Title
Social learning in labor markets and in real estate brokerage

Author
Gathright, Graton Marshal Randal

Publication Date
2010

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Social Learning in Labor Markets and in Real Estate Brokerage

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Economics

by

Graton Marshal Randal Gathright

Committee in charge:

Professor Joel Watson, Chair
Professor Gordon Dahl
Professor Silke Januszewski Forbes
Professor James Fowler
Professor Dominique Lauga

2010

The dissertation of Graton Marshal Randal Gathright is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2010

DEDICATION

To V, A, and A.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I am grateful for the dedicated support and guidance of Joel Watson and Gordon Dahl. I am also grateful to Grayson Calhoon, Sam Dastrup, Silke Forbes, Nora Gordon, David Miller, and Christopher Wignall for support and advice throughout the research process.

I especially thank Christopher Wignall, coauthor of Chapter 1. It is with his permission that I include our research in this dissertation.

VITA

2001	B. S. in Economics, Brigham Young University
2005	M. S. in Economics, University of California, San Diego
2010	Ph. D. in Economics, University of California, San Diego

ABSTRACT OF THE DISSERTATION

Social Learning in Labor Markets and in Real Estate Brokerage

by

Graton Marshal Randal Gathright

Doctor of Philosophy in Economics

University of California, San Diego, 2010

Professor Joel Watson, Chair

This dissertation presents three studies of social learning in economic decisions. In the first chapter, I present an analysis of social learning among home sellers as they select real estate agents to list their homes. We are able to identify this social effect by exploiting a natural experiment that arises from the manner in which members of the Mormon Church are assigned to congregations. We argue that the assignment to congregations is essentially random conditional on observed geography. Using real estate transaction data from the largest multiple listing service in Utah, we find that the average home seller is almost twice as likely to choose the same real estate agent as his neighbor if they are assigned to the same congregation. We also present evidence that some of the social learning in this setting is due to word-of-mouth communication and not simply observation, a distinction with important welfare considerations. We show that home sellers respond more to the real estate transaction outcomes of neighbors who reside in the same congregation than to the outcomes of similar neighbors in other congregations.

In the second chapter, I investigate the hypothesis that informal job information networks exist among residential neighbors. Using data that includes

information on the timing of residence and job transitions, I investigate whether correlation in workplace and residential location can be interpreted as evidence of social learning in neighborhoods about job opportunities. I find some evidence of a neighborhood peer effect in job choice that persists even in specifications in which reverse-causality can be ruled-out. However, I also find evidence of sorting across micro-neighborhoods on employment-related characteristics.

In the third chapter, I present a simple model in which real estate transactions are more likely to feature high quality listing agents when selling agents are in a position to learn about the quality of listing agents prior to advising buyers. I test this prediction using multiple listing service data and county records data. I employ instrumental variables to identify the relationship between measures of listing agent quality and measures of the network connection between the two agents in the network of past transaction relationships.

Chapter 1

Word-of-Mouth Learning in Social Networks

1.1 Introduction

When an individual chooses among options with unknown payoffs, she can often achieve a better expected payoff by first gathering information from peers who have chosen from the same set of options. A significant identification problem is endemic to studying such peer effects: unobserved characteristics that influence behavior may also influence which relationships form (see Manski (1995)). If some omitted variable leads two people to make similar decisions and also increases the probability that they become peers, then estimates of the peer effect will be biased upward. For example, if people with tastes for risky behavior are more likely to smoke and also tend to be friends with other risk lovers, then estimates of peer influence on smoking that fail to account for risk preferences will exhibit a positive omitted variable bias.

In this paper, we investigate social learning by home owners about the quality of real estate agents as the home owners choose agents to list their homes for sale. The social networks that we investigate are congregations of The Church of Jesus Christ of Latter-day Saints (Mormon). These congregations, called *wards*, are defined geographically in a manner such that, conditional on geography, the assignment of homes to wards does not suffer from an unobserved selection process. As a result, once we control for the geographic selection of homes into wards, we can treat the assignment of residents to wards as essentially random, and we can identify the effect of social learning on the choice of real estate agents by home sellers.

Researchers have employed a variety of approaches to estimating social effects in the presence of omitted variables. For example, Duflo and Saez (2003) treat employees of academic departments at a large university as peers and use a randomization experiment to evaluate peer effects on attendance at a retirement benefits information fair. Sorensen (2006) looks at health plan choice by employees within academic departments of the University of California system. He uses the panel structure of his data to account for the unobserved heterogeneity between departments. Bayer et al (2008) and Hellerstein et al (2008) treat census blocks and tracts, respectively, as social networks where peers may learn about job opportunities.

Each Mormon ward is defined by a set of geographic boundaries, and each church member is assigned to the ward in which he resides. The exogeneity of ward assignment to real estate agent choice arises from the process by which geographic ward boundaries are specified. In localities where the concentration of church members is high, the process of ward boundary specification produces wards that typically enclose a small geographic area and whose boundaries are not coincident with significant neighborhood boundaries (such as major roads or subdivision boundaries). We focus on Utah County, Utah were the concentration of Mormons is approximately 89%[1]. Consequently, in our sample, a typical home owner will have a set of geographically close neighbors who are fairly homogeneous, some in her ward and some in other wards[2].

We measure the influence of a home owner's peers on his choice of real estate agent. We find that the average home seller is almost twice as likely to

[1] From the Religious Congregations and Membership in the United States, 2000, collected by Association of Statisticians of American Religious Bodies.

[2] The boundary discontinuity approach here is similar to the use of school district boundaries by Black (1999) to identify the value of public elementary schools to home owners.

choose the same real estate agent as a neighbor when they are both assigned to the same ward. The importance of the ward social network for real estate agent choice can be described in terms of the change in geography that will offset a ward relationship. For example, to be as influential as a ward neighbor that is 400 feet away, a neighbor assigned to a different ward must be 30% closer.

We also present evidence that home sellers respond to peers' private information about the quality of real estate agents, suggesting that at least some of the social learning that we find arises from word-of-mouth communication rather than from simply observing peers' behavior. This distinction has important welfare implications since pure observational learning faces a higher probability of an information cascade and inefficient herding[3]. Furthermore, direct communication between consumers concerning personal experience with real estate agents can provide reputational incentives to agents to please each client. These incentives may mitigate possible agency problems in real estate brokerage (see Levitt and Syverson, 2005).

In the next section, we outline our conceptual framework and predictions. In Section 3, we present background detail on both real estate brokerage and Mormon wards, and we describe the data that we employ. In Section 4, we discuss our approach to estimation and identification. We present evidence of social learning in Section 5. In Section 6 we present evidence of social learning in wards through direct communication. Section 7 concludes.

1.2 Conceptual Framework

Our objective is to evaluate how individuals are influenced by members of their social network in selecting a real estate agent to help sell a home. If a home seller learns about real estate agent quality through her social network, then her choice of real estate agent is more likely to be influenced by the choices of neighbors who belong to her social network.

This prediction can arise from both observational learning and direct com-

[3]For an overview of the literature on information cascades, see Bikhchandani et al (1998)

munication. In the case of purely observational learning, a home seller may notice the real estate agent choice of a neighbor and infer that the peer has private information that the chosen real estate agent is a high quality agent. If the home seller can more easily observe the choice of neighbors who belong to her social network, then she is more likely to choose the same agent as a neighbor if that neighbor belongs to her social network.

Social learning about the quality of real estate agents may also arise through direct communication between peers about personal experience with real estate agents. Depending on the content of the reports from neighbors about real estate agent quality, direct communication may increase or decrease the likelihood that a home seller chooses the same agent as a peer. If such reports tend to be positive, then, on average, home sellers will be more likely to choose the same real estate agent as neighbors who belong to the same social network.

The direct communication hypothesis provides a second prediction. If social learning occurs through direct communication about personal experience with agents, then a home seller is more likely to choose the same agent as a neighbor when the neighbor's experience with the agent was positive. The home seller is less likely to choose the same agent as a neighbor whose experience with his agent was negative. If direct communication is more likely to occur between neighbors who belong to the same social network, then the effect of a neighbor's experience on the home seller's choice of agent will be stronger if they belong to the same social network.

1.3 Background and Data

Real Estate Brokerage

Nationwide, most home sellers employ a real estate agent to list their home. The contract between a home seller and her real estate agent is called a listing agreement. These contracts typically stipulate that the real estate agent will market the home in exchange for a payment, due at closing, that is expressed as a percentage of the sales price.

Virtually all real estate agents who list homes in Utah County belong to the only multiple listings service operating in the county, the Wasatch Front Regional Multiple Listings Service (WFRMLS). WFRMLS requires that its member agents add their new listings to the WFRMLS database within 72 hours of signing a listing agreement. We use data on all listings in the WFRMLS database of single family residences in Utah County from 1997-2006.

The data from WFRMLS for each listing include home characteristics (square footage, number of bedrooms, street address, etc.) and identifying information for the agents involved in the transaction. Each record also includes the date the property was listed and the asking price. For properties that resulted in a sale, we also have the sales price.

Because of the large number of listings in our sample, it is not computationally feasible to evaluate the relationship between every pair of listings. We limit our attention to pairs of listings that are located within one quarter mile and listed within five years of each other, and we call such pairs *neighbors*.

We employ several measures of geographic location of listings to account for the spatial relationship between properties. Based on street address, we place each listing on a map and calculate the distance between each pair of neighbors. Second, using geographic data from the Utah County Department of Information Systems, we determine whether each pair of homes is assigned to the same county-defined neighborhood. The county's neighborhood definitions correspond to contiguous parcels of land that were developed contemporaneously. Finally, we use data from the U.S. Census Bureau (TigerLine) to determine whether neighbors belong to the same census block.

In Table 1.1, we present summary statistics on the characteristics of the homes in our sample of listings. The mean list price in our sample is \$216,065. Fifty-nine percent of the listings result in a sale, and the mean sales price is \$192,833.

Mormon Wards as Social Networks

Mormon wards are well-suited as a setting for investigating peer effects. Wards are important social networks to those who belong to them, and the assignment of neighboring church members to wards is essentially random, conditional on the spatial relationship between homes.

Regular participation in one's assigned ward involves frequent personal interaction with co-congregants. Since the Mormon Church has no paid clergy at the ward level, the wide array of leadership, teaching and other position are performed by individual lay members. For example, all adults are assigned a list of families that they are expected to visit on at least a monthly basis. According to a 2008 Survey by the Pew Research Center[4], 75% of (self-reported) Mormons attend religious services at least once a week and 92% of them are formal members of their congregations (wards). In addition, 77% participate at least monthly in non-worship activities at church, including 63% participating in social activities at church at least monthly.

There are at least two important reasons that, virtually without exception, practicing Mormons participate in the ward to which they are assigned. First, as mentioned above, the vast majority of church responsibilities are fulfilled by individual members. A church member is not typically eligible to perform any of these duties in a ward to which she is not assigned, and holding such a position is a hallmark of full fellowship. Another reason for participation in the assigned ward stems from the two levels of church worship in Mormon Theology. The first and most basic form of church worship is the weekly Sunday service, held in local chapels and open to the public. Each ward has its own set of meetings that are managed by the ward members and leaders. The second type of worship occurs in Mormon temples. Participation in temple worship is limited to members that are in good standing and approved by their ward leaders. One of the requirements for good standing is regular participation in the Sunday services of the ward to which they are assigned. It would be difficult to overstate the importance of temple worship in Mormon theology.

