance to the nearest downtown—either Tempe or Phoenix—should be negative, indicating that the closer to downtown, the higher the property price. The sign of DistFree could be either positive or negative depending on whether proximity to freeways is considered a nuisance or benefit.

### 2.3 RESULTS

#### 2.3.1 Cluster Analysis

The cluster analysis produced five discrete types of land-use mix. The values for each cluster center are shown in Table 2.1. Two neighborhood types—amenity-rich areas with (type 3) and without (type 1) large amounts of vacant land—had very little residential land use (2% and 3% respectively). One set of neighborhoods—type 5—was dominated by residential, comprising two-thirds of all land use by area. The remaining two neighborhood types were mixed-use, dominated by residential (type 2) and amenities (type 4). The results suggest that separate hedonic analyses need to be conducted in neighborhood types 4 and 5 for single-family houses and condos.

<table>
<thead>
<tr>
<th>Land Use</th>
<th>Amenity-rich (1)</th>
<th>Residential-dominated mixed-used (2)</th>
<th>Amenity-rich with vacant land (3)</th>
<th>Amenity-dominated mixed use (4)</th>
<th>Residential (5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>26</td>
<td>4</td>
<td>5</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>TOD-allowed</td>
<td>79%</td>
<td>29%</td>
<td>46%</td>
<td>42%</td>
<td>21%</td>
</tr>
<tr>
<td>TOD-prohibited</td>
<td>3%</td>
<td>14%</td>
<td>6%</td>
<td>16%</td>
<td>7%</td>
</tr>
<tr>
<td>Vacant</td>
<td>15%</td>
<td>16%</td>
<td>45%</td>
<td>12%</td>
<td>5%</td>
</tr>
<tr>
<td>Residential</td>
<td>2%</td>
<td>42%</td>
<td>3%</td>
<td>31%</td>
<td>67%</td>
</tr>
</tbody>
</table>

#### 2.3.2 Neighborhood type and overlay zoning

Overall, 9% of single-family houses within walking distance of LRT stations were subject to overlay zoning, but there were statistically significant differences at the 0.00 level in the average percentage by neighborhood type (Table 2.2). Few single-family houses were present in neighborhood types 1, 2, and 3 (suggesting again that these types of neighborhoods and the
houses within them may skew the results if used in a “global” model). Single-family houses in neighborhood type 5 were less likely (6.7%) than neighborhood type 4 (10.5%) to have overlay zoning. Condos in neighborhood type 4 were four times as likely to have overlay zoning as condos in neighborhood type 5 (43.9% versus 11.9%), and four times as likely as a single-family house in the same type of neighborhood to have overlay zoning.

Table 2.2: Percentage of housing units subject to overlay zoning by Neighborhood Type

<table>
<thead>
<tr>
<th>Neighborhood Type</th>
<th>Single-family houses</th>
<th></th>
<th>Condos</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>% of housing units subject to overlay zoning</td>
<td>N</td>
<td>% of housing units subject to overlay zoning</td>
</tr>
<tr>
<td>1</td>
<td>11</td>
<td>9.1%</td>
<td>265</td>
<td>67.5%</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>0.0%</td>
<td>31</td>
<td>0.0%</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>33.3%</td>
<td>277</td>
<td>70.8%</td>
</tr>
<tr>
<td>4</td>
<td>2,001</td>
<td>10.5%</td>
<td>1,220</td>
<td>43.9%</td>
</tr>
<tr>
<td>5</td>
<td>2,027</td>
<td>6.7%</td>
<td>674</td>
<td>11.9%</td>
</tr>
<tr>
<td>Total</td>
<td>4,048</td>
<td>9.0%</td>
<td>2,467</td>
<td>40.0%</td>
</tr>
</tbody>
</table>

2.3.3 Neighborhood type and PAR/WAR

There were statistically significant differences at the 0.00 level in the type of neighborhood that was treated with a PAR station. Seven out of the eight PAR stations were located in residential neighborhoods, and one located in an amenity-dominated mixed-use area (see Table 2.3). The remaining 19 WAR stations were distributed across the various neighborhood types.

Table 2.3: Park-and-Ride (PAR) by Neighborhood Type

<table>
<thead>
<tr>
<th>Neighborhood Type</th>
<th># of PAR stations</th>
<th>% PAR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>12.5%</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>87.5%</td>
</tr>
<tr>
<td>Total</td>
<td>8</td>
<td></td>
</tr>
</tbody>
</table>

2.3.4 Capitalization effects

The transaction data were partitioned according to the neighborhood types obtained in the cluster analysis. Descriptive statistics for single-family houses and condos in amenity-dominated mixed-used neighborhoods (type 4), and residential neighborhoods (type 5), are in Table 2.4, divided by property type and whether or not transaction took place “Before” or “After” overlay zoning was adopted. Of note is that overlay zoning is more frequent in amenity-dominated mixed-use neighborhoods than residential neighborhoods, for both single-family houses (3% versus 1%), and condos (13% versus 3%).
Two sets of hedonic models were run for neighborhood types 4 and 5, for “Before”/”After”. Individual coefficients are shown in Tables 2.5 (single-family houses) and 2.6 (condos)\textsuperscript{xvi}. 
Table 2.4: Descriptive Statistics of housing transactions "Before" overlay zoning was announced (1995-1999) and "After" (2001-2007)

(a) Single-family houses in amenity-dominated mixed-use neighborhoods (type 4)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Before (n = 732)</th>
<th>After (n = 1,838)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lot</td>
<td>7,817</td>
<td>7,687</td>
</tr>
<tr>
<td>Living Area (sq.ft)</td>
<td>1,350</td>
<td>1,306</td>
</tr>
<tr>
<td>Pool*</td>
<td>0.14</td>
<td>0.14</td>
</tr>
<tr>
<td>Distance to freeway</td>
<td>4,512</td>
<td>4,328</td>
</tr>
<tr>
<td>Distance to nearest downtown</td>
<td>3,718</td>
<td>3,918</td>
</tr>
<tr>
<td>BinaryWD*</td>
<td>0.34</td>
<td>0.36</td>
</tr>
<tr>
<td>BinaryTOD*</td>
<td>0.03</td>
<td>0.03</td>
</tr>
</tbody>
</table>

(b) Single-family houses in residential neighborhoods (type 5)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Before (n = 1,725)</th>
<th>After (n = 4,030)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lot</td>
<td>8,738</td>
<td>8,522</td>
</tr>
<tr>
<td>Living Area (sq.ft)</td>
<td>1,485</td>
<td>1,477</td>
</tr>
<tr>
<td>Pool*</td>
<td>0.23</td>
<td>0.24</td>
</tr>
<tr>
<td>Distance to freeway</td>
<td>5,162</td>
<td>5,044</td>
</tr>
<tr>
<td>Distance to nearest downtown</td>
<td>3,224</td>
<td>3,222</td>
</tr>
<tr>
<td>BinaryWD*</td>
<td>0.19</td>
<td>0.17</td>
</tr>
<tr>
<td>BinaryTOD*</td>
<td>0.01</td>
<td>0.01</td>
</tr>
</tbody>
</table>

(c) Condos in amenity-dominated mixed-use neighborhoods (type 4)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Before (n = 259)</th>
<th>After (n = 882)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lot</td>
<td>1,184</td>
<td>1,128</td>
</tr>
<tr>
<td>Living Area (sq.ft)</td>
<td>1,090</td>
<td>1,035</td>
</tr>
<tr>
<td>Pool*</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Distance to freeway</td>
<td>4,814</td>
<td>5,226</td>
</tr>
<tr>
<td>Distance to nearest downtown</td>
<td>4,591</td>
<td>4,255</td>
</tr>
<tr>
<td>BinaryWD*</td>
<td>0.46</td>
<td>0.37</td>
</tr>
<tr>
<td>BinaryTOD*</td>
<td>0.13</td>
<td>0.10</td>
</tr>
</tbody>
</table>

(d) Condos in residential neighborhoods (type 5)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Before (n = 644)</th>
<th>After (n = 2,427)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lot</td>
<td>1,250</td>
<td>1,128</td>
</tr>
<tr>
<td>Living Area (sq.ft)</td>
<td>1,078</td>
<td>1,025</td>
</tr>
<tr>
<td>Pool*</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Distance to freeway</td>
<td>4,393</td>
<td>4,896</td>
</tr>
<tr>
<td>Distance to nearest downtown</td>
<td>2,884</td>
<td>2,819</td>
</tr>
<tr>
<td>BinaryWD*</td>
<td>0.11</td>
<td>0.13</td>
</tr>
<tr>
<td>BinaryTOD*</td>
<td>0.03</td>
<td>0.03</td>
</tr>
</tbody>
</table>

Notes: *= Binary variable
Table 2.5: Key hedonic model coefficients for single-family houses

(a) Single-family houses in amenity-dominated mixed-use neighborhoods (type 4)

<table>
<thead>
<tr>
<th></th>
<th>Before (Adj-$R^2 = 0.76$)</th>
<th></th>
<th>After (Adj-$R^2 = 0.74$)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>t-statistic</td>
<td>Coefficient</td>
<td>t-statistic</td>
</tr>
<tr>
<td>Constant</td>
<td>11.197**</td>
<td>111.296</td>
<td>11.965**</td>
<td>149.045</td>
</tr>
<tr>
<td>Lot</td>
<td>0.008**</td>
<td>4.354</td>
<td>0.006**</td>
<td>2.482</td>
</tr>
<tr>
<td>LivSqFt</td>
<td>0.373**</td>
<td>12.172</td>
<td>0.398**</td>
<td>16.500</td>
</tr>
<tr>
<td>Pool*</td>
<td>0.157**</td>
<td>5.884</td>
<td>0.059**</td>
<td>2.739</td>
</tr>
<tr>
<td>DistFree</td>
<td>-0.027**</td>
<td>-2.431</td>
<td>-0.027**</td>
<td>-21.650</td>
</tr>
<tr>
<td>DistAmen</td>
<td>-0.004**</td>
<td>-0.660</td>
<td>-0.004**</td>
<td>2.891</td>
</tr>
<tr>
<td>BinaryWD*</td>
<td>0.059**</td>
<td>2.197</td>
<td>0.057**</td>
<td>2.664</td>
</tr>
<tr>
<td>BinaryTOD*</td>
<td>-0.034</td>
<td>-0.596</td>
<td>0.019</td>
<td>0.381</td>
</tr>
</tbody>
</table>

(b) Single-family houses in residential neighborhoods (type 5)

<table>
<thead>
<tr>
<th></th>
<th>Before (Adj-$R^2 = 0.69$)</th>
<th></th>
<th>After (Adj-$R^2 = 0.66$)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>t-statistic</td>
<td>Coefficient</td>
<td>t-statistic</td>
</tr>
<tr>
<td>Constant</td>
<td>10.722**</td>
<td>241.636</td>
<td>11.378**</td>
<td>287.238</td>
</tr>
<tr>
<td>Lot</td>
<td>0.012**</td>
<td>6.017</td>
<td>0.016**</td>
<td>9.026</td>
</tr>
<tr>
<td>LivSqFt</td>
<td>0.362**</td>
<td>24.087</td>
<td>0.341**</td>
<td>24.901</td>
</tr>
<tr>
<td>Pool*</td>
<td>0.071**</td>
<td>6.022</td>
<td>0.071**</td>
<td>6.858</td>
</tr>
<tr>
<td>DistFree</td>
<td>0.023**</td>
<td>5.181</td>
<td>0.016**</td>
<td>4.337</td>
</tr>
<tr>
<td>DistAmen</td>
<td>-0.017**</td>
<td>-3.643</td>
<td>-0.009**</td>
<td>-2.244</td>
</tr>
<tr>
<td>BinaryWD*</td>
<td>0.010</td>
<td>0.676</td>
<td>-0.013</td>
<td>-0.966</td>
</tr>
<tr>
<td>BinaryTOD*</td>
<td>-0.042</td>
<td>-0.979</td>
<td>-0.118**</td>
<td>-2.724</td>
</tr>
</tbody>
</table>

Notes: (1) * Binary Variable; (2) ** p-value < .05; (3) coefficients for Lot, LivSqFt, DistFree and DistAmen have been multiplied by 1,000 (4) dummy variables for each quarter of each year, Census Tract, and age of housing are not shown.
Table 2.6: Key hedonic model coefficients for condos

