

Word-of-Mouth Learning in Social Networks*

Graton Gathright[†]

and

Christopher D. Wignall[‡]

November 24, 2008

Abstract

We analyze social learning by members of the Church of Jesus Christ of Latter-day Saints (Mormons) as they select real estate agents to list their homes. Mormon congregations are particularly useful for social analysis because they are important social networks to their members and are defined geographically. We argue that, conditional on observable characteristics, the assignment of church members to congregations does not suffer from an unobserved selection process. Using real estate transaction data from the largest multiple listings 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 home sellers respond more to the private information of neighbors who reside in the same congregation, suggesting that some of the social learning in this setting is due to word-of-mouth communication.

*We thank Eli Berman, Gray Calhoun, Gordon Dahl, Sam Dastrup, Silke Forbes and Joel Watson for their advice and helpful comments. We are grateful to the Wasatch Front Regional Multiple Listings Service for data access. Financial support from the California Association of Realtors is gratefully acknowledged.

[†]UC San Diego. gathright@ucsd.edu. <http://gathright.ucsd.edu>.

[‡]UC San Diego. wignall@ucsd.edu. <http://wignall.ucsd.edu>.

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 studies of peer effects: unobserved characteristics that influence behavior may also influence which relationships form¹. 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 (2005) and

¹Manski (1995) calls this the *reflection problem*.

Hellerstein et al (2008) treat census blocks 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 where the concentration of Mormons is approximately 89%². 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 wards³.

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 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 herding⁴. Furthermore, direct communication between consumers concerning personal experience with

²From the Religious Congregations and Membership in the United States, 2000, collected by Association of Statisticians of American Religious Bodies.

³The 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.

⁴For an overview of the literature on information cascades, see Bikhchandani et al (1998)

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 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.

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 communication. 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 so-

cial 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.

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, 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 Center⁵, 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.

⁵US Religious Landscape Survey, Pew Forum on Religion and Public Life

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.

The geographic boundaries 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 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 boundaries⁶. 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 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

⁶Based on the authors' private conversations with church leaders.

sample using the Church’s online ward assignment lookup tool⁷. In Table 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 3, we present summary statistics 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.

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:

$$P(i \text{ and } j \text{ choose the same agent}) = \alpha + \beta W_{ij} + \gamma_1 D_{ij} + \gamma_2 D_{ij}^2 + \gamma_3 N_{ij} + \epsilon_{ij} \quad (1)$$

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

⁷Meetinghouse Locator, www.lds.org.

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 discontinuities in house characteristics include 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 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:

$$\sigma_i(\text{Bedrooms}) = \sqrt{\frac{1}{J-1} \sum_{j \in J} [(\text{Bedrooms}_i - \text{Bedrooms}_j) - \text{Average Deviation}_i]^2}$$

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 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.

5 Evidence of Social Learning in Wards

The results from estimating equation 1, reported in Table 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 characteristics⁸. 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 5 shows, however, our estimate of the

⁸See Table 4 for the list of characteristics that are included.

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 block⁹. 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 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 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 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 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.

⁹This description of the census block definition is based largely on information available from the US Census Bureau, www.census.gov

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 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 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 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 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 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.

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 similarities¹⁰. 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.

¹⁰See Manski (2000).

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.

References

- BAYER, P., S. ROSS, AND G. TOPA (2005): “Place of Work and Place of Residence: Informal Hiring Networks and Labor Market Outcomes,” *NBER Working Paper*.
- 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 (2008): “Measuring the Importance of Labor Market Networks,” Discussion paper.
- 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.

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

YAVAS, A. (1994): “Economics of Brokerage: An Overview,” *Journal of Real Estate Literature*, 2(1), 169–95.

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

Table 1: Summary Statistics of Homes - Data are from the Wasatch Front Regional Multiple Listing Service and include single family residences listed in Utah County between 1997 and 2007.

