Seeing is Believing:

The Impact of Buyers' Onsite Viewing Activities on Housing Transactions[1]

Maggie Rong Hu
The Chinese University of Hong Kong
maggiehu@cuhk.edu.hk

Weida Kuang
Renmin University of China
kuangweida@rmbs.ruc.edu.cn

Xiaoyang Li
Jinan University
lixiaoyang@jnu.edu.cn

Yang Shi
University of Melbourne
shiy7@student.unimelb.edu.au

February 2022

[1] We thank Yuming Fu, Zhenyu Gao, Chongyu Wang, Su Wang, and Sisi Zhang for their valuable comments and suggestions. We also thank the anonymous referees, discussants, and participants at the AREUEA, AsianFA, and Jinan-SMU-ABFER Conference for their helpful comments. All errors remain our own.

Seeing is Believing:

The Impact of Buyers' Onsite Viewing Activities on Housing Transactions

Abstract

Buyers' onsite viewing is an important process of house transaction and is the direct measure of buyer search. Yet empirically we know little about the information revealed through a buyer's onsite viewing, and neither do we know about the impacts of buyers' onsite house viewings on transaction outcomes. Using a unique proprietary dataset which includes 4,397,652 onsite viewing records and 621,040 transaction outcomes from the largest real estate agency in China, we find that buyers who are more active in onsite viewings are associated with larger deal likelihoods, as well as higher purchase prices and greater chances of making over budget payment in completed deals. The findings suggest that a buyer's onsite viewing is a reflection of his or her housing demand. Buyers achieve improved deal likelihood from active searching. However, as they reveal their stronger demand to sellers through active onsite house viewings, they lose bargaining power and end up paying higher prices. To establish causality, we perform instrument variable regressions exploiting the exogenous variations in onsite house viewings caused by national basketball games that increase the opportunity cost of searching, and find consistent results.

Keywords: Housing search; Buyer search; Onsite house viewing; Housing demand; House price

JEL code: R20, R21, R30

1 Introduction

The housing market is a typical search market where both sellers and buyers search for each other for matches based on their conditions and requirements (Han and Strange, 2015). In searching for ideal properties, buyers visit and see their interested houses in person, before negotiating with the sellers. Onsite house viewing is the last step in the house searching stage, and it is the most direct measure of buyers' searching activities. Unfortunately, due to data limitations, our knowledge is limited when it comes to the impacts of buyers' onsite house viewings on transaction outcomes and the information revealed through onsite house viewings. In particular, how number of onsite house viewings affect deal likelihood? Will more onsite house viewings increase or decrease transaction price? And more importantly, what are the economic implications of onsite house viewings? In this research, we empirically answer the above questions using highly granular visit-level house transaction proprietary data.

In housing markets, buyers actively search for houses that meet their requirements and budgets, and they play more active roles than sellers in searching (Baryła and Zumpano, 1995). However, most of the prior research on housing search focuses on the seller side (see Cubbin, 1974; He et al., 2020, among others). The biggest challenge of empirical investigation into buyer search is data availability (Han and Strange, 2015). Our proprietary dataset from the largest real estate agency in China consists of 4,397,652 unique onsite property viewing records that reflect buyer search and 621,040 housing transactions in the city of Beijing. Thanks to the high granularity of data, which contains the complete onsite viewing records of each buyer, detailed features of the properties viewed, deal outcome after each viewing, and the purchase prices in completed deals, we are able to accurately measure buyer search (Anglin, 1997; Genesove and Han, 2012) and examine how buyers' onsite viewings affect transaction outcomes.

