CORE
Metadata, citation and similar papers at core.ac.uk
Provided by ScholarWorks @ Georgia State University

Georgia State University
ScholarWorks @ Georgia State University

Economics Dissertations
Department of Economics

Spring 5-5-2012

Essays on Social Interaction and Urban Outcomes

Zackary B. Hawley
Georgia State University

Follow this and additional works at: https://scholarworks.gsu.edu/econ_diss

Recommended Citation
Hawley, Zackary B., "Essays on Social Interaction and Urban Outcomes." Dissertation, Georgia State University, 2012.
https://scholarworks.gsu.edu/econ_diss/81

This Dissertation is brought to you for free and open access by the Department of Economics at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Economics Dissertations by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.

PERMISSION TO BORROW

In presenting this dissertation as a partial fulfillment of the requirements for an advanced degree from Georgia State University, I agree that the Library of the University shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to quote from, to copy from, or to publish this dissertation may be granted by the author or, in his or her absence, the professor under whose direction it was written or, in his or her absence, by the Dean of the Andrew Young School of Policy Studies. Such quoting, copying, or publishing must be solely for scholarly purposes and must not involve potential financial gain. It is understood that any copying from or publication of this dissertation which involves potential gain will not be allowed without written permission of the author.

Signature of Author

NOTICE TO BORROWERS

All dissertations deposited in the Georgia State University Library must be used only in accordance with the stipulations prescribed by the author in the preceding statement.

The author of this dissertation is:
Zackary Bartholomew Hawley
400 17th Street NW Unit 1306
Atlanta, GA 30363

The director of the dissertation is:
Dr. Jonathan C. Rork
Department of Economics
Reed College
3203 SE Woodstock Boulevard
Portland, OR 97202-8199

Users of this dissertation not regularly enrolled as students at Georgia State University are required to attest acceptance of the preceding stipulations by signing below. Libraries borrowing this dissertation for the use of their patrons are required to see that each user records here the information requested.

Name of User	Address	Date	Type of use (Examination only or copying)

ESSAYS ON SOCIAL INTERACTION AND URBAN OUTCOMES
BY
ZACKARY BARTHOLOMEW HAWLEY

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree
of
Doctor of Philosophy
in the
Andrew Young School of Policy Studies
of
Georgia State University

GEORGIA STATE UNIVERSITY
2012

Copyright
Zackary Bartholomew Hawley
2012

ACCEPTANCE

This dissertation was prepared under the direction of the candidate's Dissertation Committee. It has been approved and accepted by all members of that committee, and it has been accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Economics in the Andrew Young School of Policy Studies of Georgia State University.

Dissertation Chair:	Dr. Jonathan C. Rork
Committee:	Dr. Geoffrey K. Turnbull
	Dr. Andrew R. Hanson
	Dr. Cathy Yang Liu

Electronic Version (Approved):
Mary Beth Walker, Dean
Andrew Young School of Policy Studies
Georgia State University
May 2012

ACKNOWLEDGMENTS

I am able to complete this dissertation only due to the guidance, help, care, and support of many teachers, friends, and family members. I wish I had the space to acknowledge everyone here.

I want to first thank Jon Rork. He was always there; be it professional development or just someone to talk with, Jon provided wonderful guidance. Jon continually found a way to realign my thoughts and energy into productivity. I still turn to him when I find myself without direction. I know with confidence I would not be in the position I am without Jon.

I thank Geoffrey Turnbull. His rules have consistently shown me the right path. G.T. taught me an unmeasurable amount about being an economist and good colleague.

I owe a great amount of gratitude to Andrew Hanson. He is my professor turned coauthor and mentor; I am in debt to Andrew for teaching me so much about 'the process'.

I thank Shelby Frost for opening my eyes to graduate studies in economics. I thank Cathy Liu, Dave Sjoquist, and Sally Wallace for their support and advice. My research benefited greatly from their helpful insights. I also want to thank Bess Blyler and Caroline Griffin. Both Bess and Caroline made the difficulties of being a graduate student much easier.

Graduate school would not have been the best time of my life, if it were not for my friends. I thank Andrew Balthrop, Gustavo Canavire-Bacarezza, Merlin Hanauer, and Kelly Wilkin, in particular, for all the great memories from our collective struggle and ultimate success. I also want to thank Juan Jose Miranda Montero, Juan Sun, Guanlin Gao, and Bo Liu for providing the best lunch time entertainment. I am sure these friends I have made for life.

I want to more than thank Danyang Li. I appreciate all that she has done for me. I am the person I want to be because of her. Danyang showed me dreams can come true, and I have enjoyed sharing those dreams with her.

I am so very grateful to have the most understanding, supportive, and thoughtful parents, Larry and Marie Hawley. I do not think it is possible to tell you exactly how important you are to me. I appreciate so much the life you have given me and the love and support through my endeavors. I truly feel extremely honored to be your son.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

ESSAYS ON SOCIAL INTERACTION AND URBAN OUTCOMES

BY

ZACKARY BARTHOLOMEW HAWLEY

MARCH 2012

Committee Chair: Dr. Jonathan C. Rork

Major Department: Economics

This dissertation consists of two essays. The connecting theme for this body of work is social interaction. I define social interaction, for this dissertation, as non-market based face-to-face interaction between individuals where each is interested in a response. Non-market based means, for example, the individual is not trying to negotiate a price of a good or a labor contract, but is merely interested in the other's input, company, or friendship. Within the dissertation, I examine the relationship between social interaction and local area population density and the effect of social interaction on the urban consumer's location and housing consumption decision.

The first contribution investigates the relationship between social interaction and neighborhood population density. Understanding the causal impact on social interactions from a change in neighborhood density is a challenging empirical question; the choice of location for the individual may include a desire for population density or social interaction, or, most likely, both. I address endogeneity or simultaneity using an instrumental variables technique. I use the exogenous variation in a set of geological instruments to predict the neighborhood population density. The instruments include seismic hazard, landslide hazard, and the presence of sedimentary rock beneath the soil. These instruments effect population density through the types of structures that are placed on top of the land. The results suggest that

an increase in neighborhood population density increases some types of social interaction. Friends interaction increases as population density rises, but neighbor interaction falls with increases in population density. Group involvement appears to have no consistent significant effect from a change in population density. This implies a need for theoretical models involving social interaction to be flexible to differing types of social interaction.

The second contribution brings social interaction into a standard urban consumer theory model, the Alonso-Muth model. I start by outlining the canonical model and providing intuition for the inclusion of social interaction. Social interaction is allowed to take multiple forms in the model; interaction can be exogenously or endogenously determined across an area, increase or decrease with population density, or be a substitute or complement to housing consumption. This flexibility provides multiple cases and sub-cases of how social interaction could appear in reality. I capture the effects of changes in social interaction on housing demand and location choice with a number of comparative statics. Lastly, I examine the predictions of the endogenous social interaction model with empirical data. The empirical results buttress the theoretical model's predictions.

1 INTRODUCTION

Face-to-face interaction among individuals residing relatively close within space is a central component to the existence of cities. The gathering of individuals in space provides many advantages. Alfred Marshall proposes "if one man starts a new idea, it is taken up by others and combined with suggestions of their own; and thus it becomes the source of further new ideas"; the ideas he discusses get moved from one neighborhood to another "as it were in the air" (Alfred Marshall (1890), IV.X.3). The mechanism for individuals exchanging these ideas or creating knowledge spillovers within cities is very likely face-to-face interaction.

While Marshall was mostly interested in firm location choice and innovations, the notion of social interaction also plays an important role in the lives of urban residents. Social interaction is face-to-face interaction among individuals that is non-market based; this type of interaction is not determining the price of a good or negotiating a labor contract but the exchange of social pleasantries or friendship, for example. The importance of social interaction is not overlooked in economics; during the last twenty years, the economics literature has examined social activity, be it peer effects[1], neighborhood effects[2], social capital[3], etc. Social interaction, however, is unique since it is likely the mechanism that derives these social activities. The link between social interaction and economic outcomes, particularly urban growth or urban consumer decision making, is not yet obvious or concrete in the literature. The difficulty of analyzing social interaction is mainly the lack of data on individuals preferences for social involvement or interaction, a problem stated by Manski (2000) and eluded to by Putnam (2000).

Within this dissertation, I examine the relationship between social interaction and local area population density and the effect of social interaction on the urban

[1] See Bayer, Hjalmarsson and Pozen (2009) for a review of peer effects.
[2] For example, see Ioannides and Topa (2010).
[3] For example, see Sobel (2002).

consumer's location and housing consumption decision. In the first chapter, I use variation in local population density, predicted by exogenous changes in a set of geological variables, to predict effects on social interaction. The *a priori* expectation of this relationship is that social interaction and neighborhood population density are positively related. As the number of people increases within a close space around the individual, the prevalence of social interaction is expected to rise. This positive relationship assumption is proposed in the urban economics literature and leads to various outcomes such as urban sprawl (Brueckner and Largey, 2008) or increased interactivity within the economy as a whole (Helsley and Zenou, 2011). The empirical results show the type of social interaction is critical in determining the relationship with population density. Friends interaction increases with population density, as expected; however, neighbor interaction decreases and group involvement is constant with changes in local population density. These differing outcomes suggest that not all social interaction should be modeled in the same way.

In the second chapter, I formulate a new social interaction model which builds onto the Alonso-Muth model, a canonical model in urban consumer theory. The social interaction model provides insight to the literature on the effect of various forms of social interaction on the decision of location and housing consumption. Social interaction is allowed to take multiple forms in the model; interaction can be exogenously or endogenously determined across an area, increase or decrease with population density, and be a substitute or complement to housing consumption. The formulation of the model proceeds in steps allowing the assumptions surrounding social interaction to be more flexible, thus retaining the ability to produce predictions of the effects from changes in various forms of social interaction on the urban consumer's location and housing demand choices. The culminating theme of the predictions of the social interaction model is that when social interaction is a substitute to housing consumption, individuals move closer in toward the center of the city and demand less housing, but if social interaction is a complement to housing, then individuals will locate further from the center of the city and consume more housing.

Lastly, I empirically test these predictions of the model. The results yield strong support for the theoretical model. The main source of social interaction data is the Social Capital Community Benchmark Survey. The structure of the survey is such that it is collected through communities, so the data include a national sample and 41 separate community samples. Information is collected on general social interaction, trust, religious involvement, civic participation, diversity of friendship, happiness, feelings of safety, among others.

2 DOES URBAN DENSITY PROMOTE SOCIAL INTERACTION?

Introduction

Over the last two decades, the amount of land used in U.S. urban areas has increased by 2.76 times that of population growth.[4] This staggering difference is an indication of urban sprawl–city boundaries expanding while population density decreases. Urban sprawl is often described in the literature as a response to income growth or reduced transportation costs (Brueckner and Fansler, 1983; Brueckner, 2000, 2001; Glaeser and Kahn, 2003; Nechyba and Walsh, 2004). Sprawl can also be seen as an inefficient individual response to a market action;[5] this portion of the literature provides evidence that shows increased pollution, failing public services, congestion, obesity, and even exacerbated drought problems are consequences of sprawl (Squires, 2002).[6] A vast literature addresses which type of policy is best for curbing sprawl (Anas and Rhee, 2007; Anas and Pines, 2008; Irwin and Bockstael, 2004) and examines indirect effects of other policies on sprawl (Su and DeSalvo, 2008; Brueckner and Kim, 2003; Brueckner, 2005; Song and Zenou, 2006). Despite mounting evidence, urban sprawl's underlying causes and social efficiency are debatable.

I examine a causal link between a density externality, created by social interaction, and urban sprawl. The density externality is derived from the effect of lot size choice on neighbor's social interaction. An individual chooses a lot size to maximize utility but fails to incorporate the effect on his neighbors' change in social

[4] See O'Sullivan (2007).

[5] These inefficiencies are occasionally argued to be allocated to reduced housing prices for the individuals who reside in the affected areas. Housing prices are thought to adjust to the surrounding 'externalities' and, internal to the housing market, solve the inefficiencies (Li and Brown, 1980).

[6] The Urban Institute released an 11 paper review of urban sprawl and several issues including those mentioned and more. Plantinga and Bernell (2007) and Eid et al. (2008) both discuss the relationship between obesity and sprawl.

interaction from his consumption of land. Brueckner and Largey (2008) theoretically model this density externality and find social interaction causes urban sprawl. The existing empirical results, however, do not support the model's assumption; therefore, they conclude social interaction does not cause urban sprawl. I start with a similar theoretical framework that also illustrates the same result–choice of lot size creates a density externality that ultimately leads to urban sprawl. I use a more refined set of instruments, and I find evidence of social interaction causing urban sprawl, in contrast with the previous findings.

Using data from the Social Capital Community Benchmark Survey (SCCBS), which provides information on social interaction, I estimate the relationship between local population density and social interaction. The causal direction of the relationship between population density and social interaction is confounded since density and interaction are likely to be simultaneously determined. To solve this endogeneity issue, I instrument for local population density using earthquake and landslide hazard rates in addition to the presence of hard rock beneath the soil. The identification strategy assess the validity of the theoretical model's assumption of a positive relationship between social interaction and local population density that is required for social interaction, working through a density externality, to cause urban sprawl.

Additionally, I use factor analysis to create social interaction index measures which allow for a bundle of interactions to be examined simultaneously. These indexes are new to the literature and map several components of social interaction into a single index measure, allowing for variation within separate estimates to be condensed into one factor measure. I use the indexes as measures of social interaction to empirically test the theoretical model's assumption.

I find the assumption that social interaction strictly increases with population density only holds for particular types of social interaction, but not all. Social interaction involving friends increases with population density, but neighbor and group social interaction declines with increased density. Additionally, I find a negative relationship between index measures of social interaction and population density.

The remainder of the paper continues as follows. Section 2 describes a theoretical link between social interaction and urban sprawl. Section 3 states the identification strategy for the empirical methods. Section 4 discusses the data used. Section 5 presents the empirical findings. The final section concludes.

Theoretical Model of Social Interaction and Urban Sprawl

Social interaction or other terms used by economists such as peer effects[7], neighborhood effects[8] or social capital, lack a solid economic definition (Manski, 2000). I define social interaction as non-market originating, reoccurring, yet unique, face-to-face interaction where both parties are interested in a response. In urban economics, social interaction is a pillar for the existence of cities. Glaeser and Gottlieb (2006, p. 1275) state "cities make it easier to interact, and one of the main advantages of dense, urban areas is that they facilitate social interactions." There is also a growing literature on the effect of social interaction on the shape of cities. Gasper and Glaeser (1998) suggest cities will no longer exist if technological advancements remove the need for face-to-face interaction; the authors, however, find cities will continue to exist, since face-to-face interaction and technological advancements are not strong substitutes. More recently, Helsley and Zenou (2011) theoretically describe how the shape of cities is determined by social networks and their

[7] Peer effects are generally thought to be changes in behavior due to exposure to other ideas put forth by co-workers, friends, fellow students, etc. Social interactions may be the channel in which peer effects occur but the measurement of these two ideas are different. See Bayer, Hjalmarsson and Pozen (2009) for a review of peer effects.

[8] I argue that there is a distinct difference between social interaction and neighborhood effects–neighborhood effects are exogenous after location is chosen, where social interaction is still endogenously determined even after location is chosen.

central players.

Along these veins, I present a theoretical model in which social interaction is tied to the actions of others, particularly, lot size choice. In this model, the decision-maker is the individual who must decide, among other things, how much land to consume. The range of lots available to the individual is expansive. Each lot is arbitrarily clustered together in space within a neighborhood that surrounds a central business district (CBD). These separate neighborhoods are defined by a distance and direction away from the CBD. The individual consumes land directly, with housing capital suppressed, therefore as lot size increases utility increases. This abstracts from the notion of buying a smaller lot yet consuming nicer housing materials in order to gain a higher level of utility.

Let each decision-maker i's lot size be denoted as x_i. The individual also cares about non-land consumption, y_i, and social interaction, I_i, and maximizes a well behaved utility function $U(x_i, y_i, I_i)$. Within each neighborhood, preferences for all individuals are identical. This can be supported by a Tiebout type sorting model where each neighborhood provides the exact bundle of public goods to satisfy residents' preferences. Thus, the preference for social interaction, I, is also uniform across the neighborhood. I assume the level of social interaction is determined by the average population density in the neighborhood since the surrounding population increases the availability and relative closeness of possible social interaction.

$$I_i = f(\frac{n}{A}), \tag{1}$$

where, n is the number of neighbors and A is the area of the neighborhood. The theoretical model assumes that f is a monotonic function of density. However, it is likely that at some level of extremely high density social interaction is not made better by adding more people. One can think of search costs to finding a good social interaction match as reason for this phenomenon. This scenario is only likely when density is sufficiently high; therefore, the majority of the neighborhoods will not experience this type of reversal. The neighborhood's area is defined as the total amount of land consumed in a neighborhood

$$A = \sum_{j=1, j \in \mathcal{N}}^{n} x_j. \tag{2}$$

Land is available at a fixed price, P, within a neighborhood. The consumer then faces a budget constraint

$$Px_i + y_i = m, \tag{3}$$

where the price of non-housing consumption is normalized to unity and m is gross income. Each individual then solves

$$\max_{x_i, y_i, I_i} U(x_i, y_i, I_i) \ s.t. \ Px_i + y_i = m; \ \ I_i = f(\frac{n}{A}) \tag{4}$$

holding neighbors' decisions constant and known. Using the knowledge about the area of the neighborhood and the social interaction constraint, the problem is simplified to a choice of only lot size,

$$\max_{x_i} U[x_i, m - Px_i, f(n/\sum_{j=1; j \in \mathcal{N}}^{n} x_j)]. \tag{5}$$

The first order condition yields the optimal amount of land to consume

$$-PU_y + U_x - U_I \frac{nf'}{(\sum_{j=1; j \in \mathcal{N}}^{n} x_j)^2} = 0, \tag{6}$$

where the subscripts denote partial derivatives. Since all consumers are identical, the first order condition is symmetric for all individuals in the neighborhood. This results in an Nash equilibrium that is symmetric, with $\sum x_j = nx$, where x is the common lot size. Using the symmetry requirement, the first order condition is re-written as

$$\begin{aligned} -PU_y + U_x - U_I \frac{nf'}{(nx)^2} &= 0 \Longrightarrow \\ -PU_y + U_x - U_I \frac{f'}{nx^2} &= 0 \Longrightarrow \end{aligned}$$

$$\frac{U_x}{U_y} = P + \frac{U_I}{U_y}\frac{f'}{nx^2}. \tag{7}$$

This defines the optimal solution of x for the individual consumer.

If, however, the efficient or socially optimal choice of x is desired, the assumption of symmetry must be made at the onset of the problem. Thus, maximizing the utility of a representative consumer,

$$\max_x U(x, m - Px, f(\frac{1}{x})), \tag{8}$$

which gives a first order condition of

$$\frac{U_x}{U_y} = P + \frac{U_I}{U_y}\frac{f'}{x^2}. \tag{9}$$

When comparing the result of equation (7) to that of equation (9), the difference becomes a factor of $1/n$. In most cases, this factor implies the social optimum choice of lot size to be smaller than the "market" solution; therefore, if left to decide on their own, individuals over consume land. The solution of inefficient lot size choice is assured if the marginal rate of substitution between x and y is assumed to not be affected by social interaction. The over consumption of land leads to the neighborhood being inefficiently large, but urban sprawl occurs when the result is replicated across all neighborhoods within the urban area. If the assumption of a positive relationship between social interaction and density, $f' > 0$, doesn't hold, the social optimum could be exactly equal to or much larger than the "market" solution. I test this restriction on the model with empirical data.

Brueckner and Largey (2008) use a similar theoretical model. They examine the urban area as one single unit and makes city-wide assumptions such as nonexistence of a central business district (CBD), identical preferences, and uniform housing price. I alter their model to ease the acuteness of these restrictions which make the transition to the empirical model easier and more closely represents an actual urban housing market. In particular, I adjust the level of spatial geography implemented in the analysis from city to neighborhood. The movement to neighborhoods allows the model to rely on a CBD as the glue to hold the city together; without one there is no reference from which to expand away. Further, sorting by neighborhoods relaxes the city-wide identical preference and equal land (housing) price restriction. The narrowing in scope is especially helpful since the empirical framework also uses a more restrictive geographical definition (the census tract) which I refer to as a neighborhood.

