Title
Essays in the Value of Intermediaries in the Real Estate Market

Author
SHUI, XI

Publication Date
2016

Peer reviewed|Thesis/dissertation

Essays in the Value of Intermediaries in the Real Estate Market

by

Xi Shui

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Nancy Wallace, Co-chair
Professor Stefano DellaVigna, Co-chair
Assistant Professor Amir Kermani
Professor Ulrike Malmendier

Spring 2016

Essays in the Value of Intermediaries in the Real Estate Market

Abstract

Essays in the Value of Intermediaries in the Real Estate Market

by

Xi Shui

Doctor of Philosophy in Economics

University of California, Berkeley

Professor Nancy Wallace, Co-chair

Professor Stefano DellaVigna, Co-chair

The thesis consists of two chapters on real estate economics.

In the first chapter, I study the impact of intermediaries in the real estate transactions. In many markets, intermediaries collect a substantial amount of commission in exchange for their expertise. Real estate is a prominent example—Americans paid more than $60 billion for real estate brokerage services in 2014. In this paper, I find a significant positive relationship between listing agents with greater recent experience and sales price, which is entirely driven by the sorting of more experienced agents into better houses. Once both observed and unobserved house characteristics are controlled for, there is no significant effect of experienced listing agents on average sales price. However, I present a novel finding that there is a significant negative relationship between recent listing agent experience and the variance of the sales price, which means that experienced listing agents add value to risk-averse sellers. Moreover, I investigate the mechanism that drives this negative relationship. Looking at individual performance, I show that a listing agent's past performance predicts his future performance, and that good past performance leads to more listings in the future. The relationship is driven by the survivorship of better agents over time rather than the accumulated expertise of the agents. The channel is consistent with survivorship bias and decreasing returns to the number of listings for listing agents, similar to Berk and Green (2004)[8].

In the second chapter, I geocode a rich real estate repeated sales dataset and map each property to its school district and neighborhood. I study how big data algorithms differ from OLS regression in predictive power and how robust those algorithms are to data stratification. I find that it is computationally expensive for the random forest algorithm to use step functions to approach the linear data generating process. Once there are fewer predictors, the RF algorithm outperforms other algorithms. This is robust to different model specifications. In addition, the random forest algorithm provides similar results under different stratifications. I also study the effect of keywords on sales price and how informative they are in predicting sales price. I find that certain keywords can be valuable in explaining varia-

tion in the data but have insignificant impact on the average sales price, suggesting that the interaction between such keywords and other house features together should be considered when we specify our models. Lastly, I am able to exploit cross time variation in school academic performance index to identify the effect of school quality on house prices controlling for neighborhood fixed effect. I find school quality has a robust significantly positive effect on property sales price.

To My Family

Contents

List of Figures

List of Tables

Acknowledgments

I am indebted to Nancy Wallace, Stefano DellaVigna and Amir Kermani for their invaluable guidance, support and encouragement. I am grateful to my family for being supportive and encouraging all the time, and to Marco A. Schwarz for his inspirations and helpful suggestions. I also thank Philipp Strack, Christopher J. Palmer, Christopher Walters, Richard Stanton, Han Lu, Hoai-Luu Q. Nguyen, Patrick Kline, Ulrike Malmendier, Ben Handel, Heidi Abramson, Sheng Li, Yury Yatsynovich, Weijia Li, David Echeverry, Sheisha Kulkarni and the participants of real estate seminar and pre-seminar at UC Berkeley for valuable comments, as well as the Oakland/Berkeley Association of Realtors for the valuable support while this research was conducted.

Chapter 1

Does Experience Matter? — The Impact of Agent Experience on Real Estate Transactions

1.1 Introduction

Do some intermediaries outperform their peers? I investigate this question using the real estate industry as a laboratory. The previous literature on listing agents provides mixed evidence. Munneke and Yavas (2001)[42] suggest that more skilled agents obtain more listings and commissions but achieve similar sales outcomes since they are busier and exert less effort per listing.[1] Recent research presents contradictory findings that more skilled agents sell houses faster and at higher sales prices (Allen et al. (2003)[2]; Johnson et al. (2008)[32]). In this paper, I study the relationship between listing agent experience and transaction outcomes. I find that more experienced agents do not sell houses faster or for a higher price[2] but merely sort into houses with better characteristics. This raises the question of whether experienced agents add value to sellers through a different channel. I address this question with a novel finding that there is a significant negative relationship between experience and sales price variance, suggesting experienced agents add value to risk-averse sellers, which is driven by the survivorship of better agents over time.

The same puzzle exists in the finance literature, which has not analyzed the relationship between intermediary experience and the variance of performance. Instead, the previous literature on mutual fund managers focuses on the average performance, finding that managers with different levels of skill generate returns similar to index funds, and also that their performance is not persistent (Berk and Green (2004)[8]).[3] The more recent literature in finance,

[1]The previous literature in real estate finds that listing agents are capable of adding value to sellers by comparing agent performance between clients' houses and their own houses (Rutherford, Springer, and Yavas (2005)[48]; Levitt and Syverson (2008)[38]).

[2]This is consistent with the results in Munneke and Yavas (2001)[42].

[3]Since Jensen (1968)[31], there is overwhelming evidence that active mutual fund managers do not

however, documents a subgroup of managers has superior skills and persistent performance over time (Wermers (2000)[52]; Kosowski et al. (2006)[35]; Kacperczyk et al. (2014)[33]). More active and experienced managers (Cremers and Petajisto (2009)[15]; Porter and Trifts (2012)[43]) have significantly better and persistent performance, and those who underperform their peers are more likely to lose their jobs (Porter and Trifts (2014)[44]).

In this paper, I use a novel data set and compare the differences in sales price and time on market of the same house sold at different times by agents at different levels of experience measured with the log number of houses sold in the last year plus one.[4] I find at first a positive relationship between listing agent experience and sales price. However, the relationship is entirely driven by the sorting of more experienced agents into better houses. Once both observed and unobserved house characteristics are controlled for, I find that there is no significant relationship between listing agent experience and either the average sales price or the average length of time on the market.

How are the experienced listing agents able to self-select into better houses given their apparent inability to outperform less experienced agents? I present a novel finding addressing this question. I document a negative relationship between listing agent experience and the variance of the sales price, suggesting that experienced agents have more predictable sales price and thus add value to risk-averse sellers. In a market with competitive commission rates, commission rates and seller's risk aversion will clear the market as experienced listing agents can simply charge higher commissions. However, because of the lack of variation in commission rates, one will see that experienced agents self-select into better houses seeking higher returns per listing. If experienced agents have more predictable performance, all else being equal, they have the power to pick houses as risk-averse sellers prefer to work with them.

How much more are risk-averse sellers willing to pay for experienced agents than inexperienced ones? I present a simple calibration assuming a risk aversion value in the range as in Chetty (2006)[14]. Under this assumption, calibration reveals the willingness to pay is roughly one thousand dollars more for agents in the top experience decile than those in the bottom decile, at least. If one assumes a higher risk aversion rate (as implied for example by the observed equity premium), the implied value would be higher.

There are two possible explanations for experienced listing agents having more predictable sales price — learning (Barwick and Pathak (2015)[4]) and selection (Malkiel (1995)[39], Kosowski et al. (2006)[35]). As agents gain experience, they gradually learn and get a

outperform passive benchmarks (Carhart (1997)[12]; Fama and French (2010)[22]). However, this does not mean they lack skill. When there is diminishing return in deploying skill, self-interested skilled managers will increase their fund size to the point where their returns equal to their passive benchmarks in equilibrium (Berk and Green (2004)[8]; Berk and Binsbergen (2015)[7]).

[4]In the rest of the paper, I use "experienced agents" to refer to "agents with more recent experience (sold listings) in the last year" rather than how long they have been in the market. Barwick and Pathak (2015)[4] and Barwick et al. (2015)[5] measure last years number of sold listings and use it as a proxy for agent skills. Barwick and Pathak (2015)[4] show that measuring by an agents number of transactions last year is qualitatively similar to measuring by the number of years an agent has been in the market.

better sense of accurate pricing. Alternatively, only agents with good past performance become experienced in the long run since sellers infer high agent ability from good past performance and high transaction volumes. To find out which explanation is more likely to be the underlying mechanism, I examine within-agent performance and find that both the average performance and its variance persist over time, and that agents who perform better in the last year have more listings in the current year. I also find more heterogeneity in agent fixed effects among the group of inexperienced agents than among the group of experienced ones. Together this evidence suggests that selection is more likely to be the underlying mechanism than learning.

I address several concerns about the empirical analysis. First, the relationship between experience and sales price may be driven by the boom-bust cycle. This concern affects both the results about average sale price and about the variance of sale price, since experienced listing agents may be especially able to outperform less experienced ones when the market is good, and boom market may be associated with larger sales price variance as there are simply more transactions. To address this concern, I control for year fixed effects and report specifications separately for boom and bust periods, finding similar results for each. Second, agents with more experience may have larger impacts on expensive houses than inexpensive ones. To investigate this possibility, I divide listings into quartiles by predicted sales price on basic house characteristics and report results for houses in different quartiles separately. Similarly, expensive houses may have a larger variance in general compared to inexpensive ones. To investigate this, I add logarithm listing price as a control in the MLE analysis to account for agent sorting into houses with more predictable sales outcomes. Third, more experienced agents may have a larger impact when the market is less competitive. To look into this, I use Zillow's turnover rate to divide the sample into quartiles and regress each quartile separately. Overall, the results are robust.

In addition to the above robustness checks, I use listing price as a proxy for sales price and investigate the relationship between experience and both listing price mean and variance. I also investigate the impact of experience on another transaction outcome — days on market. Using similar analysis, I find there is no relationship between experience and either the average or the variance of days on market.

However, just as sellers may care more about the uncertainty of sales outcomes rather than the average, they may care more about selling within a certain time period rather than the speed of sale within that period. Thus, I use whether a listing was sold within 30 days as a dependent variable and find a significant positive relationship between experience and fast sales.

To check the robustness of the results on different measures of experience, I construct three additional measures of experience. I divide listing agents into two groups (experienced and inexperienced) and pair them with similarly divided buyer agents. I create categorical measures of experience by dividing the linear measure into quartiles. I calculate another linear measure based on an agent's number of listings in the previous two years, weighted by the recency of the listing. To check the robustness of the results on different samples, I restrict the analysis to agents with more than 30 transactions in the 14 years of the study

and extend the analysis to all agents in the repeated sales sample instead of just focusing on agents with at least 20 transactions. Overall, I find the results robust to the boom-bust cycle, market competitiveness, price range differences, different measures of experience, different samples and the proxy of sales price.

The remainder of the paper is organized as follows. In Section 2, I review the previous literature. Section 3 presents the data and the summary statistics. In Section 4, I describe the empirical approach and summarize the main findings. Section 5 presents the robustness checks. Section 6 concludes the study.

1.2 Literature Review

Agent Heterogeneity and Performance

Past literature on mutual fund managers suggests that active managers as a group do not outperform passive benchmarks (Jensen (1968)[31]; Fama (1965, 1970)[21][40]; Sharpe (1991)[49], Malkiel (1995)[39]; Gruber (1996)[26]; Carhart (1997)[12]). More recently, Berk and Green (2004)[8] propose a model with managerial skill heterogeneity and decreasing returns for managers in deploying their skill as fund size grows. Their model predicts that self-interested fund managers increase their fund size and compensation to the point where the returns to investors are similar to the returns of passive benchmarks in the long-run equilibrium, suggesting that there is similar performance across managers with different levels of skills and their passive benchmarks. However, recent finance literature documents that a subgroup of managers have superior skills and persistent performance over time. Wermers (2000)[52] presents that there is a significant proportion of skilled funds beating their benchmarks prior to 1996. Kosowski et al. (2006)[35] find that star managers have superior skills in stock picking, while Kacperczyk et al. (2014)[33] show that some managers have superior skill at market timing. Cremers and Petajisto (2009)[15] find more actively managed funds outperform their benchmark with strong performance persistence. Porter and Trifts (2012)[43] show experienced managers have significantly better and persistent performance, and recently they show that managers who underperform their peers are more likely to lose their jobs (Porter and Trifts (2014)[44]).

Similarly, the literature on real estate agents suggests that listing agents are capable of adding value to sellers (Rutherford et al. (2005)[48]; Levitt and Syverson (2008)[38]), and yet skilled agents achieve similar sales outcomes and get higher commissions because of diminishing returns to number of listings.

Munneke and Yavas (2001)[42] propose a model for real estate agents with similar logic as Berk and Green (2004)[8]. More skilled listings agents with higher commission split have more listings and are able to increase the probability of sale because of higher skills. However, they exert less effort per listing as they obtain more listings and get busier. In equilibrium, skilled listing agents obtain more listings to the point where their probability of sale is similar to unskilled ones. They discuss that this prediction can be extended to sales

price by allowing the sales price to be an increasing function of agent skill. However, later literature presents contradictory findings that more skilled agents sell houses faster and at higher sales prices (Allen et al. (2003)[2]; Johnson et al. (2008)[32]).

The previous real estate literature measures agent heterogeneity by commission split[5], experience and recent experience. More recently, studies have measured agent heterogeneity based on their number of transactions in the last year and present descriptive evidence that agents with more transactions have higher sales probabilities and earn higher commissions (Barwick and Pathak (2015)[4]; Barwick et al. (2015)[5]).[6]

Why does experience matter? Barwick and Pathak (2015)[4] conclude from the interviews with real estate agents that there is significant on-the-job learning, including familiarizing oneself with professional regulations, property inventories and local amenities, and networking with potential clients. Thus, experienced agents potentially have a better sense of local market and accurate pricing. Another reason why it matters is selection. Just as more successful funds are more likely to survive (Malkiel (1995)[39]), real estate agents who sold more listings last year are more likely to remain in the market. Recent literature finds that real estate agents' number of transactions in the last year is significantly positively correlated with this year's number of transactions (Barwick and Pathak (2015)[4]), and that top mutual fund managers demonstrate persistent performance over time (Kosowski et al. (2006)[35]).

In this paper, I present a simple seller side search model with assumptions on the distribution of the maximum bids for agents with different levels of experience in Appendix B.[7] The logic behind this model is very similar to Berk and Green (2004)[8] and Munneke and Yavas (2001)[42]. Listing agents with more sold listings in the last year are more productive. But as they become busier, their average level of effort per listing decreases. Thus their average sales price does not differ from the average sales price of the inexperienced agents. The model predicts my first hypothesis test that there is no significant relationship between experience and sales price. My empirical evidence is consistent with the prediction. However, I find there is a significant negative relationship between experience and the variance of sales price, suggesting that experienced listing agents add value to risk-averse sellers. To investigate the underlying mechanism, I compare within-agent performance and find that agents have consistent average performance and variance of performance over time, and that there is more heterogeneity among experienced listing agents. These findings indicate that selection is more likely to be the underlying mechanism than learning.

[5]Munneke and Yavas (2001)[42], Allen et al. (2003)[2] and Johnson et al. (2008)[32] compare the difference in sales outcomes between agents with higher commission split (RE/MAX agents, principle agents with similar commission split as RE/MAX agents) and those with lower commission split. I do a similar exercise on RE/MAX agents with my main sample. I find that there is no difference in sales outcomes between RE/MAX agents and their counterparts in other brokerages.

[6]In 2014, NAR members in business for more than 16 years earned $70,200 and made 15 transactions a year, while those had been in the residential real estate market for three to five years earned less than half of that amount ($30,100) and had only 10 transactions on average. In addition, those who had been in the market for two years or less had a median gross income of only $8,500.

[7]Note that MLS does have a column for the record of number of offers. It is a pity that there are almost no entries in this column.

Agent Experience and Sorting

Even though the previous literature finds little evidence that listing agents add value to sellers on average, it finds a strong pattern that agents chase after high house prices, resulting in fewer transactions per agent and unchanged average real wages given the fixed commission rate and low barriers to entry in the real estate market (Hsieh and Moretti (2003)[30]). Shelef and Nguyen-Chyung (2015)[50] find that experienced agents sort into firms offering a higher commission split and invest less in productive resources. In this paper, I find strong evidence that more experienced listing agents self-select into houses with better attributes.

Moral Hazard

The previous literature explores whether incentive or competition induces agents to sell houses faster and for a higher price. Levitt and Syverson (2008)[37] find that flat-fee agents have less likelihood of sale and sell slower but at similar sales price compared to full-commission agents, if the listings is sold eventually. Another way to study the effect of agent incentives on transaction outcomes is to study in-house[8] and dual-agent[9] transactions, in which a listing agent receives promotion bonus from his agency (Han and Hong (2013)[27]). Heisler et al. (2007)[28] find that in-house transactions significantly reduce the sales price, while Kadiyali et al. (2014)[34] and Evans and Kolbe (2005)[46] find that they have no impact on the sales price but a significant negative impact on days on market. Given that there is little information on listing agent commission rate, previous research on agent incentive has focused on buyer agents. Both Barwick and Pathak (2010)[4] and Barwick et al. (2015)[5] show evidence that a higher buyer side commission rate is associated with a higher likelihood of sale, a modest impact on the days on the market and overall no effect on the sales price. For the effect of competition on sales outcomes, Barwick and Pathak (2015)[4] find evidence that increased competition is associated with a very small increase in sales price, but overall it has no impact on either the likelihood of sale or days on market. In this paper, I find no effect of in-house transaction or of dual-agent transaction on either sales price or days on market.

Network

Garmaise and Moskowitz (2004)[24] show that informed brokers are more likely to trade with each other, especially when information asymmetry is severe. Di Maggio et al. (2015)[17] provide evidence that core dealers and their connections become more valuable when there is high uncertainty in the market. In the empirical analysis, I incorporate these findings into the robustness checks. I use experience by agent pairs to check whether the results are

[8]Sellers and buyers are represented by the same agency.

[9]Sellers and buyers are represented by the same listing agent. Whether the in-house and dual agent transactions are legal is subject to state law. California state law requires listing agent to submit a disclosure for in-house or dual-agent transactions.

robust to agent network. I control for time fixed effects and check the results separately for boom and bust periods. Overall, results are robust.

1.3 Data and Summary Statistics

Data

The data for this study consists of sold listings of single-family detached residential properties listed between 01/01/2000 and 12/31/2013, collected from Multiple Listing Service (MLS)[10] for Alameda County in California.[11] The initial data set consists of 132,550 observations. After sample cleaning[12], there are 96,438 observations with detailed information on house characteristics, information on listing and buyer agents and other listing and transaction features[13] from 2000 to 2013.

Construction of Experience Index

I am left with 96,438 observations after sample cleaning. I use these observations with listing and buyer agent information to construct a linear measure of their experience. I separately count the numbers of listings and sale for each agent in the last year. I calculate the experience index by $log(1 + N)$, where N is the number of listings the listing agent sold in the last year or the number of sales the buyer agent completed in the last year. The benchmark measure of experience uses only the most recent sales. As a robustness check, I extend the measure of experience to sales in the past two years. The log form captures diminishing return to experience. Since most dependent variables in this study (log sales price, log listing prices and log days on market) also take the log form, the coefficients can be interpreted as the (approximate) elasticity of number of sold listings in the last year on transaction outcomes. I use dummy variables to flag in-house and dual-agent transactions, as the previous literature does.

Main Sample and Repeated Sales Sample

In this step, I put two restrictions on the data set to construct the main sample. Firstly, I restrict the data to listings from 2001 (inclusive) to 2013, since I only have information starting from 2000. In the robustness checks, I restrict the data to listings from 2002 (inclusive) to 2013 when I measure experience by past two years sold listings. After this step there are

[10] A database exclusive to local real estate agents with compiled information on all properties listed by them.

[11] Alameda County is located on the east side of the San Francisco Bay. According to the 2010 census, it is the 7th-most populous county in the United States. The most heavily urbanized areas are the cities of Oakland and Berkeley. See details at http://www.alameda.courts.ca.gov/pages.aspx/about-alameda-county

[12] Detailed sample construct steps are available in the online appendix.

[13] For example, finance information (cash, conventional loans, FHA loans, etc.), listing types and etc.

87,291 observations left, of which 26,680 are repeated sales - the same property sold at least twice over the 14 years of the study. I call this the "repeated sales sample", which I use for robustness checks.

In the benchmark specification, to exclude part-time and temporary agents, I focus on agents with 20 or more listings over the 14 years in the repeated sales sample.[14] This leaves me a main sample with 10,868 observations.