(a) Condos in amenity-dominated mixed-use neighborhoods (type 4)

<table>
<thead>
<tr>
<th></th>
<th>Before (Adj-$R^2 = 0.74$)</th>
<th>After (Adj-$R^2 = 0.53$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient   t-statistic</td>
<td>Coefficient   t-statistic</td>
</tr>
<tr>
<td>Constant</td>
<td>9.730**       45.601</td>
<td>10.900**       66.634</td>
</tr>
<tr>
<td>Lot</td>
<td>0.034**       3.450</td>
<td>0.061**       2.362</td>
</tr>
<tr>
<td>LivSqFt</td>
<td>0.828**       10.818</td>
<td>0.678**       11.591</td>
</tr>
<tr>
<td>Pool*</td>
<td>N/A           N/A</td>
<td>N/A           N/A</td>
</tr>
<tr>
<td>DistFree</td>
<td>-0.083**      -5.762</td>
<td>-0.080**      -7.037</td>
</tr>
<tr>
<td>DistAmen</td>
<td>0.032**       4.349</td>
<td>0.052**       8.880</td>
</tr>
<tr>
<td>BinaryWD*</td>
<td>0.282**       6.015</td>
<td>0.164**       4.657</td>
</tr>
<tr>
<td>BinaryTOD*</td>
<td>0.001         0.008</td>
<td>0.373**       3.236</td>
</tr>
</tbody>
</table>

(b) Condos in residential neighborhoods (type 5)

<table>
<thead>
<tr>
<th></th>
<th>Before (Adj-$R^2 = 0.73$)</th>
<th>After (Adj-$R^2 = 0.60$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient   t-statistic</td>
<td>Coefficient   t-statistic</td>
</tr>
<tr>
<td>Constant</td>
<td>9.804**       104.299</td>
<td>10.726**       156.768</td>
</tr>
<tr>
<td>Lot</td>
<td>0.023**       3.349</td>
<td>0.007**       2.671</td>
</tr>
<tr>
<td>LivSqFt</td>
<td>0.693**       16.816</td>
<td>0.628**       20.592</td>
</tr>
<tr>
<td>Pool*</td>
<td>N/A           N/A</td>
<td>N/A           N/A</td>
</tr>
<tr>
<td>DistFree</td>
<td>0.031**       4.023</td>
<td>0.007**       2.863</td>
</tr>
<tr>
<td>DistAmen</td>
<td>0.000**       3.202</td>
<td>0.004**       2.527</td>
</tr>
<tr>
<td>BinaryWD*</td>
<td>-0.130**      -4.602</td>
<td>0.034**       2.514</td>
</tr>
<tr>
<td>BinaryTOD*</td>
<td>0.038         0.762</td>
<td>-0.111**      -2.685</td>
</tr>
</tbody>
</table>

Notes: (1) * Binary Variable; (2) ** p-value < .05; (3) coefficients for Lot, LivSqFt, DistFree and DistAmen have been multiplied by 1,000 (4) dummy variables for each quarter of each year, Census Tract, and age of housing are not shown.

All coefficients were significant at the 0.05 level except those shaded. The two LRT-related variables are binary, and as such can be interpreted roughly as change in percentage price in the presence of that binary variable. For single-family houses in amenity-dominated mixed-use neighborhoods, being within walking distance of an LRT station resulted in a 6% premium in house price, and overlay zoning was statistically insignificant. For single-family houses in residential neighborhoods, the coefficient for BinaryWD was statistically insignificant, and the coefficient for BinaryTOD was significant at the .05 level, and strongly negative. Being within walking distance of an LRT station in a residential neighborhood—which are primarily PAR versus WAR—produces no capitalization benefit, and in fact, houses in these types of neighborhood that are subject to overlay zoning sold at a 12% price discount to those without overlay zoning.
The coefficients of BinaryWD and BinaryTOD for condos in mixed-use amenity-dominated neighborhoods and residential neighborhoods showed a strong divergence. In mixed-use amenity neighborhoods, the coefficients for BinaryWD were 0.28 “Before” and 0.16 “After” indicating that condos that were treated with LRT sold at a premium of 28% and 16% higher than condos in similar neighborhoods without LRT. A difference of means test confirmed that the BinaryWD coefficients in these two regressions were statistically significantly different at the 0.02 level (Paternoster et al., 1998). The 0.37 coefficient for BinaryTOD in the “After” period indicates that overlay zoning leveraged the capitalization benefits afforded by LRT-related accessibility by a further 37%. In contrast, in residential neighborhoods, being treated with an LRT station resulted in a 13% discount for condo prices in the “Before” period, while the negative and statistically significant coefficient for BinaryTOD indicates that overlay zoning created an additional discount. The signs of the control coefficients were as expected, and their magnitudes appeared reasonable. The average values of the adjusted-R² for the single-family and condo regressions were 72% and 68% respectively, at the lower end of the range for similar studies.

2.3.5 Spatial Autocorrelation

Standard tests of spatial autocorrelation in each of the model’s residuals were carried out (Cliff and Ord, 1973). The Moran’s I was found to be statistically significant in two cases, but small (at 0.01 for single-family houses in neighborhood type 5 “Before”, and 0.24 for condos in neighborhood type 5 “Before”), and deemed not to be problematic.

2.4 EVALUATION AND CONCLUSIONS

The analysis suggests that land-use mix or neighborhood setting shapes whether or not an LRT station is a WAR or a PAR, and whether or not land parcels are subject to overlay zoning, and that these two factors influence the subsequent process of location sorting. The capitalization benefits associated with LRT and overlay zoning are separately identifiable, and vary according to housing type and neighborhood setting.

Single-family houses in amenity-dominated mixed-use neighborhoods that are treated with WAR stations sell at a premium of approximately 6% “Before” overlay zoning is announced, and 6% “After”. These levels are within the range of existing studies that have shown that LRT results in a modest premium (of between 0% and 10%) for single-family house prices (Duncan, 2008), and comparable to the 3% capitalization benefits quantified by Kahn (2007). The appearance of capitalization benefits as soon as stations were public knowledge is consistent with the findings of a study by (Knaap, Ding and Hopkins, 2001), who concluded that the markets respond to plans to build transit... In contrast, the absence of capitalization benefits for single-family houses in neighborhoods that are primarily residential and where most of the stations have a parking lot (PAR), and the detection of a 12% discount if they are subject to overlay zoning, suggests a different dynamic is taking place in these neighborhoods. The lack of capitalization benefits could be due to the fact that primarily residential neighborhoods do not attract the demographic associated with new urbanism (Danielsen et al., 1999; Farris, 2001; Lang and
Danielsen, 2005; Lang et al., 1997), or that the benefits of improved accessibility associated with LRT are offset by the perception of the parking lot as a disamenity (Kahn, 2007). The discount on single-family house prices associated with overlay zoning suggests that those attracted to this type of housing in primarily residential neighborhoods have a preference for low-density development and perceive the higher density development associated with overlay zoning as negative.

Condos in amenity-dominated mixed-use neighborhoods—which as Duncan (2008) argued may better represent the density preferences of those wishing to live near LRT—experience much higher capitalization effects than single-family houses. The more than four-fold capitalization effect for condos is higher in absolute terms than the 17%/6% difference identified by Duncan (2008), but similar in relative terms. The results reinforce Duncan’s argument that existing studies that use data on single-family houses to measure the capitalization impacts of LRT may have understated its impacts. Overlay zoning—and the promise of TOD-conducive land uses in the future—adds an extra 37% premium to the price of condos compared to other units in similar land-use settings.

In sharp contrast to these price rises, condos that are located in neighborhoods that are primarily residential and that have parking lots experience a price discount of 13%, likely from the increased traffic, noise, and congestion associated with the parking lot. Overlay zoning in these neighborhoods depresses prices a further 11%. This outcome is not consistent with the assumptions about density preferences, and warrants further investigation. One possibility raised by Kahn (2007) is the fear that transit will attract lower-income captive riders, and that overlay zoning may be used to produce high-density apartment housing for lower-income populations.

Overall, the results suggest that in the process of location sorting, neighborhoods that experience the strongest capitalization benefits are those most likely to evolve into TOD communities that focus on walkable neighborhoods and mixed-use, and that condos within these neighborhoods may be more sought-after than single-family houses. The implications of this study for public policy are that decisions to build LRT and adopt overlay zoning have a highly uneven effect on different types of housing in different locations across a given system. Overlay zoning—when used in the appropriate type of neighborhood—can enhance the capitalization effects generated by LRT-related accessibility improvements. However, overlay zoning—or perhaps the high-density development associated with it—can also produce a discount in residential capitalization in some neighborhoods. Further investigation into this differential valuation of overlay zoning is needed.

The methodology used in this study, which was motivated by the emerging literature on TOD typologies, suggests that neighborhood type (as determined by land-use mix) may be a useful way to partition datasets for hedonic analyses. The theoretical implications are that hedonic modeling of amenity-oriented treatments in urban settings needs to be tied more closely to the emerging literature on types of neighborhood and the density preferences associated with them, and deal more directly with endogeneity issues.
Suggestions for future research are adding variables to the model that can capture processes and features that are fundamental to LRT-inspired neighborhood change. In this model, as in many other hedonic models, neighborhood characteristics are captured by census variables at a particular point in time. While the models presented in this paper also have dummy variables to capture the effects of time, future research could investigate whether the fit of the model could be improved by including a variable with both spatial and temporal characteristics that acts as a proxy for neighborhood change. Similarly, the general model presented in this paper contains no variable that acts as a proxy for urban design, that researchers have shown to be an important determinant of price (Matthews and Turnbull, 2007; Plaut and Boarnet, 2003; Tu and Eppli, 1999). Variables such as these may be useful for studies of LRT because of the fundamental premise that such policies are intended to provide the catalyst for the construction of new TOD communities that are designed to attract a target demographic and that both the new construction and the socio-demographic changes will likely affect the price of existing properties.

2.5 NOTES

(i) An increasingly popular way to change zoning is to add overlay zoning, which is a regulatory tool that creates a special zoning district, placed over an existing base zone(s). The overlay zoning identifies special provisions in addition to those in the underlying base zone to guide development within a special area (City of Portland, 2008).

(ii) Research findings on the effects of proximity to transit on land values are not very consistent in part because impacts vary depending on the severity of traffic congestion, local real-estate market conditions, swings in business cycles, and other factors (TCRP, 2004). Other sources of inconsistency include the use of different model specifications and the fact that the quality of data vary across studies. The average premium for single-family house prices in existing studies ranged from 0% to 10% (Duncan, 2008).

(iii) Defined as being within a one-mile buffer of the LRT alignment, a commonly-used zone of influence (Parsons-Brinckerhoff, 2001).

(iv) In the Assessor Office dataset condos are individually-owned and separately taxable entities, and are treated separately from multi-family housing units that tend to be rented, are taxed as a unit (i.e. all the housing units within the development), and classified as commercial properties.

(v) Mesa, the municipality that contains the most eastern station, did not enact overlay zoning. Municipal dummy variables were not needed because dummy variables for each Census Tract were included in the models.

(vi) The data come from sales affidavits and are considered to be accurate for individually-owned properties as they are used by the Maricopa County Assessor’s Office to calculate property taxes.

(vii) Fish-netting is a commonly-used tool in geographic analysis that is used to produce grid cells. It has long been accepted that the process is sensitive to the placement, orientation, and size of the grid cell (Muller, 1977), but it does offer a means of aggregating parcel level data.
(viii) Approximately 1% (113) of observations was deleted from the dataset: one transaction for over $5 million that appeared to be incorrect based on the parcel characteristics, and transactions under $20,000, reducing the dataset to 10,571 observations.

(ix) In this particular dataset, the designation of condo reflects a form of ownership and includes both town homes and apartment-type complexes. The models could be improved by separating out the types of units more finely should such data be available.

(x) Approximately 3.6% (200) of observations were deleted for groups of nearby condos in the same development on the same sale date where the transaction price appeared to be for the entire group sale. Excluding these data left 5,285 observations.