Identification Strategy

The theory indicates social interaction is determined by the local population density. At the same time, the choice of density is likely to be determined by the demand for social interaction. The question is whether the individual chooses a level of social interaction in response to the density surrounding them; or, the individual selects a density that matches the social interaction desired? The answer to both is probably "yes." This is the classic simultaneity problem, and the resulting inference on an Ordinary Least Squares (OLS) model leads to inconsistent results. Additionally, the potential for omitted variables to occur in explaining social interaction, such as the desire to be friendly or innate friendliness, is high. I use instrumental variable estimation to obtain consistent estimates of local population density's effect on social interaction, as well as to correct for any omitted variables bias.

The set of instruments includes a measure of seismic hazard, landslide hazard, and the presence of sedimentary rock under the soil. The seismic hazard instrument measures how potentially strong and likely seismic activity is for a location. More precisely, the instrument measures the maximal horizontal acceleration of the earth as a multiple (or percentage) of gravity, or the potential severity of an earthquake; there is correlation between potential severity and probability of occurrence. The intuition for this instrument is that, as the seismic hazard increases, building becomes more complicated and costly from fear of instability, lowering the population density. The variable's range is from 0 to 100, and does not correspond directly with census tracts. Thus, the census tract measure is the area weighted average of the scores located inside the tract.

One complication with the seismic hazard instrument is that all of the high (30 or over) scores are located on the west coast. I create two additional instrumental variables using the seismic hazard rate: a tract level dummy variable indicating a high value for the hazard rate[9], and an interaction with the dummy variable and the original seismic variable–to avoid the seismic hazard rate being confounded with a regional effect. This allows the trend for building in the west coast to vary from the overall effect on building as the hazard increases.

The landslide hazard measure takes values of low, medium, and high (1, 2, and 3, respectively) and represents how likely a landslide is to occur. The story supporting this instrument is if the area is prone to landslides, the inhabitants are less likely to build upon that land for fear of losing their homes. The landslide hazard instrument is calculated at the census tract level as an area weighted average of its

[9] The dummy equals unity if the area weighted seismic hazard score is above or equal to 30, zero otherwise.

values.

The last geological instrument is the percentage of area within a census tract in which sedimentary bedrock is found. The type of rock found deep in the earth is important for determining the type of structure that is placed on top of the ground. The harder the rock the higher and larger the structure's potential. Since high density areas need large and tall structures to house masses of people, the type of rock can determine the feasibility or ease of such housing. Rosenthal and Strange (2008) use a similar set of instruments to obtain consistent estimates of the effect density has on labor market outcomes.

The instrumental variable technique relies on the instruments, themselves, not directly influencing the level of social interaction except through their effect on population density. The instruments are naturally occurring and it is difficult to link social interaction with them except through their effect on density. This exogeneity is what makes them excellent instruments. Further, the instruments have both inter and intra urban area variation. If one thinks the level of social interaction in an urban area depends upon its natural structure, then the inter-area variation within the instruments confound the estimated relationship. The instruments, however, describe intra-area variation and account for local density differences within the urban area.

The previous literature lacks exogenously determined instrumental variables. Brueckner and Largey (2008) use urban area density, metropolitan area density and urban fringe ruggedness to instrument for density to estimate the relationship between social interaction and population density. These instruments do not seem plausibly exogenous because the population density of the urban area and MSA are affected by the local density externality. If it is assumed the level of social interaction is determined by the local density, and that density creates an externality, when replicated throughout the urban area or MSA the city-wide population density is also affected. Further, if the argument is made that census tract density is a determinate of social interactions, then a similar connection could be drawn between an urbanized area or MSA density and social interaction. The previous literature also relies solely on inter-area variation in the instruments. The individuals within the same urban area, therefore, all have identical values for the instruments, with no variation to explain intra-area differences. This implies sprawl is identified as an urban area phenomenon; however, the theoretical and empirical strategy assumes it to be more localized and replicated throughout the MSA or urban area.

I consistently test the relationship between social interaction and local population density using an instrumental variables approach. If a positive relationship is found, this supports the theoretical model's prediction–urban sprawl is inefficient city growth caused by social interaction through a density externality. The instrumental variables technique is performed in two stages. I instrument for density in the first stage,

$$Tract\ Density_i = \alpha + \mathbf{\Omega GIs}_i + \mathbf{\Gamma} \boldsymbol{X}_i + \varepsilon_i \tag{10}$$

where **GIs** is the set of geological instruments. The coefficients of interest in the first stage are represented by the vector $\mathbf{\Omega}$, or the effect of the instruments on tract density. I expect a negative relationship with density for the seismic hazard, seismic interaction and landslide hazard and a positive relationship for the high seismic indicator and sedimentary rock instrument.

A vector of other control variables, $\boldsymbol{X}_i$, includes socioeconomic variables such as age, income, marital status, children, race, and others.[10] Regional dummies are also used in the specification and are included in $\boldsymbol{X}_i$. The remaining estimates, $\mathbf{\Gamma}$ and α, are the marginal effects of the other control variables and a constant, respec-

[10] The full list of additional controls is as follows: gender, age, age squared, marital status, number of kids, income, education, race, ethnicity, employment, citizenship, tenure in neighborhood, MSA murder rate, MSA dissimilarity index, and census region dummy variables.

tively.

Assuming the first stage is identified correctly, the second stage of the estimation consistently describes the link between social interaction and density.

$$SocInter_i = \alpha_1 + \beta_1 * Tract\ \hat{D}ensity_i + \mathbf{\Gamma}_1 \mathbf{X}_i + \varepsilon_1 \tag{11}$$

The coefficient of interest is β_1 which tests the assumption made in the theoretical model. The theoretically driven expectation is a positive relationship–as density increases the social interaction of individuals also increases. Again the control variables, $\boldsymbol{X}_i$, are included in the full specifications with their marginal effects represented by the vector of coefficients $\mathbf{\Gamma}_1$. The dependent variables are friends and neighbor, group involvement, and index measures of social interaction which are described below.

Data

Social interaction is ubiquitous within the urban area, but hard to measure empirically. I use data from the Social Capital Community Benchmark Survey (SC-CBS). The survey took place in 2000 and is distributed by the Roper Center for Public Opinion Research at the University of Connecticut.[11] The survey was designed by the Saguaro Seminar at the John F. Kennedy School of Government, Harvard University with the intention of formulating a uniform measure of social capital.

These data include a national sample as well as 41 separate "community" samples. The national sample includes just over 3000 respondents, which were chosen at random across the Continental U.S. with over-sampling of blacks and Hispanics.[12] As for the 41 "communities," each separate community is defined different

[11] Access to the unrestricted data can be found through www.roper.uconn.edu.

[12] Blacks and Hispanics account for 500 respondents each which resulted in an additional 288

across space; some "communities" capture a single city, while others encompass entire states.[13] Because the community samples are sponsored by different institutions or individuals, the sampling techniques are not uniform across the communities.[14] The total number of observations in the data set is just under 30,000.

I concentrate on individuals that reside within an urban area, since the objective is analyzing social interaction and urban sprawl. In the restricted use SCCBS, each individual's location is defined by the census tract for which the home is located. An urbanized area is defined by the Census Bureau and can include all or any size portion of the census tracts that make up the urban area; therefore, a respondent is in the sample if the tract in which he resides intersects at any point with an urbanized area. About 21,500 observations satisfy these conditions, the samples vary slightly for the different social interaction measures because of missing values.[15] The observations fall within 311 separate urbanized areas.

The SCCBS collects data through phone interviews where respondents share experiences about trust, friendship and social activities. I focus on the questions that address social information or non-market exchange. These select questions measure how interactive the respondents are within their respective community. The ten social interaction dependent variables, $SocInter_i$, are broken into two categories, friendship and group involvement.

I use six different measures of friendship. First, how often the respondent socializes with his neighbors (NEISOC). Second, if the respondent talks with a neighbor more than once a week (NEITALK). Third, how many people in which the

---

blacks and 294 Hispanics in the sample than otherwise would have been under random sampling.

[13] As an example, Atlanta is defined as a community which includes counties that are in the MSA where as Indiana is defined as a community which includes respondents selected random across the entire state.

[14] There are several different sampling techniques used across the communities. One community may sample heavier in Hispanics while another oversamples Native Americans. The sponsor of the community sample may have been interested in a particular group of people and required more information from that group.

[15] There is one exception, NEICOOP, or working together with neighbors to get something fixed or improved, which is missing a significant amount of responses.

respondent can confide (CONFIDE). Fourth, how many "close" friends the respondent has (FRIENDS). Fifth, how often the respondent "hangs out" in public places with friends (SOCPUBLIC). Last, how often friends visit the respondent's home (SOCHOME).

Four variables capture group involvement. First, if the respondent is involved in a neighborhood cooperation (NEICOOP). Second, if the respondent participates in a hobby-oriented club (GRPHOB). Third, how often the respondent went to any club meeting (CCLUBMET). Last, the number of non-church groups in which the respondent is involved (GRPINVLV).

Additionally, the survey includes several index variables. These indexes are generated from the information provided by the respondents found in the survey, not necessarily including the social interaction measures above. I use four indexes: social trust (SOCTRUST), faith-based interaction (FAITHBAS), civic participation (CIVICPART), and friendship diversity (DIVERSITY).

The social trust index attempts to capture the broad notion of trust in others. General interpersonal trust, trusting neighbors, trusting police and others help comprise this index. This index may be the furthest from social interaction as I defined it, but trust is pivotal in the determination of interaction. The respondent who does not trust his neighbors is probably much less prone to interact with them. Faith-based interaction includes church membership, and attending church service. The civic participation index comprises various interactions dealing with local politics and community activities–attending a political meeting or rally, work on a community project, and demonstrating, protesting, boycotting or marching. This index is most in line with the types of individual forms of social interaction used in this paper. Friendship diversity measures how many different types of people a respondent considers friends. The types of friends include several categories based on race, ethnicity, employment status, being a community leaders, etc.

I also create my own indexes from the ten social interaction components described above. The method used to create the indexes condenses the variation across several variables into several factor loadings. I use these loadings to predict a single factor measure which becomes the index measure. Instead of looking directly at specific means of interacting, all of the index measures provide an alternative way to examine how general social interaction is related to tract population density. The index (FACTOR) implements factor analysis to create one index from all ten social interaction components. The second created index (FACTOR2) excludes the respondent's involvement in a neighborhood cooperation since this variable has a large number of missing observations.

The instrument data come from the United States Geological Survey (USGS). The information is readily found through their website as boundary files and is downloadable in shapefile format. These data are overlaid with census tract boundary shapefiles to calculate the area weighted scores for each instrumental variable at the tract level. Figures 1-3 show examples of each map from which the instruments are derived. Figure 1 shows a seismic hazard map of Los Angeles, California. This map demonstrates the variation in the seismic hazard instrument (the range is from 0 to 100). The seismic gradients do not follow exactly the census tract boundaries, so I create an area weighted average of the seismic score for each census tract which constitute the seismic hazard instrument. The values of the instrument measure take the same range as the seismic hazard rate. The instrument, however, is not an integer scale. Figure 2 presents a landslide hazard map of Atlanta, Georgia. In this map, the landslide hazard displays its full range. The landslide hazard instrument constructed is the area weighted average of the values. Again, the range for the landslide hazard instrument is identical to the landslide hazard (1 to 3), but the instrument is no longer in integer form. Figure 3 displays a sedimentary rock map for Boston, Massachusetts. As shown, the presence of sedimentary rock does not coincide with the drawing of census tracts. This map provides better insight into how the instruments are constructed. Each census tract is subdivided into two categories–sedimentary rock and non-sedimentary rock. Then the area of each is calculated and the ratio to the complete census tract area is taken. The construction leads to a range of values for the sedimentary rock instrument from 0 to 1, with non-integers values possible.

Table 1 provides a description of the full set of variables. The variables of interest are the social interaction terms described above as well as the population density of each tract, in natural log form. The social interaction measures are categorical or binary responses; this complicates the intuition for the magnitudes. For example, the average level of socializing at home is 21.95 which corresponds to around twice a month. Alternatively, almost 54 percent of the individuals in the sample talks with or visits immediate neighbors at least once a week.

The descriptive table shows respondents in the sample are 41 percent male, 56 percent married, 37 percent college graduates, 15 percent black, 15 percent retired and on average 44 years of age. The average seismic hazard rate for the sample is just over a score of 11. Respondents reside in a high seismic hazard area (score of 30 or above) in about 16 percent of the sample. The average landslide hazard rate is 1.5, and the average tract in the sample has 83 percent of its area with sedimentary bedrock beneath the soil. The additional controls include some data from other sources. The murder rate is taken from the Federal Bureau of Investigation's Uniform Crime Reports and is recorded at the MSA level. The dissimilarity index captures how racially segregated the population is within a MSA which comes from Cutler, Glaeser and Vigdor's data website.[16]

[16] The dissimilarity index was found at http://trinity.aas.duke.edu/~jvigdor/segregation/index.html.

Figure 1: Seismic Hazard Map of Los Angeles, California

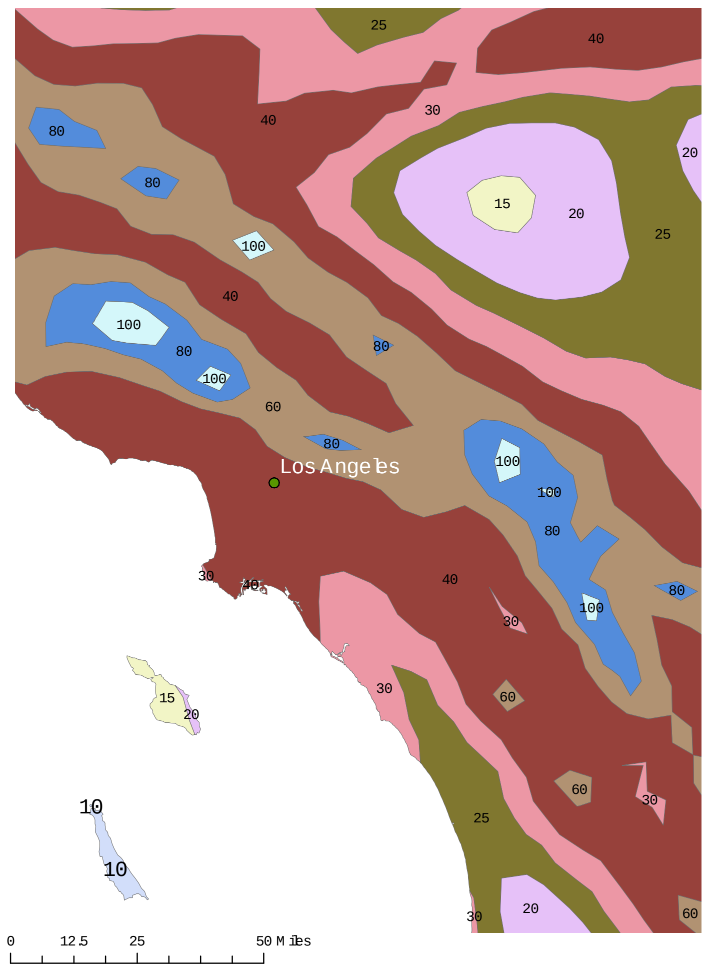


Figure 2: Landslide Hazard Map of Atlanta, Georgia

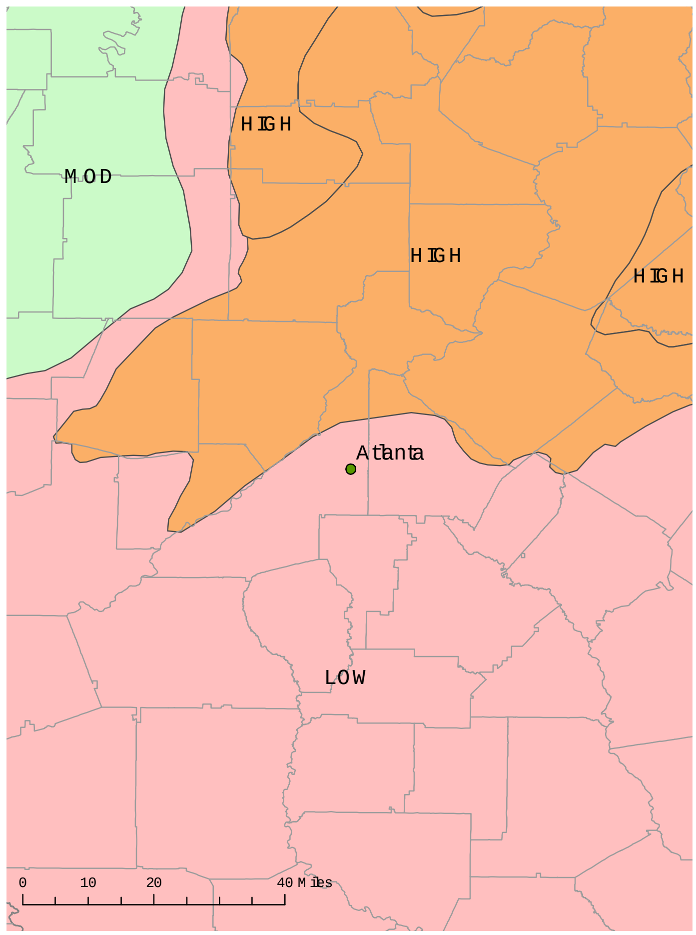


Figure 3: Sedimentary Rock Map of Boston, Massachusetts

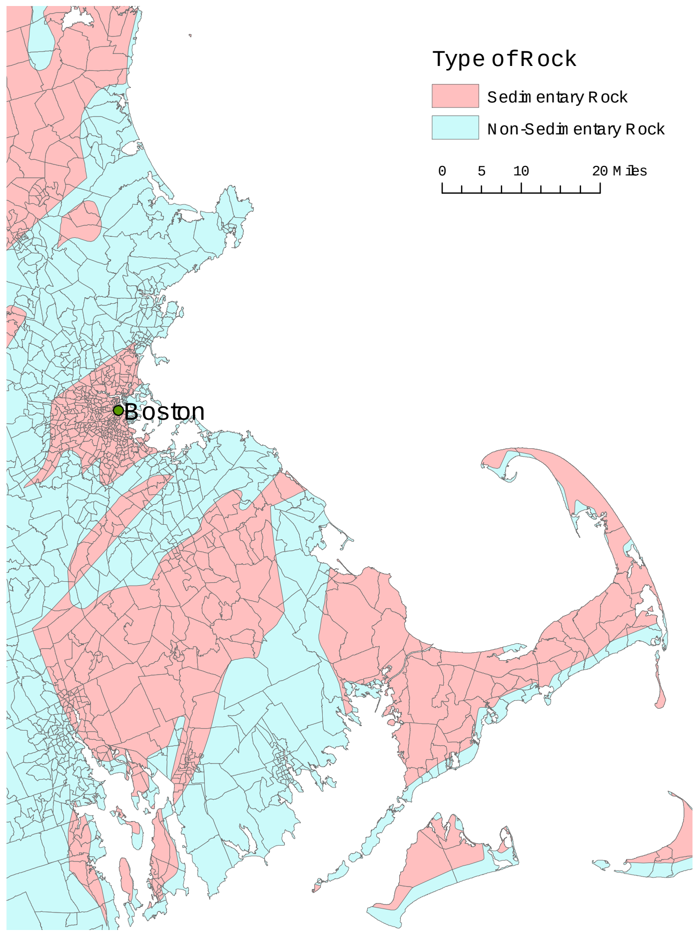


Table 1: Summary Statistics and Variable Definitions

Variable	Definition	Mean	Std. Dev.	Min	Max
Social Interaction Variables					
#NEITALK	How often respondent talks with or visits immediate neighbors.	5.061	1.848	1	7
NEITALK	=1 if respondent talks or visits immediate neighbors at least once a week, 0 otherwise.	0.539	0.498	0	1
CONFIDE	Number of people in which respondent can confide.	3.550	0.787	1	4
FRIENDS	Number of close friends.	3.317	1.063	1	5
SOCPUBLIC	Number of times per year respondent hangs out with friends in a public place.	16.968	20.147	0	60
SOCHOME	Number of times per year respondent visits with friends at home.	21.975	21.214	0	60
NEICOOP	=1 if respondent has worked with neighbors to get something fixed or improved.	0.318	0.466	0	1
HOBBYCLUB	=1 if respondent participates in a hobby, investment or garden club.	0.262	0.440	0	1
CLUBMTGS	Number of times per year respondent attends club meetings.	6.195	12.163	0	60
#GROUPS	Number of types of non-religious organizations to which respondent belongs.	3.041	2.620	0	17
Social Interaction Indexes					
FACTOR	Factor analysis of the ten social interactions above.	0	0.890	-1.7	1
FACTOR2	Factor analysis of nine social interactions, leaving out NEICOOP due to missing values.	0	0.888	-1.6	1
SOCTRUST	Index of social trust.	-0.011	0.697	-2.6	1
FAITHBAS	Index of faith-based interactions.	-0.087	0.770	-1.1	1.7
CIVICPART	Index of political and community involvement.	1.763	1.276	0	5
DIVERSITY	Index of diversity in friendships.	6.378	2.669	0	11
Population Density Measure					
TRACT DENSITY	Natural log of census tract population density: people per square mile.	8.029	1.343	0.9	12.3

Continued on Next Page. . .