Based on prior research, we propose four possible indications of buyers' onsite house viewings and analyze their implications on deal likelihood and deal price. Seller side evidence shows that sellers with stronger motives are more active in searching, leading to quicker sales at lower prices (Anglin, Rutherford, and Springer, 2003; Turnbull and Zahirovic-Herbert, 2011). Thus, the first reasonable indication of buyers' onsite house viewings should be their stronger demand, which positively predicts transaction probability and price. Anenberg (2016) shows that sellers learn about market demand during the searching process. Therefore, the second indication of buyers' onsite house viewing is buyer learning. A buyer reduces his or her information disadvantage (Bian et al., 2021; Hayunga and Munneke, 2021) by onsite seeing more houses, allowing him or her to increase the chance of purchasing a house and paying a lower price. The third possibility is that more onsite house viewings are the reflection of fierce market competition. When multiple buyers chase one house, the buyers who failed in purchasing the house need to visit other houses, leading to a higher number of onsite house viewing. The direct implication of the supply and demand conditions in housing markets is that more onsite house viewings predict lower transaction probabilities but higher sales prices (Van Dijk and Francke, 2018). The last indication is bargain hunting, where the buyer has a very tight budget relative to the housing features required. Focusing on the seller's constraint, the unrepaid mortgage, Genesove and Mayer (1997) finds that a tighter budget lowers the deal likelihood. Similarly for bargaining hunting buyers, while they pay bargaining prices conditional on deals, their chance of purchasing houses is lower. These four indications predict different combinations of the impacts of onsite viewings on deal likelihood and deal price. By empirically examining the impacts of buyers' onsite house viewing, we are able to reveal not only the influence of buyer search on transaction outcomes, but also the indication of buyers' onsite house viewings.

The empirical analysis starts from relating buyers' onsite house viewings to deal likelihoods using the full sample of onsite viewing records. We find that buyers with more onsite house viewings have higher chances of purchasing a house. On average, 1% increase in onsite house viewings result in 6.9 percentage points higher purchasing probability or 47.6% purchasing odds, respectively.

We also analyze the impact of buyers' onsite house viewings on transaction prices for those buyers who eventually make a home purchase after their onsite viewings. Note that transaction prices are only available for a subsample of buyers who purchase the house after viewing. As buyer characteristics may be related to the probability of closing a deal, the decision to buy a home may not be random, which gives rise to potential sample selection issues. To address the sample selection bias, we use the Heckman Selection Model, besides OLS analysis. We obtain consistent estimation results using both models where buyers with more onsite viewings prior to their home purchase pay higher transaction prices. Specifically, 1% more onsite viewing, on average, increases the total transaction price by RMB 92,690 (=0.026*3,565,000) and increases the probability that a buyer makes an over budget payment by 16.6 percentage points.

Our baseline findings reveal twofold impacts of active onsite house viewings by buyers. On one hand, buyers that are more active in searching achieve higher deal likelihood. On the other hand, they pay for additional price as compared to other buyers who do not actively onsite view houses. The pattern supports buyers' stronger demand as the key indication of their active onsite house viewings.

Next, we discuss the possible endogeneity associated with omitted variables. For example, buyer characteristics that are not captured in our data could affect both buyers onsite house viewings and the transaction outcomes. Motivated by Hu and Lee (2020) which uses Melbourne Cup horse races as an exogenous shock on real estate transactions, we use the Chinese Basketball Association (CBA) playoff seasons as the instrumental variable for the total number of onsite house viewings to establish causality. While CBA playoff matches are not related to housing transaction outcomes and (possibly) omitted homebuyer characteristics, they raise the opportunity cost of searching and depress buyers' onsite viewing activities. Empirically, we construct our IV as a dummy variable that indicates whether a buyer's onsite house viewing period overlaps with the CBA playoff seasons. The variable negatively predicts buyers' onsite house viewings as expected, and the main findings hold in the instrumental variable approach.

Note that CBA playoff seasons are fixed in February and March of each year, the impacts of seasonality could confound our IV analysis result. To address this issue, we construct a finer measure of the IV that reflects whether a buyer's onsite house viewing period overlaps with the game period of Beijing's basketball team, i.e., before Beijing is eliminated from the playoffs. Within the whole playoff season, the searching cost is higher when for potential buyers in Beijing when Beijing is still one of the participating teams. We restrict our data to a subsample of buyers whose searching period falls into the CBA playoff seasons of each year, i.e., February and March, and use this refined IV to provide a clearer cut identification. We find consistent evidence that while more active onsite viewings improve deal likelihoods, they also lead to higher transaction prices and more above budget payments.

Lastly, we carefully discuss the other three conjectures, namely buyer learning, buyer competition, and bargain hunting. We present empirical evidence that they are not likely to be the key indication of buyers' onsite house viewings. We also perform a series of robustness checks using an alternative sample excluding unserious buyers who only have single onsite viewing records, as well as alternative model settings that replace the business zone fixed effects of each house with district fixed effects and complex fixed effects, respectively. The results are consistent in all settings, proving the robustness of our findings.