Summary Statistics

Table (1.1) presents the summary statistics. From Panel A, there are 10,868 observations in the main sample. A typical house is listed around $699,300 and sold around $710,000 and stays on the market for around 23 days. On average, the property listed is 53 years old, and includes 3 bedrooms, 2 bathrooms, 7 total rooms, 1.73 garages and has about 1900 square feet. The listing has on average 6 photographs and the average buyers agency earns approximately 2.7% of the sales price as commissions. There is an average 2% sales premium (the difference between the sales price and the listing price divided by the listing price). Listing agents in the sample have on average more than 6 listings in the last year and buyer agents have on average have about 3 transactions in the last year. The linear measure is 1.83 for the average listing agent experience and 1.13 for the average buyer agent experience.

Comparing the summary statistics, most of the observed house characteristics are comparable among the main sample, the repeated sales sample and the large sample. A typical listing in the main sample is listed and sold for more than the one in repeated sales sample, which in turn is listed and sold for more than the one in the large sample. This suggests that the main sample is representative of both the repeated sales sample and the large sample in observed house characteristics, and that the unobserved house characteristics may differ across samples, which leads to the difference in the sales price and listing price.

Panel B presents the matching between listing agents and buyer agents. Listing and buyer agents are divided into quartiles by their experience. Listing agents are more likely to pair with buyer agents at a similar level of experience. I use this categorical measure of experience in the robustness checks.

Figure (1.1) presents sample house characteristics by agent experience quartile. The results suggest that listing agent experience is positively correlated with the number of bedrooms. The pattern holds for other characteristics like the built year, square feet and lot size of the house, meaning more experienced listing agents sort into larger and newer houses.

This data set has multiple strengths. First of all, it contains almost every sold listing in which a listing agent was hired, and it documents the specific agents for each sale and their agency information. This is crucial to my identification of agent experience. Second, it covers

[14]There are a large number of part-time and temporary agents in the sample—about 60% of the agents only had 1 to 3 transactions in the 14 years. The top 10% of the agents control 60% of the transactions. Consistent with Munneke and Yavas (2001)[42], I focus on the agents who are relatively more active in the market.

the business cycle from 2001 to 2013, which allows me to estimate the effect of experience on transaction outcomes separately for the boom and the bust periods. Third, the data set has detailed information of a wide range of house characteristics and listing features. This allows me to control for a wide range of observed and unobserved house characteristics. Appendix A describes the detailed house characteristics included in the analyses.[15]

1.4 Empirical Analysis

In this section, I examine the impact of listing agent experience on different transaction outcomes, specifically log(sales price), log(listing price), days on market, log(days on market) and whether the listing is sold within thirty days.

Let y_{ijt} be a measure of one transaction outcome for property i in zip code j with year-quarter interaction t. The empirical specification takes the following form:

$$\begin{aligned} y_{ijt} = \beta * AgentExp_{it} + \lambda_1 * Inhouse_{it} + \lambda_2 * Dualagent_{it} \\ + \alpha_1 * X_i + \alpha_2 * K + \zeta_j \ (or + \eta_i) + \kappa_t + \epsilon_{ijt}, \end{aligned} \tag{1.1}$$

where β is the parameter of interest, $AgentExp_{it}$ is the index for experience measured by the $log(1+N_{it})$, where N_{it} is the number of listings the listing agent sold in the last year prior to the current listing of property i with year-quarter interaction t. The second set of parameters of interest is λ_1 and λ_2. $Inhouse_{it}$ indicates that the transaction is carried out with the help of agents in the same agency. $Dualagent_{it}$ indicates that the transaction is carried out with the help of the same agent. X_i is a set of house characteristics[16]. K is additional controls of this listing, for example the type of finance and the type of listing. κ_t is a year-quarter interacted fixed effect. ζ_j is a zip code fixed effect and ϵ_{ijt} reflects unobservables.

Experience and Average Sales Price

In this subsection, I examine the relationship between experience and average sale price. I find a significant positive relationship, without controlling for house fixed effects. Once observed and unobserved house characteristics are controlled for, there is no correlation between listing agents with more experience and the average sales price. This suggest that the positive relationship is driven by listing agents with more experience sorting into houses with better unobserved characteristics.

Table (1.2) presents a series of regression results with added controls progressively moving from column 1 to column 8.

The standard errors are clustered at zip code level in parenthesis from columns 2 to 8. The coefficient of listing agent experience in column 1 shows that a listing agent who sold

[15]The attributes I observed and controlled for closely follow previous literature such as Levitt and Syverson (2008)[38], Barwick and Pathak (2010)[3], Hendel et al. (2009)[29], Bernheim and Meer (2013)[9], etc.

[16]The house characteristics include indicators for missing values for each category.

7 listings in the last year is more likely to sell his current listing by 1.7% percent than an average agent who sold 6 listings, if all else is equal.

The coefficients of agent experience increase substantially in column 2, which includes year-quarter fixed effects. This correlation of experience and time can be explained with the following logic. Building upon Hsieh and Moretti (2003)[30], the wages for agents drop in bust period and those with less experience quit, so the average experience increases. Building upon Di Maggio et al. (2015)[17], even though nobody does well in sales price in bust years, experienced agents have relatively more transactions than inexperienced ones because of better networks. In the main sample, the number of listings does not decrease as much for experienced listing agents as for inexperienced ones in the bust years. This leads to a robustness check of separately estimating the relationship of listing agent experience and transaction outcomes in the boom and bust periods.

In column 3, I add zip code fixed effects and house characteristics. Both coefficients go down substantially. This suggests that the positive correlation between listing agent experience and sales price in column 3 is driven by more experienced agents sorting into houses with better characteristics.

In column 5, I add several controls. First, I add property upgrades. About 10% of listings in the main sample are advertised with remodeled features. Using the descriptive paragraph in each listing, I specifically control for the extent of remodeling at five different levels: slightly remodeled, bathroom remodeled, kitchen remodeled, both bathroom and kitchen remodeled and completely remodeled. Although previous literature does not distinguish the level of remodeling, it is important to control for property upgrades since they significantly influence the final sales price.[17]

Second, I include the type of finance of the buyer and a flag on whether the real estate agency offers a loan service to buyers. In the data set, about 85% of the transactions are financed by conventional loans. Conventional loans need bank approval and usually take longer than cash payments. Since cash payments are hassle free and are processed faster, sellers prefer cash offers if everything else is equal, and sometimes even if a cash offer is slightly lower than another loan offer. Thus, including controls for the type of finance accounts for the variation of sales prices due to buyers financial constraints.[18] Column 7 includes the commission rate paid to the buyer agency. In Barwick et al. (2015)[5], the authors find evidence that buyer agents self-select into listings with higher commission rates, and that lower buyer agent commission rates are associated with longer times on market. To take into account buyer agent sorting incentives, I control for the commission rate to buyer agency.

Third, I include the type of the listing and a flag on whether the property was listed within the last three months. In the data set, exclusive right listings[19] are 92.5% of the

[17]Results show that completely remodeled properties sold for 4.4% more than non-remodeled properties. Properties with kitchen remodeling sold for 2.1% more compared to non-remodeled properties.

[18]The results show that houses sold with conventional loans are 5.35% higher in sales price than houses sold with cash. Those sold with VA loans are 4.55% higher in sales price compared to those sold with cash.

[19]There are four different types of listings - exclusive right, exclusive agency, net listing and open listing. Exclusive right to sell listing means that the listing agent will get the listing commission upon sale no

observations, exclusive agency listings are 7%, and the rest are open listings and net listings. Different listing agreements offer different incentives to agents. For example, even though listing agents are usually reluctant to sign open listings, they may have more incentive to bring in buyers and close the deal as soon as possible due to the competition once the open listing agreement was signed.[20]

Fourth, I include controls for listing agent incentives to serve as dual agents. The additional variable is called dual variable in the MLS data set. If the listing agent double-ends the deal, he is able to get 6% of the sales price as commission in total if the commission rates to both sides are 3%. His firm may also provide a bonus to him. Thus, the agent has more incentive to serve as a dual agent so he can get dual commission. A typical seller and listing agent can negotiate beforehand and agree on the commission rate when the listing agent represents both sides. The "dual variable" is an indicator, meaning that there is an agreement, in which the seller agrees to pay a total commission rate at discount, for example 5% instead of the full 6%, if the listing agent represents both sides and the listing agent is willing to discount his commission in this circumstances. To double-end the deal, he needs to exert more effort to attract buyers who are currently not working with any agent. The coefficient of this control is negative in column 5 but not statistically significant. I also include the indicator of in-house transaction and dual agent as explanatory variables while still keeping the "dual variable" as an incentive indicator. In column 5, in-house transactions are sold on average 1% lower compared to non-in-house transactions. But once dual agent is included in column 6, the coefficient of in-house transaction is not significant anymore.

Overall, I find that listing agent experience has a positive effect on the average sales price. From column 6, these variables explain 87.2% of the variation in sales price.

However, one explanation for the effect of experience on sales price is sorting on unobservables. Experienced agents are likely to get more and better listings. At the same time, houses with better attributes are more likely to sell faster and with higher prices. Even though I control for observed house characteristics, it is possible for agents to sort into unobserved house characteristics. In that case, there is a positive correlation between listing agent experience and sales prices. Many relevant unobserved house characteristics are time-invariant, such as view, the structure of the house and the neighborhood. To control for these attributes, I include time-invariant house fixed effects. Adding such fixed effects renders the coefficient on experience essentially zero and statistically insignificant. The same results hold when I include the interaction of listing and buyer agent experience in column

matter who buys the property. Exclusive agency listing means that the sellers will not pay listing agents any commission if they sell the property by themselves. If the listing agents or other agents bring the buyers to the sellers, then the sellers have to pay the listing commission since any agent with a buyer is presumed to be procured by the marketing effort of the listing agent. Net listing means that the sellers get specified net amounts from the sales prices, and the agents get as commission the difference between the sales prices and the net amounts. Open listing means that there is non-exclusive listing agreement between sellers and agents, whoever sells the property collects the commission. Agents usually ask for a flat fee up front for advertising open listing properties.

[20]I find that there is no evidence of the listing type on the transaction outcomes, and that the re-listed properties stay on the market for 15% longer compared to new listings, if all else is equal.

8. The results indicate that hiring an listing agent who sold more listings in the last year is not different from hiring one with fewer sold listings in terms of sales price.

Experience and Sorting

In this section I present evidence of sorting of experienced listing agents into better house characteristics.

First, I regress basic house characteristics on agent experience. Results in Table (1.3) suggest that experience is significantly correlated with better house characteristics.

Second, it is very likely that listing agents sort into houses by combined characteristics rather than a single one. The left graph in Figure (1.2) shows that the sale price is positively correlated with experience without controlling for house characteristics. To investigate agent sorting on combined characteristics, I regress log sales prices on basic house characteristics and predict the house sales prices, then plot them by agent experience in quartiles. The right graph in Figure (1.2) shows that listing agents with more experience self-select into houses with higher predicted log sales prices. The sorting is even stronger in the repeated sales sample including all agents.[21]

Experience and Sales Price Variance

The above results indicate that listing agents with more sold listings last year are able to sort into houses with better characteristics. Yet, experience agents do not increase the sale price. This raises the question as to why the sellers of houses with better characteristics would seek experienced agents. Are they making a mistake?

I present a novel finding that there is a significant negative relationship between listing agent experience and the variance of sales price. If more experienced agents have more predictable performance, they have the power to pick houses as risk-averse sellers prefer to work with them, all else being equal. I take the residual sales price from the OLS regression in column 8 of Table (1.2). The residuals demonstrate heteroskedasticity. I plot the residuals against listing agent experience in Figure (1.3). It shows a very strong pattern that experienced listing agents have less variance in log sales price. However, at the same time, experienced listing agents sort into houses with better features that sell for higher prices. If the model captures houses in the higher price range better than other price ranges, then one would see a pattern similar to Figure (1.3) substituting the x axis to logarithm listing price. Thus, I plot the residuals against log listing price in Figure (1.4). The different pattern illustrates that the result is not driven by the different power of the model in different sales price ranges.

[21] I also find similar pattern for sorting on unobservables. The change in coefficients from column 6 to 7 in Table (1.2) reflects that agents are not only sorting on observables but also unobserved house features. I extract house fixed effects from the regression result in Table (1.2) column 8 and plot the fixed effects by experience in quartiles. It demonstrates the same sorting pattern.

To parameterize and estimate the effect of experience on the variance of sales price, I fit the residual sales price to a normal distribution where the variance is a function of listing agent experience. Specifically, I assume $\epsilon_{it} \sim N(0, \sigma_{it}^2)$, where $\sigma_{it}^2 = exp(\gamma * AgentExp_{it} + \kappa_t + e'_{it})$. κ_t is a year fixed effect. $AgentExp_{it}$ is the linear measure of listing agent i's experience before time t. This gives the best estimates of γ by solving the following minimization problem[22]:

$$\underset{it}{\text{minimize}}\ \gamma * AgentExp_{it} + \kappa_t + e'_{it} + \frac{\epsilon_{it}^2}{exp(\gamma * AgentExp_{it} + \kappa_t + e'_{it})}. \tag{1.2}$$

One concern is that the variance of the transaction outcomes is correlated with market trends. To address this, I report the estimation results with and without controls for year fixed effects.

Table (1.4) panel A presents the regression results. In column 1, there is a significant negative correlation between listing agent experience and the variance of residuals sales price. The result is robust to controlling for year fixed effects (column 2) and even stronger in the repeated sales sample (columns 3 and 4). Figure (1.5) shows the average predicted variance of residuals on log sales price with and without year fixed effects by listing agent experience in deciles.

Another concern is that the variance of the transaction outcomes may be driven by the sorting of experienced agents into houses sold with less variance. To investigate this possibility, I take residual sales price from a similar regression to the one in Table (1.2) column 8 with one additional control—logarithm listing price. I fit the residuals to the MLE controlling for log listing price and year fixed effects. I report the estimation results in Table (1.4) Panel B. Even columns include both controls. Columns 3 and 4 include listing and buyer agent experience interaction in addition to columns 1 and 2. One can see the robust negative correlation between listing agent experience and the variance of residual sales price. Note that the sign for buyer agent experience flips from negative to positive, meaning that controlling for experienced buyer agent sorting into expensive listings, there is a positive correlation between buyer agent experience and the residual sales price. The reason may be that experienced buyer agents are more likely to successfully negotiate the final sales prices than inexperienced ones who more likely to accept the listing price. The negative sign in front of buyer agent experience in the benchmark results from Panel A columns 1 and 2 is driven by experienced buyer agents self-selecting into expensive houses listed by experienced listing agents. The effect of experienced listing agents' better pricing strategy is picked up by experienced buyer agents. Once the sorting of buyer agents is controlled for, the better pricing strategy is attributed to listing agents (Panel B, columns 1-4).

Overall, even though there is no significant relationship between listing agents with more experience and the average transaction outcomes, I find that there is a significantly negative

[22]This is very similar to the OLS regression that $y_{it} = \gamma * AgentExp_{it} + \kappa_t + e'_{it}$, where y_{it} is the logarithm of standard deviation of residuals. However the OLS regression has fewer observations after calculating the standard deviation of residuals for each level of experience.

correlation between them and the variance of sales price. Hence, experienced listing agents add value to risk averse sellers.

Inspecting the Underlying Mechanism

In this subsection, I inspect the underlying mechanism driving the above results. I explore two potential explanations - learning and selection. Learning means that agents get a better sense of the right price as they gain experience. Selection means that only good agents become experienced in the long run, since sellers infer high agent ability from good past performance and high transaction volumes.

As mentioned in the literature review, one of the elements in the model of Berk and Green (2004)[8] is diminishing returns to mutual fund managers in deploying their ability. They also find there is a high level of skills in active managers. Consider a case where there is less heterogeneity in ability and higher average ability for experienced listing agents. Experienced listing agents will obtain listings to a point where the average sales price is similar to inexperienced ones. Since they also have less heterogeneity, they can even have a lower sales price to a point that risk-averse sellers' utilities are similar between hiring experienced and inexperienced agents. However, they will not obtain listings or deploy their ability as much, when agent ability is not perfectly observable to sellers, sellers need time to update their beliefs, or listing agents have capacity constraints so that listings are not perfectly mobile.

In the appendix model, listing agent ability is mapped onto the number of buyers they attract. Due to survivorship of agents over time, there is more uncertainty in the number of buyers an inexperienced agent attracts than an experienced agent does, so one will see a larger sales price variance for inexperienced agents.

Learning

I examine the within-agent average performance and variance of performance over time.

First of all, I find that within-agent average performance persists. I summarize and collapse the data into the average performance per agent per year. Within-agent, I regress this year's average residual on last year's average residual weighted by the number of listings this year. The regression takes the form:

$$y_{it} = y_{i,t-1} + \kappa_t + \epsilon''_{it}, \tag{1.3}$$

where y_{it} is year t's average residual for agent i. κ_t is a year fixed effect. Table (1.5) presents the regression results for both the main sample and the large sample restricted to agents with 20 or more listings in the 14 years of the study. On average, a listing agent performance this year is significantly positively correlated to his performance in the last year. The residuals for the first two columns in Table (1.5) are from the regression in Table (1.2), column 8. The

residuals for the third and fourth columns are from regressing log sales price on listing agent experience and the full set of controls except, substituting zip code fixed effects to house fixed effects. The results are robust to controlling for year fixed effects.

Second, I find that the variance of agent performance persists over time. If learning is the underlying mechanism, one will observe that the listing agents have less variance of performance in later listings than those early in their career. Thus, I focus on the 269 listing agents who entered the market after 2002 and have more than 20 listings over the 14 years of the study. I calculate their average residuals and the confidence interval of residuals for their first 10 listings and second 10 listings and plot them on the left side of Figure (1.6). On the right side of Figure (1.6), I do the same analysis with 78 listing agents who entered the market after 2002 and have more than 40 listings over the 14 years. I calculated their average residuals and the confidence interval of residuals for their first 20 listings and second 20 listings. Both results suggest that the average and variance of agent performance are persistent along their career. Overall, learning is not very likely to be the reason for the findings in this study.

Selection

To explore the selection explanation, I first examine the between-agent heterogeneity in unobserved quality.

I estimate agent fixed effects with a full set of controls as in Table (1.2), column 8. I plot the estimated fixed effect on the left side of Figure (1.7). I correct for the high dimensional fixed effect estimation and infer the best unbiased estimates from the fixed effect estimation. Specifically, I define μ_j as the fixed effect for listing agent j. $\hat{\mu_j}$ is the estimated fixed effect for listing agent j from the high dimensional fixed effect estimation. Thus,

$$\hat{\mu_j} = \mu_j + e_j, \tag{1.4}$$

where $\mu_j \sim N(\mu_0, \sigma_u^2)$. Estimates of the mean and variance of μ_j are given by

$$\begin{aligned} \hat{\mu_0} &= \frac{1}{J} \sum_j \hat{\mu_j}, \\ \hat{\sigma_u^2} &= \frac{1}{J} \sum_j [(\hat{\mu_j} - \hat{\mu_0})^2 - SE(\hat{\mu_j})^2]. \end{aligned} \tag{1.5}$$

The inferred fixed effect for listing agent j is

$$E(\mu_j|\hat{\mu_j}) = \mu_j^* = \lambda_j * \hat{\mu_j} + (1 - \lambda_j) * \hat{\mu_0}, \tag{1.6}$$

where $\lambda_j = \frac{\hat{\sigma_u^2}}{\hat{\sigma_u^2} + SE(\hat{\mu_j})^2}$. I plot the inferred (corrected) fixed effect estimation on the right side of Figure (1.7). There is a wider spread of agent fixed effects for less experienced listing

agents than for experienced ones, suggesting more heterogeneity in unobserved quality among less experienced ones.

Second, I inspect the relationship between agent past performance and number of listings in the current year. In Table (1.5) Panel B, I regress within-agent the number of listings this year on his average residual sales price last year. I control for last year's number of listings, since sellers may infer agent ability from volume in addition to the average performance. I find there is a significant positive correlation between past performance and this year's number of listings and also between last year's number of listings and this year's number of listings. It suggests that agents are more likely to be successful if they have more listings and have performed better in the past.

Third, I inspect past performance and the possibility of losing business in the current year. I define an agent as losing business if he had less than half of the listings this year compared to the previous year. In Table (1.6) Panel C, I report the results from logistic regression. There is a negative sign in front of the average performance coefficient. It is not significant in the main sample, while marginally significant for the large sample. This provides suggestive evidence that a listing agent is less likely to lose business if he has better past performance.

Overall, I find supporting evidence for selection and little evidence for learning.

1.5 Robustness Checks

In this section, I examine the robustness of the results to the proxy of sales price, boom-bust cycle, market competitiveness, price ranges, different measures of experience, different samples, and marketing strategy. I explore the relationship between listing agent experience and fast sales.