Table 1 Summary Statistics and Variable Definitions – Continued

Variable	Definition	Mean	Std. Dev.	Min	Max
Instruments for Density					
SEISMIC HAZARD	Hazard rate of seismic activity in potential and magnitude.	11.610	19.372	0	90.6
HIGH SEISMIC	=1 if seismic hazard rate is 30 or above.	0.165	0.371	0	1
SEISMIC INTER.	Interaction term of high seismic and seismic hazard.	8.672	20.304	0	90.6
LANDSLIDE	Hazard rate of landslide activity.	1.573	0.693	1	3
SED. ROCK	Percentage of sedimentary rock found below the surface of the earth.	0.833	0.359	0	1
Respondent Characteristics					
MALE	=1 if respondent is male, 0 otherwise.	0.414	0.493	0	1
AGE	Age in years.	44.042	16.592	18	99
AGE2	AGE squared.				
MARRD/PARTN	=1 if respondent is married or living with partner, 0 otherwise.	0.564	0.496	0	1
KIDS6	Number of children in household under six years old.	0.266	0.685	0	8
KIDS6_17	Number of children in household between six and seventeen.	0.489	0.920	0	8
INC2	=1 if \$20k $<$ annual household income $<$ \$30k.	0.123	0.328	0	1
INC3	=1 if \$30k $<$ annual household income $<$ \$50k .	0.217	0.412	0	1
INC4	=1 if \$50k $<$ annual household income $<$ \$75k .	0.173	0.378	0	1
INC5	=1 if \$75k $<$ annual household income $<$ \$100k.	0.102	0.302	0	1
INC6	=1 if annual household income $>$ \$100k.	0.117	0.322	0	1
SOMECOLL	=1 if respondent has completed some college education, 0 otherwise.	0.323	0.468	0	1
COLLGRAD	=1 if respondent has a college degree, 0 otherwise.	0.374	0.484	0	1
BLACK	=1 if respondent is black, 0 otherwise.	0.150	0.357	0	1
HISPANIC	=1 if respondent is Hispanic, 0 otherwise.	0.092	0.288	0	1
ASIAN	=1 if respondent is Asian, 0 otherwise.	0.031	0.173	0	1
UNEMP	=1 if respondent is unemployed, 0 otherwise .	0.026	0.160	0	1
HOMEMAKER	=1 if respondent is a home-maker, 0 otherwise.	0.062	0.241	0	1
RETIRED	=1 if respondent is retired, 0 otherwise.	0.149	0.356	0	1
CITIZEN	=1 if respondent is a US citizen, 0 otherwise.	0.935	0.246	0	1
LIVING5	=1 if respondent has lived in his/her community for more than five years, 0 otherwise.	0.649	0.477	0	1

Continued on Next Page...

Table 1 Summary Statistics and Variable Definitions – Continued

Variable	Definition	Mean	Std. Dev.	Min	Max
Regional Controls. Omitted Category: Pacific					
NEWENGL	=1 if census region is New England, 0 otherwise.	0.043	0.204	0	1
MIDATLAN	=1 if census region is Mid Atlantic, 0 otherwise.	0.086	0.282	0	1
EANOCENT	=1 if census region is East North Central, 0 otherwise.	0.220	0.414	0	1
WENOCENT	=1 if census region is West North Central, 0 otherwise.	0.067	0.250	0	1
SOUTHATL	=1 if census region is South Atlantic, 0 otherwise.	0.213	0.410	0	1
EASOCENT	=1 if census region is East South Central, 0 otherwise.	0.041	0.198	0	1
WESOCENT	=1 if census region is West South Central, 0 otherwise.	0.053	0.223	0	1
MOUNTAIN	=1 if census region is Mountain, 0 otherwise.	0.069	0.254	0	1
Urban Area Controls					
MURDER	Number of murders per 100,000 people.	6.123	4.017	0	20.4
DISSIMILARITY	Dissimilarity index for segregation of blacks versus non-blacks.	0.566	0.141	0.2	0.8

Results

In each of the social interaction measures examined, I estimate two different specifications. The base model is a bare-bones approach that only utilizes the instruments in the first stage and then tests the relationship between social interaction and tract density with no other controls included. The full model incorporates the control variables, $\mathbf{X}_i$, in the first-stage and second-stage. In both models, I use clustered standard errors at the MSA.

First Stage Results

Table 2 shows the first stage results for the geological instruments. For each of the instruments, the first stage results largely yield the expected relationship with tract density and are highly significant. The landslide hazard instrument yields the expected negative relationship in the base model, but in the full model, the relationship is not significantly different from zero. The seismic hazard instruments, comprised of the rate, dummy, and interaction variables mostly follow intuition. In the base model, the relationship between tract population density and the rate of seismic hazard is positive, while controlling for a significantly positive west-coast effect with the dummy for high hazard values. The interaction term is negative signifying that as the hazard increases in a high hazard environment, tract population density declines. In the full model, the results are more precise. The hazard rate is now significantly positively related to density, and the interaction coefficient is significantly negative. The relationship of sedimentary rock with density is intuitive and consistent across both specifications with a positive and significant coefficient.

I use a number of tests to estimate the strength of the instruments. In every specification, the first stage F-statistic is well above 10, the rule-of-thumb for instrument relevance in the first stage. The minimum F-statistic is 19.13 for the clustered-base model with a maximum of 742.37 for the base model. The first stage is not underidentified or weakly identified as it passes all tests. The Anderson-Rubin test shows the instruments strongly predict local population density.

Friends and Neighbors Results

Table 3 presents the relationship between census tract population density and social interactions involving friends and neighbors in the base model specification. The majority of the estimates indicate the coefficient on tract density is positive, as expected. This provides some support for the assumption that as population density increases so does social interaction. The number of people in which an individual confides, the number of friends, and socializing with friends in public places are all significantly positively related to tract density. To give an example

Table 2: Density and Geological Instruments: First Stage Results

Specification	Base Model	Full Model
Dependent Variable: Tract Population Density	Coeff.	Coeff.
Landslide Hazard	-0.047	0.008
	(0.139)	(0.075)
Seismic Hazard	0.002	0.070***
	(0.030)	(0.025)
High Seismic Hazard	1.059***	2.086***
	(0.389)	(0.599)
Seismic Hazard Interaction	-0.001	-0.068***
	(0.030)	(0.025)
% Sedimentary Rock	0.658***	0.361
	(0.163)	(0.267)
Constant	7.363***	7.122***
	(0.315)	(0.510)
N	22052	21405
Number of Clusters	311	311
Partial R-squared	0.1441	0.0526
Instrument Tests		
F-Statistic	19.13	48.26
Kleibergen-Paap rk underidentification	72.17	21.90
p-value	0.0000	0.0005
Kleibergen-Paap rk weak indentification	14.38	4.36
p-value	-	-
Anderson-Rubin weak instrument	4.81	2.52
p-value	0.0003	0.0297

Notes: The Full Model includes various other control variables, such as age, race, income, education, marital status, etc. The full list of other controls can be seen in other tables, for example, Table 4. *,**,*** denotes statistical significance at the .10, .05, and .01 level. (Standard errors) on the size of this effect, an increase in local population density by 10 percent increases the amount of socializing with friends in public from around once a month to just under three times a month at the average. Talking with neighbors and the number of neighbors an individual knows, however, are both shown to significantly decrease with density. This tells and interesting story. If the individual resides in a lower density area, the suburbs for example, he seems to interact more with his neighbors. When the individual, however, locates in a dense urban setting, he interacts less with neighbors. This may be affected by the use of the words "immediate neighbor" when the survey was conducted. The person located down the hall or in the next building across the street may not constitute as an "immediate neighbor" in a dense urban setting.

In the full model, most of the significance is lost. In Table 4, the neighbor interactions still yield negative but not significantly negative results. The confide interaction is the only significant positive relationship remaining. Density is negatively related to both the number of friends and how often individuals socialize at home but not significantly different from zero. Last, hanging out with friends in public places is positively related to local density.

The problem with including more control variables is that the variables themselves are not exogenously determined in reality. Density choice may depend upon income, marital status or other controls, and if the inclusion of those covariates is also correlated with social interaction, it creates a bad control problem bias or proxy control bias (Angrist and Pischke, 2008). The inclusion of these controls may increase the likelihood of causal interpretation by removing some omitted variable bias, but since they could also be considered outcomes that depend upon social interaction and/or population density, they may confound the result of interest. If, however, the instruments are effective, the omitted variable bias and endogeneity is purged in the base and clustered-base model, leaving a consistent coefficient of

Table 3: Tract Density and Friendship-Oriented Social Interaction: Base Model

Specification	Base Model					
Dependent Variable	#NEITALK	NEITALK	CONFIDE	FRIENDS	SOCPUBLIC	SOCHOME
Technique	(2SLS)	(IVPROBIT)	(2SLS)	(2SLS)	(2SLS)	(2SLS)
	Coeff.	Coeff.	Coeff.	Coeff.	Coeff.	Coeff.
TRACT DENSITY	-0.109***	-0.079***	0.065***	0.063***	1.712***	0.677
	(0.039)	(0.022)	(0.022)	(0.023)	(0.454)	(0.479)
CONSTANT	5.937***	0.730***	3.027***	2.811***	3.227	16.549***
	(0.312)	(0.180)	(0.173)	(0.189)	(3.653)	(3.830)
OLS or Probit Result						
TRACT DENSITY	-0.031**	-0.012**	-0.008	-0.026**	1.045***	0.189
N	21902	22052	21968	21973	21971	21946
Number of Clusters	310	311	311	310	311	311
Uncentered R-squared	0.883	-	0.952	0.906	0.417	0.518
Centered R-squared	0.005	-	0.016	0.012	0.003	0.001
Instrument Tests						
First Stage F-Statistic	19.06	-	19.39	19.12	19.23	18.98
Kleibergen-Paap rk underidentification	11.73	-	11.78	11.73	11.735	11.72
p-value	0.0386	-	0.0379	0.0387	0.0386	0.0389
Kleibergen-Paap rk weak identification	14.38	-	14.46	14.48	14.41	14.32
Stock-Yogo (10%)	10.83	-	10.83	10.83	10.83	10.83
Hanson-J over ident. test statistic	8.61	-	0.71	0.46	6.62	6.08
p-value	0.0716	-	0.9500	0.9774	0.1575	0.1935

Notes: The instrument tests are not valid for the IV probit specification. *,**,*** denotes statistical significance at the .10, .05, .01 level. (Standard errors)

interest.

The friends results capture the majority of the positive relationships which corresponds with the assumption made in the theoretical model. The neighbor interactions, however, are consistently found to have a negative or significantly negative relationship with density. These results may indicate suburbs provide a better environment for neighbors to interact, and are counter-intuitive to the model's prediction.

The OLS results in the base and clustered-base model show the importance of the instruments. In all cases the magnitudes of the estimated coefficients is pulled to zero for the OLS results. In two of the six estimations, the OLS results differed in sign, and overall the results are more precise with the instrumental variables technique.

The friends and neighbors results tables also report the instrument tests. The first stage F-statistic, as discussed earlier, is high in all specifications. Other tests including the Kleibergen–Paap test statistics are also large signifying underidentification and weak instruments is not a concern. The Stock and Yogo (2005) critical values for weak instrument tests are used since no critical values are available in the literature when clustering is applied. In almost all cases, I fail to reject the overidentifying restrictions, as the Hansen-J statistics are close to zero. This provides evidence that the geological instruments are exogenously determined and sufficiently strong.

Group Involvement Results

Table 5 reports the results for group involvement interaction and population density in the base model. The results for social interaction involving groups, clubs or a neighborhood cooperation yield a negative relationship with local population density. The number of club meetings respondents attend is the only statistically

Table 4: Tract Density and Friendship-Oriented Social Interaction: Full Model

Specification	Full Model					
Dependent Variable	#NEITALK	NEITALK	CONFIDE	FRIENDS	SOCPUBLIC	SOCHOME
Technique	(2SLS)	(IVPROBIT)	(2SLS)	(2SLS)	(2SLS)	(2SLS)
	Coeff.	Coeff.	Coeff.	Coeff.	Coeff.	Coeff.
TRACT DENSITY	-0.054	-0.035	0.035**	-0.003	0.377	-0.780
	(0.052)	(0.043)	(0.015)	(0.015)	(0.530)	(0.521)
MALE	0.079***	0.039*	-0.106***	0.063***	-1.983***	0.136
AGE	0.022***	0.007*	-0.008***	-0.021***	-1.325***	-1.193***
AGE2	-0.000	0.000	0.000**	0.000***	0.009***	0.008***
MARRD/PARTN	0.165***	0.092***	0.022*	-0.009	-4.655***	-1.707***
KIDS6	0.129***	0.091***	-0.048***	-0.062***	-1.528***	-0.885***
KIDS6_17	0.057***	0.044***	-0.031***	-0.047***	-0.891***	-0.437**
INC2	-0.022	-0.059*	0.063***	0.010	0.513	0.266
INC3	0.081**	-0.027	0.116***	0.079***	1.833***	0.893**
INC4	0.115***	-0.033	0.156***	0.110***	2.547***	1.777***
INC5	0.126***	-0.012	0.155***	0.105***	3.232***	1.313**
INC6	0.169***	0.007	0.201***	0.248***	4.599***	3.841***
SOMECOLL	0.185***	0.094***	0.154***	0.153***	1.994***	2.179***
COLLGRAD	0.172***	0.034	0.236***	0.267***	1.386***	0.737**
BLACK	-0.418***	-0.140***	-0.218***	-0.429***	-2.652***	-3.085***
HISPANIC	-0.481***	-0.172***	-0.309***	-0.224***	-2.210***	-4.597***
ASIAN	-0.350***	-0.180***	-0.166***	-0.176***	-3.184***	-3.474***
UNEMP	-0.060	-0.014	-0.122***	-0.119***	0.549	0.065
HOMEMAKER	0.127**	0.091**	0.025	0.024	1.351**	0.682
RETIRED	0.207***	0.136***	0.060***	0.169***	1.603***	2.129***
CITIZEN	0.509***	0.303***	0.161***	0.032	1.125	3.856***
LIVING5	0.266***	0.112***	0.026**	0.110***	1.500***	1.973***
PACIFIC	0.193**	0.085	-0.000	0.151***	1.627***	1.230*
MIDATLAN	0.070	0.065	-0.046	0.002	0.103	-0.641
EANOCENT	0.140*	0.051	0.002	0.113***	1.229	-0.345
WENOCENT	0.332***	0.190***	0.010	0.131***	2.219***	0.408
SOUTHATL	0.134	0.046	-0.035	0.081***	0.211	-1.214
EASOCENT	0.110	0.033	-0.045	0.134**	1.071	-1.752
WESOCENT	0.240**	0.029	-0.034	0.115***	1.116	-0.456
MOUNTAIN	0.319***	0.105	0.017	0.174***	2.998***	2.498***
MURDER	-0.006	0.001	-0.003	-0.005**	-0.126***	-0.046
DISSIMILARITY	0.412*	0.181	-0.076	0.168	4.179*	1.570
CONSTANT	3.418***	-0.599	3.271***	3.378***	49.018***	56.595***
OLS or Probit Result						
TRACT DENSITY	0.016	0.006	0.000	-0.009	0.332***	-0.022
N	21405	21539	21483	21486	21487	21466
Number of Clusters	307	308	308	307	308	308
Uncentered R-squared	0.892	-	0.957	0.914	0.508	0.563
Centered R-squared	0.074	-	0.077	0.075	0.156	0.091
Instrument Tests						
First Stage F-Statistic	47.20	-	48.24	49.86	48.78	48.64
Kleibergen-Paap rk underidentification	3.28	-	3.37	3.33	3.40	3.34
p-value	0.6567	-	0.6426	0.6496	0.6387	0.6473
Kleibergen-Paap rk weak indentification	4.36	-	4.33	4.41	4.38	4.38
Stock-Yogo (10%)	10.83	-	10.83	10.83	10.83	10.83
Hanson-J over indent. test statistic	4.15	-	4.13	1.20	7.70	5.18
p-value	0.3858	-	0.3892	0.8778	0.1032	0.2690

Notes: The instrument tests are not valid for the IV probit specification. *,**,*** denotes statistical significance at the .10, .05, .01 level. (Standard errors)

significant relationship. This is not the expected result. Perhaps, the cost of gathering for these meetings or establishing these groups increases with higher density. Nonetheless, the hobby club interaction measure is positively but not significantly correlated with density.

The full model results for group involvement interaction are shown in Table 6. The estimation yields association with a hobby club, attending club meetings, and the number of groups all decrease with increased population density. The effect is significantly different from zero for club meetings and number of groups. The remaining estimate is a positive relationship between neighborhood cooperation and density.

The friends and neighbors and group involvement results yield a differing view of the relationship between social interaction and local population density; this is in contrast to recent findings in the literature on the relationship. Brueckner and Largey (2008) examine the same set of social interaction measures (the ten interaction measures from above) using an instrumental framework as discussed earlier. Nine of the ten estimated coefficients reported are negative, seven of which are statistically significant. Their results imply that as density in an area increases the social interaction of the inhabitants' declines. These results are contrary to their hypothesis. The last estimate is positive but insignificant. These results, however, may be driven by inadequate instruments.

The literature has not discussed, however, the importance of the measurement or type of social interaction in the realization of the relationship with density. The friends results link social interaction and density positively, while the neighbors and group involvement interaction is shown to decrease with higher density. These social interaction measures are narrowly defined and this could be a drawback of analyzing ten different specifications.