(xi) Pre-regression tests were run to identify whether there were multicollinearity problems among the independent variables. BinaryWD and BinaryTOD had a statistically significant correlation, but at 0.23 for all single-family houses and 0.46 for condos, they were below the standard threshold of concern of ±0.70 (Clark and Hosling, 1986). Pair-wise correlations for variables ranged from a high positive correlation of 0.67 for Lot and LivSqFt for condos, to -0.34 for BinaryTOD/DistAmen for condos, also below the general threshold of concern.

(xii) For condo complexes, the Assessor’s Office assigns a pro-rated proportion of the building footprint to each individual unit. The data are limited in that no information are available on views, number of floors, or building height.

(xiii) Modelers have used numerous variables to represent distance to an LRT station, ranging from straight-line distance to the station, to a series of buffers, network distance based on the street configurations, and mathematical transformations of distance such as the reciprocal distance to convey some distance decay (see Parsons-Brinkerhoff, 2001 for a detailed summary). The variable used in this study, BinWD, is a hybrid of network distance and buffering because it includes areas that are within ½-mi walking distance of an LRT station using the pre-existing street network.

(xiv) This binary variable implies that treatment with overlay zoning is dichotomous, an assumption that only holds true if similar geographic settings are considered. For example, a downtown area that already has higher density development, may not need overlay zoning to stimulate TOD.

(xv) Freeways have long been considered a potential disamenity in urban areas because of the noise, air pollution, and congestion generated by vehicular traffic (Lash, 1965). In some locations within the study area, the freeways are sunken and have noise buffers to mitigate their impact on local communities; in other places, they transect communities. In this particular study, therefore, freeways may therefore be considered a locally-unwanted land use in some areas while in others their disamenity is outweighed by the mobility benefits provided to the community. If freeways are considered a benefit, then the expected sign of the coefficient would be negative, indicating that the further a property is from a freeway, the lower the price; if freeways are considered a nuisance, the reverse would be true.

(xvi) Initial model results showed signs of heteroscedasticity. The paper presents results from LIMDEP corrected for heteroscedasticity (Greene, 2002). LIMDEP uses a revised and robust OLS covariance matrix based on White (1978) who developed a consistent estimator of the standard errors under heteroscedasticity. The correction does not change the coefficient estimates, but does change the standard errors and t-ratios.

(xvii) Coefficients for binary variables in models with a semi-logarithmic functional form approximate, but do not translate directly, into elasticities. Following the work of Halvorsen and Palmquist (1980), the approximate influence of the binary variable, \( b \), on price, \( P(H) \) is as
follows: A more accurate approximation for the estimator was suggested by Kennedy (1981), as follows: 

\[ P(H) = \exp \left( \beta_i - \frac{1}{2} \text{VAR}(\beta_i) \right) - 1 \]  

(2-4)

(xviii) Estimates for coefficients for the LRT-related binary variables calculated using both Halvorsen and Palmquist’s estimation and Kennedy’s modification showed little difference.

(xix) The major difference between this study and that conducted by Knaap et al., 2001, is that they focused on the Blue Line, which was an extension of the system that opened in 1986, rather than the starter segment, which is the case for Phoenix. The major similarity is that the analysis focused on “pre-service impacts” using data for 1992-1996, ahead of 1998 opening.

(xx) Average Odds Ratio for Transportation costs variable in quarters six and seven= \exp(0.82*X). For every 0.1% increase in Transportation costs Odds Ratio = \exp(0.82 * 0.1) = 1.08. For every 1% increase in Transportation costs Odds Ratio = \exp(0.82 * 1) = 2.3
3.0 THE GEOGRAPHY OF FORECLOSURE

3.1 INTRODUCTION

Subprime loan originations, fraud, inexperienced borrowers, and exuberant lending institutions came together to create ‘the perfect storm’ that swept across the United States housing market (Carr 2007). The resultant economic downturn in the United States and subsequent global recession provided a stunning reminder of the extent to which places have become interconnected in the contemporary world and of the risks associated with financial globalization (Crump et al. 2008; Lee et al. 2009; Aalbers 2009; Engelen and Faulconbridge 2009). Trillions of dollars have been spent by governments around the world bailing out financial institutions affected by the crisis to prevent a collapse of the global economic system (Aalbers 2009, 2009; Lee et al. 2009). An estimated eight million households in the United States were affected by foreclosure action between 2007 and 2009, over half of which concentrated into just 35 counties mostly within California, Arizona, Nevada and Florida (Heath and Merrill 2009).

The uneven effects of the foreclosure crisis support the fundamental argument made by Aalbers (2009) and reinforced by Engelen and Faulconbridge (2009) that the crisis is ‘geographic to the bone’. It also supports the point made by Kaplan and Sommers (2009) that formulating effective policy responses to the problems generated by subprime mortgage defaults, housing market implosion, and foreclosure crisis require an understanding of the geographic dimensions of these interrelated issues. A great deal of research has been directed towards understanding the causes and uneven impacts of the crisis. Surprisingly little attention has yet been paid to how foreclosures unfolded in individual metropolitan regions in the sunbelt states which experienced some of the highest foreclosure rates in the country.

Accordingly, in this paper we investigate the spatial and temporal patterns of residential foreclosures in Maricopa County, Arizona, that became real-estate owned (REO) from January 2007 through February 2009, when President Obama announced the MHA program (Squires and Hyra 2010). We use a variety of quantitative methods to understand the geography of foreclosures and how it changed over time. Our objective is to use the empirical study to link neighborhood characteristics and spatial contagion with respect to foreclosures and to illustrate how they interact. The results provide a platform for future studies into foreclosures that could, for example, investigate the effect of foreclosure on house prices across different neighborhood types, and investigate how higher gasoline prices may have exacerbated the crisis. Both studies are intended to help guide ongoing responses to the crisis especially in strongly impacted areas.

3.2 CAUSES OF FORECLOSURE

The magnitude of the crisis has elicited a great deal of debate about its causes and effects. We begin our literature review with a sub-section that describes the timeline of the crisis. After this
broad overview, we focus more specifically on empirical studies that contain quantitative models designed to explain foreclosure. This leads naturally to geographic perspectives on the crisis and studies that have used neighborhood variables to inform the geography of foreclosures as well as work that has situated the crisis within its broader social, political, and economic context. The literature review section concludes with a brief explanation of how all of these approaches can be integrated in a conceptual approach that uses geographic approaches and methodologies along with spatial contagion to explain the foreclosure crisis, and potentially offer some guidance for ongoing policy responses.

3.2.1 Background

Fallout from subprime mortgage defaults reached global proportions in February 2007 when HSBC, the world’s largest bank, revealed that it had written down $11 billion worth of mortgage-backed securities (Waples and Ringshaw 2007). In a matter of months, a procession of home loan providers including New Century Financial and American Home Mortgage filed for bankruptcy because of losses in the subprime mortgage market (McLaughlin, Stempel, and Giannone 2007; Creswell 2007). Wall Street losses associated with assets that had become ‘toxic’ were compounded by record high levels of leverage. Six of America’s largest investment banks either went bankrupt (Lehman Brothers), were acquired by other companies (Bear Stearns and Merrill Lynch), or converted to commercial banks (Goldman Sachs and Morgan Stanley) during 2008 (Labaton 2008). At the same time, government-sponsored enterprises (GSEs) Fannie Mae and Freddie Mac, who together either directly owned or guaranteed close to $5 trillion in mortgage obligations, were placed into conservatorship with the Federal Housing Finance Agency (Walker 2008). In addition, over one hundred insolvent US banks were taken over by the Federal Deposit Insurance Corporation in 2008/9 (FDIC 2010).

The federal government’s response to the crisis during this period consisted largely of channeling taxpayer money to private and quasi-governmental financial sector entities (Squires and Hyra 2010). Some commentators have remarked with consternation that the Federal Reserve Board had initially responded to the growing crisis in the subprime mortgage markets with a ‘shrug’ (Andrews 2007). It was not until February 2009 that any assistance was provided directly to homeowners, long after many people had been evicted. President Obama’s $787 billion economic stimulus plan to buttress the US economy included a $75 billion Making Home Affordable (MHA) plan to assist homeowners struggling to avoid foreclosure (Squires and Hyra 2010). In the intervening period, the wave of foreclosures spread from the rustbelt to formerly ‘hot’ markets located primarily in the rapidly-growing sunbelt states where house prices were rapidly deflating (Crump et al. 2008; Immergluck 2009).

3.2.2 Explanatory Variables of Foreclosure

Literature that pre-dates the foreclosure crisis contains two general sets of explanations about what causes foreclosures that relate to ability to pay, and willingness to pay. Homeowners who fall behind on their monthly payments—most often because of ‘income shocks’ and/or ‘trigger events’ such as loss of employment, divorce, or serious illness—have several options open to them (Elmer and Seelig 1998). They can request that the lender modify their loan payments, enter the foreclosure process or, if the housing market is rising, or if they have a sufficient
amount of equity in the house, they may be able to refinance (Coulton et al. 2009). Research has suggested that households at the lower end of the income spectrum experience foreclosure more often because they are more likely to experience income shocks and, at the same time, have less capacity to weather those shocks (Elmer and Seelig 1998). The subprime mortgage market which has long provided credit to individuals with less-than-stellar credit records and low income which has always been associated with higher levels of default (e.g. Laderman and Reid 2008 for a detailed treatment of this issue).

Other researchers have suggested that not all foreclosures would result from this ‘double trigger’ effect, and that some could just result from investor and owner-occupiers walking away from properties that were underwater, effectively treating housing as a put option (Foster and Van Order 1984). Both types of explanations emphasize the importance of house price trends in explaining foreclosure. Loan terms found to influence mortgage defaults are: (1) LTV ratio; (2) debt-to-income ratio; (3) credit score; (4) whether income was verified by the lender; and (5) adjustable-rate mortgages (ARMS) that have lower-than-market interest rates at the beginning of the repayment period.

### 3.2.3 Geographic approaches to understanding foreclosure

Explanations of foreclosure that pre-date the recent crisis have proven to be overly-simplistic because of dramatic changes in the landscape of housing finance that developed alongside the inflation of the housing market bubble. One particularly controversial issue relates to role in the crisis of the dramatic increase in securitized subprime mortgage lending and its geographic concentration in minority neighborhoods. Rather than look at factors driving the crisis in isolation, geographic explanations of the crisis that have situated housing finance issues within their broader context allow for a rich, multi-layered understanding of the interaction between borrowing decisions by individuals, lending decisions by institutions and government regulations. Forces came together to create what Carr (2007) described as a ‘perfect storm’ of subprime loan originations, fraud, inexperienced borrowers, and exuberant lending institutions. In this explanation, borrowers (especially low-income minorities) were encouraged by mortgage brokers to take out high cost loans that they could not afford to repay. The mortgages they obtained were bundled together and sold in the secondary market in the form of mortgage-backed securities (MBSs), collateralized debt obligations (CDOs) or structured products. Some argue that the government not only failed to regulate these transactions effectively but may have unwittingly helped to inflate the housing market bubble by keeping interest rates at historically low levels (Wyly et al. 2009; Mizen 2008). Others have squarely accused the federal government of failing to regulate subprime lenders (Crump et al. 2008).

A number of studies—several of them predating the foreclosure crisis—identified that subprime loans that tended to be ARMs were concentrated in low-income minority neighborhoods (e.g. Ashton 2008; Calem, Gillen, and Wachter 2004; Calem, Hershaff, and Wachter 2004; Howell 2006; Wyly et al. 2008; Newman and Wyly 2004; Crump 2005, 2007). Research has also shown that racial and ethnic minorities were more likely to hold subprime mortgages even after controlling for credit score and income (Newman and Wyly 2004; Laderman and Reid 2008; Crump 2005; Wyly et al. 2009). Considerable evidence also emerged of predatory lending practices—some of which were fraudulent—in the subprime market (Fuentes 2009; Hill and
Kozup 2007; Newman and Wyly 2004; Pennington-Cross and Ho 2008; Wyly et al. 2006; Wyly et al. 2007). It is difficult to overlook the fact that this high-cost lending—whether bona fide subprime or predatory in nature—was concentrated in the same neighborhoods in which mortgage credit was long absent (Kaplan and Sommers 2009; Mian and Sufi 2009; Wyly et al. 2009). This has caused some to go so far as to label concentrated subprime lending as housing discrimination (Howell 2006; Rugh and Massey). Some have concurred with the opinion that mainstream economics, with its ‘obsession for equilibriums’ has trouble understanding change, and that the messy environments created by change are better explained by sociologists and geographers who provide context-rich explanations grounded in multi-theoretical approaches than do the clean, abstract, and parsimonious modeling favored by economists (Aalbers 2009; Hirsch, Michaels, and Friedman 1987; Peck 2005). Extending these critiques is the important point that because economic theory dictates that mortgage pricing and foreclosure risk are determined by market forces, discrimination—if it were to exist—would be conceptualized as a market failure rather than a real-world phenomenon (Wyly et al. 2009).