Our paper makes three important contributions. First and foremost, our empirical evidence fills the literature gap on the impact of the buyer search on transaction outcomes in the housing market. In the housing search framework, buyers play a more active role than sellers who passively set asking prices and wait for potential buyers. Thus, analyzing the housing search and match from the buyer side is especially important. However, extant studies mainly focus on sellers' side activities while buyer searches are relatively scarce and insufficient as the information about buyer search is limited. Although several studies investigate the buyer search using the NAR survey data, they have limitations in the frequency, response rate and representativeness (Genesove and Han, 2012).

Using a proprietary dataset on buyers' onsite viewing records and transaction outcomes from the largest real estate agency in China, we show that buyers' search intensity increases deal likelihoods but also raises transaction prices at the same time. Apart from the research that uses the NAR survey data, a recent paper by Gargano, Giacoletti, and Jarnecic (2020) studies the buyer search using individual user's online pre-search data and provides evidence on the impacts of online broad searching. Our study differs from this paper in many aspects. Specifically, we use onsite rather than internet search records to measure the buyer search, which is one step closer to making transaction decisions for each buyer in the real estate market. More importantly, the individual transaction data allow us to observe the transaction outcomes of every single individual and relate the outcomes to the search activities of the same person. To the best of our knowledge, the real transaction records of this granularity are hitherto not used before.

Second, we prove that a buyer's total number of onsite house viewings is a direct measure of housing demand. In the housing market, there are rich data on transaction outcomes from various sources, and the equilibrium quantity and price can be easily observed. However, the housing demand cannot be explicitly observed. Due to data limitations, most of the prior research either estimates latent demands using variables such as economic and demographic factors (Marcin and Kokus Jr, 1975; Boehm and McKenzie1982), housing consumption needs (Ziegert, 1988), adjustment costs (Harmon and Potepan, 1988), etc. or using different proxies for buyer demands (for example, turnover of existing houses in Berkovec and Goodman (1996); internet searching in Van Dijk and Francke (2018), amount others). In our paper, thanks to the granularity of the data, we show that at the individual buyer level, the total number of onsite house viewings is a direct measure of his or her housing demands and valuation.

Last but not least, our findings offer a practical implication that buyers' onsite house viewing records could predict transaction outcomes, which could potentially be useful for property sellers as well as real estate agency practitioners. One of the direct implications of our findings is that although onsite house viewings improve deal likelihoods, they also reveal buyers' demand information to sellers. The revealed high demand reduces buyers' bargaining power in the housing market and could lead to higher transaction prices and above budget payments. Potential buyers could significantly benefit from our findings as the house usually constitutes the largest single asset of households (Stein, 1995; Díaz and Jerez, 2013). As big data on property search and viewing become more and more available, practitioners could also benefit from our results by utilizing the new data source to match sellers and buyers more efficiently.

The rest of the paper is organized as the following: Section 2 introduces the background of Chinses housing market and the agency we collect our data from. Section 3 develops our hypotheses. Data and sample descriptions and empirical results are presented in Sections 4 and 5. Section 6 concludes.

2 Institutional Background

China has a very large real estate market, and the market size is even greater than that of the US. According to Zillow research, the total value of homes in the US is $33.6 trillion in 2020.[2] China, as estimated by Goldman Sachs Group Inc. (GS), has a $52 trillion homes and developers' inventory in 2019.[3] Despite the differences in measurement scope and estimation year, the higher volume of the Chinese housing market is still remarkable as compared to the US market.

To study buyer search in housing markets, we use comprehensive property onsite viewing data in Beijing from a leading Chinese real estate agency, which is the largest PropTech unicorn in China and ranks the third place worldwide, following WeWork and AirBNB in valuation (Baum et al., 2020). Although there are more than 10 real estate agencies that operate in Beijing, an internal report from the studied agency shows that it facilitated more than 60% of the housing transactions in 2015. Recent data from the Beijing Municipal Commission of Housing and Urban-Rural Development show that in February 2021, the company contributes 7,132 out of 10,228 total transactions in Beijing, translating into a market share of 69.73%. The firm's dominating place ensures the representative of our data.