Experience and Other Sales Outcomes

In this subsection, I use listing price as an alternative dependent variable to sale price in the robustness check.[23] I also check the effect of experience on days on market and the logarithm of days on market.[24]

Table (1.6) presents the robustness check results with linear and categorical measures of experience in Panel A and B respectively. In both panels, the odd columns report regression results of transaction outcomes on listing agent and buyer agent experience only. Results show that listing agent experience is positively correlated with listing price and negatively correlated with days on market. The even columns report regression results of transaction outcomes on listing agent and buyer agent experience with a full set of controls as in Table

[23]Previous study shows that listing price is highly correlated with sales price.

[24]This also serves the purpose of checking whether experienced listing agents have better pricing strategy than their less experienced counterparts. If experienced ones have more accurate pricing, then they are able to sell faster.

(1.2) column 8. With a full set of controls, listing agent experience has no impact on the average listing price, days on market, or logarithm days on market. It suggests that experienced listing agents merely sort into houses that sold faster and for a higher price without achieving higher prices or shorter days on market compared to inexperienced ones.

I explore the impact of listing agent experience on the variance of other sales outcomes. In Figure (1.9), I present the graphical evidence showing experienced listing agents have a smaller variance in listing price compared to their less experienced counterparts. The first two columns in Table (1.4) panel D show that this result is significant and of similar size compared to the effect on log sales price. Investigating the relationship between experience and days on market, I find there is no significant correlation between them, as shown in columns 3 and 4 in Table (1.4) Panel D. This indicates that the variance of days on market is consistent across agents with different levels of experience.

Measure of Experience

So far, I have measured agent experience by the logarithm of last year's sold listings plus one. Now I provide three different measures of experience and show that results are robust.

Firstly, I check the robustness of the relationship between listing agent experience and the average sales price. In Table (1.8) Panel A, I pair listing and buyer agents by their experience. The results suggest that there is a positive relationship between the experience-experience pair and sales price. With a full set of controls, there is no significant correlation. In Panel B, I measure listing agent experience by the logarithm of one plus their sold listings in the last two years weighted by year.[25] The sample changes to the main sample from 2002 to 2013 because of the change of measure. Even though there is a significant positive effect of listing agent experience on sales price with a full set of controls, the effect is very small. In Panel C, I divide listing and buyer agents into quartiles by their experience and regress log sales price on listing agent quartiles controlling for buyer agent quartiles. Listing agents with 12 or more listings are not associated with higher sales price controlling for sorting on house characteristics.

Second, I present the robustness check results on the relationship between listing agent experience and the variance of sales price. Similarly in Table (1.4) panel C, I measure experience by the logarithm of one plus sold listings in the last two years weighted by year. The results are robust and of similar size as in columns 1 and 2 in Panel A. The results are even stronger in the repeated sales sample including all agents.

Other Robustness Checks

Boom and Bust Periods: I show that the results are not driven by boom-bust cycles. As discussed above, the change of coefficients from column 2 to 3 in Table (1.2) reflects

[25]Sold listings before the last year are weighted by 50% for each additional year before the last year. The result is robust for different weights.

that there is a correlation between experience and time fixed effects. Separately for boom and bust periods, I regress log sales price and log days on market on agent experience with a full set of controls as in Table (1.2) column 8. Table (1.7) Panel A shows that there is no significant correlation between listing agent experience and other sales price or days on market in any period.

Market Competitiveness: Similarly, I show that the results are not driven by differences in the behavior of agents in markets more competitive or less competitive than average. I use Zillow's turnover rate to identify the competitiveness for each zip code in Alameda County and divide the sample into quartiles. Separately for the level of market competitiveness, I regress log sales price and log days on market on agent experience with a full set of controls, as in Table (1.2) column 8. In Table (1.7) Panel B, I report the regression results for each quartile. The results are similar across different competitive markets.

Price Range: I show that the results are robust for houses in different price ranges. In Table (1.7) Panel C, I divide the listings into quartiles by predicted sales price on basic house characteristics. For expensive listings, there is a slightly positive effect of listing agent experience on sales price. However, it is not robust once I change the measure of experience to previous two years sold listings. Thus, there is a similar effect of experience on average sales price for houses in different price ranges.

Sample: Thus far, I have restricted agents to those who have more than 20 transactions in 14 years. For the robustness check of experience on the variance of sales price, I already show that the results are robust including all agents in Table (1.4) Panel A column 3 and 4. The results are also robust limiting agents to those who have 30 or more listings in the 14 years. From Figure (1.9), one can see the similar pattern demonstrated in Figure (1.3) confirming that the results are not sensitive to different samples.

Marketing Strategies: Even though the marketing strategy of listing agents are endogenous to them, one may argue that marketing is also informative of the real condition of the house. Thus I show that the results are robust controlling for listing agent marketing strategies. I adopt the method in Levitt and Syverson (2008)[38] to include the keywords in the listing text describing the house. In addition, I include the number of pictures used in the listings as a proxy for staging. In Appendix A, there is a detailed description of the list of keywords used. Table (1.9) confirms that the results are robust to marketing strategies of listing agents.[26]

Fast Sales: Sellers may not care about one or two days difference in the length of time on the market. They rather care about whether the agent can successfully sell the listing within a certain time, for example a month. Thus, I use whether a listing was sold within 30 days as a dependent variable. I call it "fast sales" if the dependent variable equals one. I use fixed effect logit regression to estimate the relationship between listing agent experience and fast sales. Table (1.10) Panel A present the regression results with adding controls. The

[26]I find that some words and phrases have significant impact on the sales price. For example, "must see" and "new" increase sales price by 2.25% and 2.84% respectively. "needs" and "tlc" reduce sales price by 8% and 6% respectively. Descriptive words like "amazing", "stunning", "fabulous ", "landscaped", "charming" and "beautiful" increase sales price by 3.02%, 2.98%, 1.82%, 1.49%, 1.44% and 0.87% correspondingly.

more experience, the more likely the listing is sold within 30 days. The results are robust with repeated sales sample (Panel B).

1.6 Conclusions

Sellers and buyers interact through intermediaries in many markets. There is mixed evidence in both the real estate and finance literature on whether a subgroup of intermediaries outperform their peers. In this paper, I focus on real estate listing agents and study the relationship between their experience and transaction outcomes. Specifically, I study whether experienced listing agents have better sales outcomes, such as higher sales price and shorter days on market compared to their inexperienced counterparts.

I find that the significant positive relationship between listing agent experience and sales prices is entirely driven by the sorting of more experienced agents into better houses. Once observed and unobserved house fixed effects are controlled for, there is no significant effect of experienced listing agents on the average sales price. Why are sellers with better houses prefer working with experienced agents given they do not achieve higher sales price?

I solve this puzzle with a novel finding. There is a significant negative relationship between experience and sales price variance, suggesting that experienced agents have more predictable sales price and thus add value to risk-averse sellers. I find supporting evidence that the underlying mechanism is the survivorship of better listing agents over time rather than accumulated expertise by the listing agents.

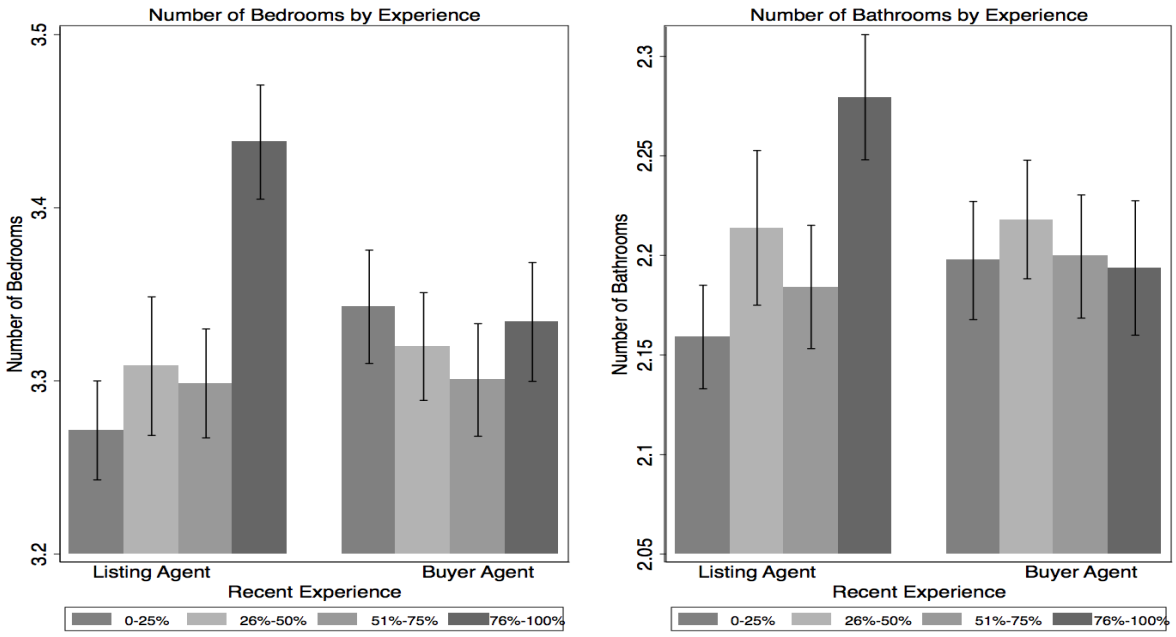
This paper is among the first to study the effect of experience on the variance of transaction outcomes. One needs to take into consideration the uncertainty of outcomes in order to calculate the value of intermediaries since it is the distribution rather than the average alone that affect sellers' and investors' utility.

Using average CRRA risk parameter estimated in Chetty (2006) [14], a seller is willing to pay one thousand dollars more to hire an agent in the top decile instead of those in the bottom decile.[27] Given that an average house is around $700k in the sample, this justifies experienced listing agent commission rate by about 5% of their earnings per listing. Note this is a lower bound estimation for the value of experienced agents as 1) I am comparing

[27]Assume sellers have CRRA utility functions $U = \frac{1}{1-\theta}[p(1-\tau)]^{1-\theta}$ for $\theta > 0$ and $\theta \neq 1$, and $U = log(p(1-\tau))$ for $\theta = 1$, where θ is the risk aversion parameter, τ is the commission rate, and p is the sales price. Sellers face different distributions of sales price when they hire listing agents with different levels of experience. To calibrate the sales price distribution for agents in the top and bottom experience deciles, I take their residuals ϵ_{it} from the regression results in Table (1.2) column 8. Since experienced agents achieve the same average sales price as inexperienced ones, they both have the same average sales price $\bar{p}$. The variance of sales price takes the form $Var(exp(log(\bar{p}) + \epsilon_{it}))$. I calculate sellers' utilities with the sales price distributions and the average risk aversion parameter in Chetty (2006) [14]. I then calculate how much the average sales price needs to increase for the sellers who hire inexperienced agents so that their expected utility equals the expected utility of those who hire experienced agents. Chetty (2006) [14] estimates the average CRRA risk parameter $\theta = 1$ with an upper bound $\theta = 2$. Using the upper bound value, a seller is willing to pay one thousand two hundred dollars to hire listing agent in the top 10% instead of one in the bottom 10%.

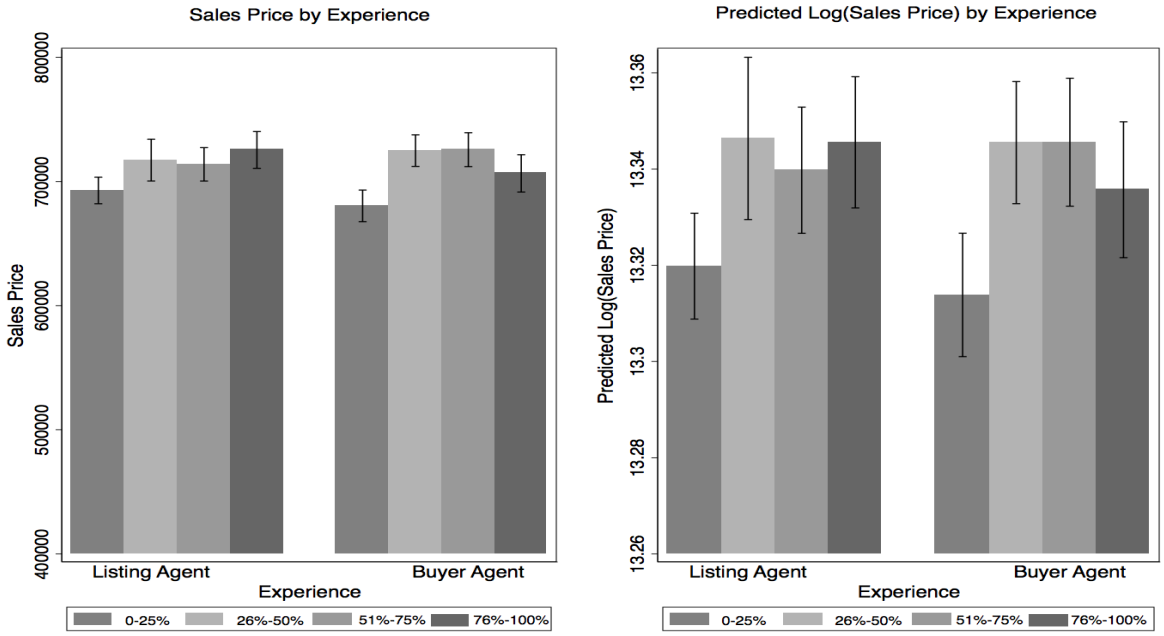
them to their inexperienced peers rather than non-agent sellers; 2) I have not taken into account the fact that experienced agents are more likely to sell listings within a month; 3) I am comparing the transaction outcomes within sold listings while it is likely that experienced agents have a higher probability of sale.

Figure 1.1: Sample House Features by Experience



Notes: Figure (1.1) shows that the listing agent experience is positively correlated with the number of bedrooms of the house, suggesting more experienced listing agents sort into houses with better features. The pattern is also strong for other house features like the square feet of the house, number of garages and the age of the house.

Figure 1.2: Sorting of Experienced Agents into Better Houses



Notes: I regress log sales prices on basic house characteristics. I take the predicted house sales prices based on basic house characteristics and plot them by agents' recent experience quartiles. There is strong evidence of listing agents with more recent experience sorting into houses with better characteristics and higher predicted sales prices.

Figure 1.3: Sample (2001 - 2013): Residual Log(Sales Price) by Agent Experience

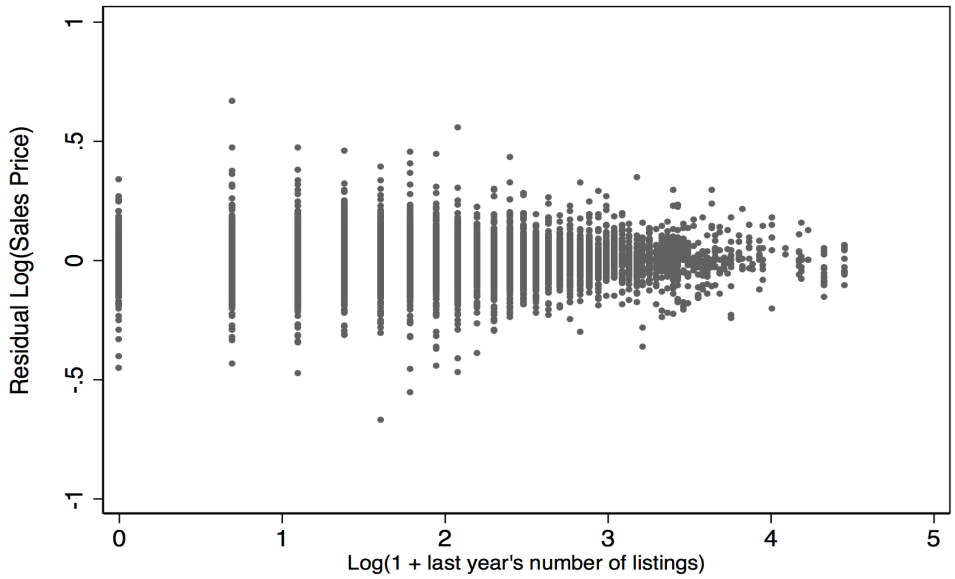
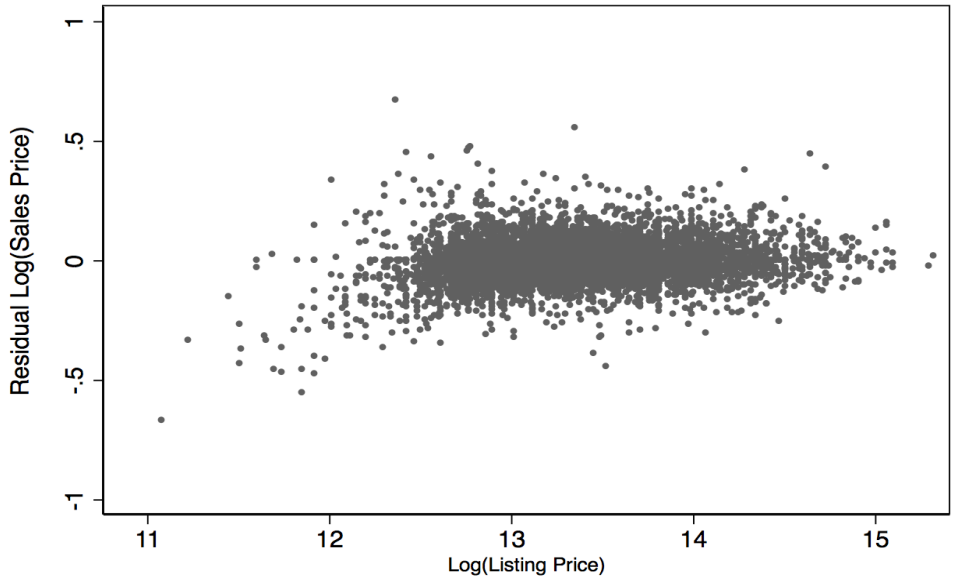
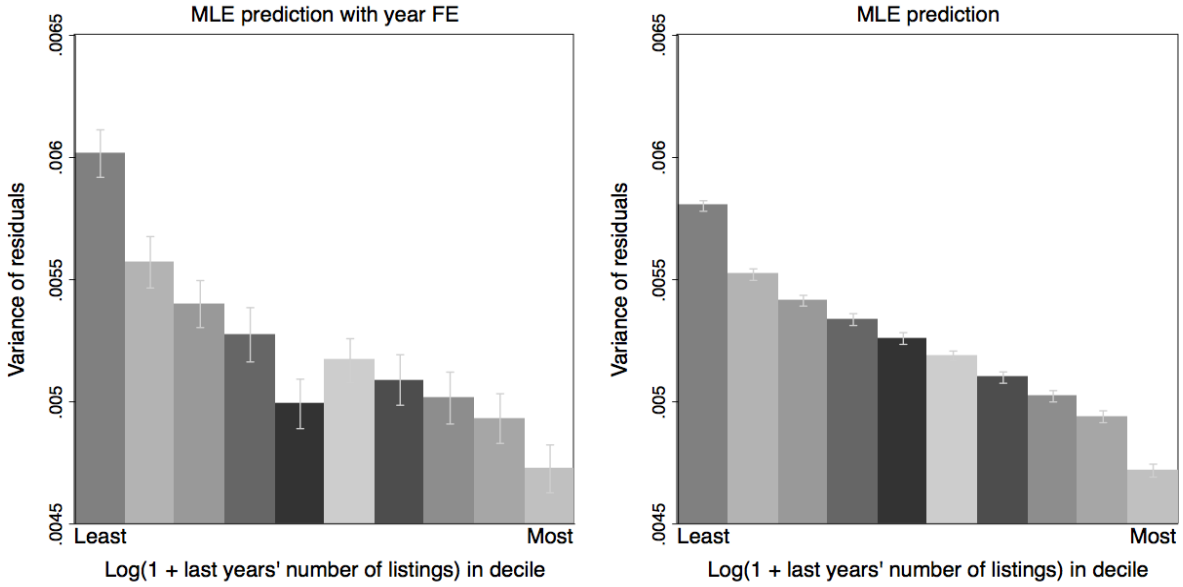


Figure 1.4: Sample (2001 - 2013): Residual Log(Sales Price) by Log(Listing Price)



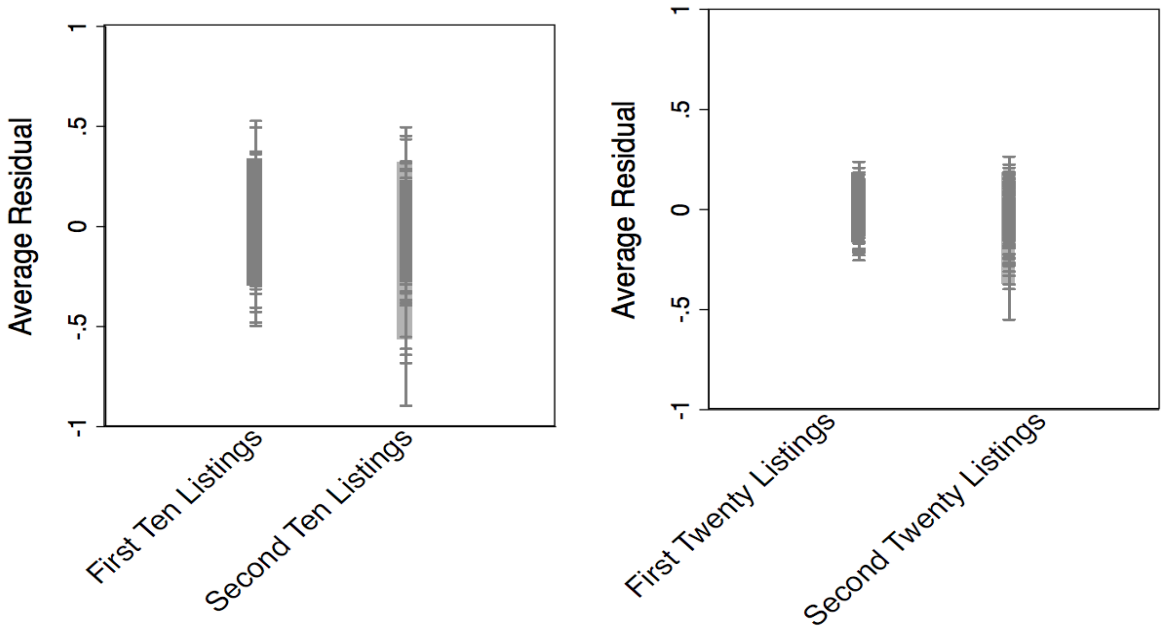
Notes: In Figure (1.3), the residuals are from regression results in Table (1.2) column 8. The scatter plot illustrates that there is a negative relationship between experienced listing agents and the variance of sales price. An alternative explanation for this effect is that more experienced agents sort into better listings and the model fits expensive listings better. Figure (1.4) shows that the alternative explanation is not likely to be true.