One very important point about the subprime mortgage market is that the stream of credit is distinct from prime mortgages and flows through less heavily-regulated independent mortgage companies via a process called ‘channel specialization’ (McCoy and Renuart 2008). Credit expansion was facilitated by the ‘originate to distribute’ business model of many mortgage finance companies separated the underwriter making the credit extension decision from exposure to the ultimate credit quality of the borrower (Ashton 2009). This compartmentalization gave mortgage brokers the incentive to maximize lending volume while shielding them from default risk (LaCour-Little 2009). Agents of subprime capital expansion strategically fought regulation and reform by arguing that the risk-based pricing of mortgage products would allow everyone to attain the American Dream of homeownership and dismissing evidence of racial discrimination in lending as anecdotal evidence of rare problems in an otherwise benevolent and what some economists theorized to be a ‘color-blind’ free-market landscape (Wyly et al. 2009).

Taken together, these studies support the major point made by Wyly et al. (2009) that the institutionalized racism that had once been responsible for ‘redlining’ in many low-income minority neighborhoods morphed into the more insidious process of ‘reverse redlining’ or ‘greenlining’ that was marketed under the guise of democratizing access to home ownership.

One of the most unsettling conclusions that can be drawn from the above-mentioned literature is that crisis may have been averted had agents understood the factors driving the foreclosure spike that started in minority neighborhoods, and taken measures to address them (Carr 2007). Instead, subprime lending continued unabated, in what Crump et al. (2008) characterized as ‘money chasing borrower’ phase, this time focusing on suburban and exurban mortgage markets. He attributed these ill-advised and poorly underwritten mortgages, problems in the broader economy, and an unexpected surge in energy costs to the rapid rise in foreclosures in formerly secure outlying metropolitan locations.

3.2.4 Economic Approaches to the Foreclosure Crisis

Several influential and comprehensive studies into the subprime mortgage crisis were undertaken by Federal Reserve Bank economists (Foote et al. 2008; Foote, Gerardi, and Willen 2008;
Gerardi et al. 2008; Gerardi and Willen 2009). Explanations included an underestimation of the default risk of subprime mortgages by market participants and rating agencies alike, a miscalculation of the likelihood of a downturn in the housing market, and a lack of understanding of the sensitivity of foreclosures to falling house prices (Gerardi et al. 2008). They accepted the argument that a major driver of the subprime crisis was the increased use of securitization made intuitive sense and was persuasive (e.g. Keys et al. 2008). Nevertheless, they concluded that the risks came from uncertainty about the future direction of house prices rather than the quality of underwriting.

3.2.5 Contagion over space and time

Another very important strand of literature is work that takes a spatial econometric approach to understanding how foreclosure reduces the price of nearby properties through the process of spatial contagion (Harding, Rosenblatt, and Yao 2009; Immergluck and Smith 2006; Lin, Rosenblatt, and Yao 2009; Rogers and Winter 2009). Immergluck and Smith (2006), using data in Chicago in 1999, found that one foreclosure resulted in a 1% decrease in value for properties within a 1/8th mile radius. Also in Chicago, Kin, Rosenblatt and Yao (2009) examined transactions from 2006 to show that foreclosures negatively affect the price of houses up to 0.9 km away for up to five years. Rogers and Winter (2009) analyzed single family home foreclosures in St. Louis between 2000 and 2007 and found that one foreclosure decreased house prices within a 1/8 mile radius by about 1%. Finally, a nationwide study by Harding, Rosenblatt and Yao (2009), estimated a contagion effect of the closest foreclosure to be less than 0.5% within 1/8th mile, and the cumulative effect of being near three or more foreclosed properties within 300 feet created a 3% discount in sale price. This research provides a much-needed spatial dimension to the issue of foreclosures but falls short of explaining why some parts of the US—and indeed different neighborhoods within cities—have been devastated by the crisis while other places have been largely unaffected.

3.2.6 The Effect on Foreclosures of the 2008 Surge in Gasoline Prices

The surge in global oil prices in 2008 was unexpected, given the sharp downwards revision in US—and therefore global—growth. Conventional theory suggests that a recession would reduce global demand for oil and lead to a price reduction. A slew of studies were undertaken to try to understand why oil prices would be rising at the same time as the US economy was going through such turmoil. Research was inconclusive, but some studies suggested speculation, the role of OPEC and resource depletion, could have played a role in the oil bubble of 2008 (Gately and Huntington, 2002, Hamilton, 2009, Smith, 2009, Tokic).

3.2.7 Evaluation of the literature

Something of a chasm separates traditional economic approaches to understanding the subprime crisis and geographic approaches that consider the institutional context in which lending took place. One way to potentially bridge this divide is through detailed spatial econometric studies that have clearly identified the importance of the role of spatial contagion in explaining how foreclosures spread. Missing from the sophisticated econometric analyses—and indeed from the models used to analyze housing market risk on Wall Street and by the Federal Reserve Board—
is any consideration of the geographic concentration of subprime lending, the geographic patterns of foreclosure, and contagion over space and time. The combination of spatial contagion and highly geographically concentrated lending are critically important factors that affect the trajectory of house prices. It is evident that geographic approaches and methods need to be incorporated into conceptual, theoretical and quantitative models to explain how and why the foreclosure crisis spread and to assist ongoing policy responses to the crisis. Evaluating these various bodies of literature studies side-by-side suggests that two important elements of any research study that attempts to integrate these approaches are the recognition that loan type and terms, income, and race/ethnicity interact; and that foreclosure creates highly localized housing price dynamics that may stimulate more foreclosure. Accordingly, we used these factors to inform the conceptual approach used in our empirical analysis that is operationalized with the empirical study detailed later in this Chapter.

3.3 EMPIRICAL ANALYSIS ON FORECLOSURES

3.3.1 Study Area

Metropolitan Phoenix has experienced rapid growth in the past several decades to become the 12th largest metropolitan statistical area in the United States with a population of over 4.3 million people spread over approximately 2,800 square miles (GPEC 2006). Housing construction plays an important role in this regional economy, predicated upon more growth (Gober 2005). Between 1990 and 2005 (before the housing market downturn) over half a million housing units were built, eighty percent of which were single family houses (Atkinson-Palombo and Gober 2010). Despite some infill and a shift towards multi-family housing, the area remains fairly low density, with very little transit, high levels of automobile ownership use, and considerable transportation costs.

The area has high levels of minorities—in this case Latino ethnicity—many of whom are concentrated into particular neighborhoods. An additional feature that may influence the geography of foreclosure is that this metropolitan area contains a large number of retirement communities located primarily at the urban fringe where lifestyle retirees are more likely than other residents to own mortgage-free homes (Gober 2005). The geography of the metropolitan area allows for an investigation of how race/ethnicity, subprime lending and foreclosure interrelate. An important point to note is that in this paper we do not directly examine house price dynamics in each neighborhood although this will be the focus of a future study.

3.3.2 Conceptual approach

Prior research suggests that race/ethnicity, subprime lending, loan-to-value (LTV) ratio and other measures of fiscal stress, interrelate with one another, and that foreclosures dampen prices of nearby housing that causes even more foreclosures. We use a variety of data and methods to empirically test how loan characteristics, race/ethnicity, LTV and measures of fiscal stress interrelate with one another; how they affect foreclosure, and how foreclosures spread over time and space. This approach adds to the literature by blending characteristics of neighborhood with spatial contagion.
3.3.3 Data

We assembled four spatial datasets into a Geographic Information System (Table 3.1). The spatial units of the data varied from the individual parcel level for housing units and foreclosure, to census tracts for data on transportation costs and on social, demographic, and economic variables collected by the Census Bureau; to the zip code level for data on loan profiles. Census tracts and zip code areas vary in their areal extent, and tend to be larger at the urban fringe. In order to correct for any cartographic distortions created by this variation in size, we fishnetted the study area creating 1 mile x 1 mile grids that lined up with census tract boundaries in downtown Phoenix. We joined all four datasets to this spatial unit. The Assessor Office data contain two important pieces of information for this analysis: first, individual land parcels and their use; and second, those land parcels that had become Real Estate Owned (REO), along with the date. We use these fields to identify raster cells that had more than 10 housing units and determine the Foreclosure Rate, defined as the number of REO foreclosures divided by the number of housing units. This operation produced a data matrix consisting of 1,560 observations with variables describing Credit Profile; Loan Type; Social, Demographic, Racial/Ethnic and Income; and Housing Stress (LTV ratios) (Figure 3.1).

<table>
<thead>
<tr>
<th>DATASET</th>
<th>SOURCE</th>
<th>GEOGRAPHIC SCALE</th>
<th>TIME PERIOD</th>
</tr>
</thead>
<tbody>
<tr>
<td>FORECLOSURES AND DETAILS OF HOUSING STOCK</td>
<td>MARICOPA COUNTY ASSESSOR'S OFFICE</td>
<td>HOUSING UNIT (POINT FILE)</td>
<td>JAN 2007- FEB 2009</td>
</tr>
<tr>
<td>LOAN CHARACTERISTICS</td>
<td>LENDER PROCESSING SERVICES (LPS)(^a)</td>
<td>ZIP CODE</td>
<td>AS OF JANUARY 2007</td>
</tr>
<tr>
<td>NATIONAL HOUSEHOLD TRANSPORTATION SURVEY</td>
<td>BUREAU OF TRANSPORTATION STATISTICS</td>
<td>CENSUS TRACT</td>
<td>AS OF 2002</td>
</tr>
<tr>
<td>GRIDS</td>
<td>DERIVED</td>
<td>1 MILE X 1 MILE</td>
<td></td>
</tr>
</tbody>
</table>

\(^a\)Data on the zip-code profile of loan characteristics, lenders, and borrowers, was purchased from Lender Processing Services, formerly McDash Analytics, which captures an estimated 60% of all lending (Foote et al. 2009).
3.3.4 Preliminary Analysis of Raw Data

During our study period (January 2007 through February 2009), a total of 52,241 housing units in Maricopa County became REO, equating to 5.81% of housing stock. As the maps in Figure 3.2-3.4 show, LTV ratios, credit scores, and subprime mortgages (all presented at the zip code level) have very specific spatial characteristics. Zip codes in the north-east of the metropolitan area have a small percentage of loans with LTV ratios of 85% and above. In contrast, the dark shading in the south-western zip codes reveal that up to one-third of loans have LTV ratios of 85% and above. A similar pattern exists for Fair Isaac Corporation (FICO) scores, with areas in the north-eastern portion of the metropolitan region having few loans originated to people with FICO scores of less than 620 (considered poor credit), and over half of the loans going to lenders with FICO scores of 700 or more (considered excellent credit). Similarly, Figure 3.4(a) shows that zip codes in the north-east have very small percentages of loans that are subprime in contrast to areas in the south-west where up to 23% of loans are subprime. Finally, Figure 3.5, which shows the cumulative foreclosure rate for each quarter from January 2007 through February 2009, shows the contrast in foreclosure rates between the north-east and south-west areas. Some locations in the north-east have zero foreclosures, while other cells in the south-western part of the metropolitan area are as high as 62.5%. Also evident in panels (f)-(i) is the increase in foreclosure rates over time in the south-eastern part of the urban fringe.
Figure 3.2: Spatial Distribution of Loan-to-Value Ratios
Figure 3.3: Spatial Distribution of Credit Scores

Figure 3.4: Spatial Distribution of Loan Types
We had a total of 22 separate variables that described three aspects of neighborhoods: (1) loan characteristics and credit profile; (2) social, race/ethnicity, and economic characteristics; and (3) transportation costs. Preliminary analysis of the data showed high levels of correlation between some of the variables. Examples of highly-correlated data pairs include: the percentage of loans
that are subprime with the percentage of loans given to borrowers with low FICO scores (0.86), education and income (0.84), and percentage Hispanic with percentage of households in poverty. This high degree of correlation precludes using these data as separate independent variables because it would result in unstable and non-robust estimated coefficients (Adair, Berry, and McGreal 1996). A further issue is that the foreclosure rate had a highly-skewed right tail and was not normally distributed. As the data did not become normal under standard transformations, we ruled out using OLS regressions.