[4]US Religious Landscape Survey, Pew Forum on Religion and Public Life

The geographic boudaries of wards are designed to include 300-500 members. In this paper, we focus our attention on Utah County, where approximately 89% of the population is Mormon. This concentration of church members leads to ward boundaries that enclose very small geographic areas, smaller than most subdivisions. For this reason, ward boundaries are not typically coincident with subdivision boundaries and major thoroughfares. Figure 1.1 illustrates the assignment of parcels to wards for a small region in Utah County.

A second reason ward boundaries are unlikely to coincide with important discontinuities in the spatial distribution of homes is that homogeneity across neighboring wards is an explicit objective of church leaders involved in the specification of ward boundaries[5]. In practice, this means that ward boundaries are likely to cut across neighborhoods. As an illustration, note that the clusters of very small parcels in Figure 1.1 are townhouses. The townhouses complexes are split and combined with neighboring detached residences to form wards similar in mix of property type.

We are able to determine the ward assignment of each property in our sample using the Church's online ward assignment lookup tool[6]. In Table 1.2, we present summary statistics on neighbors of a typical listing. The average listing has 36 neighbors in the same ward and 42 neighbors assigned to a different ward. Neighbors in different wards are, on average 40% farther away than neighbors assigned to the same ward. Without conditioning on spatial relationship, a pair of neighbors assigned to the same ward is almost three times more likely to choose the same real estate agent as two neighbors in different wards.

In Table 1.3, we present summary statistics on on wards and real estate agents. The homes in each ward are listed by a variety of real estate agents, suggesting that real estate agents do not specialize in particular wards.

The distribution of agent activity is highly skewed. More than half the agents in our sample list only one or two homes. Agents who listed more than two homes listed on average 21 homes.

[5]Based on the authors' private conversations with church leaders.
[6]Meetinghouse Locator, www.lds.org.

1.4 Estimation and Identification

Empirical Approach

To evaluate the influence of ward peers on real estate agent choice, we estimate the probability that a home seller chooses the same real estate agent as her neighbor. If ward members learn from each other, then the probability that neighbors choose the same real estate agent should differ depending on whether they are assigned to the same ward. We begin by estimating the following linear probability model:

$$P(i \textit{ and } j \textit{ choose the same agent}) = \alpha + \beta W_{ij} + \gamma_1 D_{ij} + \gamma_2 D_{ij}^2 + \gamma_3 N_{ij} + \epsilon_{ij} \quad (1.1)$$

where W_{ij} is equal to 1 if homes i and j are in the same ward, D_{ij} is the geographic distance between homes i and j, and N_{ij} equals 1 if homes i and j are assigned to the same county-defined neighborhood. We measure distance in quarter miles so that D_{ij} lies in the interval from zero to one (since we only consider pairs of homes less than a quarter mile apart).

We calculate multi-way cluster-robust standard errors using the method developed by Cameron et al (2007) for this and all specifications. Observations are on pairs of listings, and each listing in the pair belongs to many different pairs. We estimate standard errors that are robust to clustering on both listings in the pair.

The estimation results for this specification are reported and discussed in Section 5 below.

To test whether the social learning in this setting arises from word-of-mouth information transmission we estimate the following variation on the regression in equation (1):

$$P_{ij} = \alpha + \beta W_{ij} + \gamma_1 D_{ij} + \gamma_2 D_{ij}^2 + \gamma_3 N_{ij} + \delta_1 G_j + \delta_2 B_j + \delta_3 W_{ij} * G_j + \delta_4 W_{ij} * B_j + \epsilon_{ij}$$

where $G_j = 1$ if neighbor j had a good outcome (and therefore has positive information to report) and $B_j = 1$ if neighbor j had a bad outcome.

The estimation results for this specification are presented in Section 6.

Identification

We want to identify the impact of a neighbor's choice of real estate agent on a home seller's choice of real estate agent. The principal threat to identification in our setting is what Manski (1995) calls *correlated effects*. If a home owner's neighbors who are assigned to his ward have homes that are systematically different than the homes of his neighbors who are assigned to different wards, then the effect that we estimate may represent correlations in behaviour due to correlations in unobserved home characteristics. An example of such an effect is small scale geographic specialization and marketing by agents to particular neighborhoods or types of homes.

Our identifying assumption is that, after we have conditioned on geography, a home owner's intra-ward neighbors are not systematically different from his extra-ward neighbors.

One concern is that ward boundaries may coincide with unobserved discontinuities in the spatial distribution of house characteristics. Observed discountinuites in house characteristics incluce abrupt changes in house (and resident) characteristics at subdivision boundaries and geographic features like rivers, parks, and major roads.

Ward designers try to ensure homogeneity *across* ward boundaries, so adjacent neighborhoods are likely to be split into wards in a way that assigns some homes from each neighborhood to each ward. Practically, this means that ward boundaries are likely to cut across neighborhoods.

Our identification fails if, despite the planners' objectives, there are unobserved neighborhood boundaries that are correlated with ward boundaries and that affect real estate agent selection. We investigate the extent to which this may occur by calculating the absolute difference in observed characteristics for each pair of neighbors and regress out the portion of those differences that are explained by their geographic relationship (distance, distance squared, and whether they are in the same county-defined neighborhood). We then calculate the means of these orthogonalized differences for neighbors in the same ward and for neighbors assigned to different wards and perform a t-test for the equality of those means (See Table 1.4).

We reject the null hypothesis that the mean differences in observed home characteristics for neighbors assigned to the same ward are equal to the mean differences for neighbors assigned to different wards. Our large number of observations (over three million pairs of neighbors) means that we only fail to reject the null hypothesis for extremely small differences in the means.

In order to illustrate the magnitude of the difference between the means, we calculate the standard deviation of the distribution of differences of each characteristic for the set of neighbors of each individual listing. So, for example, for each listing i we calculate:

$$\sigma_i(Bedrooms) = \sqrt{\frac{1}{J-1}\sum_{j\in J}[(\mathrm{Bedrooms_i} - \mathrm{Bedrooms_j}) - \mathrm{Average\ Deviation_i}]^2}$$

where J is the set of all neighbors of i and also the number of elements in the set. We present the median individual standard deviation for each characteristic in Table 1.4. The largest difference between the means is less than 5% of the mean difference between homes and is less than 5% of the median individual standard deviation. The differences of means for the other characteristics are even smaller proportions.

1.5 Evidence of Social Learning in Wards

The results from estimating equation 1, reported in Table 1.5, suggest that a home seller is substantially more likely to choose the same real estate agent as her neighbor if they are assigned to the same ward. For an intermediate distance (one eighth of a mile, distance = .5), the probability that the home seller chooses the same agent as her neighbor is 1.4% if they are assigned to different wards and 2.4% if they are assigned to the same ward.

The distance between two neighbors is an important determinant of the probability that they choose the same real estate agent. As mentioned in the introduction, the importance of the ward social network can be described in terms of the change in geography that will offset a ward relationship. For example, to be as influential as a ward neighbor that is 400 feet away, a non-ward neighbor must be 30% closer. Similarly, a ward neighbor 900 feet away is as influential as a non-ward neighbor that is 35% closer.

As additional evidence that our specification does not suffer from an omitted variable bias, we present the results from a long regression that includes differences in observed characteristics[7]. If we have omitted a relevant neighborhood definition that is correlated with the ward definitions, then our estimate of β will be biased. A neighborhood boundary is a discontinuous change in home characteristics, so the differences in home characteristics between two neighbors should be correlated with any omitted neighborhood definition (the differences should be larger for neighbors on opposite sides of the boundary). Including these differences in our regression, then, should attenuate any omitted variable bias. As column 4 of Table 1.5 shows, however, our estimate of the ward effect in this long regression is the same as the estimate from the short regression in column 3.

Census blocks are small neighborhoods bounded by geographic features (like roads, streams, and railroad tracks) and political boundaries (like city limits and property lines). In urban areas, the census block is often the same as the city block[8]. The way the census blocks are defined means that homes in the same block are likely to be very similar in unobserved characteristics. In column 7 of Table 1.5, we present estimates of equation 1 on a subsample restricted to neighbors in the same census block. The persistence of the ward effect is additional evidence that our results are not due to bias from some omitted neighborhood definition.

Figure 1.2 illustrates that the geographic distribution of neighbors in the same ward differs significantly from that of neighbors in different wards. Neighbors in the same ward tend to be nearer to each other and the nearest neighbors are very likely to be assigned to the same ward. We are careful about how we control

[7] See Table 1.4 for the list of characteristics that are included.

[8] This description of the census block definition is based largely on information available from the US Census Bureau, www.census.gov

for the geographic relationship to ensure that the estimated ward effect is not an artifact of the spatial relationships between neighbors. We demonstrate in columns 5 and 6 of Table 1.5 that our estimation of β is not sensitive to the specification of the distance effect: it does not change when we include either a sixth degree polynomial (column 5) or a set of dummy variables representing a fine partition of the distances (0 to 33 feet, 34 to 66 feet, etc.).

We also estimate equation 1 on a subsample of our data that excludes the nearest neighbors (where the vast majority are in the same ward) and the most distant neighbors (where the majority are assigned to different wards). Column 8 presents estimates when we restrict attention to neighbors that are no less than 400 feet apart and no more than 900 feet apart (distance $\in (.3, .7)$). Our estimate of β on this subsample is slightly larger than the estimate from the full sample, offering additional evidence that the geographic relationships do not drive our estimates of the ward effect.

1.6 Social Learning Via Direct Communication

We have offered evidence of social influence on real estate agent choice within wards. We now address the source of this influence. If home sellers are learning from peers only by observing choices and making inference about private information, then the choices of peers with identical characteristics will have identical influence. If, however, peers are communicating directly, then a home seller's choices may respond to a peer's private information, including information about outcomes.

Home sellers prefer a higher sales price, all else equal. If they are learning from their peers, they are more likely to choose the same real estate agent as a neighbor if that agent sold the neighbor's house for a high price relative to the seller's ex ante expectations and less likely to choose him if he sold the house for a low price relative to expectations. For each transaction, we calculate the percent difference between sales price and list price. We categorize transactions in the top decile of the distribution of percent difference as *very high price* and transactions in the bottom decile as *very low price*. We then analyze how these outcomes affect peer influence. The results are in column 1 of table 1.6.

We find effects for peers with intermediate outcomes that are similar to the effects in the baseline specification. In addition, for neighbors in the same ward we find a large premium associated with a very high price and a large penalty for a very low price. The additional influence of ward members with very high prices is double that of those with intermediate outcomes and the penalty from a very low price almost cancels the ward effect completely.

We modify our categorization of outcomes by considering different cutoffs for the definitions of *very high price* and *very low price*. In column 2 of Table 1.6 we define these outcomes using the top and bottom quartiles of the distribution of deviations. In column 3 we use the ninety-fifth and fifth percentiles. We find that the impact of good information is smaller with the lower threshold and larger with the higher threshold. While the impact of bad information increases slightly with the higher threshold definition, it does not change with the bottom quartile definition of column 2.