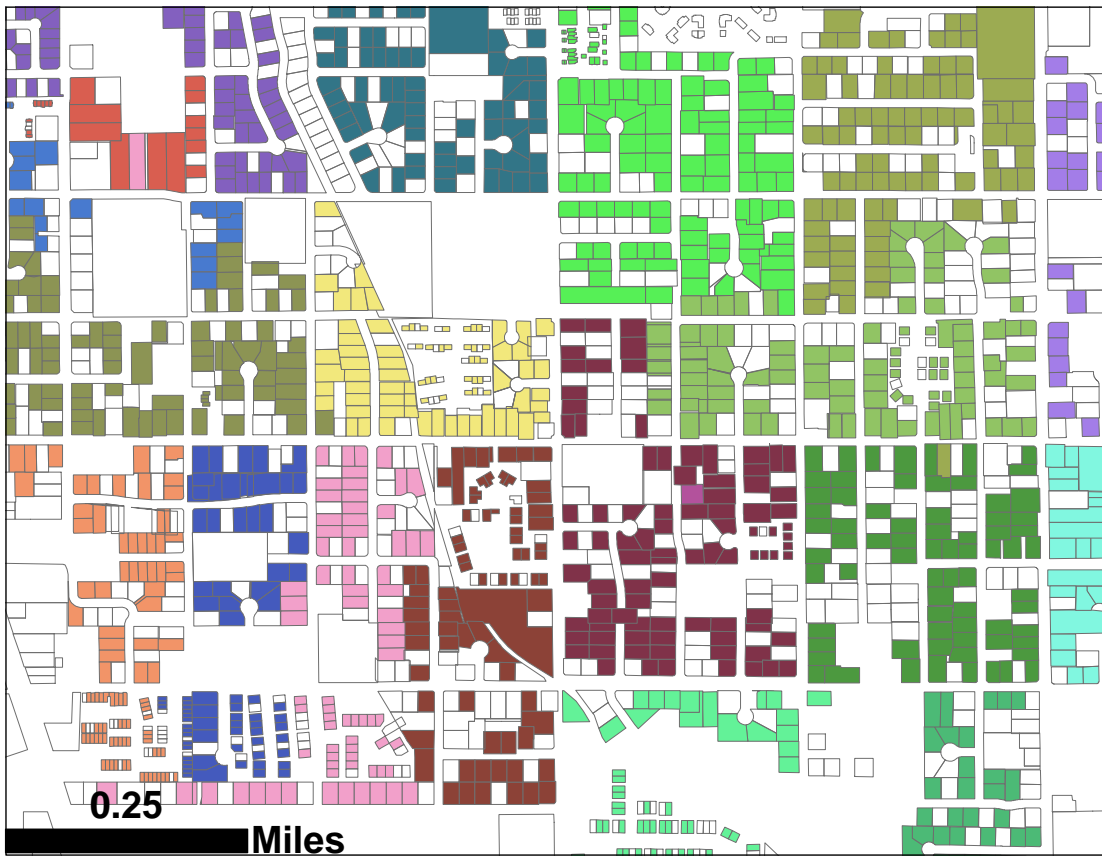


Figure 1: Ward assignment of homes in a Utah County neighborhood.