Table 5: Tract Density and Group-Involvement Social Interaction: Base Model

Specification	Base Model			
Dependent Variable	NEICOOP	HOBBYCLUB	CLUBMTGS	#GROUPS
Technique	(IVPROBIT)	(IVPROBIT)	(2SLS)	(2SLS)
	Coeff.	Coeff.	Coeff.	Coeff.
TRACT DENSITY	-0.016	0.003	-0.316*	-0.001
	(0.034)	(0.028)	(0.165)	(0.075)
CONSTANT	-0.346	-0.656***	8.732***	3.050***
	(0.271)	(0.225)	(1.330)	(0.589)
OLS or Probit Result				
TRACT DENSITY	-0.002	-0.020**	-0.144*	-0.002
N	10843	22020	22005	22052
Number of Clusters	257	311	311	311
Uncentered R-squared	-	-	0.206	0.574
Centered R-squared	-	-	0.000	0.000
Instrument Tests				
First Stage F-Statistic	-	-	19.21	19.13
Kleibergen-Paap rk underidentification	-	-	11.73	11.74
p-value	-	-	0.0386	0.0385
Kleibergen-Paap rk weak indentification	-	-	14.43	14.42
Stock-Yogo (10%)	-	-	10.83	10.83
Hanson-J over ident. test statistic	-	-	1.31	5.33
p-value	-	-	0.8601	0.2548

Notes: The instrument tests are not valid for the IV probit specification. *,**,*** denotes statistical significance at the .10, .05, .01 level. (Standard errors)

Table 6: Tract Density and Group-Involvement Social Interaction: Full Model

Specification	Full Model			
Dependent Variable	NEICOOP	HOBBYCLUB	CLUBMTGS	#GROUPS
Technique	(IVPROBIT)	(IVPROBIT)	(2SLS)	(2SLS)
	Coeff.	Coeff.	Coeff.	Coeff.
TRACT DENSITY	0.036	-0.055**	-1.020***	-0.187**
	(0.056)	(0.025)	(0.257)	(0.086)
MALE	0.061**	0.153***	-0.655***	0.077*
AGE	0.026***	0.008**	-0.258***	0.025***
AGE2	-0.000***	-0.000**	0.002***	-0.000*
MARRD/PARTN	0.078***	-0.062***	-0.931***	-0.029
KIDS6	0.028	-0.060***	-0.476***	-0.029
KIDS6_17	0.055***	-0.027**	0.448***	0.274***
INC2	-0.061*	-0.081**	-0.360	0.053
INC3	-0.008	0.063**	0.377**	0.307***
INC4	0.046	0.109***	0.716***	0.454***
INC5	0.116***	0.163***	0.414	0.604***
INC6	0.215***	0.233***	1.579***	0.993***
SOMECOLL	0.236***	0.319***	2.342***	0.948***
COLLGRAD	0.263***	0.450***	4.197***	1.789***
BLACK	0.125**	-0.021	-0.110	0.660***
HISPANIC	-0.011	-0.046	-0.590*	0.080
ASIAN	-0.213**	-0.055	-1.700***	-0.263**
UNEMP	-0.205**	-0.123**	-0.311	-0.418***
HOMEMAKER	0.067	0.021	0.511	-0.315***
RETIRED	-0.069	0.126***	0.588*	-0.093
CITIZEN	0.247***	0.175***	1.714***	0.605***
LIVING5	0.217***	0.082***	1.001***	0.385***
PACIFIC	-0.015	0.063	0.762*	0.314***
MIDATLAN	-0.244***	-0.056	-0.673	-0.187
EANOCENT	-0.106	-0.012	-0.461	-0.019
WENOCENT	0.161***	0.057	0.801**	0.260***
SOUTHATL	0.061	0.014	-0.554	0.045
EASOCENT	-0.036	-0.056	-1.301**	-0.047
WESOCENT	-0.038	-0.044	-0.078	-0.084
MOUNTAIN	0.131**	0.110***	1.341**	0.390***
MURDER	-0.013***	-0.008***	0.014	-0.011
DISSIMILARITY	0.454**	0.435***	2.514**	1.290***
CONSTANT	-2.300***	-1.153***	14.788***	0.647
OLS or Probit Result				
TRACT DENSITY	0.011	-0.014*	-0.074	0.026*
N	10592	21514	21513	21539
Number of Clusters	255	308	308	308
Uncentered R-squared	-	-	0.228	0.636
Centered R-squared	-	-	0.027	0.143
Instrument Tests				
First Stage F-Statistic	-	-	47.89	48.26
Kleibergen-Paap rk underidentification	-	-	3.36	3.37
p-value	-	-	0.6444	0.6434
Kleibergen-Paap rk weak indentification	-	-	4.37	4.37
Stock-Yogo (10%)	-	-	10.83	10.83
Hanson-J over ident. test statistic	-	-	3.61	3.24
p-value	-	-	0.4621	0.5186

Notes: The instrument tests are not valid for the IV probit specification. *,**,*** denotes statistical significance at the .10, .05, .01 level. (Standard errors)

Table 7: Tract Density and Index Measures of Social Interaction: Base Model

Specification	Base Model					
Dependent Variable	FACTOR	FACTOR2	SOCTRUST	FAITHBAS	CIVICPART	DIVERSITY
Technique	(2SLS)	(2SLS)	(2SLS)	(2SLS)	(2SLS)	(2SLS)
	Coeff.	Coeff.	Coeff.	Coeff.	Coeff.	Coeff.
TRACT DENSITY	-0.062***	-0.056***	0.033	-0.199***	0.107	0.282***
	(0.018)	(0.018)	(0.039)	(0.030)	(0.079)	(0.080)
CONSTANT	0.498***	0.451***	-0.277	1.513***	0.900	4.114***
	(0.145)	(0.143)	(0.311)	(0.245)	(0.621)	(0.644)
OLS or Probit Result						
TRACT DENSITY	-0.038**	-0.036**	-0.072***	-0.060**	0.019	0.030
N	10655	21643	21955	21839	22051	22046
Number of Clusters	254	309	311	311	311	311
Uncentered R-squared	0.977	0.977	0.977	0.977	0.977	0.977
Centered R-squared	0.145	0.145	0.144	0.145	0.144	0.144
Instrument Tests						
First Stage F-Statistic	23.09	19.21	19.35	19.33	19.12	19.08
Kleibergen-Paap rk underidentification	11.92	11.74	11.73	11.75	11.74	11.74
p-value	0.0359	0.0386	0.0387	0.0384	0.0385	0.0385
Kleibergen-Paap rk weak identification	15.91	14.35	14.52	14.68	14.42	14.40
Stock-Yogo (10%)	10.83	10.83	10.83	10.83	10.83	10.83
Hanson-J over ident. test statistic	7.23	8.08	1.99	3.88	1.40	3.86
p-value	0.1241	0.0886	0.7386	0.4228	0.8443	0.4259

Notes: *,**,*** denotes statistical significance at the .10, .05, .01 level. (Standard errors)

Table 8: Tract Density and Index Measures of Social Interaction: Full Model

Specification	Full Model					
Dependent Variable	FACTOR	FACTOR2	SOCTRUST	FAITHBAS	CIVICPART	DIVERSITY
Technique	(2SLS)	(2SLS)	(2SLS)	(2SLS)	(2SLS)	(2SLS)
	Coeff.	Coeff.	Coeff.	Coeff.	Coeff.	Coeff.
TRACT DENSITY	-0.017	-0.024	0.008	-0.098**	-0.023	0.035
	(0.025)	(0.028)	(0.013)	(0.044)	(0.053)	(0.057)
MALE	0.015	0.035***	-0.054***	-0.126***	-0.073***	-0.096***
AGE	0.008**	0.009***	0.012***	0.002	0.036***	0.056***
AGE2	-0.000	-0.000	-0.000**	0.000*	-0.000***	-0.001***
MARRD/PARTN	0.079***	0.082***	0.062***	0.059***	0.007	0.181***
KIDS6	0.054***	0.067***	-0.030***	0.043***	-0.027*	-0.062**
KIDS6_17	0.026**	0.033***	-0.001	0.079***	0.053***	0.012
INC2	-0.013	-0.024	0.022	-0.006	0.028	0.072
INC3	0.009	0.013	0.085***	0.020	0.123***	0.391***
INC4	0.036	0.018	0.130***	0.029	0.219***	0.480***
INC5	0.019	0.027	0.131***	0.032	0.247***	0.556***
INC6	0.061*	0.044*	0.162***	0.023	0.329***	0.824***
SOMECOLL	0.071***	0.074***	0.147***	0.168***	0.519***	0.938***
COLLGRAD	0.027	0.044**	0.307***	0.234***	0.910***	1.117***
BLACK	-0.158***	-0.159***	-0.497***	0.295***	0.061	-0.174***
HISPANIC	-0.122***	-0.185***	-0.400***	0.158***	-0.140***	-0.213**
ASIAN	-0.210***	-0.152***	-0.146***	0.039	-0.569***	-1.082***
UNEMP	-0.051	-0.030	-0.134***	-0.063*	-0.154***	-0.371***
HOMEMAKER	0.096**	0.065**	0.076***	0.118***	-0.085***	-0.528***
RETIRED	0.095***	0.102***	0.048**	0.067**	-0.034	-0.255***
CITIZEN	0.256***	0.236***	0.115***	0.075***	0.871***	1.302***
LIVING5	0.103***	0.108***	0.011	0.112***	0.171***	0.207***
PACIFIC	0.145**	0.080	0.056*	-0.011	0.114	0.352***
MIDATLAN	0.064	0.047	0.093**	0.012	-0.246***	-0.153
EANOCENT	0.105**	0.059	0.130***	0.044	-0.168*	-0.135
WENOCENT	0.227***	0.153***	0.120***	0.091**	0.032	0.136
SOUTHATL	0.138**	0.056	0.066*	0.096	-0.178**	-0.057
EASOCENT	0.111	0.043	0.129***	0.151	-0.188	-0.267
WESOCENT	0.146**	0.082	0.079*	0.161**	-0.286***	-0.053
MOUNTAIN	0.159***	0.117**	0.057*	-0.029	0.178*	0.360***
MURDER	-0.006*	-0.001	-0.009***	0.004	-0.010*	-0.018**
DISSIMILARITY	0.303**	0.164	-0.187***	0.278*	0.606*	0.166
CONSTANT	-0.841***	-0.721***	-0.689***	-0.129	-0.674	2.785***
OLS or Probit Result						
TRACT DENSITY	-0.002	0.003	-0.035***	-0.028***	0.035***	0.050***
N	10430	21199	21449	21373	21538	21534
Number of Clusters	252	305	308	308	308	308
Uncentered R-squared	0.059	0.062	0.982	0.982	0.982	0.982
Centered R-squared	0.059	0.062	0.328	0.329	0.328	0.328
Instrument Tests						
First Stage F-Statistic	41.27	50.34	49.35	45.85	48.21	48.43
Kleibergen-Paap rk underidentification	2.87	3.27	3.37	3.34	3.37	3.37
p-value	0.7160	0.6586	0.6436	0.6477	0.6433	0.6428
Kleibergen-Paap rk weak identification	2.85	4.39	4.33	4.41	4.37	4.37
Stock-Yogo (10%)	10.83	10.83	10.83	10.83	10.83	10.83
Hanson-J over ident. test statistic	4.99	3.93	2.61	9.80	4.39	2.07
p-value	2.8850	0.4151	0.6248	0.0440	0.3557	0.7237

Notes: *,**,*** denotes statistical significance at the .10, .05, .01 level. (Standard errors)

Index Results

Several index measures are used in the same manner as the social interaction dependent variables. All of the indexes produce a lower dimensional aspect of the social interaction data to combat one potential flaw in looking at each specific type of social interaction separately. This method reduces the number of variables of interest while keeping the existing variability, allowing for a more general definition of social interaction to be tested. The drawback for some of the indexes–SOCTRUST, FAITHBAS, CIVICPART and DIVERSITY–is they are not created from the exact 10 social interaction variables used in the rest of the study and may not exactly portray social interaction as defined here. A limitation of the other indexes, FACTOR and FACTOR2, is the assumptions that all forms of social interaction are linearly related, and that large variance implies large importance. These index measures contain several different types interaction which individually may respond differently to changes in density. The theoretical model, however, does not specify different types of social interaction, but states interaction is positively correlated with local density. The indexes are used in a similar way to the ten above estimates in the empirical model explained above.

Table 7 shows the relationship between census tract density and the social interaction indexes for the base model. The results, in general, indicate the type of social interaction matters. The FACTOR and FACTOR2 specifications consistently yield a significant negative relationship. Social interaction, when measured more generally, falls with an increase in population density. The results for the indexes from the SCCBS, however, are mostly positive. Social trust, civic participation, and diversity of friendship all have a significantly positive relationship with density. The FAITHBAS interaction index yields the remaining negative relationship. The OLS results are in line with the instrumental patterns in terms of sign, but are not as large in magnitude.

In the full model, shown in Table 8, the relationships lose magnitude and significance. The only significant result remaining in both specifications is a negative relationship between tract density and faith-based interactions. This consistent result indicates church attendance and belonging to faith-based organizations are significantly reduced with increasing local population density. As discussed above, each specification includes a series of tests for instrument strength. The tests are largely passed indicating strong instruments. The results in Tables 7 and 8 are new to the literature and supplement the earlier results.

Conclusions

This paper examines the relationship between urban sprawl and social interaction, as the result of a density externality. A theoretical model proposes individuals will not fully take into account their neighbors' social interaction when choosing lot size. This leads to a density externality and when replicated throughout the city, urban sprawl is found. The theoretical findings hinge on the assumption that social interaction and local population density are positively related. An instrumental variables approach is used to empirically estimate this relationship. The empirical evidence shows that the type of social interaction matter when determining the relationship with density. Social interaction involving friends tends to be positively related with local population density which supports urban sprawl as a result of a density externality. Social interaction with neighbors or groups largely decreases as population density increases suggesting urban sprawl is not caused by social interaction. New to the literature, index measures are used to more generally define social interaction. The empirical results yield a negative relationship between the index measures and population density.

This paper's results are important for several reasons. The relationship between social interaction and local population density is consistently estimated using exogenously determined instruments. This addresses concerns of endogeneity present in the literature. Further when measured using indexes, the relationship between social interaction and local population density suggests that urban sprawl is not caused by a density externality. This result of this paper supports the previous literature's findings. When social interaction is measured less generally, the type of social interaction is important in determining its relationship with local population density. This suggests the theory determining social interaction's effect on cities' expansion should be re-examined taking into account the different types of social interaction and how they may alter lot size choice. This might be taken one step further to allow location choice as an endogenously determined outcome within the model. This method would give insight into how social interaction, location and lot size choice interact when simultaneously optimized.

3 SOCIAL INTERACTION AND URBAN LOCATION DECISIONS

Introduction

Social interaction is a critical component of cities. The simple act of exchanging information and ideas between individuals is central to understanding why cities exist. In fact, the revival in the growth of cities over the last few decades is partially due to the improvement in social interaction of their inhabitants.[17] Social interaction is a unique aspect of cities that is fundamental for urban economic research. The effect of social interaction on location choice, however, is somewhat overlooked to this point in the literature.

This paper develops a new theoretical model to investigate the effect of social interaction on the individual's intra-urban location and housing choice. This is the first of such models to include social interaction directly into the simultaneous decisions of housing and location within a city.[18] I also preform empirical tests of the theoretical model's predictions.

I extend the standard Alonso-Muth urban consumer choice model to directly include a measure of social interaction into the individual's utility function. The derivation of the theoretical results is preformed in two stages. First, social interaction is assumed to be exogenously determined city wide; then, social interaction is endogenously determined by others' actions and allowed to vary across location. In both stages, the model yields empirically testable results which I examine using survey data on social interaction from the Social Capital Community Benchmark Survey along with census tract data for several U.S. urban areas.

[17] Glaeser and Gottlieb (2006) discuss how the reduction in crime and increased demand for interaction fueled a resurgence in urban growth.

[18] Recent theoretical work by Helsely and Zenou (2011) describes the effects of social networks and their central players on the location choice of the players involved. Additionally, Zenou (2011) presents a theoretical framework using social interaction to explain negative labor market outcomes based on the location, or distance from job, of the individual.

The exogenous social interaction model finds that location and the demand for housing are significantly altered depending on the degree of substitutability between social interaction and housing. The endogenous social interaction model's results depend upon the relationship between density and social interaction; the findings suggest either a reinforcing or offsetting effect to the exogenous model's prediction. The empirical test analyzes the effect of individuals' appetites for differing types of social interaction on the observed density across location. The empirical results offer strong support for the model's predictions.

The remainder of the paper proceeds as follows. Section 2 presents a literature review. Section 3 provides a synopsis of the Alonso-Muth model and a full explanation of how I incorporate social interaction into the framework. Section 4 empirically tests the predictions of the social interaction theoretical framework. Section 5 concludes.

Literature Review

Social capital is a concept with many definitions and applications. Economists provide a multitude of implications from social capital on economic outcomes. The issue with social capital is the lack of clarity in definition and thus measurement. The components of social capital are sometimes characterized as social norms, neighborhood effects, peer effects, obligations, or networks. These pieces are all important, yet the method of their development is, perhaps, social interaction. The actual interaction between individuals in which information, ideas, norms, and other knowledge is exchanged may be the true mechanism for the creation of social capital. Manski (2000) and Putnam (2000) charged economists to understand more about how social interaction affects outcomes and to gather more usable data on how individuals interact within their communities.

Social interaction is becoming steadily more popular in various topics within economics. The theoretical implication of social interaction is vast, for example, the understanding of capital is re-evaluated in macroeconomic models with the argument that social interaction builds different capital–human or social capital (Sobel, 2002), or the basis of how social interaction comes into existence over time (Horst, 2010) and spills over across space (Cabrales et al., 2011).

The literature shows the importance of social interaction empirically as well (Ioannides and Topa, 2010), even with the issues of econometric difficulties, measurement and lack of uniform definition (Blume and Durlauf, 2008). Social interaction affects many measurable economic outcomes: educational outcomes in the classroom and within the academic community through coauthorship (Giorgi and Pellizzari, 2011), health outcomes such as smoking (Fletcher, 2010), the prevalence of tax evasion (Fortin et al., 2004), and geographical mobility (David et al., 2010). Social interaction has even been brought into the experimental lab; the lab provides a better environment (no confounders) to measure social interaction (Hirano and Hahn, 2010). Falk and Fischbacher (2002) test how social interaction affects "crime" in the lab with the presence of social interaction and social pressure. Labor markets are also influenced by social interaction. The formation of human capital and the agglomeration of such capital is shown to increase wages (Rosenthal and Strange, 2008). Also, the way information is disseminated across individuals, through strong versus weak social ties, affects the persistence of unemployment (Zenou, 2011).

Social interaction's importance for cities is apparent immediately, cities are a direct result of individuals coming together in space. Gaspar and Glaeser (1998) discuss the future existence of cities with the advances of technology removing the need to come together in space to conduct social interaction. Social interaction affects the characteristics in terms of social and consumer amenities of cities (Glaeser and Gottlieb, 2006). The relationship between the growth of cities and social interaction is important for issues such as urban sprawl and urban redevelopment (Brueckner and Largey, 2008).

The research on theoretical implications of social interaction on urban economic outcomes is in infancy. The main application of social interaction into urban theoretical models is to allow for a central interaction location. All social interaction occurs at one point in space and there is a transportation cost to travel to the location. These costs lead to lower than optimal visits to the interaction hub and lower population density across space (Helsley and Strange, 2006). Helsley and Zenou (2011) examine the location decisions of individuals tied to social networks where again all interaction occurs at an interaction center. The model predicts individuals who are central players in the social network to reside closer to the interaction center compared to peripheral players in the social network.

This paper builds on to the urban economic modeling of social interaction in important ways. First, the individual consumes interaction at any point in space within the urban area. This removes the need for interaction to take place in one central location. Second, the model allows for the consumption of housing and the decision of location to be made simultaneous with social interaction choice. This is an advancement since location choice is not solely based on interaction value. Third, the flexibility of the model allows for the type of social interaction to matter when making location and housing consumption choices. Further, the paper provides empirical support for the predictions of the social interaction model, which is also new to the literature.

A New Theoretical Approach to Social Interactions and Urban Location Choices

This section presents a formal link of social interaction and urban location theory. I begin by introducing the basic Alonso-Muth framework, one of the canonical models in urban economics.[19] This theoretical model discusses an individual agent in a spatial world who consumes both housing and non-housing goods while simultaneously determining optimal location within a given urban area. The Alonso-Muth model leads to spatial equilibrium a central concept for urban economics. Spatial equilibrium states individuals gain no utility from relocating. Put another way, there is no other location that would make the individual happier. It is understood, given the prevalence of migrants within and across cities, that spatial equilibrium is not achieved; yet, the notion of equilibrium provides urban economics a theoretical platform to analyze how fundamental economic factors within cities determine spatial prices and land use patterns.