The markup over list price might reflect market conditions that may influence whether two neighbors select the same real estate agent. The specification in column 4 of Table 1.6 includes dummy variables for the year that each of the neighbors listed her home. We use the same definitions of *very high price* and *very low price* as in column 1 (the top and bottom deciles). Our estimate of the impact of good information outside the ward decreases substantially (0.9 to 0.3), but the impact of good information in the ward changes very little (1.4 to 1.3) and the impact of bad information in the ward doesn't change at all.

We also consider an alternative definition of outcome. We estimate a linear hedonic model of the natural log of sales price using the observed characteristics of homes and times and ward fixed effects. We then use the deviations from the predicted values (the residuals from the hedonic regression) to categorize outcomes. In column 5 of Table 1.6, we say that a home sells for a *very high price* if the difference between log sales price from the predicted log sales price is in the top decile of its distribution. We say that it sells for a *very low price* if it is in the bottom decile of its distribution. Column 7 of Table 1.6 uses the ninety-fifth and fifth percentiles. The qualitative results are not sensitive to the benchmark used - the estimates from column 5 are very similar to those in column 1.

1.7 Conclusion

We have presented evidence of word-of-mouth learning in a social network. The principal challenges to identification of this type of social effect are distinguishing which individuals are peers and separating social effects from correlations in behavior of peers that arise from unobserved similarities[9]. We address both of these issues by taking Mormon wards as our setting. The ward constitutes a social network for which the group composition is known and for which we can construct a control group of individuals who differ essentially from ward members only in their ward assignment.

The results that we present suggest that social learning plays a role in a home seller's selection of an agent to assist in an important transaction. We have also presented evidence that personal referrals are part of this social learning. It seems unlikely that this phenomenon is particular to Mormon wards. Congregations in other denominations and other types of social networks (school, athletic, service, etc) may function in similar ways.

We have exploited the geographic assignment of Mormons to congregations to identify social effects in real estate agent choice. There are many other decisions of economic importance that may be subject to social effects. The natural experiment investigated here holds promise for identifying social effects in many such decisions.

I thank Christopher Wignall, coauthor of the research presented in this chapter. It is with his permission that I include our research in this dissertation.

[9] See Manski (2000).

Table 1.1: Summary Statistics of Homes

Variable	Mean	Standard Deviation	Median
List Price	216065	125632	174900
Square Feet	2675	1255	2416
Acres	0.35	1.0	0.22
Bedrooms	3.9	1.2	4
Bathrooms	2.5	1.0	2
Garage Capacity	1.5	0.80	2
Patios	0.46	0.50	0
Decks	0.28	0.45	0
Wet Bars	0.40	0.52	0
Fire Places	0.64	0.76	0
Year Built	1984	23	1994
Sold Indicator	0.59	0.49	1
Sold Price	192833	99903	163000

Note – Data are from the Wasatch Front Regional Multiple Listing Service and include single family residences listed in Utah County between 1997 and 2007.

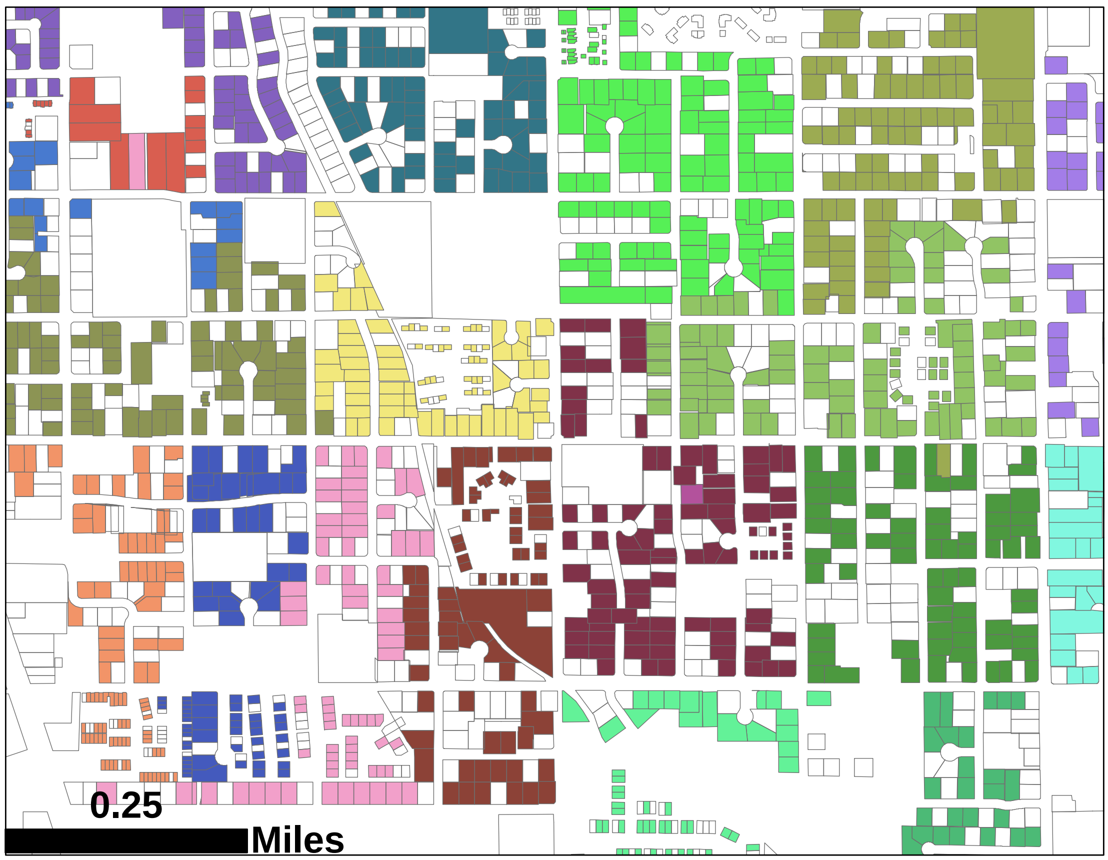


Figure 1.1: Ward assignment of homes in a Utah County neighborhood.

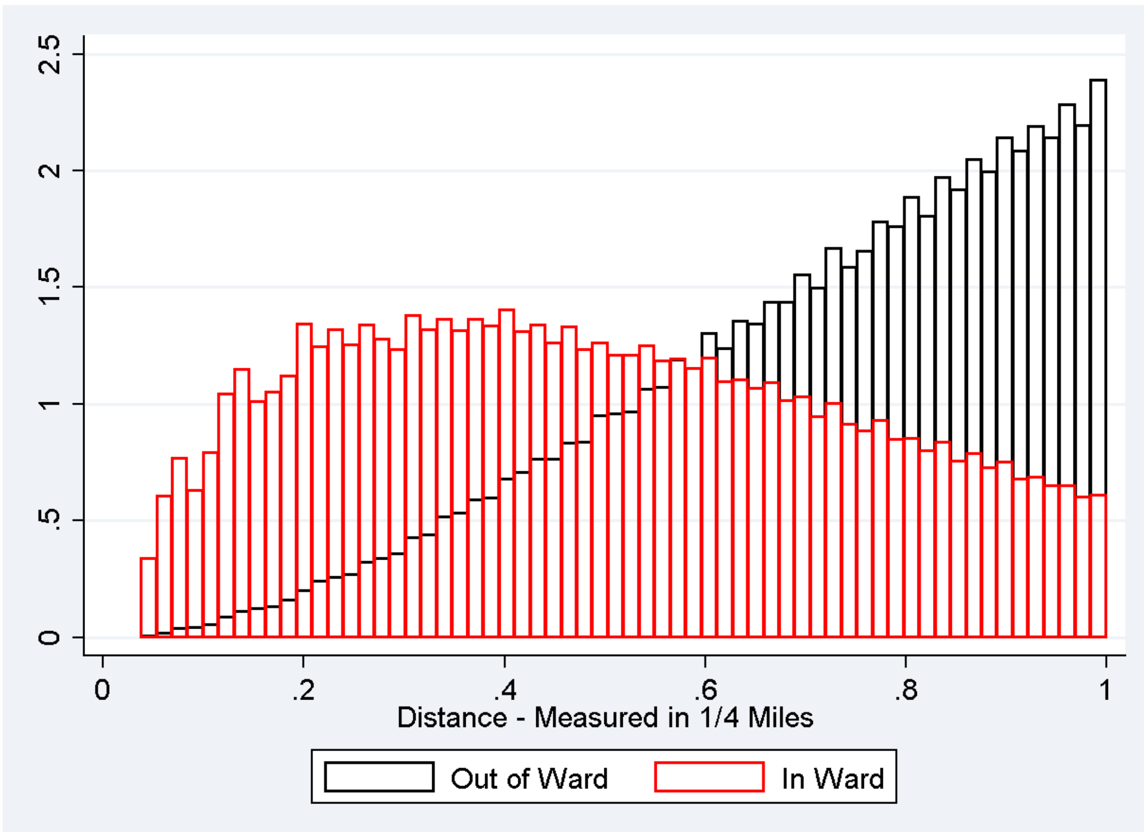


Figure 1.2: Histogram of Distances between neighbors in the same and in different wards.

Table 1.2: Summary Statistics of Neighbors

Mean (Standard Deviation) Median	In Ward	Out of Ward	In Census Block	In Neighborhood
	36	42	20	52
Number of Neighbors	(29)	(34)	(37)	(44)
	29	35	8	43
	49	51	23	64
Percent of Neighbors	(23)	(23)	(25)	(30)
	45	55	13	73
	0.49	0.71	0.48	0.58
Distance Between Neighbors	(0.25)	(0.21)	(0.26)	(0.26)
	0.47	0.75	0.46	0.59
	2.8	1.0	3.0	2.5
Percent that Chose the Same Agent	(9.9)	(4.6)	(11)	(9.8)
	0	0	0	0

Note – Geographic location of properties from Google Maps. Ward affiliation from www.lds.org. Neighborhood definition from the Utah County Assessor's Office. Census Block data from the US Census Bureau (TigerLine). Two Homes are "neighbors" if they are located within a quarter mile and sold within five years of each other. Distance is measured in quarter miles.