3.3.5 Inferential Statistics

We ran three separate sets of analysis on the data. First, we used Principal Components Analysis (PCA) to extract components from the highly correlated data and used these as the dependent variables in a Binary Logistic Model to predict what characteristics of a neighborhood affected whether or not a neighborhood experienced foreclosure over the entire time period of the analysis. Second, we used our 22 independent variables to create a set of neighborhood types, and ran an ANOVA test to determine whether there were statistically significant differences in the average foreclosure rates between neighborhood types for the entire period as well as nine consecutive quarters. Third, we used the neighborhood types in another Binary Logistic Model to predict whether the different types of neighborhood had a higher or lower chance of experiencing foreclosure, again with the data divided into nine separate quarters to understand the temporal aspects. We extended this third set of analysis to include a spatial contagion variable defined as the number of foreclosures that had taken place within a one-mile radius of the centroid of the raster cell within the prior quarter. This was designed to test for spatial contagion—that is, whether foreclosure in the prior time period created additional foreclosures.

3.3.6 Predicting foreclosure

Prior research has shown that the level of foreclosures in a neighborhood is affected by the loan profile, borrower attributes, and amount of fiscal stress. One of the strongest predictors has been whether or not a loan is subprime. The high degree of correlation in some of our independent variables motivated us to carry out PCA to extract distinct components from the data that were then used as independent variables in our model (Dunteman 1989; Joliffe 1986). Our dependent variable is defined as a binary variable indicating if the neighborhood experiences any foreclosures. The model can be represented by:

\[
\text{Foreclosure} = f(\text{Components } 1, 2, \ldots, n) \tag{3-1}
\]

The model equation can be represented by

\[
\ln \left( \frac{p_{FC}}{1-p_{FC}} \right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n \tag{3-2}
\]

Where \( p_{FC} \) is the probability that a neighborhood experiences any foreclosures, \( X_i \) is the \( i \)th component.
ln \( \left( \frac{p_{FC}}{1-p_{FC}} \right) \) is the logarithm of the odds ratio, hence in the estimated model, \( e^{\beta(aX)} \) will give the increase in odds ratio or likelihood for an ‘a’ increment in X.

### 3.3.7 Cluster analysis of neighborhood types.

We used k-means clustering to classify the 1,560 grid cells into types of neighborhood along three dimensions: (1) neighborhood credit profile and loan characteristics; (2) social/demographic/race/ethnicity; (3) transportation.

### 3.3.8 ANOVA tests of foreclosure rates.

Having clustered the data into homogenous groups based on various dimensions, we carried out an ANOVA test to determine if the mean foreclosure rate was statistically different across the groups for the entire time period and for the nine individual quarters of the period of analysis.

### 3.3.9 Modeling foreclosure by neighborhood type with and without spatial contagion.

Modeling neighborhood foreclosure propensity through a Logistic Regression setup involves a dependent variable indicating whether or not a neighborhood had any foreclosures. We ran two separate sets of models: (1) where the independent variables consisted of binary variables describing the neighborhood types; and (2) where an additional variable was included—spatial contagion—that quantified the number of foreclosures that had taken place within a one mile radius of the centroid of the neighborhood raster cell in the prior quarter. The coefficients of the binary cluster variables (covariates) represent the comparative likelihood of foreclosure for each neighborhood type over time. The coefficient of the spatial contagion variable would show whether or not nearby foreclosures (per grid cell) had an effect on surrounding properties (adjoining grid cells) (Immergluck and Smith 2006; Harding, Rosenblatt, and Yao 2009).

The two models can be represented by:

\[
P(\text{Foreclosure in time quarter } t) = f(\text{Neighborhood types } 1, 2, \ldots, 5) \quad (3-3)
\]

\[
P(\text{Foreclosure in time quarter } t) = f(\text{Neighborhood types } 1, 2, \ldots, 5, [\text{Count}]_{(t-1, r)}) \quad (3-4)
\]

The lag variable reflects the number of foreclosures that occurred in the prior quarter within a specified radius ‘r’. Hence the probability distribution of number of foreclosures during time period ‘t’ is a function of the neighborhood types and the Count of foreclosures during time period ‘t-1’ and within a ‘r’ mile radius:

Modeling the odds ratio of \( P(\text{Foreclosure in time quarter } t) \) in a Binary logistic regression model translates to:

\[
\ln \left( \frac{p_{(F,t)}}{1-p_{(F,t)}} \right) = \beta_0 + \beta_{NT1} X_{NT1} + \beta_{NT2} X_{NT2} + \beta_{NT3} X_{NT3} + \beta_{NT4} X_{NT4} + \beta_{NT5} X_{NT5} \quad (3-5)
\]
\[ \ln \left( \frac{p_{(F,t)}}{1-p_{(F,t)}} \right) = \beta_0 + \beta_{NT1} X_{NT1} + \beta_{NT2} X_{NT2} + \beta_{NT3} X_{NT3} + \beta_{NT4} X_{NT4} + \beta_{NT5} X_{NT5} + \beta_{(C,t-1)} X_{(C,t-1)} \]  

(3-6)

Where \( p_{(F,t)} \) is the probability \( P(\text{Foreclosure in time quarter } t) \)

\( X_{NTi} \) is a binary variable indicating the different neighborhood types. Clustering the data based on several dimensions of the neighborhood, to obtain binary variables indicating unique neighborhood types enables us to comprehend the chances of foreclosure for each of them.

\( X_{(C,t-1)} \) : Variables representing the count of foreclosures during the prior time period, within a ‘r’ mile radius from the centroid of the spatial grid.

If the contagion theory holds up, the lag variable will be significant in the Binary Logistic Regression Model (BLRM) after controlling for neighborhood types. Significance of the lag variable would indicate that foreclosure is determined by the number of foreclosures that took place in the prior period.

### 3.4 RESULTS

#### 3.4.1 Predicting foreclosure.

PCA revealed five components which account for 84% of the data variation. The components describe (1) loan characteristics, credit score, and LTV ratios; (2) family structure of household; (3) transportation costs and race/ethnicity; (4) education level; and (5) loan profile/income and transportation costs. Table 3.3 shows the weights of the individual variables and how they loaded onto each component.
Table 3.2: Results of Principal Components Analysis: Weights of Individual Variables and how they load onto various components

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>COMPONENT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>PERCENT PRIME</td>
<td>0.90</td>
</tr>
<tr>
<td>PERCENT SUBPRIME</td>
<td>-0.80</td>
</tr>
<tr>
<td>PERCENT ALT A</td>
<td></td>
</tr>
<tr>
<td>PERCENT INTEREST: ARM</td>
<td></td>
</tr>
<tr>
<td>PERCENT FICO: &lt; 620</td>
<td>-0.70</td>
</tr>
<tr>
<td>PERCENT FICO: &gt; 700</td>
<td>0.99</td>
</tr>
<tr>
<td>PERCENT LTV: &lt; 40</td>
<td>0.68</td>
</tr>
<tr>
<td>PERCENT LTV: 40 TO 80</td>
<td>0.72</td>
</tr>
<tr>
<td>PERCENT LTV: 80 TO 89</td>
<td>-0.52</td>
</tr>
<tr>
<td>PERCENT LTV: &gt; 90</td>
<td>-0.63</td>
</tr>
<tr>
<td>PERCENT BACHELORS DEGREE</td>
<td></td>
</tr>
<tr>
<td>PERCENT GRADUATE DEGREE</td>
<td></td>
</tr>
<tr>
<td>PERCENT BLACK</td>
<td></td>
</tr>
<tr>
<td>PERCENT WHITE</td>
<td></td>
</tr>
<tr>
<td>PERCENT HISPANIC</td>
<td></td>
</tr>
<tr>
<td>MEDIAN HOUSEHOLD INCOME ($)</td>
<td></td>
</tr>
<tr>
<td>HOUSEHOLDS WITH CHILDREN UNDER 18</td>
<td>0.83</td>
</tr>
<tr>
<td>HOUSEHOLDS WITH CHILDREN UNDER 6</td>
<td>0.80</td>
</tr>
<tr>
<td>PERCENT POVERTY RATE</td>
<td></td>
</tr>
<tr>
<td>PERCENT OWNER OCCUPIED</td>
<td></td>
</tr>
<tr>
<td>PERCENT RETIRED</td>
<td></td>
</tr>
<tr>
<td>AVERAGE TRANSPORTATION COST ($)</td>
<td></td>
</tr>
</tbody>
</table>

The Binary logistic regression model predicting whether or not a neighborhood had any foreclosures had a prediction success rate of 92% (see Table 3.4 for the model results). Four of the five components were statistically significant, with the exception being Component 4, which reflected level of education. All of the signs were as expected. For Component 1, which comprised loan characteristics, credit score, and LTV ratios, the higher the percentage of prime mortgage, lower the percentage of subprime mortgages, the higher the FICO scores, and the lower the LTV ratios, the less likely a neighborhood is to experience any foreclosures. For
Component 2, a positive sign indicates that the higher the percentage of households with families and lower the percentage of households with retired people within a neighborhood, the higher the chance that it will experience foreclosure. For Component 3, a positive coefficient indicates that the higher the percentage of minority race/ethnicity, the lower the percentage of White, the higher the percentage of households in poverty and the lower the percentage of units that are owner occupied, the more likely a neighborhood is to experience foreclosure. For Component 5, the positive coefficient indicates that the higher the transportation cost, the higher the percentage of loans that are Alt-A, the higher the percentage of ARMs, the greater the chance that a neighborhood will experience foreclosure. All of these results are consistent with prior literature.

The downside of using components rather than individual variables to predict foreclosure is that it makes interpretation of the relative impact of individual factors difficult. This method provides insight into the effect that various characteristics of a neighborhood such as loan profile have on foreclosures but does not show how the various components actually combine to produce neighborhoods. We now turn to evaluating how those individual characteristics interact to produce various types of neighborhood in order to quantify how foreclosure rates vary across these groupings.