Figure 1.5: Predicted Variance of Residuals on Log(Sales Price)



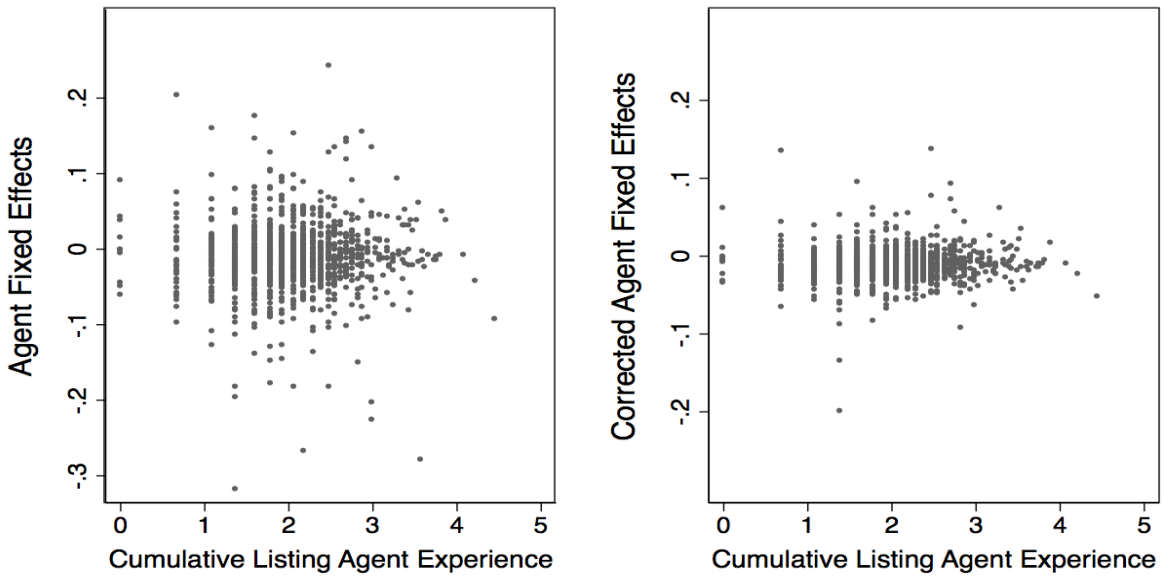
Notes: Figure (1.5) presents the predicted variance of residuals of log sales price from the maximum likelihood estimation in Table (1.4) Panel A by experience deciles with and without year fixed effects controls.

Figure 1.6: Large Sample: Within-agent Performance



Notes: The left graph shows both the average and the confidence interval of residuals sales price for the first and the second 10 listings of 269 listing agents who entered the market after 2002 (inclusive) and have more than 20 listings over the 14 years of the study. I do similar exercise for the 78 listing agents who entered the market after 2002 (inclusive) and have more than 40 listings over the 14 year. The graph on the right shows the average and the confidence interval of residuals sales price for their first and second 20 listings.

Figure 1.7: Main Sample: Agent Fixed Effects



Notes: The left graph shows the agent fixed effects (unobserved agent quality) against experience. The agent fixed effects are taken from regressing log sales price on agent experience controlling for house fixed effects and including other controls as in Table (1.2) column 8. The right graph presents the corrected agents fixed effects inferred from the estimated fixed effects from the high dimensional fixed effect regression.

Figure 1.8: Sample (2002 - 2013): Residual Log(Sales Price) by Agent Experience

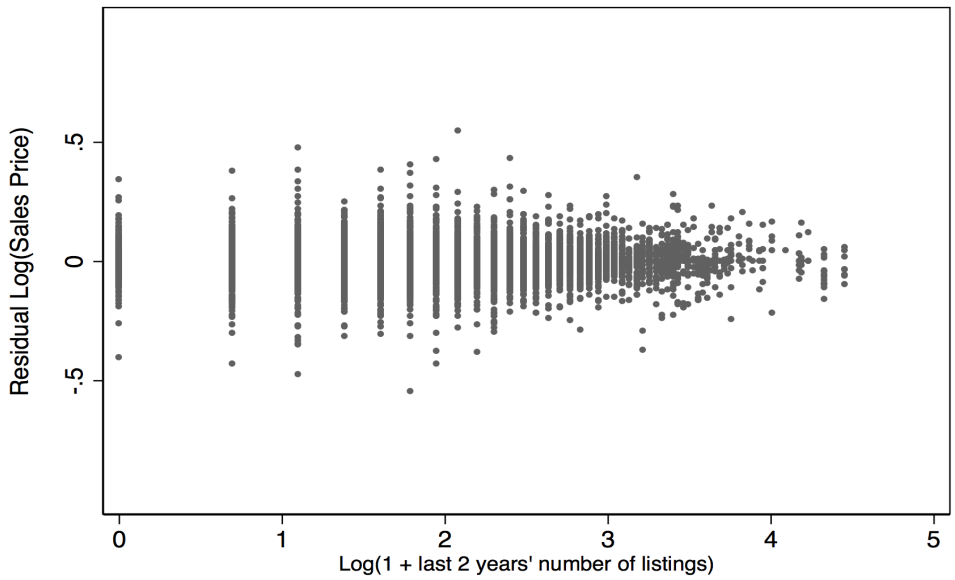
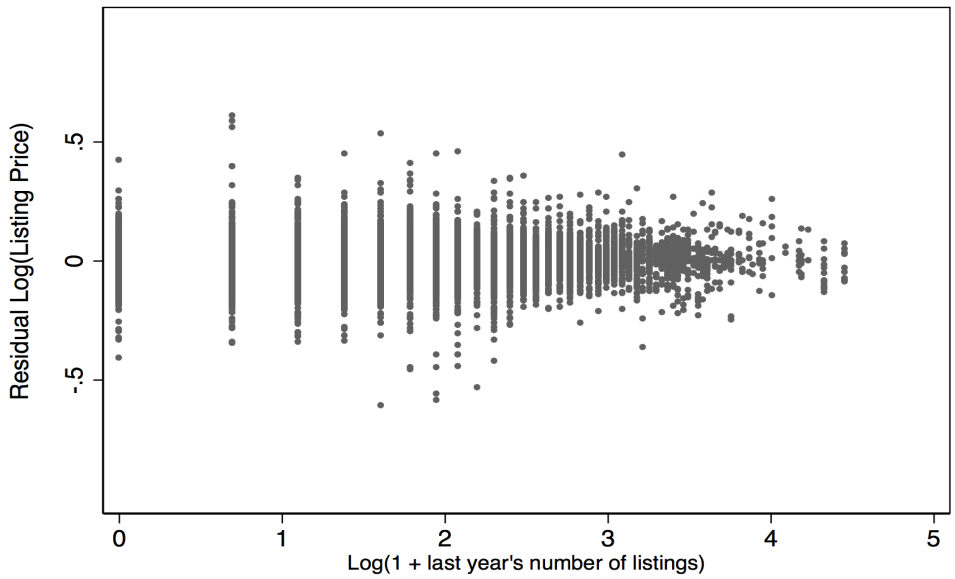


Figure 1.9: Sample (2001 - 2013): Residual Log(Listing Price) by Agent Experience



Notes: These are robustness checks of the results in Figure (1.3) limiting the sample to agents with 30 or more transactions in the 14 year and using listing price as an alternative dependent variable as sales price. Results are robust.

Table 1.1: Summary Statistics

Panel A	Main Sample		Repeated Sales Sample		Large Sample	
	Mean	SD	Mean	SD	Mean	SD
Sales Price	709967.82	358387.34	643226.67	317067.37	626948.3	316410.47
Listing Price	699286.98	361874.54	634811.18	320710.42	620392.46	320989.92
Days on Market	23.39	25.25	24.86	26.34	25.92	27.57
Number of Bedrooms	3.32	0.87	3.26	0.86	3.29	0.88
Number of Bathrooms	2.04	0.76	1.96	0.75	1.97	0.77
Total number of rooms	6.99	1.7	6.79	1.68	6.82	1.72
GarSp	1.73	0.86	1.65	0.85	1.67	0.84
Square feet	1908.09	820.92	1772.66	759.78	1781.91	774.91
Lot size in sqft	7146.07	4890.47	6655.1	4393.34	6910.54	5025.73
Number of Pictures	6.67	6.9	6.39	6.73	6.23	6.71
Age	53.41	28.84	55.19	28.55	55.14	27.95
Commission rate	2.69	0.27	2.68	0.28	2.68	0.29
Change in Price	0.02	0.07	0.02	0.07	0.02	0.07
Listing Agent: Log(1 + Last Year's # Listings)	1.83	0.85	1.28	1.01	1.27	1.03
Buyer Agent: Log(1 + Last Year's # Sales)	1.13	0.85	0.99	0.86	0.97	0.87
Observations	10868	10868	26680	26680	87291	87291

Panel B	Main Sample					Repeated Sales Sample				
					Buyer Agent					
Listing Agent	0-25%	26%-50%	51%-75%	76%-100%	Total	0-25%	26%-50%	51%-75%	76%-100%	Total
0-25%	1059	1140	905	584	3688	2761	2036	893	1060	6750
26%-50%	463	542	486	363	1854	2739	2838	1526	1764	8867
51%-75%	655	777	748	600	2780	1406	1567	978	1365	5316
76%-100%	568	630	626	722	2546	1439	1513	1030	1765	5747
Total	2745	3089	2765	2269	10868	8345	7954	4427	5954	26680

Table 1.2: Main Sample (2001-2013), List Agent Experience on Log(Sales Price)

	Dependent Variable: Log(Sales Price)							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Listing Agent:	0.0101**	0.0191***	0.00839***	0.00787***	0.00803***	0.00782***	0.00459	0.00488
Log(1 + Last Year's # Listings)	(0.00503)	(0.00653)	(0.00259)	(0.00291)	(0.00287)	(0.00281)	(0.00328)	(0.00323)
Buyer Agent:	0.00458	0.0400***	0.0103***	0.00994***	0.0109***	0.0123***	0.00294	0.00432
Log(1 + Last Year's # Sales)	(0.00500)	(0.00784)	(0.00242)	(0.00245)	(0.00240)	(0.00254)	(0.00322)	(0.00367)
Agree to lower commission					-0.00139	-0.000513	0.00746	0.00740
if dual-agent					(0.00537)	(0.00513)	(0.00610)	(0.00607)
In-house					-0.00946*	0.00894	-0.00501	-0.00497
					(0.00517)	(0.00585)	(0.00670)	(0.00671)
dual-agent						-0.0403***	-0.0141	-0.0131
						(0.0101)	(0.0106)	(0.0101)
Listing and Buyer Agent								-0.00217
Interactions (demean)								(0.00313)
Controls:								
Year Quarter FE	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Zipcode FE	No	No	Yes	Yes	Yes	Yes	No	No
House Characteristics	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Remodel	No	No	No	Yes	Yes	Yes	Yes	Yes
Type of Finance	No	No	No	Yes	Yes	Yes	Yes	Yes
Loan services from Realtor Agency	No	No	No	Yes	Yes	Yes	Yes	Yes
Buyer Agent Commission	No	No	No	Yes	Yes	Yes	Yes	Yes
Listing Types	No	No	No	Yes	Yes	Yes	Yes	Yes
Flag for Re-listed Home	No	No	No	Yes	Yes	Yes	Yes	Yes
Property FE	No	No	No	No	No	No	Yes	Yes
Observations	10868	10868	10868	10868	10868	10868	10868	10868
R-squared	0.001	0.175	0.868	0.871	0.871	0.872	0.973	0.973
Adjusted R-squared	0.000	0.171	0.867	0.870	0.870	0.870	0.941	0.941

Notes: This table reports the OLS regression results on the effect of listing agent experience on log sales price. Experience is measured by log(1+ last year's number of sold listings) per agent. In columns 3-8 standard errors are clustered at zip code level. The sample is the main sample with 10,868 repeated sales listings listed from 2001 to 2013 by listing agents with 20 or more listings. Appendix A provides detailed information on the observed house attributes included in the analysis. Significant Levels: * p<0.1, ** p<0.05, *** p<0.01

Table 1.3: Main Sample (2001-2013), Sorting of Listing Agents and Buyer Agents into Better House Characteristics

	Sorting of experienced listing agents into better house and listing attributes							
	Bedroom	Bathroom	Total Rooms	Age	Garage Space	Sqft	Lot Size	Fireplaces
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Listing Agent:	0.0678***	0.0475***	0.0426***	-0.0872***	0.0646***	0.0352**	0.0444***	0.0326**
Log(1+Last Year's #Listings)	(0.0113)	(0.0113)	(0.0113)	(0.0113)	(0.0114)	(0.0120)	(0.0109)	(0.0112)
Observations	10868	10868	10868	10868	10868	10868	10868	10868
	Sorting of experienced buyer agents into better house and listing attributes							
	Bedroom	Bathroom	Total Rooms	Age	Garage Space	Sqft	Lot Size	Fireplaces
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Buyer Agent:	-0.00866	-0.0176	0.00484	0.0307**	-0.0282*	0.0156	0.00397	0.00336
Log(1+Last Year's #Sales)	(0.0113)	(0.0113)	(0.0113)	(0.0113)	(0.0114)	(0.0120)	(0.0109)	(0.0111)
Observations	10868	10868	10868	10868	10868	10868	10868	10868

Notes: This table reports the OLS regression results of listing agent experience on basic house attributes at listing level. Experience is measured by log(1+ last year's number of sold listings) per agent. The sample is the main sample with 10,868 repeated sales listings from 2001 to 2013 listed by listing agents with 20 or more listings. Significant levels: * p<0.1, ** p<0.05, *** p<0.01.

Table 1.4: Maximum Likelihood Estimation: Impact of Experience on the Variance of Transaction Outcomes

Panel A	Variance of Residuals of Log(Sales Price)			
	Main Sample (2001-2013)		Repeated Sales Sample (2001-2013)	
	(1)	(2)	(3)	(4)
Listing Agent: Log(1 + Last Year's # Listings)	-0.0295*** (0.00801)	-0.0193** (0.00827)	-0.0685*** (0.00428)	-0.0638*** (0.00434)
Buyer Agent: Log(1 + Last Year's # Sales)	-0.0436*** (0.00796)	-0.0142* (0.00816)	-0.0418*** (0.00500)	-0.0144*** (0.00515)
Year FE Control	No	Yes	No	Yes
Observations	10868	10868	26680	26680
Panel B	**Main Sample**			
	Variance of Residuals of Log(Sales Price)			
Listing Agent: Log(1 + Last Year's # Listings)	-0.0438*** (0.00819)	-0.0444*** (0.00838)	-0.0448*** (0.00828)	-0.0439*** (0.00846)
Buyer Agent: Log(1 + Last Year's # Sales)	0.0338*** (0.00847)	0.0484*** (0.00857)	0.0329*** (0.0104)	0.0525*** (0.0106)
Listing and Buyer Agent Interactions (demean)			-0.000365 (0.00926)	-0.00667 (0.00925)
Year FE Control	No	Yes	No	Yes
Log(Listing Price) Control	Yes	Yes	Yes	Yes
Observations	10868	10868	10868	10868
Panel C	Variance of Residuals of Log(Sales Price)			
	Main Sample (2002-2013)		Repeated Sales Sample (2002-2013)	
Listing Agent: Log(1 + Last 2 Years' # Listings)	-0.0353*** (0.00995)	-0.0227** (0.0104)	-0.0737*** (0.00417)	-0.0650*** (0.00423)
Buyer Agent: Log(1 + Last 2 Years' # Sales)	-0.0305*** (0.00945)	-0.00661 (0.00961)	-0.0447*** (0.00489)	-0.0170*** (0.00499)
Year FE Control	No	Yes	No	Yes
Observations	7704	7704	24366	24366
Panel D	**Main Sample (2001-2013)**			
	Variance of Residuals of Log(LP)		Variance of Residuals of Log(DOM)	
Listing Agent: Log(1 + Last Year's # Listings)	-0.0365*** (0.00791)	-0.0226*** (0.00818)	-0.00225 (0.00789)	0.00113 (0.00809)
Buyer Agent: Log(1 + Last Year's # Sales)	-0.0475*** (0.00775)	-0.0206*** (0.00798)	0.00559 (0.00809)	0.00689 (0.00833)
Year FE Control	No	Yes	No	Yes
Observations	10868	10868	10868	10868

Notes: Panels A - C report the MLE results of listing agent experience on the variance of residuals sales price from regression results in Table (1.2) column 8. Panel D presents results of similar exercise on logarithm listing price and days on market. Significant levels: * $p<0.1$, ** $p<0.05$, *** $p<0.01$.

Table 1.5: Within-agent Performance

Panel A	Dependent Variable: Average Residual This Year			
	Main Sample (2001-2013)		Large Sample (2001-2013)	
	(1)	(2)	(3)	(4)
Average Residual Last Year	0.118***	0.118***	0.315***	0.315***
	(0.016)	(0.022)	(0.01)	(0.021)
Year FE Controls	No	Yes	No	Yes
Observations	4027	4027	10046	10046
R-squared	0.016	0.016	0.091	0.093
Adjusted R-squared	0.013	0.013	0.189	0.092

Panel B	OLS: # of Listings This year	
	Main Sample (2001-2013)	Large Sample (2001 - 2013)
Average Residual Last Year	0.093**	0.082*
(normalized)	(0.055)	(0.028)
Number of Listings Last Year	0.738***	0.727***
	(0.028)	(0.024)
Year FE Controls	Yes	Yes
Observations	4915	10046
R-squared	0.581	0.560
Adjusted R-squared	0.580	0.559
Panel C	Logit: 1(# of Listings This Year ¡ 50% # of Listings Last Year)	
	Main Sample (2001-2013)	Large Sample (2001 - 2013)
Average Residual Last Year	-0.032	-0.038*
(normalized)	(0.038)	(0.021)
Year FE Controls	Yes	Yes
Observations	4915	10046

Notes: Panel A reports the regression results of listing agent average performance in year t on his average performance in year $t-1$ with and without controls for year fixed effects. Panel B presents the effect of agent last year's average performance on his number of listings this year. The weight is the number of listings this year for the listing agent in the sample. The samples are collapsed from the main sample and the large sample from 2001 to 2013 restricted to listing agents with 20 or more listings. Standard errors are clustered at agent level. Significant levels: * p<0.1, ** p<0.05, *** p<0.01.