Table 1.3: Summary Statistics of Wards and Agents

Variable	Mean	Standard Deviation	Median
Ward			
(N = 832)			
Number of Listings	72	42	63
Number of Agents	51	26	48
Listings per Agent	1.3	1.0	1
Agent			
(N = 5904)			
Number of Listings	10	31	2
Number of Wards	7	19	2
Listings per Ward	1.3	1.0	1
Agent with at least three listings			
(N = 2649)			
Number of Listings	21	44	7
Number of Wards	14	27	6
Listings per Ward	1.5	1.5	1.2

Table 1.4: Differences in Characteristics Across Neighbors

Variable	Out	In	Difference	t-stat	Standard Deviation
Log Square Feet	0.261	0.255	0.006	17.5	0.201
Bedrooms	1.005	0.989	0.016	13.9	0.796
Bathrooms	0.707	0.700	0.008	8.6	0.601
Year Built	8.373	8.025	0.348	20.9	7.538
Log List Price	0.258	0.261	-0.003	-10.2	0.163
Log Acres	0.562	0.566	-0.004	-6.5	0.292
Garage Capacity	0.490	0.471	0.019	21.8	0.532
Fireplaces	0.404	0.398	0.006	8.8	0.507
Wet Bars	0.420	0.418	0.002	3.3	0.489
Deck	0.299	0.296	0.003	5.5	0.430
Patio	0.450	0.441	0.009	15.7	0.494

Note – We calculate the absolute difference in observed characteristics for each pair of neighbors. We regress out the portion of this difference that is explained by their geographic relationship (distance, distance squared, and whether they are in the same county-defined neighborhood). We present the mean of these orthogonalized differences in characteristics for neighbors in different wards and neighbors in the same ward. We present the t-statistic for a test that the means are equal. We also calculate the standard deviation of the differences for each individual property. For example, $\sigma_i(\text{Beds}) = \text{sqrt}(\sum_{\text{j}\in\text{J}}[(\text{Beds}_\text{i} - \text{Beds}_\text{j}) - (\sum_{\text{j}\in\text{J}}[\text{Beds}_\text{i} - \text{Beds}_\text{j}])/\#\text{J}]^2)$ where $J = \{\text{All neighbors if } i\}$. We present the median standard deviation for each characteristic to illustrate the variation between a typical listing and its neighbors.

Table 1.5: Linear Probability Model - Social Effects

Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
In the Same Ward	1.2**	1.9**	1.0**	1.0**	1.0**	1.0**	1.5**	1.1**	0.8**
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.12)	(0.06)	(0.04)
In the Same		0.6**	0.3**	0.03	0.3**	0.3**	0.9**	0.4**	0.2**
Neighborhood		(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.14)	(0.07)	(0.04)
Distance	-17.9**		-15.9**	-13.7**			-15.9**	-7.5**	-14.3**
(in 1/4 Miles)	(0.52)		(0.50)	(0.46)			(0.74)	(1.40)	(0.49)
Distance^2	11.6**		10.5**	9.2**			10.1**	4.3**	9.5**
	(0.38)		(0.36)	(0.34)			(0.62)	(1.33)	(0.36)
In the Same									1.3**
Census Block									(0.07)
Difference in Observed Characteristics				Yes					
Distance Polynomial					Yes				
Distance Dummies						Yes			
Constant	7.5**	0.6**	6.4**	8.2**	13.2**	11.2**	6.4**	3.7**	5.7**
	(0.18)	(0.04)	(0.17)	(0.20)	(0.57)	(0.54)	(0.26)	(0.36)	(0.17)
N	3,604,384	3,029,737	3,029,737	2,861,739	3,029,737	3,029,737	871,981	1,322,795	3,029,737
R^2	0.014	0.005	0.012	0.017	0.012	0.012	0.012	0.003	0.013

Linear Probability Model - Social Effects Parameter estimates from $y_{ij} = X_{ij}\beta + \varepsilon_{ij}$ where $y_{ij} = 1$ if neighbors i and j choose the same real estate agent to list their homes. Specification (4) includes the differences in all the observed characteristics listed in Table 1.4. (5) and (6) allow for a more flexible specification of distance by considering, respectively, a sixth degree polynomial and a set of 40 distance categories (0 to $\frac{1}{40}$ of a quarter mile, $\frac{1}{40}$ to $\frac{2}{40}$, etc.). We also consider the sub-samples of (7) only neighbors in the same census block and (8) only neighbors that are between 400 and 900 feet apart. All standard errors are robust to clustering on the identity of each listing in the pair of neighbors.

Table 1.6: Linear Probability Model - Information

Variable	(1)	(2)	(3)	(4)	(5)	(6)
In the Same Ward	1.2**	1.3**	1.2**	1.2**	1.1**	1.074**
	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)	(0.051)
Distance (in 1/4 Miles)	-17.8**	-17.8**	-17.8**	-17.2**	-17.9**	-17.9**
	(0.52)	(0.52)	(0.52)	(0.49)	(0.52)	(0.52)
Distance^2	11.6**	11.6**	11.6**	11.2**	11.6**	11.6**
	(0.37)	(0.37)	(0.37)	(0.36)	(0.37)	(0.37)
Very High Price	0.9**	0.2**	0.3*	0.3**	0.6**	0.7**
	(0.09)	(0.06)	(0.13)	(0.09)	(0.08)	(0.10)
Very Low Price	0.1	0.1	0.2	0.1	-0.03	-0.2*
	(0.09)	(0.05)	(0.15)	(0.09)	(0.07)	(0.08)
Ward * Very High Price	1.4**	0.5**	2.5**	1.3**	1.3**	1.6**
	(0.21)	(0.14)	(0.33)	(0.21)	(0.17)	(0.18)
Ward * Very Low Price	-1.0**	-1.0**	-1.1**	-1.0**	-0.8**	-1.0**
	(0.14)	(0.10)	(0.18)	(0.13)	(0.16)	(0.18)
Time Dummies				Yes		
Constant	7.4**	7.4**	7.4**	14.9**	7.4**	7.4**
	(0.17)	(0.18)	(0.17)	(0.98)	(0.17)	(0.17)
N	3,604,384	3,604,384	3,604,384	3,604,384	3,604,384	3,604,384
R^2	0.014	0.014	0.014	0.031	0.015	0.015

Results from linear probability model: $y_{ij} = X_{ij}\beta + \varepsilon_{ij}$ where $y_{ij} = 1$ if neighbors i and j chose the same real estate agent. We consider various definitions of "very high price" and "very low price." See text for details.

Chapter 2

Neighborhood Job-Information Networks and Residential Sorting

2.1 Introduction

In cross-sections of US workers, pairs of workers who live close together are more likely to work close together and even at the same workplace. This correlation has been interpreted in the literature as evidence of social learning in neighborhoods about job opportunities.[1] In this paper, I use employee-level data to investigate alternative explanations for this correlation. I exploit the timing of job and residence transitions to isolate a potential neighborhood peer effect from any reverse causation. In addition, I use data on employee characteristics and employment outcomes to explore the extent of employment-related neighborhood sorting.

A large literature in sociology and economics has employed survey data to document the importance of personal contacts as a source of job information; (see Ioannides and Loury (2004) for an extensive survey). Most recently, researchers have attempted to validate these findings by investigating patterns of place-of-employment across sets residential neighbors. (See Bayer et al., Hellerstein et al. and Schmutte, 2009).

[1]See Bayer, Ross, and Topa (2008) and Hellerstein, McInerney, and Neumark (2009).

A novel empirical approach for studying neighborhood peer effects in job choice is introduced in Bayer et al. Using long-form 2000 Census data for the Boston MSA, these authors find that a pair of neighbors living in the same census block *group* are more likely to share the same workplace census block if they share the same census *block* of residence. They argue that unobserved employment-related worker characteristics are uncorrelated across census blocks within census block groups. Under this identifying assumption, the observed correlation suggests the presence of a peer effects in job choice.

One specific threat to the identification in Bayer, Ross, and Topa (2008) is the possibility of reverse-causation: workers residential location choices may be influenced by social learning in the workplace about neighborhoods. However, even if reverse-causation can be ruled out, other more general forms of employment-related sorting may still be present.

The two datasets that I employ in this paper permit a detailed investigation of these alternative interpretations of spatial correlations between workplace and neighborhood of residence. The first dataset contains data on university employees at five state universities. I construct this dataset from university administrative records obtained in response to formal public records requests. The records include precise information on the timing of employment periods for each employee. To construct residential location histories for each employee, I match each employee in the data to historical telephone listings.

The precise timing of residence and job transitions in the university employment dataset permit me to isolate the potential neighborhood peer effect in job choice from a workplace peer effect in residential location choice. I estimate the impact of residential census block co-location prior to beginning a new position at the university on the probability that the new position is in the department of the university where a neighbor works. I cannot reject the hypothesis of a neighborhood network effect in this specification. The test is strong since it requires that the informal job-information network lead to same-department employment not just same-employer employment.

The second dataset that I use is assembled from restricted-use data from the

Survey of Income and Program Participation and administrative earnings records from the Social Security Adminstration. The survey data provide detailed worker demographics and information on the timing of residence transitions. The earnings records allow me to construct employment histories for the survey respondents. Using these data, I find evidence that block group neighbor pairs are more likely to have the same employer if they reside in the same census block. Again, the effect persists even when the sample is restricted to pairs that cannot have experienced a workplace residential location choice effect.

These specifications validate the findings by Bayer et al and go beyond the reverse-causation robustness tests that are possible using Census data. The Census long-form does have an indicator for whether a respondent had lived in the residence for more than five years and a measure of labor force attachment in the preceding year. This permits the estimation of the potential peer effect for a sample of pairs who were neighbors for the preceding five years and where one worker was only recently attached to the labor force. This furnishes an important robustness check, but doesn't completely rule out the possibility of reverse-causation.

A study by Hellerstein, McInerney, and Neumark (2009) considers the Census tract clustering of employees at particular establishments. This study has the advantage of an enormous nation-wide sample and the actual establishment of employment for each worker. However the timing information is the same as in Bayer et al. A very recent working paper by Schmutte (2009) uses quarterly administrative records of unemployment insurance coverage of all UI covered employees from several states. These data allow the type of timing-detailed investigation that I undertake in this paper. Schmutte (2009) only considers worker transitions from one employer to another. One contribution of the present work is to perform a timing-detailed analysis that includes transitions from unemployment to employment.

In addition to isolating potential neighborhood peer effects in job choice, I also estimate specifications designed to test for more general forms of residential sorting on employment-related characteristics. Using the university employment data, I find that block group neighbors employed at the same university *but not in* *the same department* who share the same employee classification are more likely to reside in the same census block relative to other same-group employee pairs. In the SIPP data, I find that block group neighbors who work in the same industry *but for different employers* are more likely to reside in the same census block of residence.

Finally, I also estimate a workplace peer effect in residential location choice. I do so by focusing on new university faculty. New university faculty are likely to come from outside the universities laborshed because of the national character of the labor market for university faculty. I find some weak evidence that a new assistant professor is more likely to reside in the same census block as a fellow-employee block-group-neighbor if the neighbor is a department co-worker.

In the next section I describe the construction of my two datasets. The subsequent section describes my empirical specifications and results. I conclude this chapter in Section 2.4.

2.2 Data

I use two datasets in this paper. The first dataset contains data on university employees and is constructed from public records and historical telephone listings. The second dataset is built from restricted-use data from the Survey of Income and Program Participation and administrative earnings records from the Social Security Adminstration. I will describe each dataset in turn.

2.2.1 University Employment Data

I construct the university employment dataset from employment records for five public universities: University of Washington (UW), Utah State University (USU), Utah Valley University (UVU), Washington State University (WSU), and Weber State University (WeSU). These records were obtained directly from the universities in response to formal public records requests. The records include precise information on the timing of employment periods for each university employee.

For UW and USU, I have a record of every payment from a university department to an employee during the period and the records include the date of the payment, the department making the payment and the name of the payee with her position title and classification. From these records, I infer employment start and end dates for each payee.

For the other three universities, I have employee-level data which include a date of first employment at the university, an employment termination date where applicable and the date that the employee was hired into her current position. The data also include the employee's name and the department, position title, and classification of her most recent position within the university. For WSU and WeSU, I have a record for all employees who worked at anytime from January 2005 to April 2009. For UVU, I have a record of all employees who worked at the university anytime from January 2006 to April 2009.