After presenting the canonical model, I include social interaction into the urban consumer model and examine the changes to basic urban decision making. The results of interest are the location and consumption decisions, and how those choices are influenced by the introduction of social interaction. This is a novel idea, yet, tractable. The framework yields empirically testable hypotheses regarding observable relationships between social interaction and land use patterns. Social interaction is allowed to take multiple forms in the model and, depending on the assumptions made about each form, the spatial equilibrium is altered in distinct ways.

Alonzo-Muth Model

Consider an urban area on a featureless plane. Residents travel to the central business district (CBD), a point in space where all employment takes place. The agent in this model is the utility maximizing individual, who is constrained by a budget. The individual makes choices about the level of housing and non-housing consumption, as well as, a residence location inside the urban area. The individ-

[19] The Alonso-Muth model of urban consumer theory is the first to allow for the simultaneous decision of location and consumption decisions in an urban context. See Alonso (1964), Mills (1967) and Muth (1969).

ual regularly travels along ubiquitous routes to the CBD and incurs transportation costs, T, which is increasing in distance at a non-increasing rate, and increasing at a positive, but less than identity, rate with an increase in income.[20] The individual's choices are represented as

$$\max_{x,y,k} U(x,y) \quad s.t. \quad I = P(k)x + y + T(k,I) \tag{12}$$

where x = housing consumption, y = non-housing consumption, k = distance from CBD, I = gross income, $P(k)$ = housing price function, $U(x,y)$ is a well-behaved utility function and $T(k,I)$ is a transportation cost function. The housing price function, $P(k)$, is allowed to vary across location within the urban landscape. The housing price function is not specified, however, its characteristics are deduced after equilibrium is achieved.

This model, while certainly solvable, is quite cumbersome since location is so intertwined with housing price and transportation costs. This problem is made more tractable by Muth's innovation to assign an arbitrary distance and exclude distance as a choice variable. Once this is done, the individual can make the consumption decisions optimally conditional on the arbitrary location. Lastly the individual chooses the location that yields the greatest conditional utility.[21]

Conditional on a given location, k, the individual's optimal consumption bundle is the solution to

$$\max_{x,y} U(x,y) \quad s.t. \quad I = P(k)x + y + T(k,I) \tag{13}$$

[20] In other words, it is assumed $T_k > 0,\ T_{kk} \leq 0,\ 0 < T_I < 1$. The last assumption may seem counter-intuitive: as income rises the individual can afford a faster mode of transportation. The notion of transportation costs used here, however, includes opportunity costs. In other words, the individual's time cost of traveling increases as income rises.

[21] Turnbull (1995) expands on how this simplification eases the problem and provides great insight to how both approaches are equivalent.

where k is now suppressed as a choice variable. The first order conditions for this constrained maximization problem are:

$$U_x - \lambda P(k) = 0 \tag{14}$$

$$U_y - \lambda = 0 \tag{15}$$

$$I - T(k, I) - P(k)x - y = 0. \tag{16}$$

From these conditions, the implicit Marshallian demand functions for housing and non-housing consumption, at a given location k, are defined as

$$[x(P, I^o),\ y(P, I^o)] = \arg\max[U(x, y)\ \ s.t.\ \ I^o = Px + y], \tag{17}$$

where $P = P(k)$ and $I^o = I - T(k, I)$ is income net of transportation costs. The demand functions above state, at any given k, the choices of housing and non-housing consumption are determined by the bundle $[x(P, I^o),\ y(P, I^o)]$. These demand functions help derive some useful relationships that are employed later to find the spatial equilibrium. The first relationship places the demand functions back into the budget constraint and differentiates the budget constraint with respect to price,

$$-x(P, I^o) = P(k)(\frac{\partial x}{\partial P}) + \frac{\partial y}{\partial P}. \tag{18}$$

The second relationship is found by differentiating the budget constraint with respect to net income to get[22]

$$1 = P(k)(\frac{\partial x}{\partial I^o}) + \frac{\partial y}{\partial I^o}. \tag{19}$$

These relationships are important to finding the key spatial equilibrium condition.

[22]Remember, the demand functions in equation (17) are substituted back into the budget constraint.

With the optimal choices of housing and non-housing consumption defined at any given location k, the individual chooses an optimal distance or location that yields the highest utility from consumption. In other words, the indirect utility function, found using the demand bundles, (17), is maximized by choosing location. Indirect utility, V, is defined as

$$V[P(k), I - T(k, I)] = U[x(P, I^o), y(P, I^o)]. \tag{20}$$

The indirect utility function describes the ideal consumption bundle at any location, and by doing so, the only endogenous determinate is distance. The individual's choice is

$$\max_k V[P(k), I - T(k, I)]. \tag{21}$$

Differentiate the indirect utility function with respect to k and set equal to zero to find the optimal distance,

$$V_k = V_p P_k - V_I T_k = 0. \tag{22}$$

This first order condition is simplified by substituting in for V_p and V_I which utilizes equation (20).

$$V_p = U_x \frac{\partial x}{\partial P} + U_y \frac{\partial y}{\partial P} \tag{23}$$

$$V_I = U_x \frac{\partial x}{\partial I^o} + U_y \frac{\partial y}{\partial I^o} \tag{24}$$

The next step substitutes in for U_x and U_y with information from the first order conditions in equations (14) and (15).

$$V_p = \lambda P(k) \frac{\partial x}{\partial P} + \lambda \frac{\partial y}{\partial P} \tag{25}$$

$$V_I = \lambda P(k) \frac{\partial x}{\partial I^o} + \lambda \frac{\partial y}{\partial I^o} \tag{26}$$

After gathering a λ in both equations it becomes clear how equations (18) and (19) are useful. Using those two conditions, V_p and V_I are simplified further

$$V_p = \lambda[P(k)\frac{\partial x}{\partial P} + \frac{\partial y}{\partial P}] = -\lambda x(P, I^o); \quad (27)$$
$$V_I = \lambda[P(k)\frac{\partial x}{\partial I^o} + \frac{\partial y}{\partial I^o}] = \lambda. \quad (28)$$

Now the information supplied in equation (22) reduces to

$$-\lambda x(P, I^o)P_k - \lambda T_k = 0$$

Since it is known that $\lambda > 0$,[23] the last simplification is to divide both sides by the multiplier, λ.

$$-P_k x(P, I^o) - T_k = 0. \quad (29)$$

Thus, the optimal choice of distance, k^*, is the solution to what is known in the urban literature as Muth's equation.

The spatial equilibrium outcome is driven by Muth's equation. This equation sets equal the marginal cost of distance (MCD), T_k or the additional transportation cost incurred by moving further out, with the marginal benefit of distance (MBD), $-P_k x(P, I)$ or the savings in housing price by moving another unit of distance away, where $P_k < 0$ is required for the solution to hold.[24] By setting the marginal cost equal to the marginal benefit, there is no incentive for the individual to relocate within the urban area; spatial equilibrium is achieved by locating at distance k^*. With the optimal distance given by Muth's equation, the optimal consumption bundle is found by imputing the optimal distance into the housing and

[23] This is shown by the design of the maximization problem and can be seen in equation 15 for example.

[24] This is an intuitive result–as you move further away from the center of the city the per unit cost of housing is reduced.

non-housing demand functions;

$$x^* = x(P(k^*), I - T(k^*, I)) \text{ and } y^* = y(P(k^*), I - T(k^*, I)). \qquad (30)$$

The optimal housing consumption at k^* and Muth's equation are the central relationships for this study.

The housing price function, $P(k)$, can now be fully characterized. Muth's equation requires housing prices to fall with distance, $P_k < 0$. Additionally, the housing price function is a strictly convex function of distance. Given the problem is a maximization, the second order condition, or the total derivation of Muth's equation in implicit form with respect to k, must be negative, or

$$\frac{V_{kk}}{\lambda} = -P_{kk}x(P, I^o) - T_{kk} - P_k[\frac{\partial x}{\partial P}P_k - \frac{\partial x}{\partial I^o}T_k] < 0. \qquad (31)$$

This requires that $P_{kk} > 0$ since it is known the second and third term are positive.[25]

The housing demand gradient with respect to distance is the last piece of the puzzle. In order to derive the housing gradient, I differentiate the Marshallian demand for housing by distance;

$$\begin{aligned} \frac{\partial x}{\partial k} &= \frac{\partial x}{\partial P}P_k - \frac{\partial x}{\partial I^o}T_k \\ &= P_k[\frac{\partial x}{\partial P} + x(P, I^o)\frac{\partial x}{\partial I^o}] \\ &= P_k(\frac{\partial x^h}{\partial P}) > 0, \qquad (32) \end{aligned}$$

where Muth's relationship $T_k = -P_k x(P, I^o)$ is used and $\partial x^h/\partial P$ is the compensated own price effect of housing. Therefore, the housing demand gradient is

[25] Muth's equation is $V_k/\lambda = 0$ at k^*. The second order condition is $V_{kk} < 0$ for a maximum, so that the derivative of V_k/λ with respect to k is $D = -(\partial\lambda/\partial k)V_k/\lambda^2 + V_{kk}/\lambda = V_{kk}/\lambda$, where the second equality holds since $V_k = 0$ at k^*.

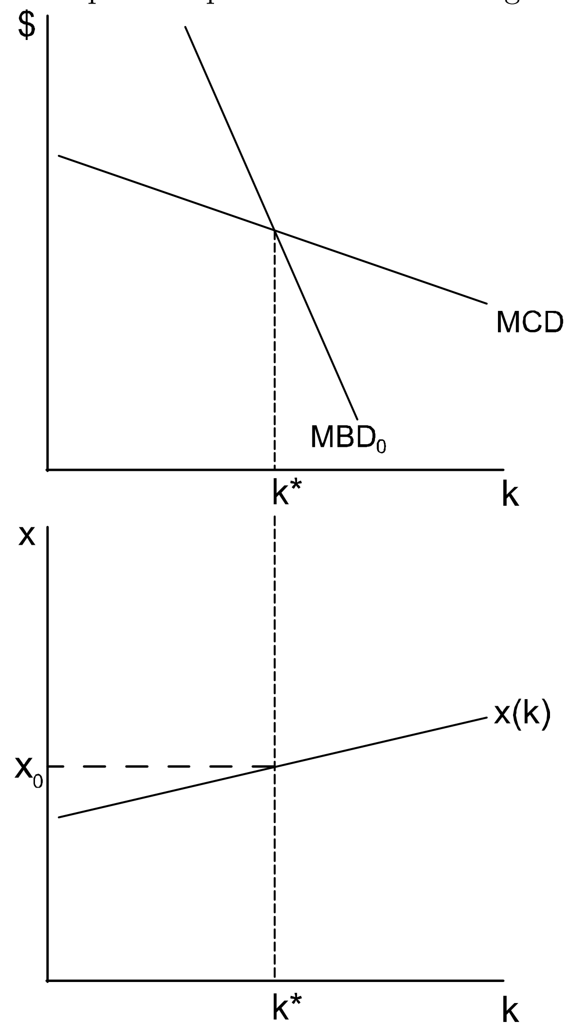
positive; housing demand increases with distance.

As shown, the optimal consumption bundle and location are all determined within the model for the individual. Following Turnbull (1995), this model is best seen in two graphs. Figure 4 shows two panels, top and bottom. The top is a representation of Muth's equation, where the MBD must cross the MCD from above since Muth's solves a maximization problem. Remember, the intersection of these two lines describes the optimal distance for an individual or the spatial equilibrium outcome–additional distance away from the CBD increases the costs more than the benefits. The bottom graph is the housing demand gradient mapped in housing–distance space. As found above in equation (32), the housing demand gradient slopes upward. Given the optimal distance, the amount of housing demanded is found by reading off the curve showing the optimal housing consumption at that distance, the bottom panel. Together, the two completely describe the urban consumers's location and housing demands.

Introducing Social Interaction into the Alonzo-Muth Model

The social interaction model extends the canonical model to incorporate socializing behaviors. The model is derived in two steps. First, social interaction enters as an exogenous parameter in the individual's utility function. This mimics an amenity for a particular city. The second step is to allow social interaction to be an endogenous choice that varies with location. The model assumes that social interaction, z, provides a benefit to the individual ($\partial U/\partial z > 0$). This additional utility may work through many channels such as information transfer or simply friendship. The model does not intend to explain why social interaction occurs; rather, the intent is to examine the behavioral responses on urban decisions when social interaction is taken into account. The set-up and notation of the model is identical to the canonical model, with the addition of a social interaction parameter, z.

Figure 4: Spatial Equilibrium and Housing Demand



Exogenous Social Interaction The individual faces a choice of housing consumption, non-housing consumption, and location. Social interaction is assumed to be exogenously determined and constant with location. The objective function faced by the individual is

$$\max_{x,y,k} U(x, y; z) \quad s.t.\ I = P(k)x + y + T(k, I). \tag{33}$$

The choice of x, y, and k are to maximize the utility function of the individual subject to a budget constraint.

In order to make the problem more tractable, I am going to optimize first with a given k value, the same technique used in the Alonso-Muth model. The simplified problem becomes

$$\max_{x,y} U(x, y; z) \quad s.t.\ I = P(k)x + y + T(k, I). \tag{34}$$

I solve for the optimal choices of housing and non-housing demand given an arbitrary distance, k. Using the Lagrangian method, the first order conditions are:

$$U_x - \lambda P(k) = 0 \tag{35}$$

$$U_y - \lambda = 0 \tag{36}$$

$$I - T(k, I) - P(k)x - y = 0. \tag{37}$$

Just as before, these conditions guide the optimal choices of housing and non-housing consumption that maximize the problem;

$$x = \phi(P(k), I - T(k, I); z) \text{ and } y = \psi(P(k), I - T(k, I); z). \tag{38}$$

Given these demand functions, I solve for the indirect utility function, V, as follows

$$V[P(k), I - T(k, I); z] = U[\phi(P(k), I - T(k, I); z),\ \psi(P(k), I - T(k, I); z); z]. \quad (39)$$

To determine what location the individual chooses, I examine the indirect utility function directly.

$$\max_k V[P(k), I - T(k, I); z] \quad (40)$$

Following the same steps as above, this gives the location equilibrium condition, or Muth's equation,

$$-P_k\phi(P(k), I - T(k, I); z) - T_k = 0. \quad (41)$$

After introducing exogenous social interaction, the solution, k^{**}, is the optimal distance where the marginal cost of distance (MCD) is equal to the marginal benefit of distance (MBD). With the optimal solution for distance, the optimal housing, x^{**}, and non-housing, y^{**}, are found by imputing k^{**} into the demand functions in equation (38).

The outcomes of interest for this model are the result of changes in the level of social interaction. The canonical model can be thought of as the social interaction model when $z = 0$. In this model, I am interested in finding the effect of introduction social interaction on location choice, k^*, and housing demand, x^*. Muth's equation is implicitly differentiated with respect to social interaction, generating the location choice comparative static,

$$\begin{aligned} 0 &= \frac{V_{kk}}{\lambda}dk + ... + \frac{\partial\phi}{\partial z}(-P_k)dz \Rightarrow \\ \frac{dk^*}{dz} &= \frac{\lambda}{V_{kk}}[P_k(\frac{\partial\phi}{\partial z})]. \end{aligned} \quad (42)$$

The sign of this relationship depends upon the sign of $\partial\phi/\partial z$.[26]

[26] It is known that $\lambda/V_{kk} < 0$ and $P_k < 0$. Thus, $sgn(dk^*/dz) = sgn(\partial\phi/\partial z)$.

The housing demand comparative static is slightly more complicated, but found similarly by implicitly differentiating the housing demand function, ϕ:

$$\begin{aligned}
\frac{d\phi}{dz} &= \frac{\partial\phi}{\partial z} + (\frac{\partial\phi}{\partial P}P_k\frac{\partial k^*}{\partial z}) - (\frac{\partial\phi}{\partial I^o}T_k\frac{\partial k^*}{\partial z}) && (43)\\
&= \frac{\partial\phi}{\partial z} + \frac{\partial k^*}{\partial z}[\frac{\partial\phi}{\partial P}P_k - \frac{\partial\phi}{\partial I^o}T_k] && (44)\\
&= \frac{\partial\phi}{\partial z} + \frac{\partial k^*}{\partial z}[(\frac{\partial\phi^h}{\partial P} - x\frac{\partial\phi}{\partial I^o})P_k - \frac{\partial\phi}{\partial I^o}T_k] && (45)\\
&= \frac{\partial\phi}{\partial z} + \frac{\partial k^*}{\partial z}[\frac{\partial\phi^h}{\partial P}P_k + \frac{\partial\phi}{\partial I^o}(-xP_k - T_k)] && (46)\\
&= \frac{\partial\phi}{\partial z} + \frac{\partial k^*}{\partial z}[\frac{\partial\phi^h}{\partial P}P_k]. && (47)
\end{aligned}$$

After combining like terms, using the Slutsky equation, and using Muth's equation, the comparative static comprises two components. First, the direct effect resulting from a change in social interaction on housing demand, $\partial\phi/\partial z$. The remaining component, $\partial k^*/\partial z[(\partial\phi^h/\partial P)P_k]$, is called the location effect. The location effect is derived from the effect of social interaction on optimal location.[27] It is known that the sign of dk^*/dz follows the sign of $\partial\phi/\partial z$; therefore, since it is also known that $P_k < 0$ and $\partial\phi^h/\partial P < 0$, the location and direct effect are always reinforcing effects. In other words, the sign of the total effect, $d\phi/dz$, follows the sign of the direct effect, $\partial\phi/\partial z$.

In order to provide intuition behind these predictions, I need to show that the sign of the housing demand comparative static, $\partial\phi/\partial z$, follows the effect of a change in social interaction on the marginal rate of substitution between housing and non-housing.

$$sgn\left(\frac{d\phi}{dz}\right) = sgn\left(\frac{dMRS_{xy}}{dz}\right)$$

This result states the degree of substitutability between social interaction and housing consumption determines the direction of the location and direct effects found

[27] See Turnbull (1995) for more discussion on this terminology.

in equation (47). In order to show this, a total differential of the first order conditions, equations (35)-(37), are used to generate the fundamental equation of comparative statics.

$$\begin{bmatrix} U_{xx} & U_{xy} & -P(k) \\ U_{yx} & U_{yy} & -1 \\ -P(k) & -1 & 0 \end{bmatrix} \begin{bmatrix} dx \\ dy \\ d\lambda \end{bmatrix} = \begin{bmatrix} -U_{xz} & \lambda & 0 \\ -U_{yz} & 0 & 0 \\ 0 & x & -1 \end{bmatrix} \begin{bmatrix} dz \\ dP \\ dI^o \end{bmatrix} \tag{48}$$

Using Cramer's rule, the sign of the housing demand comparative static is

$$sgn\left(\frac{\partial\phi}{\partial z}\right) = sgn\left(\frac{1}{|\bar{H}|}\left|\begin{bmatrix} -U_{xz} & U_{xy} & -P(k) \\ -U_{yz} & U_{yy} & -1 \\ 0 & -1 & 0 \end{bmatrix}\right|\right), \tag{49}$$

where $|\bar{H}|$ is the determinate of the bordered Hessian. Since $|\bar{H}|$ is known to be positive, it follows after simplification that

$$sgn\left(\frac{\partial\phi}{\partial z}\right) = sgn\left((-1)^6\left|\begin{bmatrix} -U_{xz} & -P(k) \\ -U_{yz} & -1 \end{bmatrix}\right|\right) \tag{50}$$

$$= sgn(U_{xz} - U_{yz}P(k)) \tag{51}$$

Now that the sign of $\partial\phi/\partial z$ is simplified, I examine the sign of $dMRS_{xy}/dz$. I use the quotient rule and that $P(k) = U_x/U_y$ (derived from the first order conditions) to find the sign of $dMRS_{xy}/dz$ as

$$sgn\left(\frac{dMRS_{xy}}{dz}\right) = sgn\left(\frac{U_{xz}U_y - U_{yz}U_x}{U_y^2}\right) \tag{52}$$

$$= sgn(U_{xz}U_y - U_{yz}U_x) \tag{53}$$

$$= sgn\left(U_{xz} - U_{yz}\frac{U_x}{U_y}\right) \tag{54}$$

$$= sgn(U_{xz} - U_{yz}P(k)). \tag{55}$$

It is seen that equation (51) and equation (55) are identical.