Table 3.3: Results of Binary Logistic Regression Model Predicting Foreclosure as a function of individual components

<table>
<thead>
<tr>
<th>VARIABLE (COMPONENT)</th>
<th>COEFFICIENT</th>
<th>STANDARD ERROR</th>
<th>P-VALUE</th>
<th>ODDS RATIO</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>-.033</td>
<td>.006</td>
<td>.000</td>
<td>0.967</td>
</tr>
<tr>
<td>C2</td>
<td>.047</td>
<td>.010</td>
<td>.000</td>
<td>1.048</td>
</tr>
<tr>
<td>C3</td>
<td>.004</td>
<td>.002</td>
<td>.017</td>
<td>1.004</td>
</tr>
<tr>
<td>C4</td>
<td>-.020</td>
<td>.019</td>
<td>.276</td>
<td>0.980</td>
</tr>
<tr>
<td>C5</td>
<td>.005</td>
<td>.002</td>
<td>.018</td>
<td>1.005</td>
</tr>
</tbody>
</table>

3.4.2 Cluster analysis of neighborhood types.

Clustering using K-means methods revealed 5 natural groupings of neighborhoods whose characteristics are shown in Table 3.5. Cluster 1 contains households that we have labeled as ‘middle-income White’. These households have a mean income of $72,410, comprise 90.3% White population, have a higher-than-average education level with 36.7%/12.1% of people with bachelor/graduate degrees respectively compared to 26.5%/8.9% for the dataset overall; LTV ratios are slightly better than the region as a whole with the percentage of loans with an LTV ratio of less than 40%, between 80 and 89%, and more than 90% at 12.0%, 10.4%, and 22.2% compared to overall rates of 11.1%, 13.0%, and 22.6% respectively.
Table 3.4: Mean Values for each cluster group and entire dataset

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>CLUSTER 1</th>
<th>CLUSTER 2</th>
<th>CLUSTER 3</th>
<th>CLUSTER 4</th>
<th>CLUSTER 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOAN CHARACTERISTICS, CREDIT PROFILE AND LOAN-TO-VALUE (LTV) RATIOS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PERCENT PRIME</td>
<td>49.8</td>
<td>49.3</td>
<td>50.5</td>
<td>47.9</td>
<td>50.3</td>
</tr>
<tr>
<td>PERCENT SUBPRIME</td>
<td>11.7</td>
<td>13.9</td>
<td>8.1</td>
<td>13.4</td>
<td>9.8</td>
</tr>
<tr>
<td>PERCENT ALT-A</td>
<td>16.4</td>
<td>13.0</td>
<td>20.5</td>
<td>14.9</td>
<td>19.6</td>
</tr>
<tr>
<td>PERCENT INTEREST: ARM</td>
<td>48.5</td>
<td>39.0</td>
<td>52.9</td>
<td>43.8</td>
<td>56.2</td>
</tr>
<tr>
<td>PERCENT FICO&lt; 620</td>
<td>7.7</td>
<td>10.4</td>
<td>5.0</td>
<td>9.7</td>
<td>4.9</td>
</tr>
<tr>
<td>PERCENT FICO&gt; 700</td>
<td>33.5</td>
<td>30.9</td>
<td>34.2</td>
<td>31.5</td>
<td>35.4</td>
</tr>
<tr>
<td>PERCENT LTV&lt; 40</td>
<td>12.0</td>
<td>10.2</td>
<td>20.3</td>
<td>10.3</td>
<td>15.4</td>
</tr>
<tr>
<td>PERCENT LTV40 TO 80</td>
<td>50.8</td>
<td>44.3</td>
<td>58.6</td>
<td>46.2</td>
<td>59.4</td>
</tr>
<tr>
<td>PERCENT LTV80 TO 89</td>
<td>22.2</td>
<td>23.0</td>
<td>13.0</td>
<td>23.7</td>
<td>17.2</td>
</tr>
<tr>
<td>PERCENT LTV&gt; 90</td>
<td>10.4</td>
<td>16.1</td>
<td>4.7</td>
<td>14.1</td>
<td>4.9</td>
</tr>
<tr>
<td>SOCIAL, RACE/ETHNICITY AND ECONOMIC CHARACTERISTICS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PERCENT BACHELORS DEGREE</td>
<td>36.7</td>
<td>15.9</td>
<td>67.6</td>
<td>21.4</td>
<td>54.6</td>
</tr>
<tr>
<td>PERCENT GRADUATE DEGREE</td>
<td>12.1</td>
<td>5.3</td>
<td>31.8</td>
<td>6.9</td>
<td>19.2</td>
</tr>
<tr>
<td>PERCENT BLACK</td>
<td>1.3</td>
<td>3.8</td>
<td>0.7</td>
<td>2.9</td>
<td>0.9</td>
</tr>
<tr>
<td>PERCENT WHITE</td>
<td>90.3</td>
<td>72.7</td>
<td>95.2</td>
<td>82.0</td>
<td>95.0</td>
</tr>
<tr>
<td>PERCENT HISPANIC (OF ANY RACE)</td>
<td>9.0</td>
<td>31.9</td>
<td>2.9</td>
<td>18.5</td>
<td>3.2</td>
</tr>
<tr>
<td>MEDIAN HOUSEHOLD INCOME ($)</td>
<td>72,410</td>
<td>35,751</td>
<td>139,326</td>
<td>51,870</td>
<td>104,245</td>
</tr>
<tr>
<td>PERCENT HOUSEHOLDS WITH CHILDREN UNDER 18</td>
<td>38.3</td>
<td>30.2</td>
<td>30.8</td>
<td>37.2</td>
<td>26.3</td>
</tr>
<tr>
<td>PERCENT HOUSEHOLDS WITH CHILDREN UNDER 6</td>
<td>11.4</td>
<td>11.2</td>
<td>7.4</td>
<td>11.7</td>
<td>8.4</td>
</tr>
<tr>
<td>PERCENT POVERTY RATE</td>
<td>4.7</td>
<td>16.6</td>
<td>2.6</td>
<td>8.1</td>
<td>2.5</td>
</tr>
<tr>
<td>PERCENT OWNER OCCUPIED</td>
<td>88.1</td>
<td>66.4</td>
<td>93.9</td>
<td>82.3</td>
<td>91.0</td>
</tr>
<tr>
<td>PERCENT RETIRED</td>
<td>16.1</td>
<td>21.3</td>
<td>18.5</td>
<td>17.2</td>
<td>20.1</td>
</tr>
<tr>
<td>TRANSPORTATION</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AVERAGE TRANSPORT COST ($)</td>
<td>13,637</td>
<td>10,434</td>
<td>16,327</td>
<td>12,472</td>
<td>13,916</td>
</tr>
</tbody>
</table>

41
Cluster 2 comprises of neighborhoods that we describe as ‘very low income minority’. The mean household income is $35,751 compared to a mean regional rate of $56,686; and the percentages of Hispanic and Black populations are 31.9% and 3.8% are the highest for all the groups at 1.7 and 1.5 times the rate for the study area as a whole. LTV ratios are the weakest for the entire group with 16.1% of all loans with LTV > 90%, and the category has the highest percentage of subprime loans at 13.9%.

Cluster 3 is classified as ‘very high income White’. The mean household income is the highest of all the groups at $139,326, which equates to 2.5 times the average income for the study area, and 3.9 times that in the lowest income neighborhoods that belong to Cluster 2. The neighborhoods are 95.2% White compared to 82.4% for the region as a whole, which is the highest concentration for all five groups. This group of neighborhoods has the largest amount of home equity, strongest LTV ratios, lowest percentage of loans that are subprime, and the highest percentage of loans that are prime.

Cluster 4 is labeled ‘low- to middle-income minority’. This group of neighborhoods has income of $51,870, which is approximately 91% of the study area average, and has the second highest percentages of both Hispanic and Black populations. This group has the lowest/second highest percentage of loans that are prime/subprime. It is the second most leveraged neighborhood grouping after Cluster 2, with 14.1% of loans with LTV ratios > 90% and 23.7% of loans with LTV ratios of 80-89%.

Cluster 5 is labeled as ‘high income White’. The mean income for this group of neighborhoods is $104,245, which is the second highest income group after Cluster 3 (very high income White), and 1.8 times the average for the study area as a whole. This group of neighborhoods has the second lowest percentage of subprime loans at 4.9% compared to a regional average of 13.0%, and the second lowest percentage of loans with LTV ratios of > 90%.

3.4.3 ANOVA tests of foreclosure rates.

One-way ANOVA tests for the mean foreclosure rate overall and for the individual nine quarters of the study period were statistically significant at the 1% level. The average foreclosure rates for each quarter are shown in Figure 3.6. ‘Very low-income minority’ and ‘low- to middle-income minority’ neighborhoods had the highest rates of foreclosure over study period. These rates reached a peak in Q8 and Q9 respectively, and then dropped off in Q9 when a foreclosure moratorium was implemented. Of note, though, is that from Q1 through Q8 foreclosure rates steadily increased in ‘middle-income White’ neighborhoods and also ‘high-income White’ neighborhoods. The lowest rates of foreclosure were consistently experienced by ‘very high income White’ neighborhoods.
Neighborhood Types Based on Socio-Economic and Racial/Ethnic Variables and Mortgage Loan Profile

Cluster Category
- Red: Middle Income Whites/Highly Leveraged/High Transport Costs
- Light Green: Very Low Income Minority/Highly Leveraged/Poor Credit/Subprime
- Dark Brown: Very High Income Whites with High Level of Home Equity
- Blue: Low Income, Racially/Ethnically Diverse, Highly Leveraged/Subprime
- Yellow: High Income Whites with High Levels of Home Equity

Figure 3.6: Neighborhood types based on social, economic, credit, and loan variables
3.4.4 Modeling foreclosure by neighborhood type without/with spatial contagion.

The estimated odds ratios (set out in Table 3.6) reveal the propensity of the different types of neighborhood to experience foreclosure over the time period of the analysis. Two sets of results are presented without/with the spatial contagion variable.
Table 3.5: Model Estimates and odds ratios for binary logistic regression of foreclosure rate (by quarter) as a function of neighborhood type without and with spatial contagion variable

BINARY LOGISTIC REGRESSIONS OF FORECLOSURE RATE VERSUS NEIGHBORHOOD TYPE

<table>
<thead>
<tr>
<th>Q</th>
<th>CLUSTER</th>
<th>CLUSTER</th>
<th>LAG</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>(A) MODEL COEFFICIENTS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>3</td>
<td>-0.40</td>
<td>.</td>
<td>-2.97</td>
</tr>
<tr>
<td>4</td>
<td>.</td>
<td>.</td>
<td>-1.70</td>
</tr>
<tr>
<td>5</td>
<td>-0.50</td>
<td>.</td>
<td>-3.55</td>
</tr>
<tr>
<td>6</td>
<td>-0.52</td>
<td>.</td>
<td>-2.48</td>
</tr>
<tr>
<td>7</td>
<td>-0.30</td>
<td>.</td>
<td>-2.44</td>
</tr>
<tr>
<td>8</td>
<td>-0.33</td>
<td>.</td>
<td>-3.09</td>
</tr>
<tr>
<td>9</td>
<td>-0.48</td>
<td>.</td>
<td>-3.28</td>
</tr>
<tr>
<td>(B) ODDS RATIOS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>3</td>
<td>0.67</td>
<td>.</td>
<td>0.05</td>
</tr>
<tr>
<td>4</td>
<td>.</td>
<td>.</td>
<td>0.18</td>
</tr>
<tr>
<td>5</td>
<td>0.60</td>
<td>.</td>
<td>0.03</td>
</tr>
<tr>
<td>6</td>
<td>0.59</td>
<td>.</td>
<td>0.08</td>
</tr>
<tr>
<td>7</td>
<td>0.74</td>
<td>.</td>
<td>0.09</td>
</tr>
<tr>
<td>8</td>
<td>0.72</td>
<td>.</td>
<td>0.05</td>
</tr>
<tr>
<td>9</td>
<td>0.62</td>
<td>.</td>
<td>0.04</td>
</tr>
</tbody>
</table>

“*” indicates that the coefficient was not significant at the 5% level.

Looking first at the set of models without the spatial contagion variable, the model fits are fairly strong ranging from 52.8% to 76.9% (see Table 3.7). Negative coefficients for a particular cluster group indicate that that group is less likely other cluster groups to experience any foreclosures. For most of the time periods, Clusters 1, 3, and 5 (‘middle income White’, ‘very high income White’, and ‘high income White’) are consistently less likely to experience foreclosure relative to the other two types of neighborhood. A comparison of the magnitude of these coefficients reveals the extent of the differences in the probability of foreclosure between neighborhoods. ‘Very high income White’ neighborhoods are 13 times less likely to experience foreclosures than any other neighborhood during the study period. The respective figures for ‘high income White’ neighborhoods and ‘middle income White’ neighborhoods are 3.3 times and 1.5 times less likely respectively. In contrast, in quarters 4 and 7, ‘low- to middle-income minority’ neighborhoods were 1.45 and 1.46 times more likely to experience foreclosure than any other neighborhood type.
Table 3.6: Model fits for binary logistic regression model without/with spatial contagion variable

<table>
<thead>
<tr>
<th>QUARTER</th>
<th>WITHOUT SPATIAL CONTAGION VARIABLE</th>
<th>WITH SPATIAL CONTAGION VARIABLE</th>
<th>% INCREASE IN MODEL FIT</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>52.8</td>
<td>72.1</td>
<td>19.3</td>
</tr>
<tr>
<td>3</td>
<td>59.8</td>
<td>75.3</td>
<td>15.5</td>
</tr>
<tr>
<td>4</td>
<td>65.0</td>
<td>78.7</td>
<td>13.7</td>
</tr>
<tr>
<td>5</td>
<td>70.5</td>
<td>79.0</td>
<td>8.5</td>
</tr>
<tr>
<td>6</td>
<td>73.7</td>
<td>80.7</td>
<td>7.0</td>
</tr>
<tr>
<td>7</td>
<td>76.4</td>
<td>84.0</td>
<td>7.6</td>
</tr>
<tr>
<td>8</td>
<td>76.9</td>
<td>83.3</td>
<td>6.4</td>
</tr>
<tr>
<td>9</td>
<td>65.3</td>
<td>80.3</td>
<td>15.0</td>
</tr>
</tbody>
</table>
The introduction of the spatial contagion variable had a strongly positive impact on the model fits, increasing them to between 72.1% and 84% respectively (an average improvement of 11.6%). In all of the time periods, the spatial contagion variable had a positive coefficient, meaning that foreclosures within and close to the neighborhood that had occurred during the prior period had a statistically significant effect on predicting whether or not a neighborhood experienced foreclosures in the current period. In these models, the neighborhood types were not statistically significant except on three occasions where ‘very high income White’ and ‘high income White’ neighborhoods still had a lower propensity to experience foreclosure. In time period 2, ‘high income White’ neighborhoods were 3.4 times less likely to experience foreclosure, even after controlling for spatial contagion. In time periods 5 and 8, ‘very high income White’ neighborhoods were 7.7 and 4.8 times less likely to experience foreclosure than other neighborhoods.