From these records, I construct a dataset of non-student university employees which includes the employing university, the employee's full name and most recent position, and the start date, end date, classification and department of the employee's most recent position and an indicator for whether the start date is also the original hire date for the employee.

I eliminate from my sample those employees who work at satellite campuses of Washington State University. The data for other universities may include records for personnel at satellite locations. These are a concern because they can produce false matches with telephone directories. That would make it more difficult to find an effect.

Table 2.1 presents non-student employment levels by year for each university in sample in the top row of each cell. The criteria for classifying an employee as employed in a particular year is discussed later in this subsection.

Annual telephone listings

In order to construct residence histories for these university employees, I match each employee by name to the residential telephone directory for the area where the employing university is located and then link the matched listing to earlier editions of the directory.

In order to perform this matching, I first gathered digital versions of the telephone directory for each relevant region for all years available. For most of the locations, directories are available from 2005 to present. These digital directories are available from etear.dexmedia.com as collections of flash video frames. I decompiled these video files using a perl module called SWF::Parse. I parsed the decompiled video files to extract the text elements of the telephone listings and construct a database of telephone listings.

The pages with the listings for the telephone company (Qwest) are in a different format that I have not been able to decompile into text elements; there are one or two such pages per directory.

Even if the parsing process were perfect, telephone listings are not a perfect record of address histories. Some people do not have a land-based telephone or choose to be unlisted. Also the directories represent a point-in-time snapshot of residents with listed land-lines, while this group is surely very dynamic. Increasingly, new move-ins may not even obtain a land-line but use a cell-phone exclusively. In particular, new hires may be more likely to be matched falsely if there is a growing trend to use cell phones instead of land-lines upon relocating or if new hires tend to be younger and younger people are the ones opting for no land-line.

County property records would provide an alternative way to obtain addresses for employees based on employee name. Unfortunately, county records were readily available for only one of the universities in my sample.

Matching employees to telephone listings

I match each employee with a telephone listing in the book corresponding to the employee's latest year of employment. I match these records using record linking software called Link Plus, available from the Center for Cancer Research. The matching process uses the name frequency in the directory.

There exists the possibility of validating this record linking process. For Brigham Young University (BYU), a private university, I have the full name and home address of virtually all current employees. I can validate my record linking process by linking the BYU employees to the local phone directory and then comparing the geo-coded locations (see below) of the telephone listing address and the university records address for each employee.

Linking phone directories across years to create address profiles

For each employee who matches to a telephone listing with a match score above the minimum cutoff of 7.8, I create a telephone listing profile. The construction of profiles proceeds as follows. For all listings with exactly the same listing name, I group them across years by phone number. I take the phone number group of the listing to which the individual was matched by the record linking as the base of the profile. Next if the profile does not span all available years, I check to see if any other phone number based profile ends right before the base profile begins. If so, and there is only one such profile, then this phone number group is added to the base profile. Similarly, if there is a single phone-number profile that begins in the year after the base profile ends, the profiles are combined.

An alternative construction in which each employee is matched to each annual telephone directory without regard to consistent linking across years obtains similar results for specifications that don't rely on the profiles.

Geocoding

Following Bayer et al, I employ Census geography in my empirical approach. Census geography is maintained by the U.S. Census Bureau to facilitate the U.S. Decennial Census, the administration of other Census Bureau surveys and the reporting of Census and survey in a systematic and cross-comparable way. The aspect of US Census geography that is relevant here includes state, county, tract, block groups and blocks (tabulation blocks). Within counties, census tracts are mutually exclusive and collectively exhaustive. Within tracts, census blocks are mutually exclusive and collectively exhaustive. Census blocks are identified by a 4 digit number. Blocks for which the identifier begins with the same digit are treated as belong to a census block group. Blocks in a group are also contiguous.

I gather census geography for the matched employee-listings using the online Census geography lookup tool on the Census Bureau's American Fact Finder (AFF) website. Figure 2.1 presents a histogram of the number of Census blocks per block group in the sample of matched and geo-coded employees.

The geo-coding process produces some additional sample attrition. Address information in telephone listings is not perfectly accurate, and some addresses cannot be geocoded because of missing elements or because the given address is invalid. Further, the AFF geo-coding database only includes addresses enumerated in the 2000 Decennial Census, therefore, any residence built since that time is not geo-coded.

In Table 2.1, the bottom row of each cell gives the employment level by year for each university after the sample attrition due to failure to match to the telephone directories.

Variable definitions

Some constructed variables are used across several specifications are defined here.

Since residential location is observed only as an annual snapshot, it makes sense to work with variables indicating whether a worker was employed in a particular 'telephone directory year', or TDYEAR. All years reported in this paper are TDYEAR's. The difference between TDYEAR and calendar year depends on the cutoff month for the directory local to a particular university.

I treat an employee as employed during a TDYEAR if the current hire date precedes cutoff date for inclusion in the telephone directory published in that year by 1 month and follows it by 2 months.

I take an employee's hire year to be the first TDYEAR he is considered employed by the definition above.

I take an employee to have moved when the geo-coded census block of residence changes from one year to the next for the employee's constructed residence history profile.

Table 2.2 presents the number of new staff hires by university and year.

A pair of employees is treated as being of the same household if they are matched to the same telephone listing.

Each entity provided an employee classification that included faculty as a category. I coded employees as staff if they were not classified as faculty or student.

I code an employee as an assistant professor if the provided job title is "assistant professor," "assistant research professor," "clinical assistant professor."

2.2.2 Data from the Survey of Income and Program Participation

I also construct a dataset with similar structure from the Survey of Income and Program Participation (SIPP) matched to administrative earnings records from the Social Security Administration. In particular, I select all worker from the the first waves of each of the 1996, 2001, 2004 and 2008 SIPP panels. For each of these workers, I obtain migration history information from the SIPP Migration Topical Module (administered in the second wave of each SIPP panel).

To create a partial employment history for each worker, I also match these workers to the Detailed Earnings Record (DER) maintained by the Social Security Administration. The DER contains a record of all Internal Revenue Service Form W-2 filed for the worker. Employers in the US are required by law to submit a W-2 form to the IRS at the beginning of each year for each employee who has worked for them during the preceding year. These W-2 records include an Employer Identification Number (EIN) that is designed to be unique to employers across years. Using these data, I construct a record of unique employers (EIN) for each worker and infer the beginning and ending year of employment for each worker-employer pair.

Since the 2008 SIPP panel has not yet been linked to the DER, I created alternate job history for each worker in the 2008 panel. For these workers, I focus on their employers at the time of the first wave SIPP interview. For these jobs, I identify the employer by the worker's report of the name of the employer and I take the worker's report of the date the job began as the employment start date.

Using block level census geography for workers in my sample, I build a dataset of same-panel worker pairs who reside in the same census block *group*. For each pair, I determine whether the two workers reside in the same census *block* and whether they have any employer in common. I also determine if the two workers in a pair work in the same industry and whether they have the same occupation.

Workers in a pair are taken to have the same employer if they have an EIN in common in their in-scope employment histories. Which part of the history in in-scope depends on the specification. In some specifications, I use the details on the timing of residence and job tenures to focus attention on worker pairs where they were neighbor before they were co-workers.

2.3 Empirical specifications and results

2.3.1 Neighborhood social networks and job choice

I begin with specifications and results using the university employment data. My first specifications replicates as closely as possible the empirical approach of Bayer et al. This allows me to confirm the principal finding from literature and to establish the relevance of the data used here for examining alternative explanations.

I estimate the probability that a pair of employees at the same university who reside in the same census block will be observed working in the same department of the university relative to other pairs who reside in the same census block group but in different census blocks. The fundamental difference here is that I am trying to predict employment in the same university department whereas Bayer et al. predict employment in the same census block.

Again, in an attempt to replicate Bayer et al as closely as possible, I take a cross-section of employees at a university-specific moment in time. I take as my cross section all employees employed on the directory cutoff date for the most recent directory available for the area where the employing university is located. As in Bayer et al, I exclude pairs who belong to the same household.

I estimate the following specification:

$$samedept_{ij} = \rho_g + \beta sameblock_{ij} + \epsilon_{ij}, \tag{2.1}$$

where ρ_g represents the group-specific intercept for census block group g.

I include block group fixed effects to control for unobserved heterogeneity in block-group patterns for pairs of employees to work in the same university department. In this way, I compare same-block pairs with different-block pairs within block groups but not across them.

I find that residing in the same census block increases the probability of working in the same department by 0.4 percentage points, an increase of more than 20% over the probability that a pair of neighbors living in different blocks will work in the same university department (1.5%). These results are presented in Column 1 of Table 2.4. The effect is statistically significant based on cluster robust standard errors clustered at the block group level.

But this cross-section of workers includes faculty employees who are unlikely to have learned about their jobs through neighbors. This is because the labor market for faculty is primarily national, even international. The estimates in Column 2 of Table 2.4 show that the effect persists even when faculty-faculty pairs have been excluded. We also see evidence of the same-block effect for faculty-faculty pairs. This may be due in part to residential sorting of new faculty based on learning in the workplace about neighborhoods, an effect I attempt to isolate in a specification discussed later in this section.

University staff are a group for which the neighborhood network story is more plausible. When staff-staff pairs are isolated, as in the estimation presented in Column 3 of Table 2.4, we find an important same-block effect for this group as well.

The reference group in the specification presented in Column 3 of Table 2.4 is the set of faculty-staff pairs. The estimate for this group is not statistically significant (but appears to be smaller). For cases in which such a pair represents a faculty hire, a neighborhood networking is not expected. When the the new hire in a pair is a staff member, we might expect an important neighborhood networking effect. It may be that the small and insignificant effect arises from combining these opposing effects.

A cross-sectional approach cannot separate the effects for new faculty and staff hires. However, in the no-reverse-causation specifications discussed below I can estimate the neighborhood effect operating for staff hires and their faculty neighbors.

In my no-reverse-causation specifications, I re-estimate Equation (1) focusing on new hires and their at-the-time-of-hire neighbors. This allows me to address the possibility that correlation between residence and workplace arises from workplace learning about residential locations. It also allows me to isolate the potential peer effect effects for staff hires and their faculty neighbors.

I take each new employee and pinpoint his or her residential location prior to being hired. Then I estimate whether block co-location increases the probability that the new hire works in the same department as his pre-hire block group neighbors. Thus, the estimation uses both a different sample and slightly different geographic information for workers (residence prior to hire instead of residence after hire) than the baseline specification above.

The results for these estimations are in the last three columns of Table 2.4. The neighborhood effect seems to persist even in the no-reverse-causality sample. The estimated effect of residing in the same block is virtually unchanged from the baseline specification. However, the estimates are statistically significant at a lower level.

In Column 6 of Table 2.4, I present results for a specification which separates the effects by the the classification of the neighbor. The effect for staff new hires and their faculty neighbors is not statistically significant.

Using the SIPP data, I estimate specifications similar to Equation 1, only I endeavor to predict whether a pair of neighbors will share an employer in common. In Columns 1 and 2 of Table 2.5, I present estimates of the increase in probability of sharing an employer in common based on residential block co-location for all SIPP workers and their neighbors at the time of SIPP interview. The probability increases by 1.2 percentage points which is an 8% increase over the probability that a pair of different-block neighbors in the sample will have an employer in common. The estimated effects drop by nearly 50% in the no-reverse-causation sample but are still statistically significant, though only at the 10% level.