In the US, house sellers and buyers are served by listing agents and cooperating agents separately. But the real estate agencies in China are two sided. For the studied agency, it assigns different agents to serve both seller clients and buyer clients at the same time. The agency has a commission sharing mechanism that encourages agents to actively acquire seller clients and buyer clients. That is, after facilitating a deal, the commission will be partitioned between the two agents who serve the seller client who offers the house sold and the buyer client who purchased the house.

Agents who acquire a seller client who has a house to be sold will sign a contract with the seller that authorizes the company to match his or her house with the company's buyer clients. The agent will then input the house characteristics and asking price into the company's system, which is available to all agents in the company. In many cases, a house is delegated to multiple real estate agency companies.

---

[2] Retrieved from https://www.worldpropertyjournal.com/real-estate-news/united-states/los-angeles-real-estate-news/real-estate-news-zillow-housing-data-for-2020-combined-housing-market-value-in-2020-us-gdp-china-gdp-rising-home-value-data-11769.php.

[3] Retrieved from https://www.wsj.com/articles/china-property-real-estate-boom-covid-pandemic-bubble-11594908517.

Delegation is free of charge, and only the agency company that sells the house gets commission fees. To secure the commission fee, the agent has to make sure the house is purchased by the company's own buyer clients before being sold by other real estate agency companies. The competition gives the agent strong incentives to persuade the seller client into reducing the asking price on the company's system so that the house can be matched with more buyer clients and sold with higher probabilities. A seller client has the right to adjust the asking price upward or downward at his or her own will. The agency company also keeps these records of the price adjustments of each delegated house.

When a buyer side agent acquires a buyer client, the agent will first ask the buyer about his or her budget range, as well as the ranges of house areas and the number of bedrooms the buyer is interested in. The agent will ask buyer clients some questions such as *What is your maximum/minimum budget? How many bedrooms do you expect of your own house?* or *This house has an asking price of RMB 8 million, will you consider it?* This is a standardized practice that helps the agent to learn a buyer client's demand and search for matches more efficiently. The buyer information is also recorded in the company's system, thus available in our data. The agent will search in the company's database for delegated houses that meet the buyer client's requests. Next, the agent recommends possible matches to the buyer client, who will browse the information of those houses on the company's website and select the potential targets. Then, with the buyer's permission, the agent will take the buyer to onsite view the selected houses. The company keeps clear onsite viewing records, as they are important for the internal commission sharing process. When a deal is closed, only two agents who either serve the seller client of the house sold or accompany the buyer client to onsite view the selected house are eligible of sharing the commission.

After onsite house viewings, a buyer first expresses interest in a certain house and makes an offer to the house owner through his or her agent. Seller receiving the offer has the right to accept, make a counteroffer or directly reject the offer. The buyer and seller can make unlimited rounds of offers and counteroffers until the bid and ask prices are close enough. Then, the two agents that serve the seller client and buyer client will help arrange a face-to-face meeting. If both sides can agree on a final transaction price, they close the deal. In this case, the seller and buyer pay 1% and 2.2% of the transaction price as a commission to the company, respectively.

3 Literature Review and Hypothesis Development

To understand buyers' search effort through onsite viewing activities, we start by analyzing the uniqueness of the housing market. Different from efficient goods and financial market, which clear through price adjustments. The housing market is far from perfectly efficient (Case and Shiller, 1990). Due to the heterogeneities among houses, information frictions, the lack of central exchange, among other reasons, the housing market can be characterized as a typical search market, where both buyers and sellers need extra effort to search for each other for matches (Wheaton, 1990). The theoretical analysis starts from a one-sided search, which studies seller search (Stull, 1978) or buyer search (Courant, 1978) separately. Following research further models the process of search and match between sellers and buyers under the framework of random search, focused search, segmented search and directed search.[4] Recent theory papers calibrate search models using microdata to explain the time series dynamic of housing prices and housing markets (Díaz and Jerez, 2013; Head et al., 2014), focusing on foreclosures (Guren and McQuade, 2020), broad housing searchers (Piazzesi, Schneider, and Stroebel, 2020), joint buyer-sellers (Anenberg and Bayer, 2020), mortgages (Garriga and Hedlund, 2020), etc.