Table 1.6: Main Sample (2001-2013), Impact of Listing Agent Experience on Other Transaction Outcomes

Panel A	Log(Listing Price)		DOM		Log(DOM)	
	(1)	(2)	(3)	(4)	(5)	(6)
Listing Agent:	0.0102**	0.00422	-1.647***	-0.688	-0.0649***	-0.025
Log(1 + Last Year's # Listings)	(0.00504)	(0.00386)	(0.287)	(0.663)	(0.0108)	(0.0266)
Buyer Agent:	0.000469	0.00308	-1.592***	-1.287*	-0.0648***	-0.0546**
Log(1 + Last Year's # Sales)	(0.00502)	(0.0035)	(0.285)	(0.736)	(0.0107)	(0.024)
Listing and Buyer Agent		-0.00211		0.571		0.0116
Interactions (demean)		(0.0026)		(0.619)		(0.0293)
Controls:						
No Control	Yes	No	Yes	No	Yes	No
Full set of Controls	No	Yes	No	Yes	No	Yes
Observations	10868	10868	10868	10868	10868	10868
R-squared	0	0.973	0.007	0.603	0.008	0.6
Adjusted R-squared	0	0.94	0.007	0.133	0.008	0.127

Panel B	Log(Listing Price)		DOM		Log(DOM)	
	(1)	(2)	(3)	(4)	(5)	(6)
Listing Agent : Quartile 2	0.0266**	0.00293	-1.469**	-0.986	-0.0354	-0.0106
(5 to 6 Listings in the Last Year)	(0.0126)	(0.00626)	(0.716)	(1.456)	(0.0269)	(0.0551)
Listing Agent : Quartile 3	0.0155	0.000465	-1.516**	0.0745	-0.0432*	0.0169
(7 to 11 Listings in the Last Year)	(0.0111)	(0.0089)	(0.633)	(1.056)	(0.0237)	(0.0271)
Listing Agent : Quartile 4	0.0345***	0.00792	-4.385***	-2.132	-0.180***	-0.0917
(12+ Listings in the Last Year)	(0.0115)	(0.00756)	(0.653)	(1.317)	(0.0245)	(0.0553)
Controls:						
No Control	Yes	No	Yes	No	Yes	No
Full set of Controls	No	Yes	No	Yes	No	Yes
Buyer Agent Recent Experience	Yes	Yes	Yes	Yes	Yes	Yes
Observations	10868	10868	10868	10868	10868	10868
R-squared	0.003	0.973	0.009	0.604	0.011	0.602
Adjusted R-squared	0.003	0.94	0.008	0.135	0.01	0.129

Notes: This table reports the OLS regression results on the effect of listing agent experience on log listing price, days on market and log days on market. Panel A and B show results with linear and categorical measure of experience respectively. In even columns, standard errors are clustered at zip code level and a full set of controls is included (year-quarter FE, house characteristics, remodeling, type of finance, whether realtor agency provide loan service, buyer agent commission rate, listing types, flag for re-listed properties and house FE). The sample is the main sample with 10,868 repeated sales listings from 2001 to 2013 listed by listing agents with 20 or more listings. Significant Levels: * p<0.1, ** p<0.05, *** p<0.01

Table 1.7: Main Sample (2001-2013), Impact of Listing Agent Experience on Sales Price - Robustness Checks

	Dependent Variable: Log(Sales Price)			
	(1)	(2)	(3)	(4)
Panel A	2000-2001	2002-2007	2008-2011	2012-2013
Listing Agent:	0.00328	0.00234	0.0103	0.0290
Log(1+Last Year's #Listings)	(0.127)	(0.00982)	(0.109)	(0.177)
Observations	2211	5184	1977	1496
R-squared	0.998	0.992	0.997	0.999
Adjusted R-squared	0.936	0.956	0.918	0.922
Panel B	Least Competitive			Most Competitive
	Quartile1	Quartile2	Quartile3	Quartile4
Listing Agent:	0.00465	0.00662	0.00807	-0.00839
Log(1+Last Year's #Listings)	(0.00693)	(0.00678)	(0.00717)	(0.00893)
Observations	2740	3394	2340	2394
R-squared	0.972	0.974	0.981	0.977
Adjusted R-squared	0.937	0.938	0.956	0.941
Panel C	Predicted Log(Sales Price) Quartiles			
	Quartile1	Quartile2	Quartile3	Quartile4
Listing Agent:	0.00531	0.00753	0.00337	0.0125**
Log(1+Last Year's #Listings)	(0.00680)	(0.00526)	(0.00504)	(0.00486)
Observations	2685	2683	2712	2788
R-squared	0.951	0.959	0.959	0.967
Adjusted R-squared	0.875	0.893	0.895	0.924

Notes: Panels A - C presents that the results are robust to boom-bust cycles, market competitiveness and different price ranges respectively. In Panel B, I index the zip codes by competitiveness using Zillow turnover rate. I separate the sample into quartiles by market competitiveness and regress logarithm sales price on listing agent experience. In Panel C, I predict sales prices by basic house characteristics and divide the listings into quartiles by predicted price ranges. I regress logarithm sales price on listing agent experience in each price range. A full set of controls is included in all columns (year-quarter FE, house characteristics, remodeling, type of finance, whether realtor agency provide loan service, buyer agent commission rate, listing types, flag for re-listed properties and house FE). Significant Levels: * p<0.1, ** p<0.05, *** p<0.01

Table 1.8: Robustness Check on Different Measures of Experience

Dependent Variable: Log(Sales Price)		
Panel A	Main Sample (2001 - 2013)	
Listing Agent & Buyer Agent	0.0185	0.00469
Experienced * Inexperienced	(0.012)	(0.00537)
Listing Agent & Buyer Agent	0.0215*	0.00478
Inexperienced * Experienced	(0.0116)	(0.00743)
Listing Agent & Buyer Agent	0.0281**	0.00933
Experienced * Experienced	(0.0115)	(0.00733)
No Control	Yes	No
Full Set of Controls	No	Yes
Observations	10868	10868
R-squared	0.001	0.973
Panel B	Main Sample (2002 - 2013)	
Listing Agent :	0.0244***	0.00771**
Log(1 + Last 2 Years' \# Listings)	(0.00517)	(0.00377)
BuyerAgent:	0.0280***	0.00598
Log(1 + Last 2 Years' \# Sales)	(0.00475)	(0.00394)
Listing and Buyer Agent		-0.00402
Interactions (demean)		(0.00267)
No Control	Yes	No
Full Set of Controls	No	Yes
Observations	9867	9867
R-squared	0.007	0.974
Panel C	Main Sample (2002 - 2013)	
Listing Agent : Quartile 2	0.0266**	-0.0005
(5 to 6 Listings in the Last Year)	(0.0125)	(0.00622)
Listing Agent : Quartile 3	0.0185*	0.0006
(7 to 11 Listings in the Last Year)	(0.0111)	(0.00795)
Listing Agent : Quartile 4	0.0339***	0.0099
(12+ Listings in the Last Year)	(0.0114)	(0.00680)
No Control	Yes	No
Full Set of Controls	No	Yes
Buyer agent recent experience	No	Yes
Observations	10868	10868
R-squared	0.004	0.973

Notes: This table presents the robustness check with different measures of experience. The right-most columns include the full set of controls: year-quarter FE, house characteristics, remodeling, type of finance, whether realtor agency provide loan service, buyer agent commission rate, listing types, flag for re-listed properties and house FE. Significant Levels: * p<0.1, ** p<0.05, *** p<0.01

Table 1.9: Robustness Check Including Listing Agent Marketing Strategy

	Main Sample (2001 - 2013)				Repeated Sales Sample (2001 - 2013)			
	Log(SP)	Log(LP)	DOM	Log(DOM)	Log(SP)	Log(LP)	DOM	Log(DOM)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Listing Agent:	0.00383	0.00349	-0.452	-0.0184	0.000496	-0.00184	-0.892***	-0.0318**
Log(1+Last Year's #Listings)	(0.00323)	(0.00361)	(0.638)	(0.0254)	(0.00200)	(0.00189)	(0.293)	(0.0129)
Buyer Agent:	0.00543	0.00397	-1.330*	-0.0541**	0.00683***	0.00591**	-0.922**	-0.0374**
Log(1+Last Year's #Sales)	(0.00384)	(0.00352)	(0.731)	(0.0241)	(0.00246)	(0.00232)	(0.414)	(0.0152)
Listing and Buyer Agent	-0.00210	-0.00198	0.642	0.0138	-0.00256	-0.00224	0.431	0.0109
Interactions (demean)	(0.00320)	(0.00259)	(0.615)	(0.0296)	(0.00183)	(0.00175)	(0.411)	(0.0168)
Agree to lower commission	0.00758	0.00791	-0.0979	-0.0304	0.00704	0.00772	-0.130	-0.00472
if dualagent	(0.00665)	(0.00669)	(1.284)	(0.0519)	(0.00568)	(0.00559)	(0.687)	(0.0193)
Inhouse	-0.00601	-0.00428	0.651	-0.0249	-0.00455	-0.00295	-0.163	-0.0678**
	(0.00648)	(0.00549)	(1.041)	(0.0426)	(0.00532)	(0.00455)	(0.850)	(0.0294)
dualagent	-0.00725	0.00517	-0.0924	-0.0294	0.00282	0.0166*	1.568	0.0708
	(0.00951)	(0.00985)	(1.936)	(0.0678)	(0.00749)	(0.00938)	(1.116)	(0.0458)
Controls:								
Full set of Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Keywords in Description	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Number of Pictures	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	10868	10868	10868	10868	26680	26680	26680	26680
R-squared	0.976	0.975	0.612	0.611	0.965	0.965	0.589	0.591
Adjusted R-squared	0.946	0.945	0.141	0.139	0.926	0.926	0.127	0.132

Notes: This is a robustness check of the relationship between listing agent experience and different transaction outcomes including the marketing strategy of listing agent. A full set of controls and additional controls for keywords in descriptive paragraphs and number of pictures are included in all columns (year-quarter FE, house characteristics, remodeling, type of finance, whether realtor agency provide loan service, buyer agent commission rate, listing types, flag for re-listed properties and house FE). Significant Levels: * p<0.1, ** p<0.05, *** p<0.01

Table 1.10: The Impact of Listing Agent Experience on Fast Sales

	Dependent Variable: 1(DOM¡30)		
Panel A	Main Sample (2001 - 2013)		
	(1)	(2)	(3)
Listing Agent:	0.126**	0.114**	0.117**
Log(1 + Last Year's # Listings)	(0.0495)	(0.0512)	(0.0538)
Buyer Agent:	0.0930*	0.112*	0.122**
Log(1 + Last Year's # Sales)	(0.0494)	(0.0595)	(0.0617)
Listing and Buyer Agent		-0.0442	-0.054
Interactions (demean)		(0.0517)	(0.0543)
Controls:			
No Control	Yes	No	No
Full Set of Controls	No	Yes	Yes
Keywords in Description	No	No	Yes
Number of Pictures	No	No	Yes
Observations	3549	3549	3549
Panel B	Repeated Sales Sample (2001 - 2013)		
	(1)	(2)	(3)
Listing Agent:	0.127***	0.117***	0.110***
Log(1 + Last Year's # Listings)	(0.0252)	(0.0256)	(0.0264)
Buyer Agent:	0.0949***	0.0976***	0.0973***
Log(1 + Last Year's # Sales)	(0.0295)	(0.0298)	(0.0304)
Listing and Buyer Agent		-0.0261	-0.031
Interactions (demean)		(0.0264)	(0.027)
Controls:			
No Control	Yes	No	No
Full Set of Controls	No	Yes	Yes
Keywords in Description	No	No	Yes
Number of Pictures	No	No	Yes
Observations	9519	9519	9519

Notes: This table reports the fixed effect logit regression results on the relationship between listing agent experience and whether the listing is sold within 30 days. A full set of controls is included in column 2 (year-quarter FE, house characteristics, remodeling, type of finance, whether realtor agency provide loan service, buyer agent commission rate, listing types, flag for re-listed properties and house FE). Column 3 includes controls for keywords in descriptive paragraphs and number of pictures in addition to a full set of controls. Significant levels: * p<0.1, ** p<0.05, *** p<0.01.

Chapter 2

The Power of Words — A Machine Learning Approach to Predicting Real Estate Sales Outcomes

2.1 Introduction

In this chapter, I explore big data algorithms and their difference in predictive power compared to traditional OLS regressions using real estate repeated sales data. Often times in empirical research, researchers need to find a subset of predictors and they face the trade-off between computational efficiency and predictive power of the model especially when the dataset is long and wide shape. By using penalizing regression algorithms, researchers can select non-trivial features or predictors according to the relative importance of them in minimizing mean square error. These algorithms include but are not limited to random forest, lasso, ridge, elastic-net and non-negative garrote.

First of all, I geocode a rich real estate repeated sales dataset and map each property with detailed school district and neighborhood information. I stratify the data using different group variables and find that the random forest algorithm provides similar results under different stratifications. Secondly, I compare the out-of-sample predictions across different regression algorithms and find out which algorithm can explain most of the variation of the response variable. Thirdly, I use the keywords in online advertisement as a more specific example of variable selection. I show a subset of keywords remains significant and robust under both random forest and OLS. Lastly, I present regression results of school academic performance on property sales price controlling for a wide range of house characteristics and neighborhood fixed effect. I find that school quality has a significant positive correlation with property sales price. Specifically, a ten-point increase in API (Academic Performance Index) in a school district increases the property sales price by 0.7% to 0.9%. The effect of school quality on days on market is ambiguous.

2.2 Literature Review

Machine Learning Algorithm and Out-of-Sample Prediction

Kotsiantis, Zaharakis and Pintelas (2006)[36] describe various classification algorithms in relative detail and point out two key questions in machine learning. The first one is to figure out under which conditions a certain algorithm is applicable to a specific problem. The second one is the increased computation time given classification algorithms take into all possible combinations of predictors. Wu et al. (2008)[54] present the top ten data mining algorithms with detailed descriptions and examples. Castle et al. (2009)[13] compare the performance of twenty-one different model selection algorithms (MSA) based on the coefficient unconditional mean squared error and find some MSA differ substantially in variable selections. Domingos (2012)[18] summarizes twelve key lessons in machine learning practice, among which some are related to fundamental economic principles, for example "correlation does not imply causation". Some are key issues to focus on, such as "feature engineering", and pitfalls to avoid "overfitting". Einav and Levin (2013)[19] lay out the opportunities and challenges as large-scale datasets become more accessible. Varian (2014)[51] discusses decision tree and variable selection algorithms in manipulating and analyzing big data and provides several empirical examples.

Keywords and Sales Price

Rascoff and Humphries (2015) [45] suggest that based on Zillow analysis houses described as "luxurious" tend to sell at a premium of 8.2% and that certain keywords like "landscaped" and "beautiful" increase the sales prices by 4.2% and 2.3% above the listing prices. Many other papers with different focuses include keywords as controls in their analyses. Levitt and Syverson (2008) [38] include a rich set of keyword dummies as additional controls for house quality. Barwick, Pathak and Wong (2015)[5] include keywords such as "Renovated", "Remodeled", "Maintained", "Needs updating" in their analysis to control for remodeling between two repeated sales.

School quality and Property Value

Previous literature explores whether school quality has a significant effect on property price. Black (1999)[10] uses the boundary approach to study school quality and house prices. Bogart and Cromwell (2000)[11] find that disrupting neighborhood schools resulting from a school district realignment reduces house values by 9.9%. Bayer et al. (2007)[6] find substantial neighborhood heterogeneity across school attendance zone boundaries. They show that household willingness to pay increases by less than one percent when school performance increases by five percent after controlling for neighborhood demographics across the attendance zone boundaries. Their estimates are substantially lower than previous estimates with only neighborhood fixed effect. Fack and Grenet (2010)[20] construct control groups

for each transaction to control for neighborhood differences. They find that one standard deviation increase in public school performance leads to a 1.5 to 2.5% increase in house prices. Based on the inverse distance strategy in Fack and Grenet (2010)[20], Dhar and Ross (2012)[16] find fairly moderate effects of test scores on property values — one standard deviation increase in the test scores increases the sales price by at least 4.1%. Figlio and Lucas (2004) [23] present evidence that house prices respond significantly to additional public school information provided by the school report cards.

2.3 Data and Summary Statistics

Figure (2.1) shows the base map of Alameda County in California and a layer of neighborhoods consisting of 295 polygons from The Neighborhood Project.[1] There are other neighborhood data sources such as zillow.com, which provides ready-to-use shape files. Unfortunately, Zillow does not have a complete neighborhood information for the whole Alameda County.[2] Other realty websites, such as Redoak realty, provide interactive maps of neighborhood areas. It requires reserve engineering to construct the shape files from polygon coordinates in those interactive maps. Thus, I use the completed neighborhood boundaries maps showed in Figure (2.1) from the Neighborhood Project. It will be helpful to compare neighborhood information from different sources in the future.

Figure (2.2) presents the public school boundary map on top of the base map of Alameda County. The dataset contains shape files based on longitude and altitude and is ready to use for geocoding.[3] There are eighteen unified school districts, of which two are further divided into subareas.

I take the main sample (MLS dataset) in Chapter 1 and geocode the property addresses. I start with 96,438 observations that are sold listings of single family detached properties between 2000 and 2013. Among which, 1.03% listings addresses are not matched with coordinates. For the rest of the geocoded addresses, I map each property address to its corresponding unique neighborhood and school district. I double check the matching by comparing the matched school district information to the internal one in the MLS dataset. There are about 3% observations with mismatched school district and 9.44% with missing neighborhood information. Excluding those, I have a final sample of 88,240 observations, among which 24,461 are repeated sales. I use the repeated sales sample as the main sample in this chapter and the final sample for robustness checks.

In Figure (2.3), I show all the properties in the final sample on the map. From the figure, residential transactions are mostly concentrated in Berkeley, Oakland and Fremont unified school districts. Specifically, 49.56% of all transactions and 52.56% of repeated sales

[1]The data can be downloaded from http://zetashapes.com/editor/06001 (accessed on April 27, 2016).

[2]In fact, Zillow neighborhood shape files only covers 30-40% of the areas in Alameda County.

[3]School district data can be found here: https://data.acgov.org/Geospatial-Data/Unified-School-District-Boundaries/b4ug-2x6q (accessed on April 27, 2016)

transactions are in the above three districts. There are very few transactions in Sunol Glen and Mountain House Elementary school districts.

I collect all the school locations and academic performance information from the California Department of Education website. Academic performance information is represented by the Academic Performance Index (API) which measures the academic performance and progress of individual schools in California. I use the growth API which has a half-year lag. For example, the growth API released in Fall 2012 comes from the 2012 spring test scores. It would be helpful to use the base API, which has a year lag, as a robustness check in future studies.

After geocoding school addresses, I map each property to schools in its eligible school district. I calculate the distance between the property and schools based on their coordinates. I further calculate a weighted school API for each property by the following expression:

$$WAPI_{it} = \sum_j API_{jt} * \frac{\frac{1}{d_{ij}}}{\sum_j \frac{1}{d_{ij}}}, \tag{2.1}$$

where $WAPI_{it}$ is the weighted API at year t for property i. API_{jt} is the API for school j in year t. d_{ij} is the distance between property i and school j, $j \in \{1, ...j, ...n\}$, where n is the number of schools in the school district which property i is eligible for. The weight $\frac{\frac{1}{d_{ij}}}{\sum_j \frac{1}{d_{ij}}}$ is the inverse ratio of distance between each school and property i. [4]

Table (2.1) presents the summary statistics of the weighted and the average APIs in each school district. They are very comparable to each other and the distribution of both APIs are close to a normal distribution.[5]

2.4 Empirical Analysis

In this section, I examine the impact of different stratification methods on random forest estimation results. With this exercise, I hope to shed light on the question of how robust the random forest (RF) algorithm is to the selection of training dataset.

Stratification

I implement the RF algorithm using the RandomForest package in r. I set aside one fold of the repeated sales data for testing and use the rest nine folds for training and cross validation. There are four different stratifications based on random sampling (1) of the repeated sales

[4] I also calculate weighted API for the top 5 schools in the district given that people may only care about the top schools when they purchase the property in the school district. My results are robust to this.

[5] Among all districts, Albany, Castro Valley, Fremont, Piedmont and Pleasanton are the better school districts in Alameda County.

dataset, (2) within each zip code, (3) within each neighborhood area, and (4) within each year quarter interaction.

I run two model specifications. Both have weighted API, basic house characteristics, year-quarter fixed effect, other controls (e.g., type of finance, buyer agent commission) and keywords in advertisement as explanatory variables, and log sales price as the response variable. The only difference is that one has zip code fixed effects and the other has neighborhood fixed effects. All the OLS regression results in this chapter are clustered either at zip code or neighborhood levels.

Table (2.2) shows the importance of explanatory variables by different models specifications and stratifications. %IncMSE means how much the mean squared error (MSE) will increase given the corresponding variable is permuted. IncNodePurity is the average reduction of node impurity, in this case residual sum of squares (RSS) over all trees before and after the variable is used for the split. I focus on %IncMSE given it is a more robust measure compared to IncNodePurity.