2.3.2 Neighborhood sorting specifications

Despite the evidence presented above that the same-block effect persists after ruling out residential location choice based on workplace learning, more general residential sorting could still explain the observed effect. To look for evidence of neighborhood sorting on employment-related characteristics, I first investigate whether pairs of university employees within a census block group *who work in different departments* are more likely to reside in the same census block if they share the same employment classification at the university (eg. faculty, staff, administration, etc). The classifications provided by the universities are slightly finer than just faculty and staff.

I estimate the following specification:

$$sameblock_{ij} = \rho_g + sameclass_{ij} + \epsilon_{ij} \tag{2.2}$$

where ρ_g is the census block group fixed effect for group g.

I estimate this specification for the sample of pairs of neighbors employed by the same university as of the most recent telephone directory cutoff date for the area where the employing university is located. I exclude pairs in which both neighbors' employment is in the same department at the university.

I present the results for this specification in Table 2.6. The probability of residing in the same census block increases by 0.6 percentage points for employees who share the same classification but not the same employing department. This effect is a 4% increase relative to the probability for pairs with different classifications.

Using SIPP data, I estimate a similar specification in which I predict block co-location with indicators for whether the workers in a pair are employed in the same industry and whether they are employed in the same occupation. I exclude from my sample pairs of neighbors who work for the same employer. I find that same-industry increases the probability of residential block co-location by 2.6 percentage points. Neighbor pairs in the sample with different industries and different occupations have, on average, a 26% chance of residing in the same census block.

2.3.3 Social learning in the workplace and residential location choice

The no-reverse-causation specifications above were designed to isolate neighborhood peer effects in job choice from workplace peer effects in residential location choice. In this subsection, I describe specifications and estimation designed to measure the latter effect.

The distinctiveness of the labor market for university faculty provides an opportunity for a test of social learning in residential location choice. Because the labor market for new faculty is national, new faculty may have few community connections in the location of a new job and may be particularly prone to choosing a residential location based on information from their new co-workers. I look for evidence that new faculty are more likely to choose a residence in the same census block as department colleagues than they are to choose the same census block as other fellow university employees residing in the same census block group.

The results are presented in Column 3 of Table 2.6. I find that the probability that a new assistant professor resides in the same census block as a same block group department co-worker is 15% percentage points higher than for other group neighbors employed in other departments. This represents an almost 100% increase over the probability of the newly hired assistant professor residing in the same census block as a non-departmental fellow employee in his census block group. The specification does not condition on departmental fixed effects which may account for some of the magnitude of this estimate.

To explore whether workplace social learning impacts residential location choices of other university employees, I estimate the probability that, following a post-hire move, a university employee will live closer to his same-department neighbors than to his extra-department neighbors. I restrict attention to pairs of neighbors where exactly one of them changed residential locations while they were both employed for the university.

The results for this specification are presented in Column 4 of Table 2.6. The effect of working in the same department on whether a worker moves into the same census block as a fellow employee is not statistically significant. It may be that the number of moves is too small to detect this effect.

Table 2.7 presents similar results for SIPP worker pairs. In these results, worker pairs who share no employer in common have a higher probability of residing in the same census block when they work in the same industry.

2.4 Conclusion

The principal finding of the literature is robust to specifications that rule-out reverse causation as an explanation. However, there appears to be some sorting across census blocks within census block groups. Therefore, the observed neighborhood effect that remains after reverse-causation has been removed may not be due entirely to a neighborhood peer effect.

Table 2.1: Employment level by university and year

	2005	2006	2007	2008	2009
uofw	25212	28473	29856	31874	29196
	7914	8799	9280	10597	10677
usu	3159	3141	3122	3087	0
	1212	1227	1263	1311	0
uvu	2055	2262	2547	3341	0
	740	803	933	1284	0
weber	1165	1212	1282	1337	1402
	402	435	555	608	680
wsu	0	5047	5256	5353	5345
	0	942	1049	1049	1065

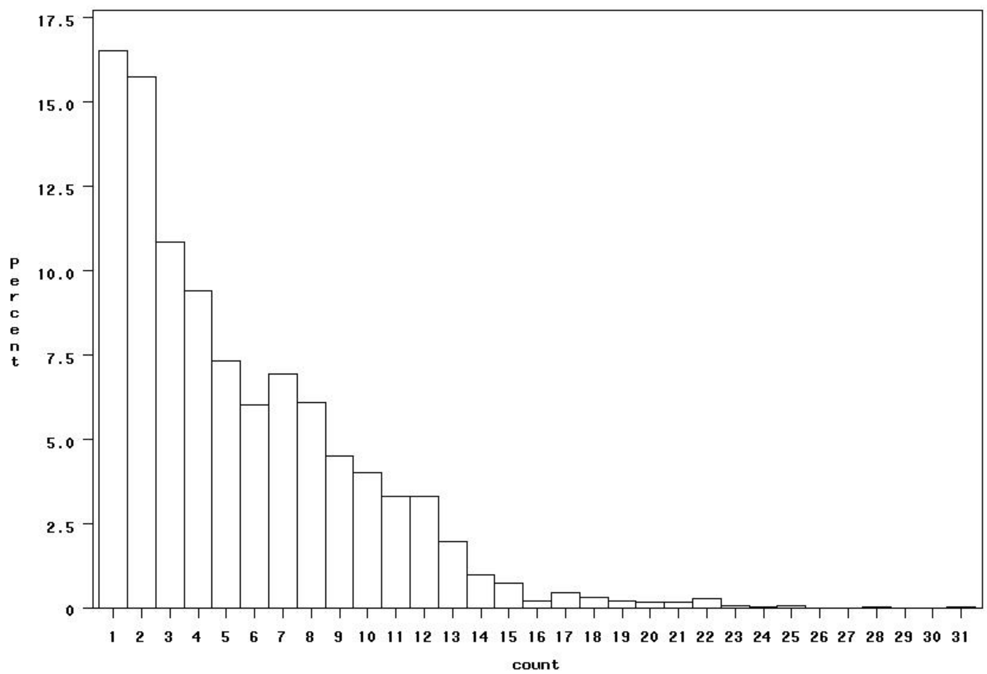
Note – The top row of each cell is the number of non-students employed by the university over the course of the year beginning one month before the cutoff dates for inclusion in the annual telephone directory where the university is located. The bottom row of each cell gives the employment level at the university that remains in my sample after attrition from record linking to the telephone directory and geo-coding.

Table 2.2: Number of staff new hires by university and year

	2005	2006	2007	2008	2009
uofw	549	1409	1297	1694	1476
usu	30	48	73	46	0
uvu	119	118	134	197	0
weber	12	11	26	21	25
wsu	0	41	35	35	2

Note – Number of staff new hires by the university after attrition from record linking to the telephone directory and geo-coding.

Figure 2.1: Distribution of census *blocks* per census block *group*



Note – Distribution of Census blocks per Census block group in the sample of employees who match acceptably to the telephone directory and whose telephone listing addresses could be precisely geo-coded.

Table 2.3: SIPP worker pairs

	Same Census Block		Different Census Blocks	
	Percent	Frequency	Percent	Frequency
Not both born in US	7.0887	1786	8.3500	6565
Both born in US	91.3435	23014	90.6605	71280
Neither born in US	1.5678	395	0.9895	778
Not both hourly	40.0956	3103	42.2041	9758
Both hourly	43.6878	3381	44.6910	10333
Neither hourly	16.2166	1255	13.1050	3030
Male–Female	49.5892	12494	49.8544	39197
Both male	23.7904	5994	22.6651	17820
Both female	26.6204	6707	27.4805	21606
Rent–Own	24.2826	6118	31.7452	24959
Both own	65.0923	16400	61.7567	48555
Both rent	10.6251	2677	6.4981	5109
Detached–attached	9.9429	1793	17.6780	10715
Both detached	81.3619	14672	77.8856	47208
Both attached	8.6952	1568	4.4364	2689
Different educ. levels	81.1907	20456	83.9525	66006
Same eduation level	18.8093	4739	16.0475	12617
Different races	14.6656	3695	19.6864	15478
Same race	85.3344	21500	80.3136	63145

Note – Pairs of respondents to the 2004 panel of the Survey of Income and Program Participation (SIPP) who reported at least one job in Wave 1 of the survey and who reside in the same Census block *group*.

Table 2.4: Probability of working in the same university department

VARIABLES	(1) same_dept	(2) same_dept	(3) same_dept	(4) same_dept	(5) same_dept	(6) same_dept
same_block	0.004***	0.003***	-0.001	0.004**	0.003*	0.004*
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
faculty-faculty		0.014***	0.017***			
		(0.002)	(0.002)			
sameblock*faculty-faculty		0.008*	0.011***			
		(0.004)	(0.004)			
staff-staff			0.009***			
			(0.001)			
sameblock*staff-staff			0.006***			
			(0.002)			
faculty						-0.010***
						(0.001)
sameblock*faculty						-0.003
						(0.003)
Constant	0.015***	0.013***	0.008***	0.014***	0.014***	0.018***
	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)
Observations	169611	169611	169611	51112	51112	51112
Number of groups	1881	1881	1881		1111	1111

Observations are pairs of same-university same-census-block-group employees. Columns (1) to (3) use a cross-section of employees. Columns (4) - (6) use a no-reverse-causality sample. Standard errors are clustered on census block groups.

Table 2.5: Probability that SIPP worker pairs have the same employer

VARIABLES	(1) cross-section	(2) cross-section	(3) No R.C.	(4) No R.C.
Same Block	0.0122**	0.0117**	0.00668**	0.00637*
	(0.00525)	(0.00477)	(0.00279)	(0.00355)
Block Group FE		yes		yes
Constant	0.154***	0.144***	0.0468***	-0.00667
	(0.00645)	(0.0299)	(0.00352)	(0.0323)
Observations	96980	96980	52778	52778
Number of Block Groups		8026		6089
R^2	0.037	0.019	0.008	0.004

*** p<0.01, ** p<0.05, * p<0.1

Note – Observations are pairs of same-panel same-census-block-group workers interviewed in Wave 1 of the 1996, 2001, 2004, or 2008 SIPP panel. The first two columns are a simple cross-section of workers. The second two columns restrict attention to pairs who were neighbors before becoming co-workers. In parentheses are Census block group-clustered robust standard errors. All specifications also include SIPP panel fixed effects.

Table 2.6: Probability of residing in the same census block: university data

	(1)	(2)	(3)	(4)
VARIABLES	same_block	same_block	AP	RLC
same_class	0.007**	0.006***		
	(0.003)	(0.002)		
same_dept	0.008	0.004	0.153*	0.005
	(0.009)	(0.008)	(0.089)	(0.027)
sdsc	0.051***	0.042***		
	(0.015)	(0.012)		
Constant	0.149***	0.149***	0.150***	0.148***
	(0.006)	(0.001)	(0.001)	(0.010)
Observations	262815	262815	2792	10261
Number of group		1914	195	

Census block group-cluster robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note – Observations are pairs of same-panel same-census-block-group workers interviewed in Wave 1 of the 1996, 2001, 2004, or 2008 SIPP panel. The first two columns are a simple cross-section of workers. The sample in the third column is new assistant professors and their same-block-group neighbors. The sample used in the fourth specification is the set of post-hire movers and their neighbors prior to the move. In parentheses are Census block group-clustered robust standard errors. All specifications also include SIPP panel fixed effects.