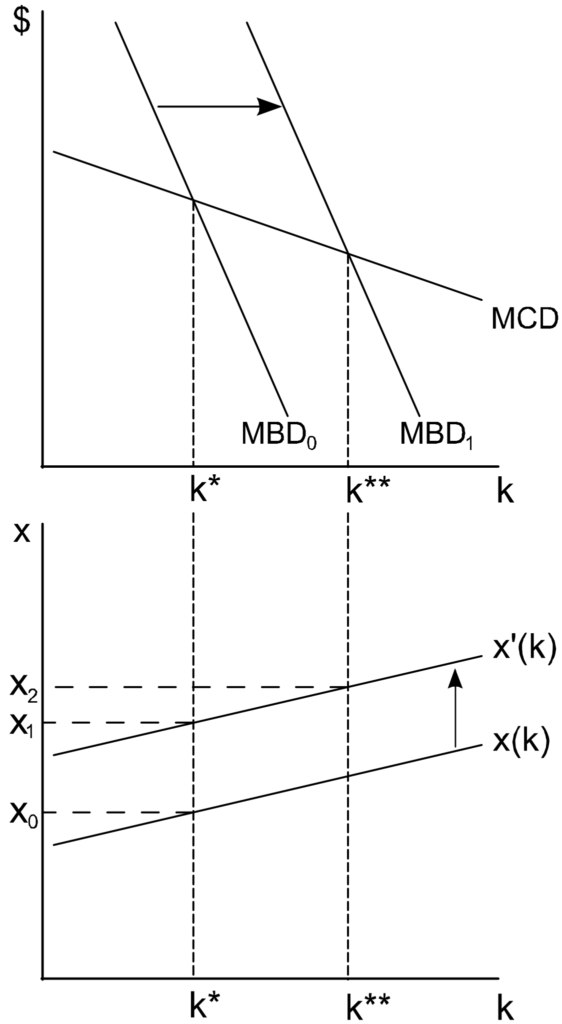
The effect of introducing exogenous social interaction on housing consumption and location choice simply depends on the degree of substitutability between social interaction and housing:

$$\frac{dx^*}{dz} \ \& \ \frac{dk^*}{dz} \gtreqless 0 \text{ as } \frac{\partial \phi}{\partial z} \gtreqless 0 \text{ or as } x, z \text{ are } \left\{ \begin{array}{c} complements \\ unrelated \\ substitutes \end{array} \right\}. \tag{56}$$

There are two interesting cases to be examined–social interaction and housing as complements or as substitutes. When social interaction and housing are complements the introduction of social interaction into the model causes the individual to move further away from the center of the city and consume more housing. This effect comprises a positive location effect, or the change in location caused by social interaction, and a positive direct effect, or the shift in housing demand caused by social interaction. Intuitively, this individual consumes social interaction and housing together, and social interaction is constant throughout the city. The individual moves away from the center of the city in order to consume more housing. Think about the family dinner parties or having friends and family visit the home as socializing and housing consumed as complements.

Figure 5 explains the introduction of social interaction as complementary to housing, Case 1. In the top panel the marginal benefit of distance is shifted to the right from an increase in housing demand. This alters the spatial equilibrium and the individual moves further out from the CBD. In the lower panel, the direct effect from introducing social interaction is represented by the housing demand gradient shifting up, a movement from x_0 to x_1. The reinforcing location effect is shown as a movement along the new housing demand gradient until spatial equilibrium is

Figure 5: Social Interaction and Housing as Complements, Case 1



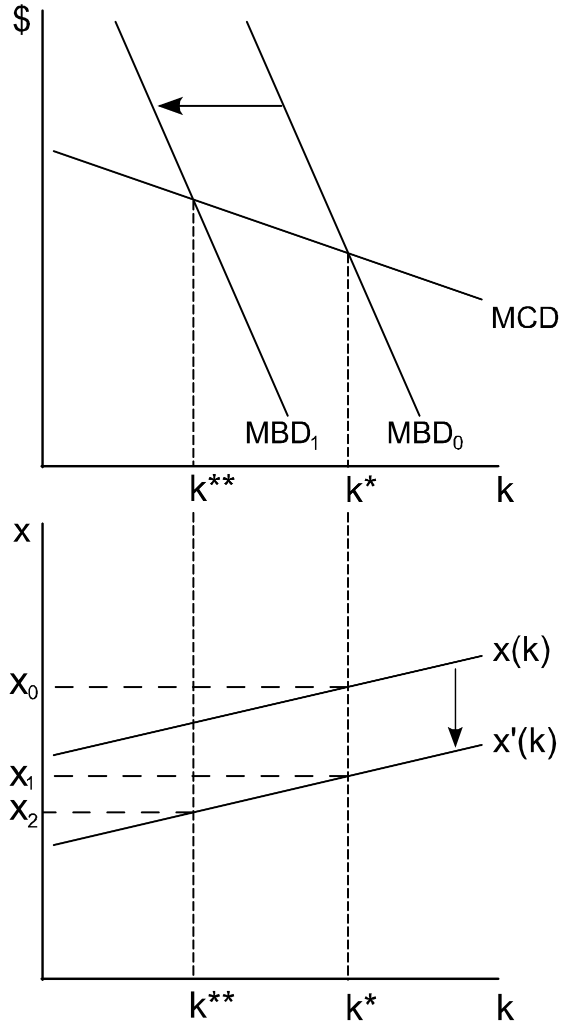
obtained, from x_1 to x_2. Figure 5 results in the individual moving further out and consuming more housing in Case 1.

When social interaction and housing are substitutes, the individual moves closer into the city and consumes less housing. The substitutes case yields a negative location and direct effect. The intuition for social interaction as a substitute for housing is the individual consumes social interaction outside the home. This may involve meeting friends for dinner at restaurants, going to bars, or joining clubs. The individual consumes less housing and instead socializes.

Figure 6 provides a graphical description of the introduction of social interaction as a substitute to housing, Case 2. The top panel shows a reduction in the marginal benefit of distance from the reduction in housing demand. This causes the individual to move closer into the city. In the bottom panel, the direct effect caused by a reduction in housing demand in response to the introduction of substitutable social interaction compounds the negative location effect. Figure 6 states the individual moves closer to the CBD and consumes less housing.

Endogenous Social Interaction While examining the effects of introducing exogenous social interaction is important, the reality is individuals choose location, housing, and social interaction together. Social interaction is interesting in that the individual does not consume interaction individually. The surroundings dictate the level of social interaction that can potentially occur. In other words, social interaction is endogenously determined along with location choice. Building on the exogenous social interaction model, the following model defines social interaction to be a function of others' choices, in particular, lot size choice. The exogenous model finds two interesting cases depending on the substitutability between housing and social interaction. For simplicity, I assume that social interaction and housing are complementary goods for the development of the endogenous social interaction

Figure 6: Social Interaction and Housing as Substitues, Case 2



model. The substitute goods case is discussed after the derivation of the model's results.

Social interaction enters the model as a constrained choice variable. The technology of social interaction is defined over the spatial structure of the neighborhood such that the constraint describes the individual's level of social interaction given the actions of a set of neighbors which reside in close proximity. The relationship between neighbors' housing choice in spatial equilibrium, $\tilde{x}$, and the individual's social interaction, $z(k)$, is not explicit in this model.

$$z(k) = f(\tilde{x}(k)) \tag{57}$$

The relationship between individual's social interaction and distance (or location) is affected in three distinct ways by others' lot size choice:

$$\frac{\partial z}{\partial k} = \frac{\partial f}{\partial x}\frac{\partial \tilde{x}}{\partial k} \Rightarrow$$

$$\frac{\partial z}{\partial k} \gtreqqless 0 \;\; as \;\; \frac{\partial f}{\partial x} \gtreqqless 0. \tag{58}$$

Since the housing demand gradient is known to increase with distance, a result seen in both the base and exogenous social interaction models, the shape of the social interaction gradient is determined by how density affects social interaction. If an increase in others' (housing choice) lot size raises (density is reduced) the social interaction enjoyed by the individual this implies social interaction increases with distance (falls with an increase in density). If an increase in neighbors' lot size reduces the social interaction of the individual, then social interaction decreases with distance (increases with an increase in density). The last possibility is simply the exogenous case; an increase in neighbors' lot size has no effect on the individual's social interaction. The endogenous social interaction model examines two important sub-cases depending on the relationship between density and social interaction.

With the sub-cases defined, the objective function faced by the individual is

$$\max_{x,y,z,k} U(x,y,z) \quad s.t.\ I = P(k)x + y + T(k,I);\ z(k) = f(\tilde{x}(k)). \tag{59}$$

The individual's choice of x, y, z and k maximize the utility function subject to a budget constraint and the social interaction constraint. In order to make the problem more tractable, I am going to optimize first with a given distance, the same technique used in the base model. The new problem becomes

$$\max_{x,y} U(x,y;z(k)) \quad s.t.\ I = P(k)x + y + T(k,I). \tag{60}$$

With location held constant, the individual finds a reaction function in response to neighbors' choices of housing. I assume individuals in neighborhoods are identical;[28] the Nash equilibrium that results is identical housing consumption choice for all individuals in the neighborhood. The Nash solution for each distance is represented by, $\hat{x}(k)$; thus, the social interaction constraint is altered to

$$\hat{z}(k) = f(\hat{x}(k)). \tag{61}$$

Now that the problem is simplified, I can solve for the optimal choices of housing and non-housing consumption given an arbitrary distance and level of social inter-

[28] This type of sorting based on social interaction can be supported by Helsely and Zenou (2011).

action. The first order conditions are

$$U_x - \lambda P(k) = 0 \tag{62}$$

$$U_y - \lambda = 0 \tag{63}$$

$$I - T(k, I) - y - P(k)x = 0. \tag{64}$$

These will guide the choice of x and y

$$x = \phi(P(k), I - T(k, I); \hat{z}(k)) \text{ and } y = \psi(P(k), I - T(k, I); \hat{z}(k)) \tag{65}$$

Remember, given the optimal solution, I can find the indirect utility function, V, as follows

$$V[P(k), I - T(k, I), \hat{z}(k)] = U[x(P(k), I - T(k, I); \hat{z}(k)),\ y(P(k), I - T(k, I); \hat{z}(k)), \hat{z}(k)]. \tag{66}$$

The individual chooses the optimal location by maximizing the indirect utility function directly.

$$\max_k V[P(k), I - T(k, I), \hat{z}(k)] \tag{67}$$

The first order condition for this maximization is

$$V_P P_k - V_I T_k + V_z \frac{\partial z}{\partial k} = 0. \tag{68}$$

The first order condition is simplified by using equations (27), (28) and that $V_z = U_z$ to find Muth's equation that defines spatial equilibrium,

$$-xP_k + \frac{U_z}{\lambda} \frac{\partial z}{\partial k} = T_k. \tag{69}$$

Muth's equation is altered from the exogenous model, but the solution to Muth's equation, k^*, is the optimal distance where the marginal cost of distance (MCD) is equal to the marginal benefit of distance (MBD). Here the MBD includes a social interaction component, $(U_z/\lambda)(\partial z/\partial k)$; the effect of this new element is determined by the sign of the social interaction gradient. The sign of the social interaction gradient follows the relationship between density and social interaction, or the assumption made on the f function in the social interaction constraint. This implies both important sub-cases need to be examined.

Muth's equation describes the spatial equilibrium; however, I still need to find the effect of endogenous social interaction on the housing demand gradient. The housing demand in Nash equilibrium is shown in implicit form as

$$\hat{x} - \phi(P(k), I - T(k); f(\hat{x})) = 0. \tag{70}$$

Using the Implicit Function Theorem, the housing demand gradient is

$$\frac{\partial \hat{x}}{\partial k} = -\frac{-\frac{\partial \phi}{\partial P} P_k + \frac{\partial \phi}{\partial I} T_k}{1 - \frac{\partial \phi}{\partial z}\frac{\partial f}{\partial \hat{x}}} \tag{71}$$

Notice, the sign follows the numerator which is the exogenous social interaction result. This is because of Lipschitz condition that others' actions alter own actions less than own actions alter own actions. In other words, the denominator is always positive. The effect on the housing demand gradient depends on the same sub-cases as the other results.

Keep in mind I have assumed that social interaction and housing are complementary goods, case 1. I examine a positive relationship between housing and social interaction as the first sub-case; in other words, sub-case A states

$$\frac{\partial z}{\partial k} > 0 \;\; by \;\; \frac{\partial f}{\partial x} > 0. \tag{72}$$

In this case, Muth's equation states the MBD shifts to the right from the previously exogenous social interaction result of spatial equilibrium, equation (69). The housing demand gradient becomes steeper than in the exogenous social interaction model, since the denominator of equation (71) is less than 1. Both of these effects lead to the individual moving further away from the center of the city, consuming more housing and consuming more social interaction.

As before this is best seen in a series of graphs, Figure 7 shows Case 1-A. There are now three panels to the figure: the top panel still describes the spatial equilibrium, the middle panel shows the housing demand gradient, and the bottom panel illustrates the social interaction gradient. The top panel shows the MBD curve shifting to the right caused by the increase in housing demand. The middle panel shows an increase in the slope of the housing demand gradient, but the exogenous model's result still lies on the new housing demand gradient. The bottom panel describes the social interaction gradient. In this case, social interaction increases with distance. Figure 7 yields the individual moves further out and consumes more housing and social interaction. In Case 1-A, the endogenous model's predictions reinforce the results of the exogenous model.

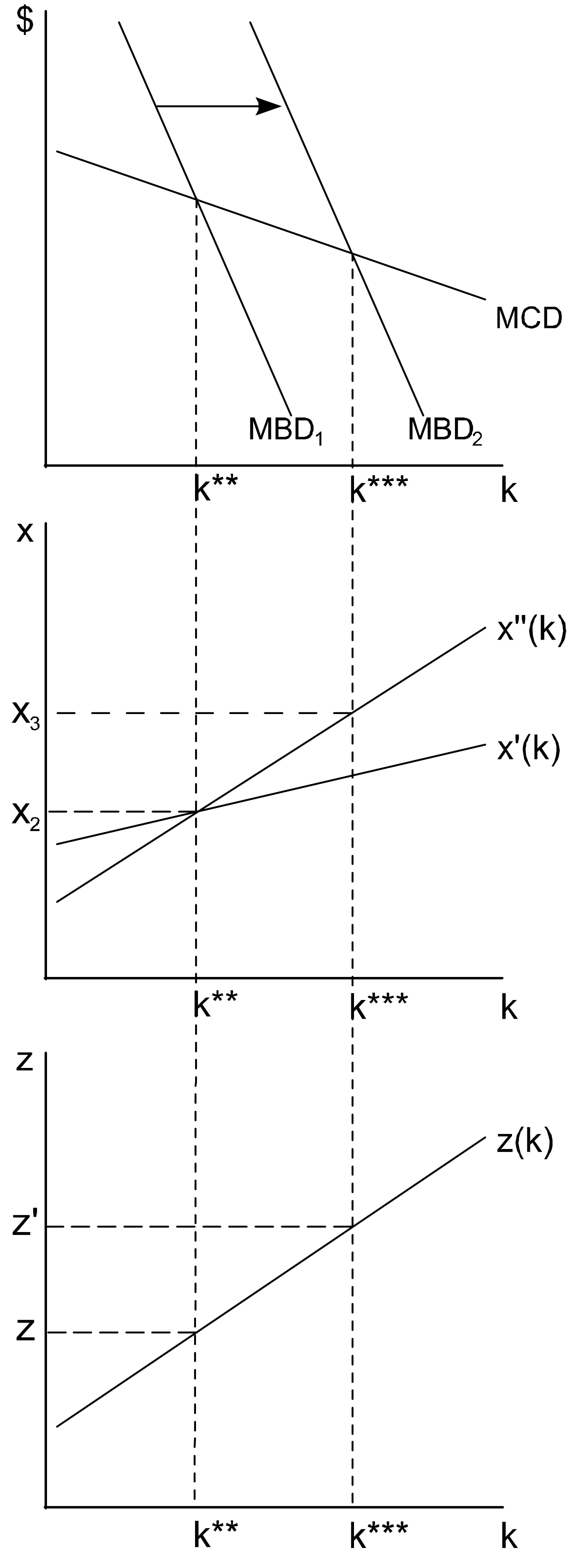
In sub-case B, social interaction and distance are negatively related.[29]

$$\frac{\partial z}{\partial k} < 0 \;\; by \;\; \frac{\partial f}{\partial x} < 0. \tag{73}$$

With this assumption, Muth's equation finds the new spatial equilibrium closer to the CBD from a shift left of the MBD curve. The housing demand gradient is flatter than in the exogenous social interaction model, and the social interaction gradient is downward slopping. Both of these effects imply the individual moves closer to the center of the city, consumes less housing and consuming more social interaction

[29] Put another way, social interaction and density are positively related–as density rises, interaction increases.

Figure 7: Social Interaction and Housing as Complements; Social Interaction Gradient Increasing, Case 1-A



compared to the exogenous case. Figure 8 shows the results for Case 1-B graphically; the results are contrasting from the exogenous model.

The remaining two possibilities deal with social interaction and housing as substitute goods, Case 2. The same two sub-cases, A and B, need to be examined. In Case 2-A, social interaction and housing are substitutes and social interaction increases with distance. In this case, the MBD curve shifts to the right, and the housing demand gradient becomes steeper. The individual moves further out from the CBD and consumes more housing and social interaction. The graphical representation is identical to Figure 7. The predictions, however, contrast the exogenous social interaction findings for Case 2-A.

In Case 2-B, social interaction and housing are still substitutes but social interaction decreases with distance. Muth's equations states the MBD curve shifts to the left and the housing demand gradient becomes flatter. This implies the individual moves closer into the city and consumes less housing but more social interaction. Figure 8 shows this graphically, but the results for Case 2-B reinforce the exogenous model's predictions.

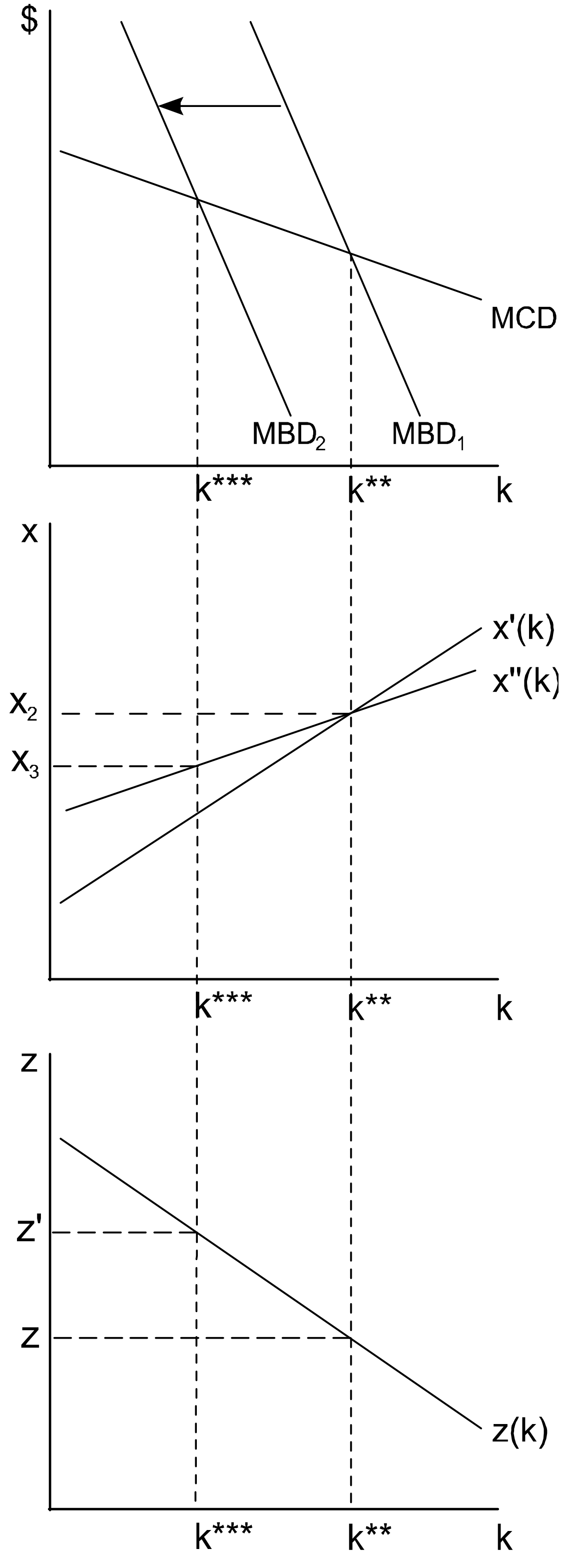
The next step is establishing, empirically, tests of these relationships. Luckily, the Social Capital Community Benchmark Survey provides data which allows me to test these predictions.