Panel A presents the results for specification with neighborhood FE and Panel B presents the one with zip code FE. In odd columns, the ranks of the importance of features are comparable across different stratifications. In Panel A, random sampling outperforms other stratifications in out-of-sample predictions. From Panel B, random sampling within zip code slightly outperforms others. From both Panels, one can see that weighted API is an important predictor and time, zip code / neighborhood fixed effects and the keywords in description are all powerful predictors.[6]

Figure (2.4) presents the out-of-sample prediction by different algorithms under four different stratifications. OLS, Lasso (glmnet package) and Elastic-Net outperform Ridge and RF in out-of-sample prediction. RF gives the worst prediction across all cases except that in Panel D RF catches up with Ridge. It is a interesting observation that RF does not outperform OLS as it usually produces better out-of-sample predictions. Thus in the next subsection, I study why RF underperforms OLS.

Out-of-Sample Prediction

I implement the Random Forrest, Lasso, Ridge and Elastic-Net algorithms. Figure (2.5) shows the rank of the out-of-sample predictions based on different algorithms for two model specifications — one with neighborhood FE and the other one with zip code FE.

In Panel A, ranking from the best fit (least prediction error) to the worst fit:

1. OLS out-of-sample and OLS in-sample predictions
2. Lasso min (min: best lambda to minimize the cross-validated error)
3. Lasso se (se: the largest value of lambda such that the cross-validation error is within one standard deviation of its minimum value)

[6]For both panels in Tabel (2.2), I omit the less important variables: Pool/Hot tub, Remodeling, Loan services from Realtor Agency, Buyer Agent Commission, Flag for Re-listed Home, Listing Types, Inhouse, Dualagent.

4. Elastic-Net min
5. Elastic-Net se
6. Ridge min
7. Ridge se
8. RF

In Panel B, OLS in-sample prediction outperforms Lasso min, followed by OLS out-of-sample prediction, Elastic-Net min and Lasso se. The rank for the other algorithms remains the same as in Panel A. More specifically, from Table (2.3) column 4, one can see that OLS, Lasso and Elastic-Net explain the variation in the data slightly better (0.5% more) than Ridge and much better (6% more) than RF.

One possible reason for RF to underperform OLS is that it is computationally expensive and inefficient for the RF algorithm when the data generating process (DGP) is linear. The RF algorithm takes into consideration all the possible interactions of predictors. It is hard to use step functions to approach the linear regression. Thus by reducing the number of predictors, I can test whether random forest outperforms OLS in out-of-sample prediction. If that is the case, then it is likely that the underperforming of RF is due to the linear DGP in the repeated sales data.

I run a two-step regression as follows:

$$y_{ijt} = \kappa_t + \zeta_j \ (or \eta_j) + \epsilon_{ijt}, \tag{2.2}$$

where y_{ijt} is the property i in zip code j (or neighborhood j) sold in year-quarter t. κ_t is a year-quarter interacted fixed effect. ζ_j is a zip code fixed effect, while η_j is the neighborhood fixed effect. ϵ_{ijt} is the residual log sales price after taking out the time and location fixed effect. In the second step I regress the residual log sale price on weighted API, house characteristics, keywords and other controls.

$$\epsilon_{ijt} = \beta * WAPI_{it} + \alpha_1 * X_i + \alpha_2 * K + \epsilon^{'}_{ijt}, \tag{2.3}$$

where X_i is a set of house characteristics, K is additional controls of this listing, for example the type of finance and the type of listing, $\epsilon^{'}_{ijt}$ reflects unobservables.

I further decompose the second regression into another two scenarios where the predictors include only house characteristics and only keywords.

Figure (2.6) Panel A shows that RF with only 100 trees outperforms OLS, Lasso, Elastic-Net and further outperforms Ridge regression in the out-of-sample prediction after taking the two-step regression to reduce the number of predictors. Panels B and C in Figure (2.6) also show that RF outperforms other algorithms when there are only house characteristics or keywords as predictors in the feature selection process. Thus computational inefficiency for RF to approach linear DGP is very likely to be the reason why RF underperforms OLS when I include the time and location fixed effects in the regression.

I calculate the proportion variation explained in the outcome of the testing data to measure the goodness of fit. Specifically, it is measured by the Mean Squared Prediction

Error (MSE) divided by the total sum of errors, $\frac{\sum(y_{it}-\hat{y_{it}})^2}{\sum(y_{it}-\overline{y_{it}})^2}$, where y is the real value and $\hat{y}$ is the predicted value in the testing data.

Columns 1-3 in Table (2.3) correspond to model specifications where I use 1) house characteristics and other variables (type of finance, weighted API, etc); 2) only keywords; 3) both as predictors.

In models 1 and 3, RF further explains about 15% more of the variation in the out-of-sample prediction compared to other algorithms, which is a sizable improvement in predictive power. In model 2, OLS explains 2.8% more of the variation than RF does.

Keywords and Average Transaction Outcomes

As discussed before, Table (2.2) shows that the keywords and weighted API are both important predictors. This is robust across two model specifications and four different stratification methods. Therefore, I estimate the size of the effect of keywords on the average transaction outcomes.

Let y_{ijt} be a measure of one transaction outcome for property i in neighborhood j with year-quarter interaction t. The OLS specification takes the following form:

$$y_{ijt} = \lambda_m * D_{imt} + \beta * WAPI_{it} + \alpha_1 * X_i + \alpha_2 * K + \eta_j + \kappa_t + \epsilon_{ijt}, \tag{2.4}$$

where $D_m = 1$ if the mth keyword in the keywords list is present in the advertisement of property i at time t, $D_m = 0$ otherwise.

Table (2.4) columns 2 and 5 show the coefficients for each keyword dummy from the OLS regression. Columns 3 and 6 present the increase in MSE given the corresponding keyword dummy is permuted in the RF algorithm. Ranking by the descending order of the increment in MSE, the most important predictors are located at the top of column 3. A keyword, for example "new", can be valuable in explaining the variation in the testing data but has an insignificant effect on the average log sales price. It is also true that the keywords, such as "fantastic" or "amazing", have little value in prediction but are significantly correlated with the log sales price.

I then flag the five most frequently used keywords in the MLS description in each neighborhood in Alameda County and sum the flags across all neighborhoods for each keyword. I take the top sixteen keywords with the most flags. In Table (2.5), these words are listed in the left most column. Column 1 shows the total number of times a certain keyword is among the top 5 most frequently used keywords in a neighborhood. Column 4 shows how frequently a certain keyword appears in a property description in the repeated sales sample. Even though keywords like "built-in" are less frequently mentioned compared to "spacious" and "close", it is more informative in prediction and significantly positively correlated with the average log sales price. Overall, there are quite some commonly used adjective words in online advertisement for listings, some are not valuable for predictions and have no significant effect on the average transaction price. This evidence suggests that some keywords are interacted with other house features and the interactions have significant impact on the

transaction outcomes and are informative in both in-sample and out-of-sample predictions. It further suggests that the interactions between certain keywords and other house features should be considered in model specifications.

School Quality and Average Sales Price

In this subsection, I estimate the correlation between school quality and transaction outcomes.

As described in the data section, there are two measurements of school quality — API and the weighted API. I run regression (2.4), in which β is the coefficient of interest.

Table (2.4) presents the results. Even columns correspond to model specification with neighborhood FE and odd columns correspond to the one with zip code FE. There is a significantly positive effect of school quality on the average transaction price. From column 4, a ten-point increase in weighted API leads to a 0.754% increase in the final sales price. Given the \$700k average sales price in Alameda County, one standard deviation increase in weighted API, which is 95 points, leads to a 7.163% increase in the transaction price, which equals to \$50k. The results are robust to both the main sample (results in columns 1 and 2) and the repeated sales sample (results in columns 3 and 4).

From columns 5 to 6, there is a significant negative effect of school quality on days on market. However this effect becomes less significant when I use the repeated sales sample, and insignificant when I include zip code FE instead of neighborhood FE. Thus, there is no clear evidence whether better school quality leads to faster sales. Table (2.7) presents the robustness check results using average API as explanatory variable. The results are robust.

2.5 Conclusion

In this chapter, I use the repeated sales data to study how big data algorithms differ from OLS regression in predictive power and how robust those algorithms are to training set selection. I find that it is computationally expensive for the random forest algorithm to use step functions to approach the linear data generating process. Once there are fewer predictors in, RF outperforms other algorithms and it is robust to different model specifications. In addition, the random forest algorithm provides similar results under different stratifications.

I also study the effect of keywords on sales price and how informative they are in predicting sales price. I find that certain keywords can be valuable in explaining variation in the data but have insignificant impact on the average sales price, indicating that the interaction between such keywords and other house features together should be considered when we specify our models in empirical research.

Lastly, I find school quality has a robust significantly positive effect on property sales price controlling for a wide range of house characteristics and neighborhood FE or zip code FE.