Table 2.7: Probability that worker pairs in SIPP reside in same census block

	(1)	(2)
VARIABLES	Same Block	Same Block
same industry	0.0583***	0.0266**
	(0.0201)	(0.0131)
same occupation	0.0481*	0.00344
	(0.0259)	(0.0194)
Block Group FE		yes
Constant	0.248***	0.258***
	(0.0118)	(0.00598)
Observations	35140	35140
R^2	0.002	0.000
Number of block groups		3864

Census block group-clustered robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note – Observations are pairs of same-panel same-census-block-group workers interviewed in Wave 1 of the 1996, 2001, 2004, or 2008 SIPP panel. In parentheses are Census block group-clustered robust standard errors. All specifications also include SIPP panel fixed effects.

Chapter 3

Reputation Transmission in Real Estate Brokerage Networks

3.1 Introduction

In this chapter, I present a simple model in which real estate transactions are more likely to feature high quality listing agents when selling agents are in a position to learn about the quality of listing agents prior to advising buyers. I test this prediction using multiple listing service data and county records data. I employ instrumental variables to identify the relationship between measures of listing agent quality and measures of the network connection between the two agents in the network of past transaction relationships.

The typical real estate transaction involves two real estate agents. The seller contracts with an agent (the listing agent) to assist in the sale of the property in exchange for a commission. It is usually a second agent who procures a buyer (the selling agent); in this case, the selling agent receives a pre-specified proportion of the listing agent's commission. Once a buyer's offer to purchase has been accepted by the seller, the escrow period begins during which the contract contingencies can be met. When all of the contingencies are met, the transaction closes and the agents are paid the commission.

If real estate agents differ in their competence or diligence at attending to the processes necessary to reach closing, then agents will be heterogeneous in the costs to others of transacting with them. Real estate agents report that they collect information from colleagues about other agents and advise clients on potential transactions based on this information.[1] For instance, a selling agent might discourage a buyer from making an offer on a particular property knowing that the listing agent has a reputation for inattention to essential escrow period details such as disclosures.

To look for evidence of this type of network-based reputation transmission, I formulate a simple social network formation model of the relationships between real estate agents. I explicitly model the transmission of information about agent quality across the network and the role of this information in decisions that impact the network formation.

In the model, real estate agents are matched at random to transact. Each agent in a match can agree to transact or reject the match in favor of transacting with another agent at random. Prior to making a decision, each of the agents in a match receives a report about the other's cost type from each of the agents with whom both agents have recently transacted. When the number of these common neighbors in the network of recent transactions is higher, the agents receive more reports and are less likely to inadvertently reject a match with a low cost agent. This conclusion arises even if agents make no inference about another agent's quality from the number of common neighbors between them. A prediction of the model, then, is that in the set of consummated transactions, the number of common network neighbors will be correlated with the agents' cost types.

In this chapter, I present a basic test of this prediction. I regress proxies for listing agent quality on the number of common network neighbors between the two agents in a transaction. I address the endogeneity of the number of common neighbors between two agents using instrumental variables which are valid under the model's assumptions. The first instrument is the number of transaction partners of the selling agent. This number will be correlated with the number of common neighbors and is a valid instrument if high quality agents acting in the capacity of

[1] Based on my own conversations with various real estate agents.

selling agents are not better at selecting high quality listing agents than are other selling agents. Under similar assumptions, the number of transaction partners of all of the listing agent's transaction partners and the number of transaction partners of all of the selling agent's transaction partners are also instruments for the number of common neighbors.

I implement this test using data on the sale of residential properties in Utah County, Utah from January 1997 to August 2008. The data include the listing and transaction details for all residential properties listed with a real estate agent during the sample period. For each month, I construct a network of recent transactions in which two agents are considered neighbors if they have participated together in a transaction in the past two years. I take the number of common neighbors between two agents as a measure of how much information is available to one of the agents about the other.

I also use these data and property records from Utah County Recorder's and Assessor's offices to generate three different proxies for listing agents' unobserved cost types. The first is the proportion of the listing agent's listings where the correct county property identifier was listed in the MLS database. The second is the proportion of the listing agent's deals where the listing agent's report of the square footage exceeded the square footage as recorded by the county. The third is the proportion of the listing agent's listings where the listed square footage matched the county's records exactly. These measures reflect agent attention to detail and/or accuracy in information disclosure and have the potential to proxy for agent attributes that matter to buyers and their agents.

I find that variation in the number of common neighbors that is due to variation in the instruments can explain variation in the proxies for listing agent cost type. The statistical significance of the effects, however, is not robust to clustering of the model errors on listing agent.

In the next section I present my model and derive the implication that I test in this chapter. In the subsequent section, I discuss the data that I use. My empirical approach and results are presented in Section 4.

3.2 Model

A proportion θ of N real estate agents are high-cost (HC) types; the remainder are low-cost (LC) types.

In each period, pairs of agents are matched at random to potentially transact. The agents can each accept or reject the match. If both agents accept the match, the transaction takes place and the agents realize the costs of transacting with the other.

If either agent rejects the match, both participate in a second round of matching in which each is paired with an agent drawn at random from the same distribution over agents as in the first round. In the second round, however, transactions take place automatically with no opportunity for agents to reject matches.

Transacting with another agent either costs zero or one units. The unit cost is realized with probability $\alpha > \frac{1}{2}$ when transacting with a high-cost agent and with probability $(1 - \alpha)$ when transacting with the low-cost type. The parameters θ and α are commonly known to the agents.

In making the accept/reject decision, agent i has access to information about agent j through network neighbors common to agents i and j in the network of recent transactions.

Agents i and j have n_{ij} common network neighbors and each of these reports to agent i the cost of his last transaction with agent j (and vice versa). The agents each take the reports as independent realizations and update their beliefs about the other agent's type. I assume that agent i makes no inference about agent j's type from the number of common neighbors.

The following proposition states a prediction of the model.

Proposition 1. *In the set of consummated transactions, the number of common network neighbors between the two agents covaries negatively with each agent's cost type.*

The proof is presented in the appendix. Intuitively, when an agent receives more reports about the type of the agent to whom he is matched, it is less likely that he will inadvertently reject a match to a low-cost agent. The sample of matches that are consummated as transactions will reflect this selection.

3.3 Data

To implement my test of the proposition, I use data on real estate in Utah County, Utah. I employ listing and transaction details from the Wasatch Front Regional Multiple Listings Service (WFRMLS) and property characteristics from the County Assessor's office. I construct measures of the relationship in the network of agent pairs. I also construct proxies for the quality or cost type of listing agents.

Real estate sales data from Wasatch Front MLS

I use data on the listing and transaction details of residential properties in Utah County, Utah that were listed with the Wasatch Front Regional Multiple Listings Service between January 1997 to August 2008. The data include identifiers for the participating agents and their firms, the transaction date (where applicable), and, according to the listing agent, the square footage, the county tax identifier for the property, and the year the property was built. I exclude listings where one of the agents or firms is not identified.

Table 3.1 presents descriptive statistics for the 3707 listing agents in my analysis sample. These agents have a mean number of listings per agent of 30.87 and a mean number of transactions per agent of 7.65.

In constructing networks of recent transactions, I take the time period of a transaction to be the calendar month of the transaction. In a given calendar month, two agents are considered *neighbors* in the network of recent transactions if they have transacted together in the two years preceding the beginning of the calendar month. Any agent with whom both of two agents have transacted in the same period is considered a *common neighbor* of the two agents. A second-degree neighbor of an agent is a neighbor of a neighbor. The network measures that I use in this chapter are the number of common neighbors between two agents and the number of second degree neighbors of an agent. Descriptive statistics for these measures for the listing agents in my sample are presented in Table 3.1.

Property records from Utah County

I also employ property data from the County Assessor of Utah County.[2] These data include property characteristics as of the county's last assessment. I also make use of the address for the property as recorded by the county, the county's tax identifier for the property, the county's record of the year the property was built, and the total square footage of the property based on the most recent tax appraisal completed by the county. I obtain the square footage measure by summing the square footage by floor as reported by the county.

Proxies for listing agent cost type I generate three proxies for agents' cost types. The first is the proportion of the listing agent's listings for which the county property identifier listed in the MLS existed in the county records and for which corresponding addresses in the MLS and county records matched. There are two reasons to believe that this measure might proxy for something of interest to selling agents and buyers. First, it may reflect the agent's attention to detail. Second, an invalid taxid makes it more difficult to verify property characteristics and other information about the property from the county. An agent might enter an invalid taxid if information from the county would be disfavorable to a sale. In what follows, I call this measure VALID.

A second possible proxy for a listing agent's cost type is the proportion of the listing agents deals where the listing agent's report of the square footage exceeded the square footage as recorded by the county. A dishonest listing agent might overstate the square footage of a property for sale in order to attract potential buyers. I call this measure EXAGERATE.

A final measure, called EXACT, is the proportion of the listing agent's listings where the listed square footage matched the county's records exactly. The county's record of the square footage is based on an appraisal and is subject to some error. The report in the MLS of the square footage that differs from the county's record may indicate that the listing agent took care to obtain his own estimate of the square footage. An exact match between MLS and county might indicate a lack of listing agent effort.

Descriptive statistics for these proxies are presented in Table 3.1.

[2]The county data were collected from the county's website: http://www.co.utah.ut.us.

3.4 Empirical specification and results

I test the prediction from Proposition 1 by estimating the following relationship:

$$\hat{\theta}_i^L = \beta_o + \beta_1 n_{ijt} + \epsilon_{ijt} \tag{3.1}$$

where $\hat{\theta}_i$ is a proxy for agent i's cost type and n_{ijt} is the number of common neighbors between agents i and j at time t.

In my empirical specification, I exclude listings for which there is only one agent participating in the transaction or where both agents belong to the same firm or for which the agents have transacted together before. I also exclude any listing for a property that was under construction at the time of listing or which resulted in a cash transaction.

The endogeneity of n_{ijt} is addressed by using instrumental variables. I use the the selling agent's degree, or number of neighbors, and the second order degree for both the selling and listing agents as instruments. These measures will be correlated with the number of common neighbors since the agents will be likely to have more neighbors in common the more neighbors the selling agent has. Similarly, if the neighbors of the two agents have more neighbors, then common neighbors will be more likely for the agent pair. First stage results (see Table 3.2) suggest that the instruments have adequate strength based on the F-stat cutoffs suggested in Stock and Yogo (2002). The F-statistic for the first stage model is 148.

These measures will be valid instruments if high quality agents acting in the capacity of selling agents are not better at selecting high quality listing agents than are other selling agents. Since there are three instruments and only one endogenous regressor, it is possible to perform some testing of identifying assumptions. The value of this is diminished by the similarity of the identifying assumptions across instruments.

Results from two stage least squares estimation with and without cluster-robust standard errors are presented in Table 3.3. Comparable results where observations are weighted by the number of listings that gave rise to the average that appears as the dependent variable are presented in Table 3.4. The results from the weighted and unweighted regressions are qualitatively similar.