Due to data availability, the majority of the empirical evidence is on the seller side, using Multiple Listing Service (MLS) data in the US. For example, research shows that seller search activities and transaction outcomes are affected by factors such as homeownership (Kang and Gardner, 1989), down payments and liquidity constraints (Stein, 1995), pricing strategies (Donald, Terry, and Daniel, 1996), etc. Cheng, Lin, and Liu (2021) also investigate the optimal search strategies of sellers in the housing market. However, the investigation into buyer search is scarce and evidence is drawn from either questionnaire data (Anglin, 1997) or National Association of Realtors (NAR) survey data (e.g., Baryła and Zumpano, 1995; Elder, Zumpano, and Baryla, 1999; Baryla, Zumpano, and Elder, 2000) that are of low frequency and response rate (Genesove and Han, 2012). Given the limited prior empirical evidence on buyer search, we analyze buyers' onsite house viewings leveraging the findings of seller search and develop our hypotheses.

Why some buyers are more active in searching than others and onsite view more houses? The first possible explanation is that more onsite viewings reflect buyers' stronger demand. Seller side evidence reveals a positive relationship between sellers' search activities and their motive to sell quickly (Springer, 1996). For example, Anglin, Rutherford, and Springer (2003) show sellers of stronger motives use lower listing prices to attract potential buyers, leading to higher deal likelihood within a shorter searching time. Focusing on owners of vacant houses and sellers with reallocation plans, Zuehlke (1987), Glower, Haurin, and Hendershott (1998), Knight (2002) and Turnbull and Zahirovic-Herbert (2011) provide consistent evidence. Similarly, more active searching should be also be related to buyers' stronger demand. Under

[4] Han and Strange (2015) provides a detailed survey on the microstructures of housing search.

the search framework, a buyer keeps searching for houses until either of these two events happen: 1) the buyer purchases a house, i.e., the buyer matches with a seller in the market, or 2) the buyer quits the market without purchasing any houses, which means the buyer fails to match with other sellers in the market. In deciding whether to visit more houses or quit searching, a buyer compares the search benefit and search cost, including both the direct cost and the opportunity cost in searching. The searching activity continues if and only if the benefit is enough to cover the cost. Buyers with stronger housing demand have a more positive assessment of owning a house, leading to a higher valuation of searching. Therefore, more onsite house viewings should be associated with buyers' stronger demand, and predict higher deal likelihoods.

Another possibility is that buyers' learning by searching. Recent literature in the search model emphasizes agents' learning through the searching activities under information asymmetry (Eaton et al., 2021). The housing market is characterized by high information asymmetry (Firoozi et al., 2006; Wong, Yiu, and Chau, 2012), and Anenberg (2016) provides theoretical and empirical evidence on sellers' Bayesian learning about the uncertain market demand. In housing markets, buyers are at information disadvantages against both sellers (Bian et al., 2021) and real estate agents (Levitt and Syverson, 2008; Hayunga and Munneke, 2021). Therefore, buyers through onsite viewing different houses, acquire more information about the housing market. They gain more accurate understandings of the market conditions, housing valuation, fair values of business zones, etc. Therefore, buyers who spend more time and energy searching should have higher chances of finding better deals.

In both of the above cases where buyers' onsite house viewings reflect their strong demand or active learning, buyers onsite visit more houses in order to improve transaction probability. Thus, we propose the first hypothesis as follows.

H1a: Other things being equal, more onsite house viewings positively predict deal likelihoods.

However, more onsite viewings could also be the outcome of fierce buyer competition. Like any other market, the transaction outcomes in the housing market are affected by the supply and demand conditions (Novy-Marx, 2009; Van Dijk and Francke, 2018). When buyer competition is fierce in the housing market, multiple potential buyers compete to buy one listed house. For each buyer, an onsite visited house could be sold to other buyers, and the buyer needs to start over the searching process again and onsite view more houses that are still on market.

Under the search framework, fierce buyer competition reduces the contract rate, leading to a lower probability of house purchasing. In other words, although buyers onsite view a larger number of houses, most of the purchasing attempts ended up unsuccessful, resulting in lower deal likelihoods.