3.5 EVALUATION AND DISCUSSION

Three main points emerge from our analysis. The first is that foreclosures have a distinct racial and ethnic dimension that layers on top of the loan/credit dimension. This is evident from the fact that loan and credit variables load onto a separate component to race/ethnicity and that both components are statistically significant. Our second point is that grouping neighborhoods by a variety of characteristics that relate to foreclosure reveals that although foreclosures did affect all types of neighborhoods, by far the highest rates were experienced in very low income minority neighborhoods and low-to middle income minority neighborhoods. This confirms the idea put forward by (Crump et al. 2008) that although foreclosures ‘democratized’ to neighborhoods across the board, their effects were still the strongest in minority neighborhoods, even in these formerly ‘hot markets’ that were the last to be affected. Our third major finding is that minority neighborhoods are orders of magnitude more likely to experience foreclosures than White neighborhoods both without and with controls for spatial contagion. This suggests the possibility that contagion may differ across different neighborhood types.

With respect to the role of transportation costs, the way that we have operationalized the variable (in a static rather than dynamic way that would capture increases in gasoline prices that occurred in mid-2008), does not allow us to make any definitive statements. The chart of increased foreclosures by neighborhood does show all foreclosure rates for all neighborhood types trending upwards in the first half of 2008, but it is difficult with our models to separate out what portion of this may be due to gasoline prices. Investigating the role that rising energy prices played in the foreclosure crisis is investigated in-depth in a separate analysis.

To summarize, the analyses reinforce that within this metropolitan area, the places and people who have been the hardest hit are precisely those who have the least capacity to cope (Rugh and Massey; Crump et al. 2008). While foreclosures in our study area did spread to non-minority and higher income neighborhoods, foreclosure rates were nowhere near the level experienced in minority neighborhoods. One important finding for policy is that in every type of neighborhood, foreclosures the prior quarter are a strong indicator of future foreclosures, which suggests that at the very least, some element of spatial dependence needs to be incorporated into risk models. At first glance, the significance of the spatial lag variable that provides evidence of spatial contagion could reinforce the drive for some people to want to segregate themselves from
minority neighborhoods. However, a geographic approach allows us to consider the foreclosure crisis in its broader context, and realize that no matter how much households may wish to isolate themselves from others, society is intricately connected through multiple mechanisms, not least of which is the global financial markets, and that other forms of contagion (such as financial contagion) exist. As scholars such as *Crump et al. (2008)* have pointed out, the issue reached a crisis point in part because people chose to ignore the increase in foreclosures that occurred in low-income, minority neighborhoods. One of the major lessons to be taken away from this crisis by society at large is that households, neighborhoods, metropolitan areas, and nations are interconnected and that public policy needs to reflect the risks associated with that interconnection.

We envisage extending this preliminary study in four ways. First, the neighborhood typology could be used in an econometric analysis to provide a case study of spatial contagion in a strongly affected metropolitan area that allows for an investigation of whether the impact of foreclosures on house prices differs across neighborhood types. A second study is to compare the incidence of foreclosure with data that show households who received assistance through the MHA program initiated in 2009 to see if the program has targeted a particular segment of homeowners. Third, as mentioned earlier, this study could provide a starting point for an indepth analysis of how rising energy prices may have contributed to foreclosures. Finally, this work could inform ground some qualitative research about what happens to people and neighborhoods after foreclosures increase and tax revenues decline, setting the stage for a reevaluation of US housing policy. As *Shlay (2006)* suggested, perhaps surrendering national housing policy to a volatile marketplace may not be the best solution.
4.0 EMPIRICAL ANALYSIS OF ADDITIONAL IMPACT ON FORECLOSURES OF HIGHER GASOLINE PRICES

4.1 INTRODUCTION

Below is the conceptual model that we created based upon the above-mentioned literature to inform our empirical study (Figure 4.1). First is that the possible role of speculation, the role of OPEC and dynamics associated with resource depletion in the oil market suggest that this variable needs to be treated as being exogenous to the US financial market. Oil affects foreclosure rates in two ways. First is in a direct manner by further pressing household budgets. And second, by dampening consumer sentiment, US stock markets, and the general economic outlook for the US because expenditure on oil represents a net capital export. The poorer economic outlook creates unemployment which in turn affects foreclosures as households’ ability to pay deteriorates. There are effectively two loops of negative feedback. One transmitted around the economic system because of banking sector failures, downward revision of growth, and a surge in unemployment associated with the subprime loan crisis. The second feedback loop has all the same transmission points as the previous one, except that rather than the subprime loan crisis, this was caused by oil price increases. The dynamic created in the general financial sector can be considered a separate effect from the vicious cycle that took hold in the housing markets as a result of foreclosure which is shown by the ellipse on the left hand side of the figure.

Figure 4.1: Conceptual Model to Understand the Impact of Oil Prices on Foreclosures

49
The research presented in the previous Section suggests two things: (1) that our analysis of gasoline price effects on foreclosure should use a sub-set of neighborhoods that are somewhat homogeneous; (2) it suggests that any model of foreclosure over time include a spatial contagion variable to account for nearby foreclosures that could affect some neighborhoods and not others. Accordingly, we focused our analysis of the possible additional impact on foreclosures of higher gasoline prices on middle-income white neighborhoods, the locations of which are shown in Figure 4.2. These neighborhoods are distributed across the metropolitan area. Descriptive statistics for these neighborhoods are shown in Table 4.1.

![Map of metropolitan Phoenix. Each grey square represents a 1 mile x 1 mile grid that contains 10 or more housing units. Areas in black represent the neighborhoods included in the study.](image)

**Table 4.1: Descriptive statistics of neighborhood variables**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Percent Prime</td>
<td>33.3</td>
<td>77.7</td>
<td>49.8</td>
<td>5.2</td>
</tr>
<tr>
<td>Percent Subprime</td>
<td>0.0</td>
<td>16.5</td>
<td>11.7</td>
<td>2.8</td>
</tr>
<tr>
<td>Percent Alt A</td>
<td>9.8</td>
<td>33.3</td>
<td>16.4</td>
<td>2.5</td>
</tr>
<tr>
<td>Percent Interest ARM</td>
<td>0.0</td>
<td>58.7</td>
<td>48.5</td>
<td>6.1</td>
</tr>
<tr>
<td>Percent FICO: &lt; 620</td>
<td>0.0</td>
<td>12.1</td>
<td>7.7</td>
<td>2.4</td>
</tr>
</tbody>
</table>
Percent FICO: > 700  |  27.0 |  59.7 |  33.5 |  5.1
---|---|---|---|---
Average bachelors degree  |  5.1 |  60.6 |  36.7 | 11.5
Average graduate degree  |  0.0 |  32.0 |  12.1 |  4.9
Average black  |  0.0 |  5.5  |  1.3  |  1.1
Average white  |  44.3 |  97.9 |  90.3 |  6.9
Average Hispanic  |  1.8  |  29.1 |  9.0  |  6.3
Median household income  |  62,094 |  88,209 |  72,410 |  7,374
HH with children under 18  |  10.5 |  59.3 |  38.3 | 12.8
HH with children under 6  |  1.5  |  27.4 |  11.4 |  5.7
Average poverty  |  1.2  |  34.8 |  4.7  |  3.1
Average owner occupied  |  58.2 |  100.0 |  88.1 |  7.9
Average percent retired  |  0.0  |  0.4  |  0.2  |  0.1
Percent LTV: <40  |  0.0  |  25.2 |  12.0 |  2.7
Percent LTV: 40 to 80  |  33.3 |  63.8 |  50.8 |  6.1
Percent LTV: 80 to 89  |  6.7  |  66.7 |  22.2 |  4.4
Percent LTV: >90  |  0.0  |  19.7 |  10.4 |  4.3

Average household income for our subset of neighborhoods ranges from $62,094 to $88,209. Fair-Isaac Corporation (FICO) scores indicate that these neighborhoods have fairly low percentages of household with poor credit (FICO<620) at a mean of 7.7%, a much higher percentage of households with good credit (FICO>700), with a mean of 33.5%. On average, the
mortality types within our neighborhoods are 50% prime, 12% subprime, and almost half (49%) have adjustable rates, and 16% Alt-A. Almost a third of mortgages in the average neighborhood are highly-leveraged, with LTV rates of 80-89% or > 90%, but there is a wide range of variation in these levels, some neighborhoods having around 85% of all loans in these two highly-leveraged categories.

4.2 LOGISTIC REGRESSION ANALYSIS

We used Binary Logistic Regression Models (BLRMs) to quantify what factors predicted whether or not a neighborhood would have a higher than median foreclosure rate for each of the 8 quarters from Q2 2008 through Q1 2009. Since a lagged foreclosure rate variable is used, a model for time period 2008-Q1 is not fit as the study period does not include data from the previous quarter. The explanatory variables that we included as controls described the percentage of mortgages that are prime, and Alt-A, percentage of mortgages with an adjustable rate (ARM), FICO scores, LTV ratios, income, a spatial contagion variable that quantifies the amount of foreclosures that took place within a 1-mile radius in the prior quarter. To this we added our variable of interest, percentage of income spent on transportation fuel, which was derived using VMT (which varied at the census tract scale), a constant measure for vehicle fuel efficiency, and the average quarterly price of US Gulf Coast Conventional Gasoline, and household income, which varied across census tracts.

\[
\text{Transportation Fuel Cost} = \frac{\text{Vehicles Miles travelled x Gasoline Price}}{\text{Fuel Efficiency}}
\]  

\[
\text{Percent of income spent on transportation fuel } X_{TC} = \frac{\text{Transportation Fuel Cost}}{\text{Household Income}} \times 100
\]  

The BLRM equation can be represented by:

\[
\ln \left( \frac{\pi}{1 - \pi} \right) = \beta_0 + \beta_{\text{Alt-A}}X_{\text{Alt-A}} + \beta_{\text{ARM}}X_{\text{ARM}} + \beta_CX_C + \beta_{\text{LTV}}X_{\text{LTV}} + \beta_{\text{HH Income}}X_{\text{HH Income}} \\
+ \beta_{\text{FC,t-1}}X_{\text{FC,t-1}} + \beta_{\text{TC}}X_{\text{TC}}
\]  

Where:
- \( X_{\text{Alt-A}} \): Percent of loans that are Alt-A
- \( X_{\text{ARM}} \): Percent loans that have adjustable rate mortgages
- \( X_C \): Credit profile
- \( X_{\text{LTV}} \): Loan Profile
- \( X_{\text{HH Income}} \): Median Household Income
- \( X_{\text{FC,t-1}} \): Foreclosure rate in the previous quarter
- \( X_{\text{TC}} \): Percentage of income spent on transportation fuel
- \( \pi \): Probability that a neighborhood will have a higher-than-median foreclosure rate.
4.3 HYPOTHESIS

We hypothesize that as gasoline prices increase through 2008 (see Figure 4.3), the additional stress on household budgets will result in more foreclosures. That is, the coefficient of the variable of interest—percentage of income spent on transportation fuel—will be positive and significant and increase over the eight quarters of our study.