I focus here on the results from the weighted regressions. An additional common neighbor between the two agents in a transaction is associated with a 0.2 percentage point increase in the listing agent proportion of listings with valid county property identifiers and addresses. An additional common neighbor is also associated with a 0.2 percentage point increase in listing agent proportion of listings with reported square footage greater than the square footage reported by the county for the same property and a 0.1 percentage point decrease in the listing agent proportion of listings with exact correspondence between MLS and county square footage reports. The statistical significance of these relationships is not robust to allowing model errors to be clustered for listing agents.

3.5 Conclusion

I have presented a model of real estate transactions in which selection of potential transactions into the set of transactions leads to low-cost-of-transacting agents appearing more frequently in transactions where the agents have more information about each other prior to choosing to transact. The empirical evidence presented here provides some evidence that variation in proxies for listing agent cost type can be explained by variation in the number of common recent transaction partners between two agents. The interpretation of the results presented here relies on the assumption that low-cost listing agents are not also better at discerning another agent's cost type.

Table 3.1: Descriptive statistics

Variable	N	Mean	Std Dev	Minimum	Maximum
Listings per agent	3707	30.87	84.30	1	1635
Transactions per agent	3707	7.65	20.99	1	418
common2	28367	0.614	2.568	0	89
degs	28367	11.288	14.879	0	120
deg2s	28367	172.052	193.961	0	1158
deg2l	28367	263.140	237.445	0	1282
validavg	28352	0.446	0.191	0	1
exagerateavg	26964	0.363	0.208	0	1
exactavg	26964	0.383	0.243	0	1

Table 3.2: First stage

	(1) common2
degs	0.074***
	(0.021)
deg2s	-0.002
	(0.001)
deg2l	0.002***
	(0.000)
F-stat	148***
Constant	-0.394***
	(0.048)
Observations	28367

Robust standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Note – First Stage of Two-Stage Least Square regression. Common 2 is the number of common neighbors between the two agents in transaction in the network of transactions in the county from the two years preceding the transaction. Definitions of degs, deg2s, deg2l.

Table 3.3: Number of common neighbors and proxies for listing agent quality

VARIABLES	(1) valid	(2) valid	(3) exagerate	(4) exagerate	(5) exact	(6) exact
common2	0.011***	0.011**	0.006***	0.006	-0.007***	-0.007
	(0.001)	(0.005)	(0.001)	(0.005)	(0.002)	(0.006)
Constant	0.440***	0.440***	0.359***	0.359***	0.388***	0.388***
	(0.001)	(0.005)	(0.002)	(0.006)	(0.002)	(0.008)
Observations	28352	28352	26964	26964	26964	26964

Standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Note – The estimate are from (unweighted) two-stage least squares estimation of the relationships. Dependent variables are various proxies for listing agent quality. VALID is the proportion of the listing agent's listings where taxid was correct. EXAGERATE is the proportion of the listing agents deals where the listing agent's report of the square footage exceeded the square footage as recorded by the county. EXACT is the proportion of the listing agent's listings where the listed square footage matched the county's records exactly. The second column for each dependent variable presents the estimates with standard errors clustered on the listing agents.

Table 3.4: Number of common neighbors and proxies for listing agent quality–weighted

VARIABLES	(1) valid	(2) valid	(3) exagerate	(4) exagerate	(5) exact	(6) exact
common2	0.002***	0.002	0.002***	0.002	-0.001**	-0.001
	(0.000)	(0.002)	(0.000)	(0.002)	(0.000)	(0.002)
Constant	0.424***	0.424***	0.393***	0.393***	0.329***	0.329***
	(0.001)	(0.014)	(0.001)	(0.019)	(0.001)	(0.022)
Observations	28352	28352	26964	26964	26964	26964

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Note – The estimates are from (weighted) two-stage least squares estimation of the relationships. Dependent variables are various proxies for listing agent quality. VALID is the proportion of the listing agent's listings where taxid was correct. EXAGERATE is the proportion of the listing agents deals where the listing agent's report of the square footage exceeded the square footage as recorded by the county. EXACT is the proportion of the listing agent's listings where the listed square footage matched the county's records exactly. The second column for each dependent variable presents the estimates with standard errors clustered on the listing agents.

Appendix

Proposition 1. *In the set of consummated transactions, the number of common network neighbors between the two agents covaries negatively with each agent's cost type.*

Proof. The result is established by modeling the selection of matches into the sample of completed transactions.

A first round match between agents i and j will result in a transaction if neither agent rejects the match. Agent i will reject the match if and only if the expected cost of the transaction exceeds the expected cost of transacting in the second round. The condition can be written as $\hat{\theta}_{ij} > \theta$.
where $\hat{\theta}_{ij}$ is agent i's posterior belief at time t about j's type.

By Bayes' Rule,

$$\hat{\theta}_{ij} = \frac{Pr(x_{ij}|\theta_j = LC, n_{ij})Pr(\theta_j = LC)}{Pr(x_{ij}|n_{ij})}$$

where x_{ij} is the number of reports of high-cost realizations for agent j from the neighbors common to agents i and j. Note that the belief is proportional to $Pr(\theta_j = LC)$ and not $Pr(\theta_j = LC|n_{ij})$ because agent i takes all reports as independent draws and makes no inference from the number of common neighbors.

Since x_{ij} is the number of negative reports from n_{ij} independent reports, it is distributed according to the binomial distribution, and we obtain

$$\hat{\theta}_{ij} = \frac{(1-\alpha)^{x_{ij}}\alpha^{n_{ij}-x_{ij}}\theta}{(1-\alpha)^{x_{ij}}\alpha^{n_{ij}-x_{ij}}\theta + \alpha^{x_{ij}}(1-\alpha)^{n_{ij}-x_{ij}}(1-\theta)}.$$

Substituting the above expression into the condition above and rearranging yields $x_{ij} > \frac{n_{ij}}{2}$.

Let $S_{n_{ij}} \equiv \frac{n_{ij}}{2}$. Then the probability that agent i rejects agent j when agent j is a low-cost type agent is given by

$$1 - Pr(x_{ij} \leq S_{n_{ij}}|\theta_j = LC, n_{ij}),$$

where $Pr(x_{ij} \leq S_{n_{ij}} | \theta_j = LC, n_{ij})$ is given by

$$F_{LC}(n_{ij}) = \sum_{k=0}^{floor(Sn)} \begin{pmatrix} n_{ij} \\ k \end{pmatrix} (1-\alpha)^k \alpha^{n_{ij}-k}.$$

When agent j is a high-cost agent, the expression becomes

$$1 - Pr(x_{ij} \leq S_{n_{ij}} | \theta_j = HC, n_{ij}),$$

with $Pr(x_{ij} \leq S_{n_{ij}} | \theta_j = HC, n_{ij})$ given by

$$F_{HC}(n_{ij}) = \sum_{k=0}^{floor(Sn)} \begin{pmatrix} n_{ij} \\ k \end{pmatrix} \alpha^k (1-\alpha)^{n_{ij}-k}.$$

It is clear that $F_{LC}(n_{ij}) \geq F_{HC}(n_{ij})$. Thus we have that for any value of n_{ij}, a high-cost type agent is more likely to be rejected then a low-cost type. It remains to be demonstrated that the difference between the two functions is increasing in n_{ij}. Suppressing subscripts for ease of exposition, let $\kappa(n) \equiv F_{LC}(n) - F_{HC}(n)$. Then

$$\kappa(n) = \sum_{k=0}^{floor(S_n)} G_{LC}(k|n) - G_{LC}(k|n), \tag{3.2}$$

where

$$\begin{aligned} G_{LC}(k|n) &= \begin{pmatrix} n \\ k \end{pmatrix} (1-\alpha)^k \alpha^{n-k}, \text{and} \\ G_{HC}(k|n) &= \begin{pmatrix} n \\ k \end{pmatrix} \alpha^k (1-\alpha)^{n-k}. \end{aligned}$$

We can write $\kappa(n+1)$ as

$$\begin{aligned} & [floor(S_{n+1}) - floor(S_n)][\alpha^{n+1} - (1-\alpha)^{n+1}] \\ + & \sum_{k=0}^{n} \frac{(n+1)}{(n+1-k)} \left[\alpha G_{LC}(k|n) - (1-\alpha) G_{HC}(k|n)\right]. \end{aligned}$$

To demonstrate that $\kappa(n+1) - \kappa(n) > 0$, it is sufficient to show that

$$\sum_{k=0}^{floor(S_n)} \left[\left(\frac{(n+1)}{n+1-k)}\right)\alpha - 1\right] G_{LC}(k|n)$$
$$-\left[\left(\frac{(n+1)}{n+1-k)}\right)(1-\alpha) - 1\right] G_{HC}(k|n) > 0.$$

Note that $G_{LC}(k|n) \geq G_{HC}(k|n)$ for all $n \in \mathbb{N}$ and $k \leq n$, and that, by assumption, $\alpha > \frac{1}{2}$. It follows that $\kappa(n+1) > \kappa(n)$.

Thus, as n increases, the probability that a high-cost agent is rejected increases relative to the probability that a low-cost agent is rejected. $\square$

Bibliography

Bayer, P., S. Ross, and G. Topa (2008): "Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes," *Journal of Political Economy*, 116(6).

Bikhchandani, S., D. Hirshleifer, and I. Welch (1998): "Learning from the Behavior of Others: Conformity, Fads, and Informational Cascades," *The Journal of Economic Perspectives*, 12(3), 151–170.

Black, S. (1999): "Do Better Schools Matter? Parental Valuation of Elementary Education," *Quarterly Journal of Economics*, 114(2), 577–599.

Cameron, A., J. Gelbach, and D. Miller (2006): "Robust Inference with Multi-Way Clustering," *NBER Working Paper*.

Conley, T., and C. Udry (2008): *Learning about a New Technology: Pineapple in Ghana*.

Duflo, E., and E. Saez (2002): "Participation and investment decisions in a retirement plan: the influence of colleagues' choices," *Journal of Public Economics*, 85(1), 121–148.

——— (2003): "The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence From a Randomized Experiment," *Quarterly Journal of Economics*, 118(3), 815–842.

Hellerstein, J., M. McInerney, and D. Neumark (2009): "Measuring the Importance of Labor Market Networks," Discussion paper.

Ioannides, Y., and L. Loury (2004): "Job information networks, neighborhood effects, and inequality," *Journal of Economic Literature*, 42(4), 1056–1093.

Levitt, S. D., and C. Syverson (2005): "Market Distortions when Agents are Better Informed: The Value of Information in Real Estate Transactions," Working Paper 11053, National Bureau of Economic Research.

Manski, C. (1993): "Identification of Endogenous Social Effects: The Reflection Problem," *The Review of Economic Studies*, 60(3), 531–542.

——— (1995): *Identification problems in the social sciences*. Harvard University Press Cambridge, Mass.

Schmutte, I. (2009): "Job Referral Networks and the Determination of Earnings in Local Labor Markets," Discussion paper, Center for Economic Studies–US Census Bureau.

Sorensen, A. (2006): "Social Learning and Health Plan Choice," *The RAND Journal of Economics*, 37(4), 929–945.