Another reason for buyers' active onsite housing viewing is bargain hunting, where a buyer has a very tight budget relative to the housing features he or she demands. These buyers only purchase houses at bargaining prices, which is similar to the indebted homeowners who only sell houses at high prices. Sellers are constrained by their remaining mortgage value when selling houses, and Genesove and Mayer (1997) finds that higher loan-to-value houses stay on market for a longer time and are harder to be sold.

Applying their findings to the buyer side, the bargain hunters are also more difficult to find matching houses. They need to visit more houses, and even go through the negotiation process with different sellers to find houses within their budget. Similar to the effects of buyer competition, bargain hunting also leads to a lower contract rate and reduced the probability of purchasing a house.

Therefore, if more onsite house viewings reflect fierce buyer competition or bargain hunting, then they should be adversely related to transaction probability, leading to the competing hypothesis

H1b: Other things being equal, more onsite house viewings negatively predict deal likelihoods.

We proposed four possible indications of buyers' active house viewings. Two of them predict higher deal likelihoods and the other two suggest the opposite. Apart from deal likelihoods, we are equally interested in the impacts of buyers' onsite viewing activities on transaction prices. These four indications also predict transaction prices differently. Combining the results on both deal likelihood and deal price, we are able to better reveal the economic implications of buyer search.

Evidence on seller search shows that stronger motives revealed through active searching could hurt a seller by reducing his or her bargaining power. The key idea of seller search is to model the tradeoff between higher transaction prices and shorter sales time (Wheaton, 1990; Yavas and Yang, 1995; Krainer and LeRoy, 2002, among others). Active seller-searchers use lower asking prices to invite more arrivals of buyers but at the cost of lower deal price (Merlo and Ortalo-Magne, 2014; Albrecht et al., 2016; Hayunga and Pace, 2019). These findings can help us understand the influence of buyer search on transaction prices.

In particular, if our first conjecture that a buyer's onsite house viewing reflects his or her stronger housing demand is true, then the onsite house viewing should be related to higher transaction prices. By onsite viewing a house, a buyer expresses interests in the property. A greater number of onsite viewings indicates that the buyer is willing to consider more houses, revealing his or her high valuation of owning a house to the market participants. Sellers, through observing the buyer's onsite viewing activities, understand the buyer has a higher reservation value, which gives them stronger market power

(Wilhelmsson, 2008), allowing them to bargain for a higher deal price (Steegmans and Hassink, 2017). From a buyer's perspective, more active onsite house viewings lead to additional payments in purchasing a house.

Apart from stronger demand, our third proposition on buyer competition also predicts an increased transaction price. Supply side evidence shows that seller competition leads to lowered transaction prices. Turnbull and Dombrow (2006) find that greater spatial concentration of sellers increases price competition. Using foreclosure as a supply shock, Anenberg and Kung (2014) prove that seller competition decreases housing prices. Their findings are robust to disentangling the disamenity effect associated with the foreclosure.

Both seller and buyer competition influence deal price in housing markets. Harding et al. (2003) describe the relationship between market power and transaction price as "weak buyers pay higher prices and weak sellers receive lower prices for their homes". That is to say, prior findings on seller competition indicate that buyers also suffer losses from buyer competition. If the more active onsite house viewing is a reflection of fierce buyer competition, then the deal price should be higher.

Both stronger demand and fierce buyer competition suggest a positive relationship between number of onsite viewings and deal prices, leading to our second main hypothesis.

H2a: Other things being equal, buyers with more onsite house viewings prior to their home purchase pay higher purchase prices.

Qiu and Zhao (2018) show that less informed home buyers in the secondary housing market pay around 1~2.3% more than those better-informed homebuyers. Fan et al., (2021) finds consistent evidence in Hong Kong that mainland buyers on average pay about 1.3% premium as compared to locals. Utilizing the disclosure of noise around the airport, Pope (2008) finds that when buyers are better informed about the noise level, they pay about 2.9% less price. One implication of their findings is that if the onsite viewing is the learning process of the buyers, then more onsite viewings reduce their information disadvantage in the housing market, allowing them to bargain for better deals. Thus, transaction price and onsite viewings should be negatively related.