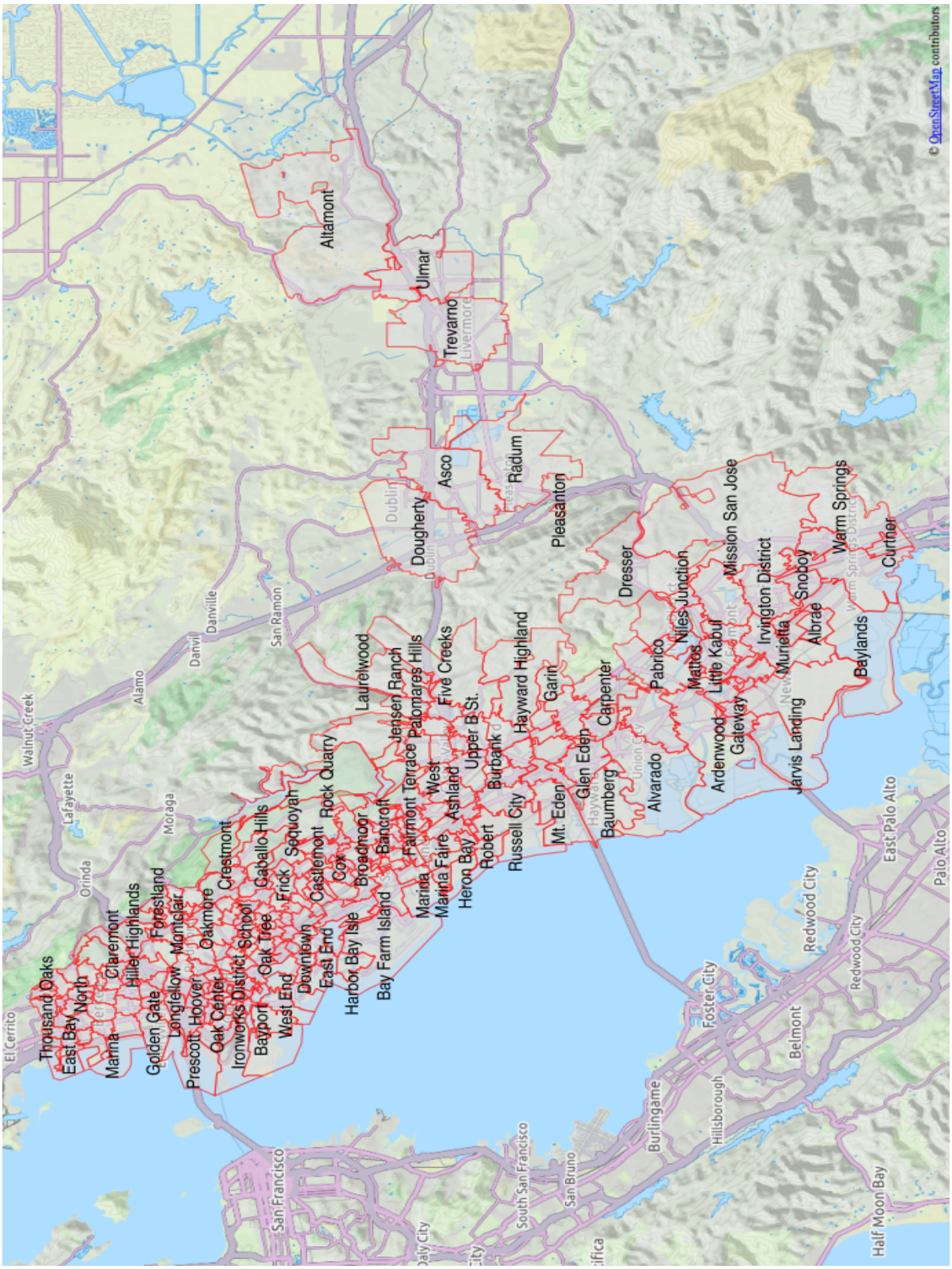


Figure 2.1: Alameda County Neighborhoods

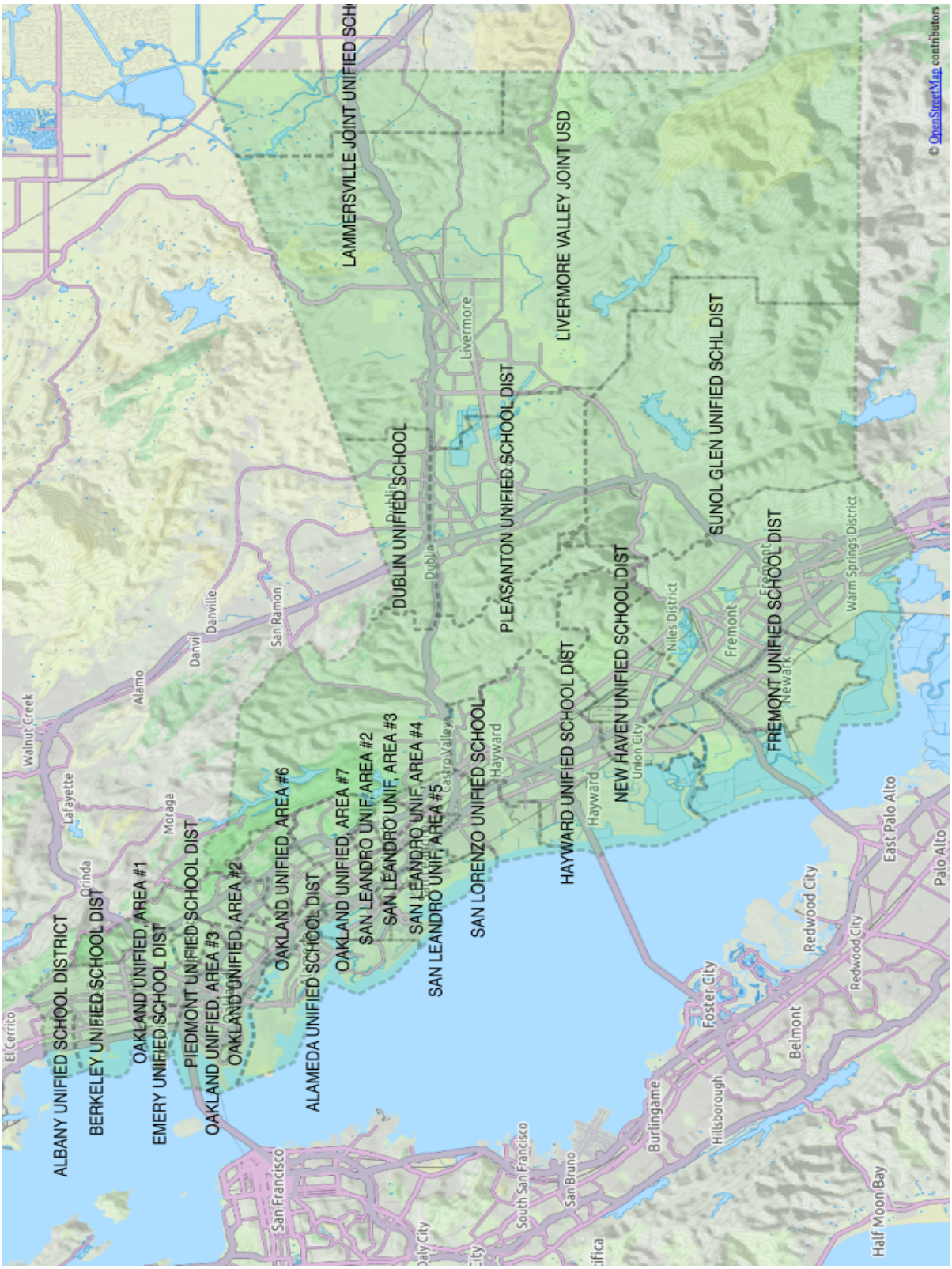


Figure 2.2: Alameda County School Districts

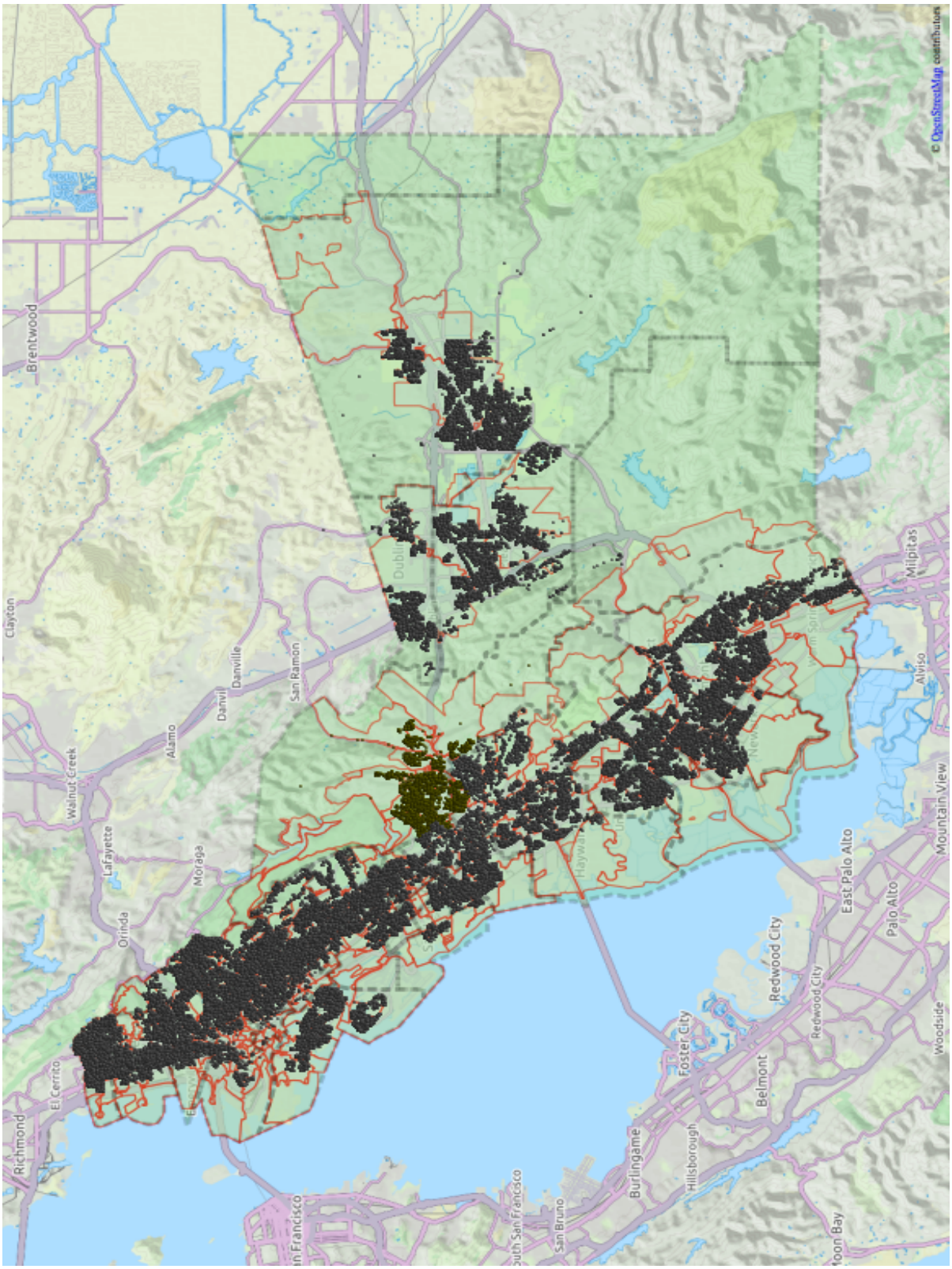


Figure 2.3: Basemap, Neighborhoods, School Districts and Properties

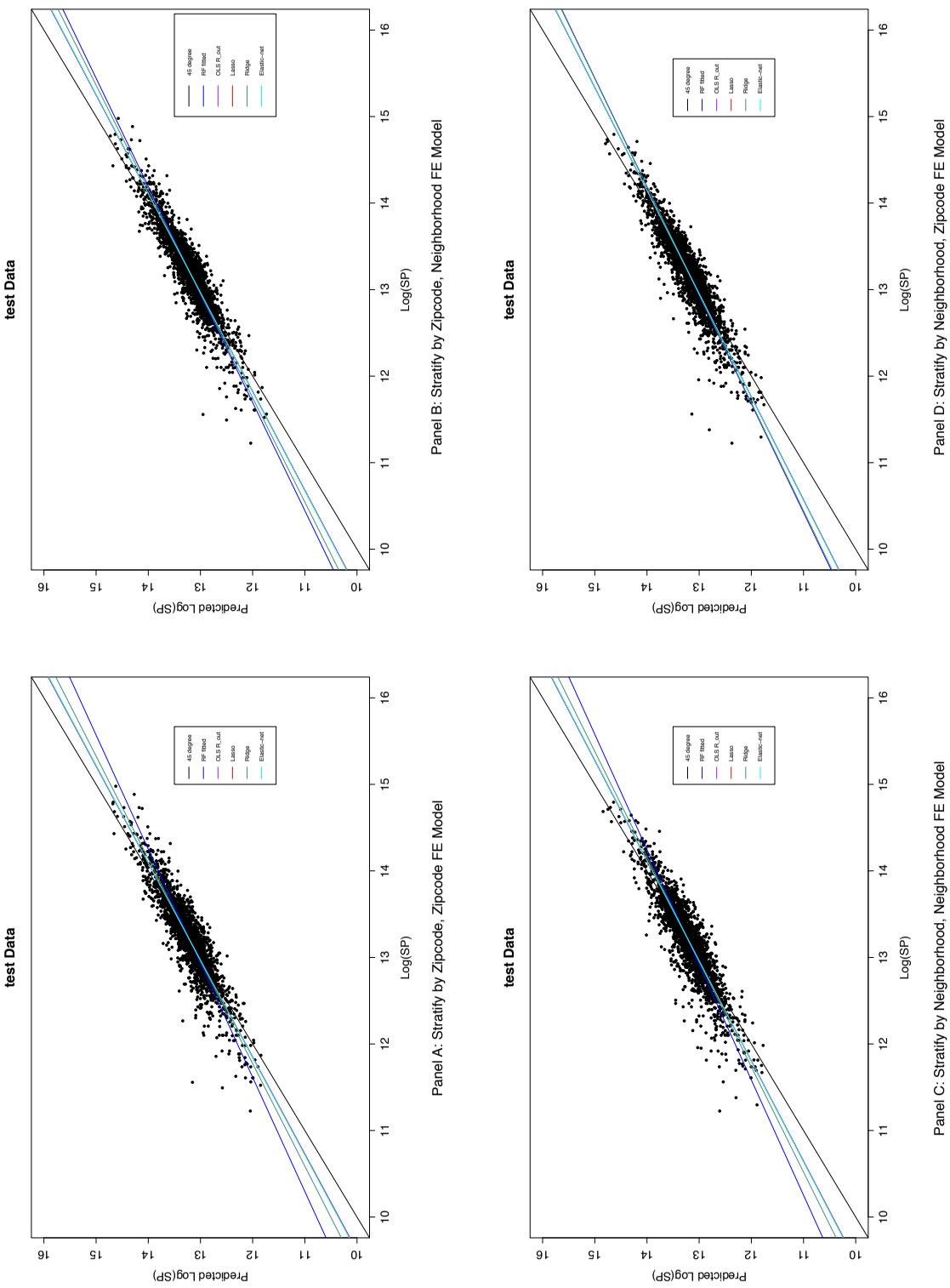
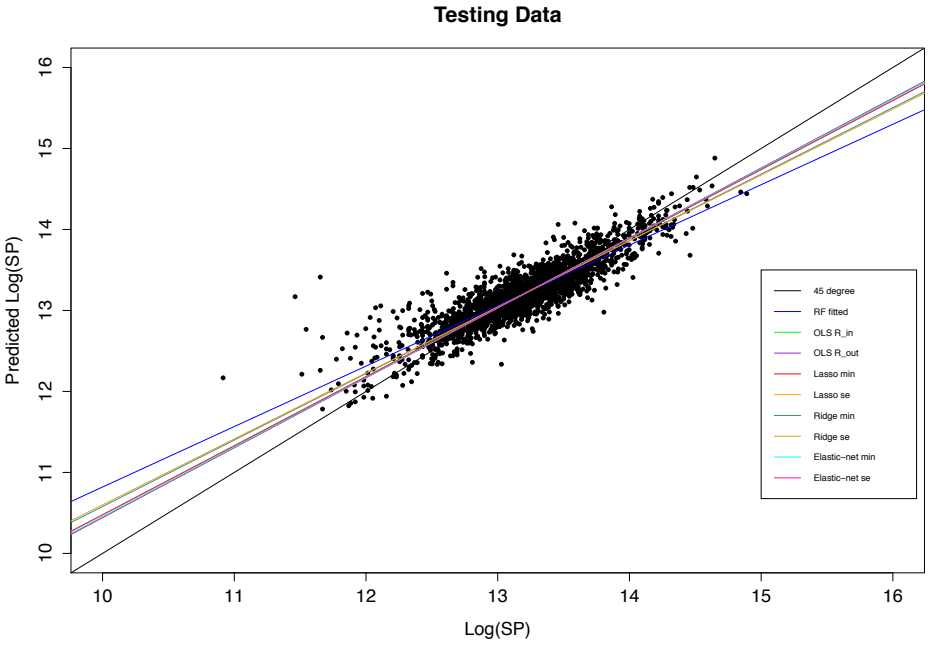
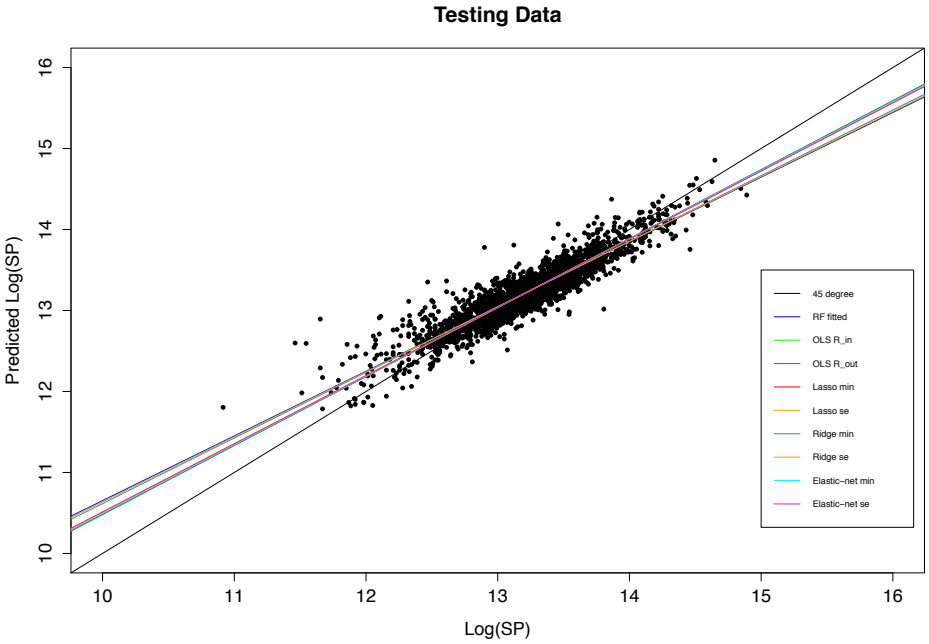


Figure 2.4: Comparison of Stratification Methods

Figure 2.5: Comparison of Out-of-Sample Predictions of Log(SP)



Panel A: Neighborhoods FE



Panel B: Zipcode FE

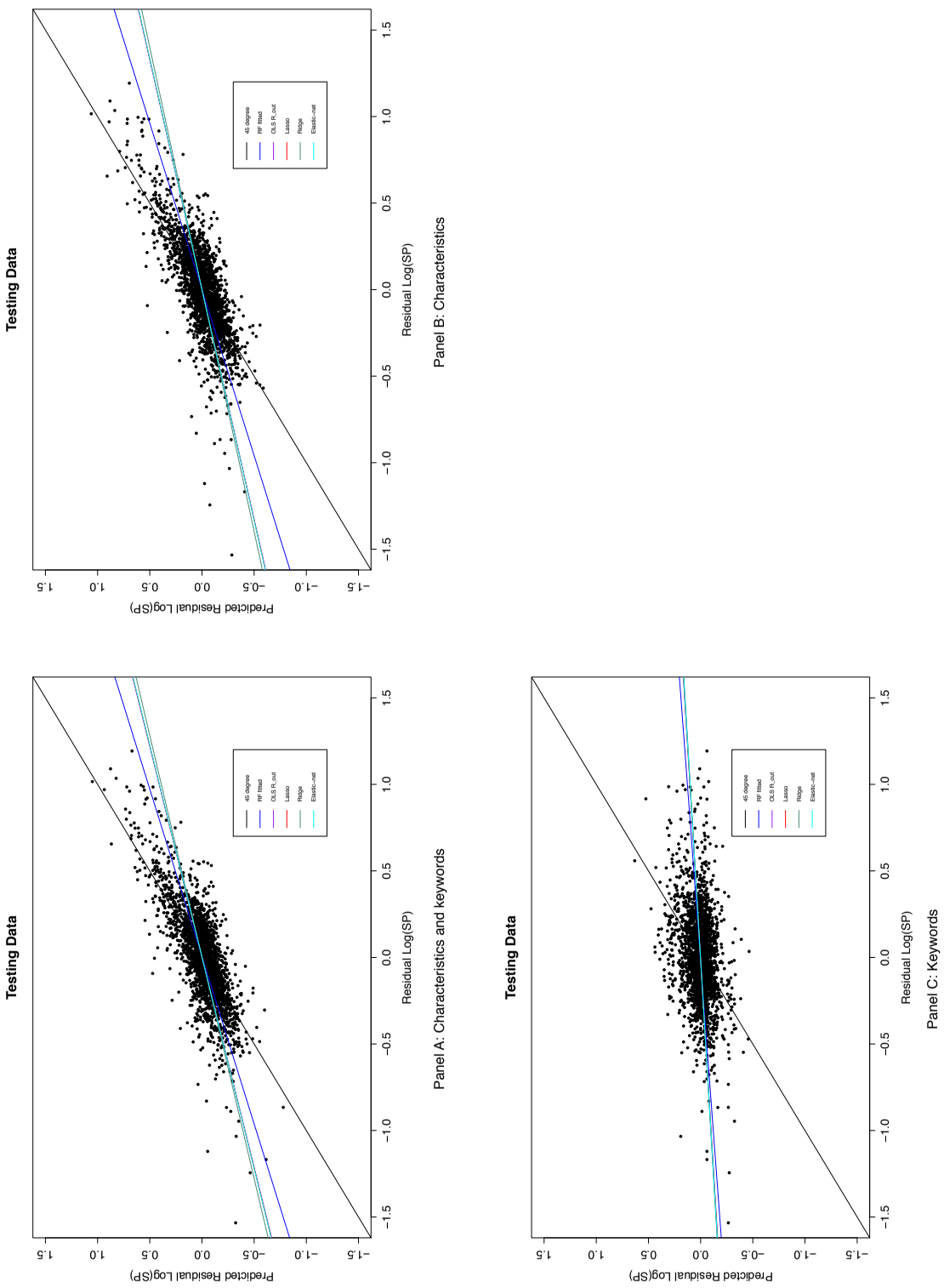


Figure 2.6: Comparison of Out-of-Sample Predictions of Residual Log(SP)

Table 2.1: Academic Performance Index by School District

	Weighted API			Average API		
District	Mean	SD	N	Mean	SD	N
ALAMEDA UNIFIED	804.1002	48.14462	1181	780.4809	33.71949	1181
ALBANY CITY UNIFIED	839.8879	37.86955	282	834.9633	28.83029	282
BERKELEY UNIFIED	764.2582	45.37008	1811	760.8994	40.03888	1811
CASTRO VALLEY UNIFIED	833.9241	38.50866	903	841.1721	18.72718	903
DUBLIN UNIFIED	810.4273	46.25383	653	812.9224	32.81593	653
EMERY UNIFIED	657.3617	69.04156	13	646	59.27486	13
FREMONT UNIFIED	832.1896	41.41217	3913	833.2958	31.24286	3913
HAYWARD UNIFIED	673.3303	40.04267	1475	672.9166	38.55988	1475
LIVERMORE VALLEY JOINT UNIFIED	775.2626	32.14954	1937	780.8619	20.12291	1937
NEW HAVEN UNIFIED	754.4606	41.11098	620	743.1995	28.30991	620
NEWARK UNIFIED	715.8033	35.52561	528	715.5297	27.83127	528
OAKLAND UNIFIED	659.2711	71.24851	7133	647.5764	61.96736	7133
PIEDMONT UNIFIED	902.4042	16.49592	350	903.0107	13.70606	350
PLEASANTON UNIFIED	878.0497	23.47724	2254	874.8368	16.28403	2254
SAN LEANDRO UNIFIED	705.5862	30.17691	862	703.5303	26.91968	862
SAN LORENZO UNIFIED	687.918	46.76241	546	686.584	38.30119	546

Table 2.2: Random Forest Feature Selection by Different Stratification Methods

Stratification Methods								
Stratification:	Random		Zipcode		Neighborhood		Year Quarter	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A	%IncMSE	IncNodePurity	%IncMSE	IncNodePurity	%IncMSE	IncNodePurity	%IncMSE	IncNodePurity
API (weighted)	97.45	686.4	95.3	680.42	107.61	640.26	85.9	692.31
Bedroom	10.86	191.38	10.56	186.26	9.73	109.73	10.8	181.78
Bathroom	10.2	417.52	11.12	433.76	11.35	478.26	9.12	353.1
Garagae Spaces	15.11	32.69	10.92	38.7	13	34.44	10.53	33.91
Fireplaces	12.62	107.44	13.23	148.35	16.4	113.01	13.78	125.27
Age of the House	51.01	172.63	59.7	164.01	41.07	168.52	31.15	178.85
Sqft	47.35	1378.68	52.93	1346.26	46.72	1419.99	53.88	1466.31
Lot size	28.42	177.14	26.9	163.19	25.38	159.08	23.91	149.36
Year Quater FE	571.38	332.89	600.22	332.75	594.52	336.38	574.49	337.57
Neighborhood FE	554.2	324.69	529.16	328.88	517.23	325.13	541.12	325.79
Type of Finance	31.58	67.39	31.79	70.94	29.26	66.73	25.74	67.65
Buyer Agent Commission	43.29	163.6	46.87	152.75	44.52	164.23	46.09	151.24
Keywords in Description	159.55	240.03	173.07	242.68	176.75	242.64	168.7	238.69
Mean of squared residuals		0.0396		0.0414		0.0410		0.0410
In-sample, Var explained		80.38		79.46		79.56		79.75
Out-of-sample, —		86.42		81.44		80.49		80.04
Panel B	%IncMSE	IncNodePurity	%IncMSE	IncNodePurity	%IncMSE	IncNodePurity	%IncMSE	IncNodePurity
API (weighted)	50.48	516.82	55.85	480.73	52.61	489.08	54.86	475.56
Bedroom	8.66	117.04	8.65	156.48	8.22	134.04	8.49	106.69
Bathroom	12.21	340.86	10.78	296.6	12.01	314.61	10.12	274.78
Garagae Spaces	12.85	34.19	17.22	34.8	10.76	33.92	13.58	36.45
Fireplaces	10.87	66.44	8.81	74.21	12.7	62.31	10.04	76.67
Age of the House	32.68	106.71	27.45	116.44	22.21	125.76	22.65	129.22
Sqft	43.59	1147.43	42.4	1179.52	44.15	1177.69	42.15	1276.19
Lot size	23.76	135.24	19.5	138.02	20.58	144.24	24.33	142.7
Year Quater FE	606.52	309.69	579.25	311.78	587.9	307.89	605.09	310.46
Zipcode FE	93.08	1161.2	77.79	1132.18	89.68	1100.59	99.55	1120.91
Type of Finance	21.23	53.63	23.49	57.36	27.42	55.34	20.38	54.84
Buyer Agent Commission	48.43	136.5	48.27	148.69	52.19	147.18	39.97	144.56
Keywords in Description	133.04	194.46	130.21	191.05	136.25	193.67	115.72	194.52
Mean of squared residuals		0.0396		0.0320		0.0318		0.0316
In-sample, Var explained		84.32		84.12		84.15		84.39
Out-of-sample, —		84.68		85.47		84.71		83.85

Table 2.3: Variation Explained by Different Specifications and Algorithms

Model Specifications:				
	(1)	(2)	(3)	(4)
	Characteristics	Words	Both	Both
Random Forest	55.000	6.720	56.060	80.380
OLS	38.220	9.581	41.370	86.421
Lasso	38.255	9.736	41.320	86.850
Ridge	38.078	9.735	41.187	85.906
Elastic-Net	38.255	9.752	41.327	86.843
Response Variable:	Residual Log(SP)	Residual Log(SP)	Residual Log(SP)	Log(SP)

Table 2.4: The Effect and Importance of Keywords

Keywords	OLS coefficients	%IncMSE	Keywords	OLS coefficients	%IncMSE
(1)	(2)	(3)	(4)	(5)	(6)
granite	0.0146***	31.51	spacious	0.00318	3.36
deck	0.0180*	30.64	spectacular	0.0501***	3.22
beautiful	0.0293***	22.83	unique	-0.000378	2.85
needs	-0.0844***	22.57	renovated	0.0468***	2.64
charming	0.0369***	20.51	quiet	0.00275	2.58
as is/as-is	-0.0390***	19.67	brand new	0.0056	2.35
landscaped	0.0309***	19.24	move-in	0.00616	2.27
tlc	-0.0451***	18.65	immaculate	0.0205***	2.24
new	0.00365	17.58	neighborhood	0.000915	1.88
wonderful	0.0276***	15.05	copper	0.0161***	1.88
built-in	0.0296***	14.8	delightful	0.0272***	1.79
block	0.0345***	12.94	estate sale	-0.0768***	1.69
stunning	0.0565***	9.85	dream	-0.00384	1.6
huge	-0.00628*	8.5	appealing	0.0126	1.46
maple	0.0147***	7.96	breathtaking	0.0294*	1
gourmet	0.0323***	7.57	maintained	0.0111***	0.95
fabulous	0.0428***	7.32	needs updating	0.0347**	0.88
custom	0.0142***	7.27	fantastic	0.0218***	0.81
!	0.00447***	6.67	tasteful	0.0432***	0.75
close	-0.00113	6.55	state-of-the-art	0.0993***	0.7
potential	-0.0593***	6.47	amazing	0.0370***	0.29
clean	-0.00896**	6.01	elegance	0.0437***	0.19
handyman	-0.0925***	5.3	bank-owned	-0.192***	0
must see	0.00946**	4.5	stately	0.0660***	-0.12
motivated	-0.0265***	4.28	mint	0.0309**	-0.2
newer	-0.00317	4.19	pride	0.0225***	-0.25
hurry	-0.0133**	3.87	corian	0.00847*	-2.25
vintage	0.0314***	3.65	leaded	0.0624***	-3.89

Table 2.5: Most Frequently Used Keywords

Keywords	N	OLS coefficients	%IncMSE	Frequency in Advertisement
	(1)	(2)	(3)	(4)
!	239	0.00447***	6.67	11551
new	228	0.00365	17.58	10259
beautiful	159	0.0293***	22.83	5185
close	144	-0.00113	6.55	4760
deck	76	0.0180*	30.64	3287
spacious	73	0.00318	3.36	3213
granite	64	0.0146***	31.51	3518
charming	44	0.0369***	20.51	1928
as is/as-is	32	-0.0390***	19.67	1479
newer	23	-0.00317	4.19	2517
huge	19	-0.00628*	8.5	2119
custom	15	0.0142***	7.27	1890
built-in	14	0.0296***	14.8	1361
potential	13	-0.0593***	6.47	736
must see	11	0.00946**	4.5	1387
motivated	11	-0.0265***	4.28	1361

Notes: N stands for the total number of times a certain keyword is among the top 5 most frequently used keywords in a neighborhood. Significant Levels: * p<0.1, ** p<0.05, *** p<0.01

Table 2.6: Weighted School Quality on Log(Sales Price) and Days on Market

	Log(Sales Price)				Days on Market			
	Main Sample		Repeated Sales		Main Sample		Repeated Sales	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
API (weighted)	0.000890***	0.000747***	0.000919***	0.000754***	-0.0152**	-0.0120***	-0.00969	-0.0101*
	(0.000282)	(0.000126)	(0.000263)	(0.000132)	(0.00586)	(0.00324)	(0.00596)	(0.00569)
Observations	82240	82240	24461	24461	82240	82240	24461	24461
R-squared	0.864	0.878	0.865	0.881	0.104	0.110	0.112	0.124
Adjusted R-squared	0.864	0.877	0.864	0.879	0.102	0.105	0.105	0.110
Controls:								
Year Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Zipcode FE	Yes	No	Yes	No	Yes	No	Yes	No
Neighborhood FE	No	Yes	No	Yes	No	Yes	No	Yes
House Characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Remodel	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Type of Finance	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Realtor Agency Loan services	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Buyer Agent Commission	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Inhouse dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dualagent dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Keywords in description	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Listing Types	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Flag for Re-listed Home	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports the OLS regression results on the effect of school quality on log sales price. School quality is measured by API per school district per year. In odd columns, standard errors are clustered at zip code level. In even columns, standard errors are clustered at neighborhood level. Significant Levels: * p<0.1, ** p<0.05, *** p<0.01

Table 2.7: Average School Quality on Log(Sales Price) and Days on Market

	Log(Sales Price)				Days on Market			
	Main Sample		Repeated Sales		Main Sample		Repeated Sales	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
API (average)	0.000903***	0.000695***	0.000929***	0.000708***	-0.00949*	-0.00781**	-0.00420	-0.00574
	(0.000273)	(0.000104)	(0.000244)	(0.000107)	(0.00543)	(0.00342)	(0.00480)	(0.00529)
Observations	82240	82240	24461	24461	82240	82240	24461	24461
R-squared	0.864	0.877	0.864	0.880	0.104	0.109	0.112	0.123
Adjusted R-squared	0.864	0.877	0.863	0.878	0.101	0.105	0.105	0.110
Controls:								
Year Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Zipcode FE	Yes	No	Yes	No	Yes	No	Yes	No
Neighborhood FE	No	Yes	No	Yes	No	Yes	No	Yes
House Characteristics	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Remodel	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Type of Finance	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Realtor Agency Loan services	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Buyer Agent Commission	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Inhouse dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Dualagent dummy	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Keywords in description	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Listing Types	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Flag for Re-listed Home	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table reports the OLS regression results on the effect of school quality on log sales price. School quality is measured by API per school district per year. In odd columns, standard errors are clustered at zip code level. In even columns, standard errors are clustered at neighborhood level. Significant Levels: * p<0.1, ** p<0.05, *** p<0.01

Bibliography

[1] James Albrecht, Pieter A Gautier, and Susan Vroman. "Directed search in the housing market". In: *Review of Economic Dynamics* 19 (2015), pp. 218 –231.