Data and Empirical Specification

I use data from the Social Capital Community Benchmark Survey (SCCBS) which took place in 2000 and is distributed by the Roper Center for Public Opinion Research at the University of Connecticut.[30] This survey was designed by the Saguaro Seminar at the John F. Kennedy School of Government, Harvard University with the intention of formulating a uniform measure of social capital. The

[30] Access to the unrestricted data can be found through www.roper.uconn.edu.

Figure 8: Social Interaction and Housing as Complements; Social Interaction Gradient Decreasing, Case 1-B



data comprises a national sample as well as 41 separate "community" samples. The national sample includes respondents who were chosen at random across the Continental U.S. with over-sampling of blacks and Hispanics.[31] As for the 41 "communities," each separate community is defined spatially different; some are attempts to capture a city, while others are entire states.[32] Because the community samples were sponsored by different institutions or individuals, the sampling techniques are not uniform across the communities.[33]

The SCCBS data include social interaction measures that describe individuals' social activity within their communities. The theoretical model proposed suggests the degree of substitutability between social interaction and housing affects location choice; however, data on individuals' preferences for social interaction and housing are not available. The SCCBS does provide information on socializing at home and away from home. These data work as a proxy for the degree of substitutability between social interaction and housing. The respondent's preference proxy is generated from two variables in the survey, $SocHome$ and $SocPublic$. The $SocHome$ variable is the frequency of having friends visit the home, while the variable $SocPublic$ is the frequency of socializing with friends in public places. These data act as a revealed preference between social interaction and housing. If the individual chooses to socialize at home more often, then the individual has revealed a preference for consuming housing and social interaction as complimentary goods. On the other hand, if the respondent prefers to interact outside the home, then housing and social interaction are consumed as substitute goods. The range for the socializing data is from zero to sixty, intuitively this implies a range of socializing

[31] Blacks and Hispanics account for 500 respondents each which resulted in an additional 288 blacks and 294 Hispanics to be included in the sample than otherwise would have been.

[32] As an example, the Atlanta "community" is defined as several counties within the Atlanta MSA where as the Indiana "community" is a random sample drawn across the entire state.

[33] There are several different sampling techniques used across the communities. One community may sample heavier in Hispanics while another oversamples Native Americans. The sponsor of the community sample may have been interested in a particular group of people and required more information from that group.

never to more than once a week.

The restricted use SCCBS provides each individual's location by the Census tract (denoted tract(s) from here forward) for which the home is located. The individual response data is then aggregated up to the tract level using the sampling weights provided. This generates a tract preference for the degree of substitutability between social interaction and housing. Since the effect on urban location decisions is desired, the sample only includes tracts that intersects at any point with an urbanized area. The Census Bureau defines an urbanized area as densely settled territory which can include all or any size portion of the census tracts that make up the urban area. Therefore, the data include tracts that lie in or significantly close to an urbanized area. This reduces the sample and brings the total observations to around 7800 tracts. These data are spread out over 276 Metropolitan Areas (MSAs) and 301 different Urbanized Areas (UAs).

The 2000 U.S. Census data are merged by tract with the SCCBS data. The Census data provides additional information at the tract level such as the number of households with children, total housing units, median age, structural age distributions and other demographics. Most importantly, the population density for each tract is supplied. This measure is used as the dependent variable in most of the specifications below; thus testing if the type of social interaction in the neighborhood desired affects population density. The population density is seen as a combination of lot size and distance choice for the individual; thus, the model's predictions are tested as the resulting relationships between both types of socializing and population density.

Additional data are used to examine other outcomes besides population density. The average time spent commuting to work data comes from the SCCBS. Straight line distance data is generated from the center of each tract to the nearest CBD. The CBD is approximated by a city hall or capital building for each urban area, and the area of each tract measured in square miles is also calculated.

Table 9 provides descriptive statistics of the data. In the sample, the average census tract's population is characterized as being 69 percent white, 17 percent black, 51 percent female, and 13 percent foreign born. The median age is just over 35 years and the median income is $49,000. The average tract has just under 2000 housing units with the vacancy rate at 6 percent, and 62 percent of units are owner occupied. Table 9 also describes the social interaction data. The average tract socializes more often at home rather than socializing in public. This gives an overall indication of how social interaction may affect city size and formation given the results of the new model. Commuting time is just under 30 minutes on average and the average distance to the CBD is around 10 miles.

In the theoretical model, housing choice and location are both altered once social interaction is included. The empirical specification uses changes in tract population density to predict both of those changes simultaneously. The relationship between population density and distance from the CBD is known to have an inverse relationship–as distance increases, population density falls. Further, housing demanded is also known to have an inverse relationship with population density–as housing demand rises (keeping amount of land available unchanged), population density decreases. Thus, the predictions of the theoretical model can be seen by realized choices of population densities from tracts with differing preferences on socializing at home or in public.

The primary empirical analysis of this paper is assessing the relationship between tract population density and two different types of social interaction activities. The general specification applied to the data is

$$Tractden_i = \beta_1 SocPublic_i + \beta_2 SocHome_i + \Phi X_i + \varepsilon_i. \tag{74}$$

Table 9: Descriptive Statistics

Variable	Mean	Standard Dev.	Minimum	Maximum	Definition
Census 2000 Data					
tractden	8.07	1.32	0.87	12.29	Natural log of the tract population density
pctwhite	0.69	0.28	0	1	Percentage of white population
pctblack	0.17	0.26	0	0.99	Percentage of black population
pctfemales	0.51	0.03	0.04	0.66	Percentage of females
pctforeign	0.13	0.14	0	0.79	Percentage of foreign born population
med_age	35.20	6.05	15.20	77.50	Median age
housingunits	1989.39	964.02	7.00	11522.00	Number of housing units
pctvacant	0.06	0.05	0	0.72	Percentage of vacant housing units
pctown	0.62	0.25	0	1	Percentage of owner occupied housing units
households	1873.13	899.63	9	11168	Number of households
pctmarr_chd	0.22	0.11	0	0.84	Percentage of married households with kids
medstructure	1964.52	15.89	1939	1999	Median age of structures
medincome	49.03	22.29	4.43	200	Median income in thousands
msaden	6.47	0.82	3.35	9.48	MSA population density
uaden	7.96	0.46	6.75	8.86	UA population density
Social Capital Community Benchmark Survey Data					
sochome	21.94	17.79	0	60	Preference for socializing at home
socpublic	16.78	16.65	0	60	Preference for socializing in public
econsat	1.09	0.53	0	2	Satisfaction of personal economic situation
tele3	0.39	0.99	0	5	Intensity of telecommuting
tvone	2.38	1.26	1	5	Television is primary form of entertainment
worktime	30.94	20.26	0	96	Average work hours per week
polknow2	2.35	1.38	1	5	Knowledge of local politics
wwwacc	0.58	0.42	0	1	Identifier for access to the internet
wwwtime	2.94	4.21	0	25	Time spend on the internet per week
relmem2	0.55	0.42	0	1	Identifier if religious member

Note: Number of observations for all variables is around 7831 varying slightly due to missing values.
All the SCCBS data only come from the respondents of the survey and are used as proxies for the tract's beliefs.

The dependent variable is the population density at the i^{th} tract. The variable $SocPublic_i$ describes the tract's desire to socialize in public, where $SocHome_i$ captures the preference of social interaction at home. A host of controls, X_i, are included in different specifications. The controls comprise racial composition, median age, gender composition, household structure (children present), median age of the housing structures, percentage of foreign born residents, and median income. I estimate equation (74) for all tracts that lie inside an urban area using ordinary least squares (OLS) framework.

Results

Table 10 presents four different specifications for determining the relationship between social interaction and population density. The first specification is the base model. The dependent variable is tract population density which is only explained by the two forms of social interaction, $SocPublic$ and $SocHome$. The relationships are exactly what is predicted by the theoretical model. As preference for interaction in public increases, so does the population density of the tract. This implies individuals move in closer to the CBD and consume less housing. The opposite effect is found between socializing at home and population density. An increase in the social interaction at home reduces the population density of the tract. This suggests individuals who prefer consuming social interaction and housing as compliments move further out and consumer more housing. The base model results are strongly statistically significant. The magnitudes may appear small; yet, as the tract desires one more public interaction over a year, the tract population density increases by 0.63 percent. At the average density, this increase corresponds to an additional 42 people per square mile. In perspective, this change is not trivial for one additional interaction at the average. The magnitude is roughly half and working in the opposite direction for an additional interaction at home. The explanatory power of the base model specification is low with an R^2 of only 0.005.

The second specification is the base model with fixed effects for Census divisions. The inclusion of these fixed effects removes unobserved regional differences from the estimated relationships. The results are again consistent with the theoretical prediction and are strongly statistically significant. The magnitude of the estimated relationship between social interaction at home and population density is increased, where the relationship for socializing in public is reduced. The resulting R^2 is much larger, 0.127.

The remaining two specifications are the plus model and the plus with fixed effects model. These models include other control variables omitted in the base models. The plus model and plus with fixed effects model results, shown in Table 10, are aligned with the theoretically predicted outcomes; however, the statistical significance is lost on the socializing at home relationship. The magnitudes for both social interaction relationships are reduced as well. The control variable results are mainly as expected. Younger residents reside in higher density areas; the presence of children in the family structure reduces the population density. The older the housing structures the greater the population density. Foreign born residents choose higher density areas, and as income increases (controlling for age and family structure) so does population density of the tract. As expected, the goodness of fit is increased with the addition of more controls: 0.557 for the plus model and 0.579 for the plus with fixed effects model.

The OLS results above suggest when there is preference for socializing at home, individuals reside in tracts that have much lower population densities, and when socializing is substitutable for housing individuals choose higher density tracts. The theoretical model's predictions do not directly state a choice of population density, so as an alternative specification other dependent variables are used. The straight-line distance between the center of the tract and the CBD, average commuting

Table 10: Effects of Different Social Interactions on Tract Population Density

Dependent Variable: Tract Population Density	(1) Base	(2) Base w/ FE	(3) Plus	(4) Plus w/ FE
Socializing in Public	0.0063***	0.0056***	0.0017**	0.0014**
	[4.162]	[4.614]	[2.209]	[2.060]
Socializing at Home	-0.0029**	-0.0034***	-0.0003	-0.0007
	[-2.593]	[-3.394]	[-0.445]	[-0.975]
Percentage of Whites†			-0.6353*	0.0294
			[-1.814]	[0.072]
Percentage of Blacks†			-0.0677	0.7472**
			[-0.188]	[2.009]
Median Age			-0.1290***	-0.1077***
			[-7.323]	[-8.285]
Median Age Squared			0.0009***	0.0007***
			[4.740]	[4.512]
Percentage of Females			7.3505***	7.6405***
			[8.962]	[9.539]
Percentage of Households with Children			-2.7872***	-2.7366***
			[-12.881]	[-10.124]
Median Structure Age			-0.0313***	-0.0296***
			[-12.991]	[-17.282]
Percentage of Foreign Born Residents			3.9099***	4.3819***
			[9.934]	[13.711]
Median Income			0.0083***	0.0073***
			[4.177]	[3.454]
Constant	8.0480***	8.0700***	69.4930***	65.0162***
	[64.520]	[92.877]	[15.833]	[19.143]
Includes Census Division Dummies	No	Yes	No	Yes
Observations	7,659	7,659	7,659	7,659
R-squared	0.005	0.127	0.557	0.579

Notes: Standard errors are clustered at the urban area (190 clusters). T-statistics in brackets.

*, **, *** denotes significance at the 10%, 5% and 1% level, respectively.

†The joint F-test is significant at the 1% level for the percentage of whites and Blacks in a tract.

Table 11: Effects of Different Social Interactions on Distance From Central Business District

Dependent Variable : Distance of Tract to CBD	(1) Base	(2) Base w/ FE	(3) Plus	(4) Plus w/ FE
Socializing in Public	-0.0133	-0.0167**	0.0037	-0.0020
	[-1.574]	[-2.391]	[0.509]	[-0.373]
Socializing at Home	-0.0037	0.0053	-0.0073	-0.0023
	[-0.586]	[0.828]	[-1.296]	[-0.434]
Percentage of Whites†			11.3102**	10.9753***
			[2.458]	[5.811]
Percentage of Blacks†			10.6193**	5.7873***
			[2.273]	[2.794]
Median Age			-0.0772	-0.1443
			[-0.718]	[-1.425]
Median Age Squared			0.0019	0.0024*
			[1.385]	[1.684]
Percentage of Females			7.3368**	6.7699*
			[2.201]	[1.713]
Percentage of Married			23.1194***	23.5264***
Households with Children			[8.282]	[10.385]
Median Structure Age			0.1107***	0.1587***
			[7.028]	[8.467]
Percentage of Vacant Units			5.1731	11.1205***
			[1.141]	[2.764]
Percentage of Foreign Born			10.6026***	-6.6896**
Residents			[3.259]	[-2.526]
Median Income			0.0163	-0.0494***
			[1.219]	[-5.076]
Constant	9.4580***	9.3191***	-228.9784***	-315.0841***
	[14.073]	[82.646]	[-7.666]	[-8.563]
Includes MSA Dummies	No	Yes	No	Yes
Observations	7659	7659	7659	7659
R-squared	0.001	0.271	0.281	0.564

Note: Standard errors are clustered at the urban area (190 clusters). T-statistics in brackets.

*, **, *** denotes significance at the 10%, 5% and 1% level, respectively.

†The joint F-test is significant at the 1% level for the percentage of whites and Blacks in a tract.

Table 12: Effects of Different Social Interactions on Average Commute Times of Census Tract

Dependent Variable : Avg. Commute Time of Tract	(1) Base	(2) Base w/ FE	(3) Plus	(4) Plus w/ FE
Socializing in Public	-0.0081	-0.0110***	-0.0050	-0.0053*
	[-1.607]	[-2.977]	[-1.289]	[-1.784]
Socializing at Home	-0.0097**	0.0010	0.0013	0.0030
	[-2.188]	[0.310]	[0.382]	[1.099]
Percentage of Whites†			2.1676	0.0468
			[0.782]	[0.027]
Percentage of Blacks†			11.6852***	5.8267***
			[3.422]	[3.096]
Median Age			0.1823	0.1960**
			[1.251]	[1.997]
Median Age Squared			-0.0017	-0.0021
			[-0.904]	[-1.545]
Percentage of Females			6.3042	6.6740*
			[1.037]	[1.673]
Percentage of Married			15.3755***	19.4652***
Households with Children			[6.283]	[7.546]
Median Structure Age			0.0079	0.0480***
			[0.389]	[4.126]
Percentage of Vacant Units			0.2815	4.8869**
			[0.089]	[2.443]
Percentage of Foreign Born			17.6908***	-1.6920
Residents			[4.387]	[-1.053]
Median Income			0.0190	-0.0578***
			[1.404]	[-4.794]
Constant	26.7376***	26.5540***	-6.6871	-78.1346***
	[38.454]	[377.629]	[-0.183]	[-3.463]
Includes MSA Dummies	No	Yes	No	Yes
Observations	7659	7659	7659	7659
R-squared	0.002	0.512	0.275	0.632

Note: Standard errors are clustered at the urban area (190 clusters). T-statistics in brackets.

*, **, *** denotes significance at the 10%, 5% and 1% level, respectively.

†The joint F-test is significant at the 1% level for the percentage of whites and Blacks in a tract.

Table 13: Effects of Different Social Interactions on Square Miles of Census Tract

Dependent Variable: Square Miles of Tract	(1) Base	(2) Base w/ FE	(3) Plus	(4) Plus w/ FE
Socializing in Public	-0.0512***	-0.0418***	-0.0190	-0.0182
	[-2.896]	[-2.766]	[-1.437]	[-1.466]
Socializing at Home	0.0358**	0.0231	0.0235	0.0168
	[2.334]	[1.430]	[1.628]	[1.129]
Percentage of Whites			-21.0659	-19.2216
			[-0.770]	[-0.723]
Percentage of Blacks			-25.1639	-21.3965
			[-0.888]	[-0.807]
Median Age			0.5947**	0.2011
			[2.181]	[0.924]
Median Age Squared			-0.0019	0.0028
			[-0.608]	[0.899]
Percentage of Females			-56.6902***	-64.9381***
			[-3.691]	[-3.686]
Percentage of Married			35.6555***	35.2218***
Households with Children			[5.717]	[5.559]
Median Structure Age			0.1691***	0.1754***
			[5.768]	[3.954]
Percentage of Vacant Units			58.3156**	53.6742**
			[2.475]	[2.439]
Percentage of Foreign Born			-37.2229	-38.4884
Residents			[-1.276]	[-1.413]
Median Income			-0.1115***	-0.0860**
			[-2.917]	[-2.038]
Constant	5.2391***	5.3596***	-298.9421***	-301.7191***
	[5.890]	[14.088]	[-6.384]	[-4.516]
Includes MSA Dummies	No	Yes	No	Yes
Observations	7659	7659	7659	7659
R-squared	0.001	0.090	0.038	0.117

Note: Standard errors are clustered at the urban area (190 clusters). T-statistics in brackets.

*, **, *** denotes significance at the 10%, 5% and 1% level, respectively.

time, and area (measured in square miles) are used as dependent variables. Table 11 shows the straight-line distance results. The only statistically significant result is a negative relationship between socializing in public and distance found in the base model with fixed effects (specification (2)). Table 12 displays the results using the average commuting times (a proxy for true distance to the CBD) for the tract as the dependent variable. The results are mostly as expected; socializing in public is negatively related with commuting time, where socializing at home leads to increased commuting times. In the base model, the relationship between socializing at home and commuting time, however, is negative and statistically significant. The relationship between social interaction and tract area is shown in Table 13. Again, the results follow expectation. The preference for socializing in public is associated with smaller tracts, and socializing with friends at home is done more often in larger tracts. The base model exhibits statistical significance for both forms of interaction, while in the base model with fixed effects, only the socializing in public result is statistically significant. The plus and plus with fixed effects models yield no statistically significant relationships for the socializing variables.

I also examine a closer look at a few particular urban areas. Table 14 shows the relationship between tract population density and both variables of interest for: Chicago, IL; Gary, IN; Grand Rapids, MI; Los Angeles, CA; and Washington D.C. Only the base and plus models are used since there is no census division variation. While not all results are statistically significant, the relationships are as the theoretical model predicts. These results suggest the location choice within urban areas also is affected by social interactions.