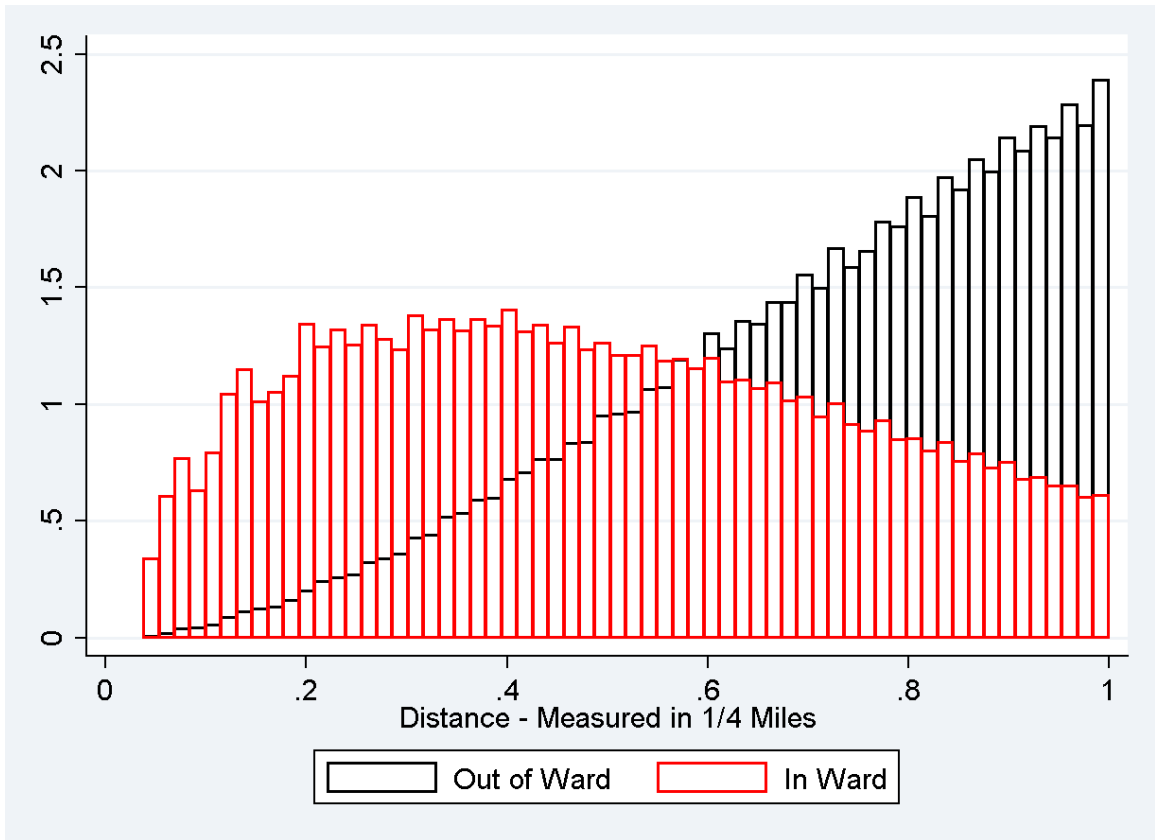


Figure 2: Histogram of Distances between neighbors in the same and in different wards.

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

Table 2: Summary Statistics of Neighbors - Geographic location of properties from Google Maps. Ward affiliation from www.lids.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 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

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

Table 4: Differences in Characteristics Across Neighbors - 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, $\sigma_i(\text{Beds}) = \text{sqrt}(\sum_{j \in J} [(\text{Beds}_i - \text{Beds}_j) - (\sum_{j \in J} [\text{Beds}_i - \text{Beds}_j]) / \#J]^2)$ 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.

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

Table 5: 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 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 subsamples 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.

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

Table 6: Linear Probability Model - Information 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.” Columns (1) - (3) use deviations from list price ($\ln(\text{SalesPrice} - \ln(\text{ListPrice}))$): column (1) defines “very high price” as a price in the top decile of the distribution of deviations from list price and a “very low price” as one in the bottom decile. Columns (2) and (3) use the top and bottom quartiles and the top and bottom five percentiles respectively. Specification (4) uses the same definition as column (1) but also includes year fixed effects for each of the neighbors. (5) and (6) use the predicted price from a hedonic regression of $\ln(\text{SalesPrice})$ on the home’s observed characteristics, including dummies for the year and ward, as the benchmark price. So, in (5) a “very high price” is $\ln(\text{SalesPrice}) - \ln(\text{SalesPrice})$ in the top decile of the distribution. Column (6) uses the top and bottom five percentiles.