![](image)

Figure 4.3: Scaled trend of prices of Crude oil, Gasoline, foreclosure rate, and Dow Jones Industrial Average during the study period 2007-Q1 – 2009-Q1

4.4 RESULTS

The fitted models for each quarter successfully predicted the outcomes from 72% to 78% indicating very good model prediction.

The estimated coefficients of all the explanatory variables are as expected: the higher the percentage of Alt-A loans, the lower/higher the percentage of FICO scores >700/<620, the higher/lower the percentage of LTV ratios 80-90 and > 90/<40 the higher the likelihood that a neighborhood would have a higher-than-median foreclosure rate. The variable representing the lagged effect of foreclosures in the previous quarter is positive implying a general trend of increasing foreclosure rates over time. The variable representing the percentage of household income spent on transportation fuel is significant in almost all the models with the exception of the second quarter. In quarter 3, for every tenth of a percent increase in transportation costs, the likelihood of the neighborhood having a higher-than-median foreclosure rate increases by around 16%. In quarter 5, for every tenth of a percent increase in transportation costs, the likelihood of the neighborhood having a higher-than-median foreclosure rate increases up to around 25%. In quarters six and seven, for every tenth of a percent increase in transportation costs the likelihood of the neighborhood having a higher-than-median foreclosure rate increases by around 8%. Equivalently, for every 1% increase in transportation costs, the neighborhood is around 2.3
times more likely to have a higher-than-median foreclosure rate. In quarters 8 and 9, for every 0.1 percent increase in transportation costs the odds ratio of a neighborhood having a higher-than-median foreclosure rate increases by 36% as compared to 8%, 16% and 25% in the previous quarters. Equivalently, a neighborhood is around 20 times more likely to have a higher-than-median foreclosure rate for every 1% increase in transportation costs. This increase in neighborhood foreclosure propensity from 2.3 to 20 times in quarters 8 and 9 which correspond to the last quarter of 2008 and first quarter of 2009, can probably be attributed to the rising prices of gas. The range of percent of income spent on transportation initially is up to 1.98%, and later in quarters 6 and 7, or 2008-Q2 and 2008-Q3, jumps up to 3.80% indicating the effect of rising fuel price on the monthly expenses of households. These results suggest that rising gasoline prices placed an additional burden on household budgets that exacerbated foreclosures.

4.5 DISCUSSION AND CONCLUSIONS

One limitation of our empirical study is that we have assumed that VMT remained constant throughout the study period. Even though the metropolitan area is almost exclusively auto-dependent, three important reasons suggest that travel is somewhat responsive to gasoline prices. First, the metropolitan transit authority reported that bus ridership (especially rapid bus transit going from the outer suburbs into downtown Phoenix) increased. Second, anecdotal evidence exists that carpooling increased, although there are no official statistics on this rather informal but effective way of increasing transportation efficiency. Third, around one-third of all VMT in the metropolitan area is for discretionary, non work-related trips (NHTS, 2002) that could be reduced in the face of higher gasoline prices. A lack of data on VMT at more discrete timescales precluded us from improving this limitation to our analysis.

Despite these limitations, higher oil and therefore gasoline prices do appear to have exacerbated residential foreclosures. However, a much more important cause of foreclosure may have been households’ high degree of leverage and the fact that the housing market bubble burst. Taken together, these two factors underscore the vulnerability of the American Dream to market forces. Just as some people have questioned the risks of leaving US housing policy in the hands of the ‘free-market’, a just as prominent risk may relate to structural exposure to and overreliance on a diminishing commodity the price of which is increasingly decoupled from US economic prospects.
5.0 LIGHT RAIL TRANSIT’S AFFECT ON LAND VALUE—HOW MUCH COMES FROM NEW CONSTRUCTION?

5.1 INTRODUCTION

This study has been completed and will be submitted to Environment & Planning A. The results were presented at a regional conference, NESTVAL, in Fall 2010.

5.2 ABSTRACT

Urban revitalization strategies usually involve multiple complementary initiatives that are either sequenced or implemented simultaneously. In these dynamic settings, it is often difficult to disentangle the discrete impact of individual projects. This is especially true for Light Rail Transit (LRT) which is increasingly being adopted alongside initiatives such as overlay zoning—an ordinance that prohibits certain future land uses and encourages others, and sets minimum density requirements—to encourage Transit-Oriented Development (TOD). Studies that quantify the impact of LRT on land values may also be capturing the effect of new construction which existing literature has shown to have a positive impact on land values. We create a variable based on the assessed value of new construction for various space/time combinations to represent neighborhood change (Nungal). We use data on condo transactions in the LRT corridor in metropolitan Phoenix, Arizona, to create hedonic models to estimate the effect of LRT and overlay zoning on the price of pre-existing condos excluding and including Nungal. The Nungal variable is statistically significant and improves the model fit. New housing construction increases the price of properties within a ¼-mile whereas new non-residential construction has a positive spatial spillover that can be detected up to two miles. The effect is pulse-like, peaking during the year in which construction is completed, before tapering off. The methodology that separates the effects of multiple initiatives is generalizable to a wide range of policies.
6.0 ADDITIONAL WORK IN PROGRESS

6.1 METHODOLOGY AND GIS

George Bently, a PhD student, is undertaking analysis on methodology relating to aggregating data across various spatial scales in GIS. The following is an abstract of a paper that he will present at the AAG meetings in April 2011.

6.1.1 Abstract: Network Interpolation of Population Values

In spatial analysis, demand is frequently aggregated into source units having an areal extent. When using such data in a flow model, distances are calculated as an average between each source unit and the set of destinations. In a network, this average distance might be the shortest path between a destination and the centroid of the source unit. However demand is never uniformly distributed within each spatial unit. In urban areas it is most likely located near the road system that bounds most areal units. It is proposed that areally aggregated data be interpolated to the adjacent segments of the bounding road network using a dasymetric approach. Analysis compares using network interpolated population versus the areal approach to ascertain the number of people within a time threshold of destinations as well as total weighted travel time to those destinations. A case study from Phoenix, Arizona illustrates the approach.

6.2 IMPACT OF FORECLOSURE ON SURROUNDING HOUSE PRICES. DOES THE EFFECT DIFFER IN DIFFERENT TYPES OF NEIGHBORHOODS?

The analysis presented in Chapter 3 shows that foreclosure rates have differed across neighborhood types. The next step to extend this research is to determine whether different types of neighborhood experience spatial contagion from foreclosures differently—that is, whether the price effect is stronger in some neighborhoods than in others. All data relating to this study have been collected and are awaiting analysis in Summer 2011.

6.3 COMPARISON OF THE PERFORMANCE OF HOUSE VALUES IN PROPERTIES IN LRT STATION AREAS IN AN UP-MARKET AND IN A DOWN-MARKET AND ITS IMPLICATIONS FOR THE CONSIDERATION OF RESILIENT CITIES

Once the analysis in Section 6.2 is completed, the question about the price performance of residential units in the LRT corridor in a down-market may be addressed. This analysis will add
to the ongoing debate about the resilience of the traditional suburbs and housing in transit-accessible neighborhoods.

6.4 EXTENSION OF ANALYSIS TO CALIFORNIA

Field research undertaken in California in December 2009 that involved interviews with real estate professionals and city planners revealed that California also had a highly varied geography of foreclosures. According to these professionals, one of the areas that had suffered the most was the Inland Empire, located around 75 miles from downtown Los Angeles. This area contained ‘cookie-cutter’ houses in subdivisions, and it was thought that residents who lived here had long commutes to work. Conducting a similar analysis to the one undertaken for Arizona would require extensive data on credit profiles and loan profiles. A proposal for a joint study with the San Francisco Federal Reserve Board (Dr. Carolina Reid) is currently pending.
7.0 EDUCATION-RELATED INITIATIVES

7.1 TEACHING

The material generated by this study has been used in a number of undergraduate and graduate classes at the University of Connecticut: GEOG2100, Economic Geography; URBN2000W, Introduction to Urban Studies; GEOG4200W, Geographic Analysis of Urban Social Issues; and GEOG5290, Advanced Urban Geography. Students have come from a wide variety of disciplines: geography, sociology, civil and environmental engineering, economics, resource economics, mathematics, public policy, and political science. The analysis has been able to drive home to students the differences.

7.2 COLLOQUIUM

The material was also shared with students at the University of Miami, Department of Geography, at a colloquium.
8.0 FIELD STUDY, DECEMBER 2009

A trip was taken to Phoenix, Arizona, in December 2009 to observe the LRT system in operation, to meet with real estate professionals working in LRT station areas, and to survey station areas for developments that may not be evident from secondary data. The main points were as follows.

8.1 REAL ESTATE MARKET DYNAMICS IN LRT CORRIDOR

Most of the mixed-use buildings that had been constructed along the LRT system—primarily those in and around the two downtowns of Tempe and Phoenix—were vacant. Some high-profile projects had been halted. Others, notably Centerpoint in downtown Tempe, had been abandoned after the developer declared bankruptcy. Zaremba Group took over the development in February 2011 and plan to open the complex in Summer 2011. Many of the apartment buildings that had been constructed near LRT station areas had managed to attract occupants by significantly reducing rents and offering incentives for lease signing such as one or more months free rent. The new, up-market rental apartment buildings retained their standards about credit scores. However, older apartment complexes with fewer amenities than the newer places, had significantly relaxed their credit requirements in order to adjust to market realities.

8.2 LAND USE DYNAMICS

Several stations, most notably along Apache Boulevard in Tempe and near Bethany Home Road in Phoenix, had experienced a considerable amount of demolition but the real estate market crash appeared to have halted reconstruction. The transit corridor therefore contained a considerable amount of not only vacant properties but also vacant land.

8.3 MULTI-MODAL SYSTEM DEVELOPMENT

The transit center at A-mountain in Tempe was busy, and the introduction of minibuses in and around Tempe could go some way to easing some of the challenges associated with using the LRT system in Phoenix’s harsh climate. A cycle shop that offered showers and bike storage in downtown Tempe was struggling for customers. The transportation center on Apache Boulevard had a lot of connecting buses, but at the time of the visit, little activity was seen.

8.4 SOCIAL AND DEMOGRAPHIC CHANGE

The field visit suggested that the downturn in the real estate market has interrupted the momentum of downtown revitalization. Newly-built properties are not attracting the market that they had anticipated, and had been forced to slash rents to keep properties occupied. Analysis of 2010 Census data would verify the extent to which demographics have changed since the system was proposed.
REFERENCES


M. Aalbers, 'Geographies of the financial crisis', *Area*, 41 (2009), 34-42.


B. Heath and D. Merrill, Confined Crisis *USA Today* (2009).


P. A. McCoy and E. Renuart, The legal infrastructure of subprime and nontraditional home mortgages In E. S. Belsky (ed.), *Borrowing to Live: Consumer and Mortgage Credit Revisited* (Washington, DC, 2008).


APPENDIX A

SELECTED PHOTOGRAPHS OF FIELD STUDY, LRT SYSTEM, ARIZONA, DECEMBER 2009
On-board rules and regulations

On board directions for bike users

Apartment housing along transit corridor

New apartment building marketed to students

Street-level retail for rent

Lobby of Grigio Metro apartment building
Lobby of Grigio Metro apartment building (2)

Marketing at Pinot Grigio

Commercial Vacancies, Tempe

Commercial Vacancies, Tempe (2)

Existing apartments in Tempe slashing prices

Tempe apartments offering ‘blow-out specials’
Vacant land alongside rail line, Tempe

More vacant land alongside LRT line, Tempe

Front view of train

Mixed-use vacancy, Tempe

Bike storage facility, Tempe

Mixed use building for lease, Tempe
Vacant land, Central Ave, Phoenix

Vacant mixed-use building, Tempe

Abandoned Centerpoint, Tempe

Billboards/gang graffiti, Centerpoint, Tempe

Centerpoint Towers, Tempe

Commercial land for lease, Tempe Town Lake
Celebrate Light Rail banners, downtown Phx

Residential lettings, Central Ave, Phoenix

Vacancies, Le Chateau on Central, Phoenix

Vacant land next to LRT track, Central Ave

Live ‘Lofty’, Pavilions on Central Ave, Phx

Vacant retail, Bethany Home Road, Phoenix