Instrumental Variables Approach

The causal relationship between social interaction and population density is hard to distinguish. Is social interaction a function of density or is density a function

Table 14: A Single MSA Approach on The Effects of Different Social Interactions on Tract Population Density

Dependent Variable :	Chicago, IL		Gary, IN		Grand Rapids, MI		Los Angeles,CA		Washington D.C.	
Tract Population Density	Base	Plus	Base	Plus	Base	Plus	Base	Plus	Base	Plus
Socializing in Public	0.0015	0.0028	0.0112	0.0059	0.0334***	0.0171**	0.0025	0.0020	0.0232*	0.0226**
	[0.440]	[1.252]	[1.575]	[1.012]	[3.239]	[2.516]	[0.813]	[1.007]	[1.975]	[2.270]
Socializing at Home	-0.0020	-0.0016	-0.0148**	-0.0106**	-0.0269***	-0.0110**	-0.0094***	-0.0021	-0.0012	-0.0039
	[-0.626]	[-0.764]	[-2.299]	[-2.160]	[-3.479]	[-2.002]	[-3.423]	[-1.166]	[-0.115]	[-0.446]
Percentage of Whites		-1.0433*		1.3960		-4.3459		0.5880*		-2.9941
		[-1.756]		[0.323]		[-1.250]		[1.760]		[-0.728]
Percentage of Blacks		-0.5082		2.0429		-3.1052		0.2980		-1.5391
		[-0.862]		[0.472]		[-0.952]		[0.738]		[-0.381]
Median Age		-0.1069*		-0.5891		0.0278		-0.2367***		-0.3239
		[-1.894]		[-1.294]		[0.156]		[-4.446]		[-1.303]
Median Age Squared		0.0009		0.0067		-0.0004		0.0023***		0.0033
		[1.145]		[1.071]		[-0.144]		[2.973]		[0.938]
Percentage of Females		6.1909***		-21.8761**		5.1958*		8.0356***		7.5090
		[3.481]		[-2.516]		[1.674]		[5.804]		[1.650]
Percentage of Married		-3.7436***		-17.3154***		-4.6815***		-4.1432***		-1.3163
Households w/ Children		[-8.169]		[-3.770]		[-2.801]		[-8.552]		[-0.611]
Median Structure Age		-0.0307***		-0.0365**		-0.0333***		-0.0138***		-0.0355***
		[-11.677]		[-2.678]		[-5.364]		[-4.254]		[-2.891]
Percentage of Vacant		-2.3231**		-19.1782***		-11.5069***		-7.8117***		2.0827
Units		[-2.100]		[-4.150]		[-3.692]		[-7.497]		[0.332]
Percentage of Foreign		2.9354***		5.8803		0.2512		2.9590***		2.2416
Born Residents		[5.587]		[0.997]		[0.090]		[8.228]		[0.931]
Median Income		0.0107***		0.0577*		0.0118		-0.0046*		0.0135
		[4.209]		[2.015]		[1.009]		[-1.893]		[1.558]
Constant	8.5959***	69.1668***	7.6160***	103.6538***	8.1024***	74.9509***	9.2385***	37.6817***	8.0602***	82.3720***
	[105.154]	[13.300]	[35.760]	[4.008]	[37.190]	[6.972]	[123.221]	[5.953]	[32.075]	[3.266]
Observations	541	541	54	54	80	80	389	389	54	54
R-squared	0.001	0.590	0.099	0.625	0.177	0.721	0.001	0.590	0.080	0.556

Note: T-statistics in brackets. *, **, *** denotes significance at the 10%, 5% and 1% level, respectively.

of social interaction? The answer is likely "yes" to both. This is the classic simultaneity issue. The results above state the simple correlations if simultaneity is ignored. The next step is to examine a causal relationship between location choice and social interaction. An instrumental variables approach is used to establish causality. In the first stage, both socializing in public, $SocPublic$, and socializing at home, $SocHome$, are determined by two instruments. The first instrument is an indicator of television being the primary form of entertainment for the respondent. The intuition is that if individuals spend time watching television they are not socializing with friends. The second instrument is the amount of time spent on the internet. Keep in mind, these data are from the year 2000, and while the internet was widely available in all urban areas at that time, the activities on the internet were limited to chatting with or e-mailing friends and family. The intuition is that the more socially active respondents are the more time is spent chatting or e-mailing on the internet Both instrument data come from the SCCBS and are aggregated in the same method as the interaction data.

The first stage estimations are shown as

$$SocInteraction_i = \delta_1 Television_i + \delta_2 Internet_i + \Phi X_i + \epsilon_i. \tag{75}$$

Since there are two endogenous regressors and two instruments, the first stage is exactly identified. The variable $SocInteraction$ represents either socializing in public or socializing at home, and the two instruments are $Television$ and $Internet$. The same control variables are used in plus and plus fixed effects model for the instrumental variables technique and are represented by the vector X. The second stage is then estimated as

$$Tractden_i = \beta_1 Soc\hat{P}ublic_i + \beta_2 Soc\hat{H}ome_i + \Phi X_i + \xi_i. \tag{76}$$

Table 15: An IV Approach on the Effects of Social Interactions on Tract Population Density: Base Model

Dependent Variable :	First Stage Sochome	First Stage Socpub	Second Stage Tract Density
Televison is Primary Form of	-1.1169***	-1.9966***	.
Entertainment	[-6.683]	[-11.062]	
Weekly Internet Usage	0.4368***	0.4150***	.
	[7.791]	[7.312]	
Socializing in Public	.	.	0.0860***
			[2.743]
Socializing at Home	.	.	-0.0730***
			[-3.040]
Constant	18.1930***	25.4309***	8.2429
	[37.053]	[48.522]	[45.011]
Includes MSA Dummies	No	No	No
Observations	7637	7637	7637
F-Test Statistic	60.98	100.63	.
Wald Chi-Squared Test Statistic			9.75
R-squared	0.022	0.033	.

Note: Standard errors are clustered at the urban area (190 clusters). T-statistics in brackets.

*, **, *** denotes significance at the 10%, 5% and 1% level, respectively.

Table 16: An IV Approach on the Effects of Social Interactions on Tract Population Density: Full Model

Dependent Variable :	First Stage Sochome	First Stage Socpub	Second Stage Tract Density
Televison is Primary Form of	-1.8867***	-1.0932***	.
Entertainment	[-10.23]	[-6.43]	
Weekly Internet Usage	0.4085***	0.4218***	.
	[7.13]	[7.49]	
Socializing in Public	.	.	0.0451**
			[2.471]
Socializing at Home	.	.	-0.0288**
			[-2.016]
Percentage of Whites	4.9415	3.4661	0.1750
	[1.40]	[1.04]	[0.375]
Percentage of Blacks	0.8064	0.3440	0.5377
	[0.22]	[0.10]	[1.180]
Median Age	-0.2345	-0.1604	-0.0781***
	[-1.27]	[-0.90]	[-5.192]
Median Age Squared	0.0004	-0.0013	0.0005***
	[0.20]	[-0.61]	[2.687]
Percentage of Females	2.3308	6.9408	6.3233***
	[0.32]	[1.01]	[7.319]
Percentage of Married	-5.6144*	-15.2620***	-2.8233***
Households with Children	[-1.93]	[-5.68]	[-6.470]
Median Structure Age	0.0184	0.0103	0.0282***
	[1.12]	[0.68]	[16.074]
Percentage of Foreign Born	-5.1462	2.4862	4.3261***
Residents	[-1.49]	[0.75]	[11.494]
Median Income	0.0034	0.0252**	0.0091***
	[0.26]	[2.00]	[3.668]
Constant	28.7916***	20.1605***	5.0806***
	[4.66]	[3.33]	[7.039]
Includes Regional Dummies	Yes	Yes	Yes
Observations	7637	7637	7637
F-Test Statistic	14.75	12.96	.
Wald Chi-Squared Test Statistic			2194.62
R-squared	0.042	0.036	.

Note: Standard errors are clustered at the urban area (190 clusters). T-statistics in brackets.

*, **, *** denotes significance at the 10%, 5% and 1% level, respectively.

Table 15 shows the first and second stage results for the base model. The instruments are statistically significantly related to social interaction in the expected ways. Television as the primary form of entertainment reduces the socializing behaviors, where internet usage increases both socializing at home and socializing in public. The second stage results are as the theoretical model predicts; tract density is positively related with socializing in public and negatively related with socializing at home. The magnitudes are much larger than in the OLS base model, shown in Table 10. The results suggest an increase in socializing in public once more during a year leads to an increase in tract population density of 8.6 percent At the average population density this implies an increase of 572 people per square mile. The socializing at home result indicates a reduction in population density by 486 people. Table 16 shows the full model results.

Conclusions

This paper develops a theoretical framework that examines the effect of introducing exogenous and endogenous forms of social interaction into the choices of housing and location. The exogenous social interaction model results suggest the degree of substitutability between social interaction and housing determines the effect on location and housing choice. The endogenous model builds upon this and finds either reinforcing or contracting effects depending on the relationship between social interaction and density. I then test these predictions using simple empirical methods which provide support for the theoretical results.

The implications of this paper, along with the growing literature on social interaction, is that informal interaction or face-to-face meetings among friends, neighbors, and family are important in determining the shape and size of urban areas. Social interaction, whether compliments or substitutes to housing and whether increasing or decreasing with density, does effect the location and housing choices.

4 CONCLUSION

This dissertation consists of two essays. The first essay examines the relationship between social interaction and population density. I recognize the endogeneity concerns with social interaction and population density and use instrumental variables to get causal inferences. The instrumental variables include geological data on seismic hazard, landslide hazard, and the presence of sedimentary rock beneath the soil. I use this exogenous variation in the geological data to predict changes in local population density through impacts on building high density structures. The instruments allow me to examine the causal link between increases in local population density and social interaction. The results suggest social interaction with friends increases with local population density where interaction with neighbors declines with increases in density. Group involvement is seen to be unaffected by changes in density. Index measures of social interaction show various results implying the type of social interaction is important for determining the relationship with density.

The second chapter looks more closely at the theoretical approach to understanding how social interaction affects the urban consumer. I add social interaction into the Alonso-Muth model, a standard theoretical model in urban economics. The new model produces several cases and sub-cases depending on the type of social interaction that is being portrayed. The model is flexible in that social interaction can be exogenously or endogenously determined, increase or decrease with density, or be a complement or substitute with housing consumption. The main finding of the model is that when social interaction is consumed as a substitute for housing, individuals move closer in to the center of the city and consume less housing; whereas, when social interaction and housing are complement goods, individuals will move further away from the center of the city and consume more housing.

Lastly, I test these predictions of the model with empirical methods and data. It is shown that the individual who reveals a preference for social interaction inside the home lives in lower density areas of the city (an implication of moving further out and consuming more housing). When the individual prefers to interact outside the home he lives in higher density areas of the city (an implication of moving closer in and consuming less housing). The results are in support of the model's predictions.

These essays have wide-ranging implications in urban economics. Understanding how social interaction occurs is fairly new to the literature, and mapping the differences in social interaction preferences across density is valuable. If social interaction is the mechanism for dissemination of information on potential job opportunities, or for crime deterrents in neighborhoods through a neighborhood watch group, then examining what the relationship between the density and interaction level becomes important. Further, policy makers and urban planners are interested in how urban areas grow and evolve. As the continual wave of technology advances the types of social interaction and method of social interaction may change. Therefore, the theoretical understanding of how individuals will relocate and alter their housing demand preferences as changes in social interaction begin to take place is quite important when developing urban renewal or revitalization plans. All in all, the studies presented in this dissertation aim at informing the significance social interaction has on economic decision making in an urban environment.

REFERENCES

Alonso, W., (1964) *Location and Land Use*. Massachusetts: Harvard University Press.

Anas, A., & Pines, D. (2008). Anti-sprawl Policies in a System of Congested Cities. *Regional Science and Urban Economics*, 38(5), 408-423.

Anas, A., & Rhee, H.-J. (2007). When Are Urban Growth Boundaries Not Second-Best Policies to Congestion Tolls? *Journal of Urban Economics*, 61(2), 263-286.

Angrist, J. D., & Pischke, J.-S. (2008). *Mostly harmless econometrics: an empiricist's companion*: Princeton University Press.

Bayer, P., Hjalmarsson, R., & Pozen, D. (2009). Building Criminal Capital behind Bars: Peer Effects in Juvenile Corrections. *Quarterly Journal of Economics*, 124(1), 105-147.

Blume, L. E., & Durlauf, S. N., (2008). *New Palgrave Dictionary of Economics-Empirics of Social Interaction*. London: Palgrave MacMillan.

Brueckner, J. K. (2000). Urban Sprawl: Diagnosis and Remedies. *International Regional Science Review*, 23(2), 160-171.

Brueckner, J. K. (2001). Urban Sprawl: Lessons from Urban Economics. *Brookings-Wharton Papers on Urban Affairs*, 65-89.

Brueckner, J. K. (2005). Transport Subsidies, System Choice, and Urban Sprawl. *Regional Science and Urban Economics*, 35(6), 715-733.

Brueckner, J. K., & Fansler, D. A. (1983). The Economics of Urban Sprawl: Theory and Evidence on the Spatial Sizes of Cities. *Review of Economics and Statistics*, 65(3), 479-482.

Brueckner, J. K., & Kim, H.-A. (2003). Urban Sprawl and the Property Tax. *International Tax and Public Finance*, 10(1), 5-23.

Brueckner, J. K., & Largey, A. G. (2008). Social Interaction and Urban Sprawl. *Journal of Urban Economics*, 64(1), 18-34.

Burchfield, M., Overman, H. G., Puga, D., & Turner, M. A. (2006). Causes of Sprawl: A Portrait from Space. *Quarterly Journal of Economics*, 121(2), 587-633.

Cabrales, A., Calvó-Armengol, A., & Zenou, Y., (2011). Social Interactions and Spillovers. *Games and Economic Behavior*, 72(2), 339-360.

David, Q., Janiak, A., & Wasmer, E., (2010). Local Social Capital and Geographical Mobility. *Journal of Urban Economics*, 68(1), 191-204.

De Giorgi, G., & Pellizzari, M., (2011). Understanding Social Interactions: Evidence from the Classroom. *IZA Discussion Paper*, No. 5624.

Eid, J., Overman, H. G., Puga, D., & Turner, M. A. (2008). Fat City: Questioning the Relationship between Urban Sprawl and Obesity. *Journal of Urban Economics*, 63(2), 385-404.

Falk, A., & Fischbacher, U., (2002). "Crime" in the Lab-Detecting Social Interaction. *European Economic Review*, 46(3), 859-869.

Fletcher, J. M., (2010). Social Interactions and Smoking: Evidence Using Multiple Student Cohorts, Instrumental Variables, and School Fixed Effects. *Health Economics*, 19(2), 466-484.

Fortin, B., Lacroix, G., & Villeval, M., (2007). Tax Evasion and Social Interactions. *Journal of Public Economics*, 91(11-12), 2089-2112.

Gaspar, J., & Glaeser, E. L. (1998). Information Technology and the Future of Cities. *Journal of Urban Economics*, 43(1), 136-156.

Glaeser, E. L., & Gottlieb, J. D. (2006). Urban Resurgence and the Consumer City. *Urban Studies*, 43(8), 1275-1299.

Glaeser, E. L., & Kahn, M. E. (2003). Sprawl and Urban Growth. *National Bureau of Economic Research, Inc, NBER Working Papers*: 9733.

Helsley, R. W., & Strange, W. C., (2007). Urban Interactions and Spatial Structure. *Journal of Economic Geography*, 7(2), 119-138.

Helsley, R. W., & Zenou, Y. (2011). Social Networks and Interactions in Cities. *IZA Discussion Paper:* 5506.

Hirano, K., & Hahn, J., (2010). Design of Randomized Experiments to Measure Social Interaction Effects. *Economic Letters*, 106(1), 51-53.

Horst, U., (2010). Dynamic Systems of Social Interactions. *Journal of Economic Behavior & Organization*, 73(1), 158-170.

Ioannides, Y., & Topa, G., (2010). Neighborhood Effects: Accomplishments and Looking Beyond Them. *Journal of Regional Science*, 50(1), 343-362.

Irwin, E. G., & Bockstael, N. E. (2004). Land Use Externalities, Open Space Preservation, and Urban Sprawl. *Regional Science and Urban Economics*, 34(6), 705-725.

Li, M. M., & Brown, H. J. (1980). Micro-Neighborhood Externalities and Hedonic Housing Prices. *Land Economics*, 56(2), 125-141.

Manski, C. F. (2000). Economic Analysis of Social Interactions. *Journal of Economic Perspectives*, 14(3), 115-136.

Marshall, A. (1890). *Principles of Economics.* London: Macmillan and Co., Ltd.

Mills, E., (1967). Transportation and Patterns of Urban Development: An Aggregative Model of Resource Allocation in a Metropolitan Area. *American Economic Review: Papers and Proceedings*, 57(2)197-210.

Muth, R. F., (1969) *Cities and Housing: The Spatial Pattern of Urban Residential Land Use.* Chicago: University of Chicago Press.

Nechyba, T. J., & Walsh, R. P. (2004). Urban Sprawl. *Journal of Economic Perspectives*, 18(4), 177-200.

O'Sullivan, A. (2007). *Urban economics*: McGraw-Hill/Irwin.

Plantinga, A. J., & Bernell, S. (2007). The Association between Urban Sprawl and Obesity: Is It a Two-Way Street? *Journal of Regional Science*, 47(5), 857-879.

Putnam, Robert D. (2000). *Bowling Alone: The Collapse and Revival of American Community.* New York: Simon & Schuster.

Rosenthal, S. S., & Strange, W. C. (2008). The attenuation of human capital spillovers. *Journal of Urban Economics*, 64(2), 373-389.

Sobel, J., (2002). Can We Trust Social Capital? *Journal of Economic Literature*, 40(1), 139-154.

Song, Y., & Zenou, Y. (2006). Property Tax and Urban Sprawl: Theory and Implications for US Cities. *Journal of Urban Economics*, 60(3), 519-534.

Squires, G. D. e. (2002). *Urban sprawl: Causes, consequences and policy responses*: Urban Institute Press.

Stock, J. H., & Yogo, M. (2005). Testing for Weak Instruments in IV Regression. In *Identification and Inference for Econometric Models: A Festschrift in Honor of Thomas Rothenberg*, D. K. W. Andrews & J. H. Stock (eds.). (pp. 80-108): Cambridge University Press.

Su, Q., & DeSalvo, J. S. (2008). The Effect of Transportation Subsidies on Urban Sprawl. *Journal of Regional Science*, 48(3), 567-594.

Turnbull, G. K., (1995). *Urban Consumer Theory.* Washington D.C.: The Urban Institute Press.

Zenou, Y., (2011). Spatial versus Social Mismatch: The Strength of Weak Ties. *CEPR Discussion Paper,* No. DP8244.

VITA

Zackary Bartholomew Hawley was born July 26, 1984 in Lithonia, Georgia, USA. He holds a Bachelor of Science in Economics (Minor: Mathematics) from Georgia State University and a Master of Arts in Economics from Georgia State University.

Zack began Georgia State's Doctoral program in 2006 to study Urban Economics. He has worked as a graduate research assistant for Dr. Jonathan C. Rork and as a research associate for the Fiscal Research Center at Georgia State University. He was the sole instructor of Intermediate Microeconomics (Spring 2011).

While at Georgia State Zack has been the recipient of several internal and external awards. He held the Andrew Young School of Policy Studies Dean's Fellowship from 2006-2009. In 2009 he received the George Malanos Scholarship. In 2010 Zack received the Dan Sweat Dissertation Fellowship and the E. D. Dunn Fellowship. In 2011 he received the AREUEA Doctoral Session Travel Grant and the Lindau Young Economist distinction with an invitation to attend the 4th Meeting of Nobel Laureates in Economic Sciences.

Zack has also received external funding for his research while at Georgia State. In 2009 he and his advisor received a grant from the Kauffman Foundation to study the impacts of state sponsored scholarship programs on interstate brain drain.

Zack's research has been published in the *Journal of Urban Economics* and the *Journal of Housing Economics*. He has presented his research at numerous universities and at conferences held by the Southern Economics Association, the Western Economic Association, and the American Real Estate and Urban Economics Association.

Zack received his Doctor of Philosophy degree in Economics from Georgia State University in May, 2012. He accepted an assistant professorship at Texas Christian University.