[2] MT Allen et al. "Salespersons compensation and performance in the housing market". In: *Journal of the Academy of Finance* 1.2 (2003), pp. 62–71.

[3] Panle Barwick and Parag A Pathak. "The impact of commissions on home sales in greater boston". In: *The American Economic Review: Papers and Proceedings* 100.2 (2010), pp. 475–479.

[4] Panle Jia Barwick and Parag A Pathak. "The costs of free entry: an empirical study of real estate agents in Greater Boston". In: *The RAND Journal of Economics* 46.1 (2015), pp. 103–145.

[5] Panle Jia Barwick, Parag A Pathak, and Maisy Wong. *Conflicts of Interest and the Realtor Commission Puzzle*. Tech. rep. National Bureau of Economic Research, 2015.

[6] Patrick Bayer, Fernando Ferreira, and Robert McMillan. *A unified framework for measuring preferences for schools and neighborhoods*. Tech. rep. National Bureau of Economic Research, 2007.

[7] Jonathan B Berk and Jules H Van Binsbergen. "Measuring skill in the mutual fund industry". In: *Journal of Financial Economics* 118.1 (2015), pp. 1–20.

[8] JonathanB̃ Berk and RichardC̃ Green. "Mutual Fund Flows and Performance in Rational Markets". In: *Journal of Political Economy* 112.6 (2004), pp. 1269–1295.

[9] B. Douglas Bernheim and Jonathan Meer. *Do Real Estate Brokers Add Value When Listing Services Are Unbundled?* Working Paper 13796. National Bureau of Economic Research, 2008.

[10] Sandra E Black. "Do better schools matter? Parental valuation of elementary education". In: *Quarterly journal of economics* (1999), pp. 577–599.

[11] William T Bogart and Brian A Cromwell. "How much is a neighborhood school worth?" In: *Journal of urban Economics* 47.2 (2000), pp. 280–305.

[12] Mark M Carhart. "On persistence in mutual fund performance". In: *The Journal of finance* 52.1 (1997), pp. 57–82.

[13] Jennifer L Castle, Xiaochuan Qin, W Robert Reed, et al. *How to pick the best regression equation: A review and comparison of model selection algorithms*. Tech. rep. 2009.

[14] Raj Chetty. "A new method of estimating risk aversion". In: *The American Economic Review* 96.5 (2006), pp. 1821–1834.

[15] KJ Martijn Cremers and Antti Petajisto. "How active is your fund manager? A new measure that predicts performance". In: *Review of Financial Studies* (2009), hhp057.

[16] Paramita Dhar and Stephen L Ross. "School district quality and property values: Examining differences along school district boundaries". In: *Journal of Urban Economics* 71.1 (2012), pp. 18–25.

[17] Marco Di Maggio, Amir Kermani, and Zhaogang Song. "The value of trading relationships in turbulent times". In: *SSRN Working Paper* (2015).

[18] Pedro Domingos. "A few useful things to know about machine learning". In: *Communications of the ACM* 55.10 (2012), pp. 78–87.

[19] Liran Einav and Jonathan D Levin. *The data revolution and economic analysis*. Tech. rep. National Bureau of Economic Research, 2013.

[20] Gabrielle Fack and Julien Grenet. "When do better schools raise housing prices? Evidence from Paris public and private schools". In: *Journal of public Economics* 94.1 (2010), pp. 59–77.

[21] Eugene F Fama. "The behavior of stock-market prices". In: *The journal of Business* 38.1 (1965), pp. 34–105.

[22] Eugene F Fama and Kenneth R French. "Luck versus skill in the cross-section of mutual fund returns". In: *The journal of finance* 65.5 (2010), pp. 1915–1947.

[23] David N Figlio and Maurice E Lucas. "What's in a grade? School report cards and the housing market". In: *American economic review* (2004), pp. 591–604.

[24] Mark J Garmaise and Tobias J Moskowitz. "Confronting information asymmetries: Evidence from real estate markets". In: *Review of Financial Studies* 17.2 (2004), pp. 405–437.

[25] David Genesove and Lu Han. "Search and matching in the housing market". In: *Journal of Urban Economics* 72.1 (2012), pp. 31–45.

[26] Martin J Gruber. "Another puzzle: The growth in actively managed mutual funds". In: *The journal of finance* 51.3 (1996), pp. 783–810.

[27] Lu Han and Seung-Hyun Hong. "In-House Transactions in the Real Estate Brokerage Industry: Matching Outcome or Strategic Promotion?" In: *Summer Real Estate Symposium. Monterey, California*. 2013.

[28] Jeffrey Heisler, Jarl G Kallberg, Crocker H Liu, et al. "The impact of dual agency". In: *The Journal of Real Estate Finance and Economics* 35.1 (2007), pp. 39–55.

[29] Igal Hendel, Aviv Nevo, and François Ortalo-Magné. "The Relative Performance of Real Estate Marketing Platforms: MLS versus FSBOMadison. com". In: *The American Economic Review* (2009), pp. 1878–1898.

[30] Chang-Tai Hsieh and Enrico Moretti. "Can Free Entry Be Inefficient? Fixed Commissions and Social Waste in the Real Estate Industry". In: *Journal of Political Economy* 111.5 (2003), pp. 1076–1122.

[31] Michael C Jensen. "The performance of mutual funds in the period 1945–1964". In: *The Journal of finance* 23.2 (1968), pp. 389–416.

[32] Ken Johnson, Leonard Zumpano, and Randy Anderson. "Intra-firm real estate brokerage compensation choices and agent performance". In: *Journal of Real Estate Research* 30.4 (2008), pp. 423–440.

[33] Marcin Kacperczyk, STIJN VAN NIEUWERBURGH, and Laura Veldkamp. "Time-Varying Fund Manager Skill". In: *The Journal of Finance* 69.4 (2014), pp. 1455–1484.

[34] Vrinda Kadiyali, Jeffrey Prince, and Daniel H Simon. "Is Dual Agency in Real Estate a Cause for Concern?" In: *The Journal of Real Estate Finance and Economics* 48.1 (2014), pp. 164–195.

[35] Robert Kosowski et al. "Can mutual fund stars really pick stocks? New evidence from a bootstrap analysis". In: *The Journal of finance* 61.6 (2006), pp. 2551–2595.

[36] Sotiris B Kotsiantis, Ioannis D Zaharakis, and Panayiotis E Pintelas. "Machine learning: a review of classification and combining techniques". In: *Artificial Intelligence Review* 26.3 (2006), pp. 159–190.

[37] Steven D Levitt and Chad Syverson. "Antitrust implications of home seller outcomes when using flat-fee real estate agents. rookings-Wharton Papers on Urban Economics". In: (2008).

[38] Steven D Levitt and Chad Syverson. "Market distortions when agents are better informed: The value of information in real estate transactions". In: *The Review of Economics and Statistics* 90.4 (2008), pp. 599–611.

[39] Burton G Malkiel. "Returns from investing in equity mutual funds 1971 to 1991". In: *The Journal of finance* 50.2 (1995), pp. 549–572.

[40] Burton G Malkiel and Eugene F Fama. "Efficient capital markets: A review of theory and empirical work". In: *The journal of Finance* 25.2 (1970), pp. 383–417.

[41] R Preston McAfee and Daniel Vincent. "Sequentially optimal auctions". In: *Games and Economic Behavior* 18.2 (1997), pp. 246–276.

[42] Henry J Munneke and Abdullah Yavas. "Incentives and performance in real estate brokerage". In: *The Journal of Real Estate Finance and Economics* 22.1 (2001), pp. 5–21.

[43] Gary E Porter and Jack W Trifts. "The Best Mutual Fund Managers: Testing the Impact of Experience Using a Survivorship Bias Free Dataset (Digest Summary)". In: *Journal of Applied Finance* 22.1 (2012), pp. 105–117.

[44] Gary E Porter and Jack W Trifts. "The career paths of mutual fund managers: the role of merit". In: *Financial Analysts Journal* 70.4 (2014), pp. 55–71.

[45] Spencer Rascoff and Stan Humphries. *Zillow Talk: The New Rules of Real Estate*. Grand Central Publishing, 2015.

[46] Evans Richard and Kolbe Phillip. "Homeowners' Repeat-Sale Gains, Dual Agency and Repeated Use of the Same Agent". In: *Journal of Real Estate Research* 27.3 (2005), pp. 267–292.

[47] Richard Rogerson, Robert Shimer, and Randall Wright. "Search-Theoretic Models of the Labor Market: A Survey". In: *Journal of Economic Literature* 43.4 (2005), pp. 959–988.

[48] Ronald C Rutherford, TM Springer, and Abdullah Yavas. "Conflicts between principals and agents: evidence from residential brokerage". In: *Journal of financial Economics* 76.3 (2005), pp. 627–665.

[49] William F Sharpe. "The arithmetic of active management". In: *Financial Analysts Journal* 47.1 (1991), pp. 7–9.

[50] Orie Shelef and Amy Nguyen-Chyung. "Competing for Labor through Contracts: Selection, Matching, Firm Organization and Investments". In: (2015).

[51] Hal R Varian. "Big data: New tricks for econometrics". In: *The Journal of Economic Perspectives* 28.2 (2014), pp. 3–27.

[52] Russ Wermers. "Mutual fund performance: An empirical decomposition into stock-picking talent, style, transactions costs, and expenses". In: *The Journal of Finance* 55.4 (2000), pp. 1655–1703.

[53] William C Wheaton. "Vacancy, search, and prices in a housing market matching model". In: *Journal of Political Economy* (1990), pp. 1270–1292.

[54] Xindong Wu et al. "Top 10 algorithms in data mining". In: *Knowledge and information systems* 14.1 (2008), pp. 1–37.

[55] Abdullah Yavaş. "A simple search and bargaining model of real estate markets". In: *Real estate economics* 20.4 (1992), pp. 533–548.

[56] John Yinger. "A search model of real estate broker behavior". In: *The American Economic Review* 71.4 (1981), pp. 591–605.

Appendix A

Observed House Attributes Included in the Analysis

Basic measures of house scale
Number of bedrooms, Number of bathrooms, Age, Number of pictures, sqft, lotsqft

Indicators of housing quality
Pool (categorical): missing, 0, 1
Fireplace (categorical): missing, 0, 1
Remodeling (categorical):
none, slightly remodeled, bathroom remodeled, kitchen remodeled, bathroom and kitchen remodeled, completely remodeled
House style (categorical): Bungalow, Cape cod, Colonial, Contemporary, Craftsman, French, Georgian, Loft, Modern, Other, Pueblo style, Ranch, Spanish, Split level, Tudor, Victorian
House roof (categorical): missing, Asphalt, Built up, Composition, Metal, Other, Shake Shingle, Slate, Tile

Presence of (dummy variable equaling to one if attribute is present in house):
House exterior: Brick, Cement concrete, Composition, Metal, Other, Shingle, Stone, Stucco, Vinyl, Wood, Wood products Heating: Baseboard, Electric, Forced, Gas, Heat Pump, None, Other, Radiant, Stove, Wall
Cooling: Central, Evaporative, Geothermal, None, Other, Refrigeration, Solar, Wall
Flooring: Carpet, Concrete, Hardwood, Laminate, Linoleum / Vinyl, Other, Slate, Softwood, Tile

Keywords used to describe home in listing (dummy variable equaling one if word/phrase or some shortened variant of it is used in the home description)

needs updating, estate sale, as is/as-is, brand new, must see, handyman, needs, tlc, motivated, potential, close, !, new, spacious, elegance, beautiful, appealing, renovated, vintage,

state-of-the-art, maintained, wonderful, fantastic, charming, stunning, amazing, granite, immaculate, breathtaking, neighborhood, spectacular, landscaped, built-in, tasteful, fabulous, leaded, delightful, move-in, gourmet, copper, corian, custom, unique, maple, newer, hurry, pride, clean, quiet, dream, block, huge, deck, mint, stately.

Appendix B

A seller side search model with risk aversion

This model is built upon Munneke and Yavas (2001)[42] and Berk and Green (2004)[8], and the real estate search literature (Yinger (1981)[56]; Wheaton (1990)[53]; Yavas (1992)[55], Rutherford et al. (2005)[48]; Genesove and Han (2012)[25]; Albrecht et al. (2015)[1]), the labor literature (a detailed review of search-theoretic models can be found in Rogerson et al. (2005)[47]) and the repeated auction literature (McAfee and Vincent (1997)[41]). It is a standard discrete time search model with risk averse sellers and risk neutral agents.

Consider the problem of a risk-averse seller who works with a listing agent and has to decide whether to take up the best offer now or continue to search for another best offer in the next period. To keep the analysis simple, assume sellers are homogeneous in risk aversion $\lambda > 0$ and search cost $c > 0$. In each search period, the seller pays a search cost c, which is the cost of waiting, inconvenience or any opportunity cost of not moving out sooner. The listing agent attracts a number of buyers N using different marketing strategies such as open houses, virtual tours and the listing advertisement. The seller observes the number of offers N and infers the future number of offers from N after the first open house. The future number of offers is an increasing function of N. To keep the model simple, I assume that the seller infers the future number of offers fully from today's number of offers ceteris paribus. Obviously, it is unrealistic to assume that the seller believes that he will receive exactly the same number of offers in the future as today. This assumption simplifies the model without changing the qualitative results. All I need for the results to hold is that seller believes that the future number of offers is positively correlated with today's number of offers ceteris paribus.

Buyers draw their match-value of the home from the same uniform distribution $U[0, 1]$ and they bid their match-value as in a second price auction. Note the match-value is buyers' private information. By order statistics, the maximum bid p follows $Beta(N, 1)$. At the end of each period, the seller optimizes over the maximum bid distribution and decides whether to take the best offer p now or continue to search with his listing agent. Following the labor literature closely, I use a CRRA utility function to analyze the problem. Note the results

would continue to hold for any given concave utility function.

Seller's Maximization problem:

$$\begin{aligned} U_{accept} &= \frac{1}{1-\theta}[p(1-\tau)]^{1-\theta} \quad for\ \theta \neq 1 \\ U_{decline} &= \frac{1}{1+r} * \big(\int \max\{U_{accept}, U_{decline}\} dG(p) - c \big), \end{aligned} \tag{B.1}$$

where U_{accept} is the utility to the seller if he accepts the maximum bid p (I use "the maximum bid" and "the best offer" interchangeably). $\frac{1}{1+r}$ is the discount factor. $G(p)$ is the cumulative distribution function of p, $G(p) = p^N$. $g(p)$ is the density function of p, $g(p) = N * p^{N-1}$. Note that for $\theta \to 1, U_{accept} \to log(p(1-\tau))$. All results would still hold.

$U_{decline}$ is the utility if he declines. There exists a unique p_R such that $U_{decline} = U_{accept}(p_R)$ with the property that seller accepts the offer if $p \geq p_R$, declines if $p < p_R$.

Then:

$$(\frac{1+r}{1-\theta} - 1)[p_R(1-\tau)]^{1-\theta} + c = \int_{p_R}^{1} \frac{1}{1-\theta} \big([p(1-\tau)]^{1-\theta} - [p_R(1-\tau)]^{1-\theta} \big) * N * p^{N-1} dp. \tag{B.2}$$

The LHS describes the cost of search, while the RHS describes the expected gain from future search.

Then:

$$(1+r)p_R^{1-\theta} - \frac{(1-\theta)}{N+1-\theta} p_R^{N+1-\theta} = \frac{N}{N+1-\theta} - \frac{c(1-\theta)}{(1-\tau)^{1-\theta}}. \tag{B.3}$$

Comparative statics:
The larger the discount factor $\frac{1}{1+r}$, the higher the continued search utility, and the higher the p_R.
The larger the search cost c, the lower the continued search utility, and the lower the p_R.
The larger the commission rate τ, the lower the continued search utility, and the lower the p_R.
The larger the number of buyers N, the higher the continued search utility, and the higher the p_R.
The larger the risk aversion θ, the lower the continued search utility, and the lower the p_R.

Sales price p_A^i (accepted price):
Similarly, p_A follows the beta distribution $Beta(N, 1)$ truncated at $p_R(N, \theta, r, c, \tau)$.
For the seller working with the listing agent with N, the average sales price is

$\overline{p_A} = \int_{p_R}^{1} p * \frac{N*(p-p_R)^{N-1}}{(1-p_R)^N} dp = \frac{N+p_R}{N+1}$.

The variance of the sales price is

$Var(p_A) = \frac{N(1-p_R)^2}{(N+1)^2(N+2)}$.

$\overline{p_A}$ is increasing in N, for any given N. $Var(p_A)$ is a decreasing in N.

Assume experienced listing agents draw $N_E \sim N(N_E, \sigma_E^2)$, inexperienced ones draw $N_I \sim N(N_I, \sigma_I^2)$, where $N(N_E, \sigma_E^2)$ and $N(N_I, \sigma_I^2)$ are the distributions of the maximum bids. Experienced agents are more productive. But as they become busier, their average level of effort per listing decreases. With perfect mobility of listings, experienced agents can even have a lower sales price in the long run equilibrium as they are preferred by risk-averse sellers. However, listing agents are capacity constrained on the number of listings. Moreover, in the short-run, sellers need time to assess agent performance and update their beliefs. Thus, the setting is slightly different from Berk and Green (2004)[8] and listing agents will not obtain as many listings as when ability is perfectly observable and listings are perfectly mobile. Thus $N_E \sim N_I$.

This leads to

Proposition 1: $\overline{p_A}_E = \overline{p_A}_I$ iff $N_I = N_E$. Listing agent experience does not have an impact on the average sales price.

Proposition 2: $Var(p_A)_E < Var(p_A)_I$ iff $\sigma_E < \sigma_I$. Due to survivorship of better agents over time, there is less heterogeneity in ability among experienced listing agents, which leads to lower variance in sales price.

Even though I cannot test the empirical distribution of the number of bids due to the lack of information to support the conditions for above propositions in this study, I can test for any difference in sales price among agents at different experience levels. This section provides a framework for future research that has information on both the empirical distributions of the number of bids and sales price.