Besides, bargain hunting also predicts lower deal prices, holding property characteristics constant. Genesove and Mayer (1997) find that houses are sold at higher prices when the homeowners have a larger amount of outstanding mortgage and are constrained by their borrowing in house selling. Ortalo-Magne and Rady (2006) provide market level evidence that house price decreases when buyers are subject to negative shocks on their budget. If more frequent onsite viewing is due to bargain hunting and limited budgets, a buyer will not purchase a house unless its price is low enough to meet his or her strict requirement. Conditional on purchasing a house, the transaction price should be lower among buyers who onsite view more houses to search for cheap deals.

The conjectures that buyers' learning through onsite house viewings and that more onsite house viewings reflect bargain hunting suggest that more active onsite viewings lead to reduced transaction prices. Thus, we prose the competing hypothesis as,

H2b: Other things being equal, buyers with more onsite house viewings prior to their home purchase pay lower purchase prices.

So far, we discussed four possible implications of buyers' active onsite viewings, namely stronger demand, buyer learning, buyer competition, and bargain hunting. These four indications predict four different combinations of the impacts of onsite house viewings on deal likelihood and deal price. Thus, by examining the influence of onsite house viewing on both dimensions, we are able to empirically reveal the dominating indications of buyers' onsite visits. The table below summarizes the prediction of the four indications.

		Deal likelihood	
		Higher	Lower
Deal price	Higher	Stronger demand	Buyer competition
	Lower	Buyer learning	Bargain hunting

4 Data and Sample Description

We empirically examine the above hypotheses using a proprietary dataset with detailed onsite viewing records and transaction outcomes from January 2013 to December 2017 in Beijing from the largest Chinese real estate agency. The total transaction volume in our sample is RMB 244,843.4 million. These large amounts of the data are from a single real estate agency, thereby having advantages of consistency in practices and marketing procedures that may vary between different real estate brokers. (McGreal et al., 2009).

Our data are of three layers. At the buyer level, each buyer is assigned a unique buyer ID, and we know the gender, age, locality, payment method (i.e., full cash payments or financing by bank loans) and budgets on total housing prices, areas in square meters, and the total number of bedrooms of each buyer. More importantly, the data include the detailed onsite house viewing records, including the house viewed with a unique property ID assigned by the agency and the viewing time. The buyer level data allow us to construct the focal variable, *TotalVisit*, which is the total number of onsite house viewings of each buyer client.

A house usually stays in the market until sold. However, it is common in our data that a single buyer makes multiple rounds of search attempts discontinuously. In particular, after consecutive searching for a period of time, a buyer may quit searching without making any purchases. However, after staying inactive for months, the same buyer could return to the search market later and start new rounds of house viewings. Directly aggregating the total number of onsite house viewings by each buyer ID assumes that a buyer keeps searching from his or her first visit until the last one. However, for many of the buyer clients, they temporally exit the market after search failures and return to the market later when new houses are supplied.[5]

Empirically, calculating the total number of onsite house viewings by each buyer from the first visit results in an overestimation of deal likelihoods and an upward bias in buyers' onsite house viewings. Instead, we alleviate this concern by treating a buyer who does not purchase any houses and seizes searching for more than one year (i.e., 360 days) as a search failure event. Then the resumed onsite house viewing activities, even from the same buyer, are treated as a new search attempt with the total number of onsite house viewings calculated separately.

At the property level, we know the basic features of each delegated property, including total areas in square meters, number of bedrooms, management fees, the floor number, years after building completion, and whether the house is located nearby a subway station or school. Each delegated house also has a unique ID that allows us to calculate the total number of visitors before the house is sold (*NumVisitor*), and we include it as a control variable. At the transaction level, we know the seller's characteristics such as gender, age, etc., as well as the transaction price.

We merge the above three layers of data with the unique buyer and property identifier and obtain the combined data for the following analyses in this paper, which includes 4,397,652 onsite viewing records from 621,040 unique buyers on 512,701 unique properties, among which 68,569 buyers purchased a house after onsite viewings.

Table 1 Panel A reports the summary statistics of buyers' onsite house viewing activities, buyer characteristics, and buyer budget measures using the buyer level full sample. Among all the buyers in our

[5] In our sample, the median time a house exists in market before sold is 32 days, meaning that under regular market conditions, around half of the houses on market will be replaced in approximated a month.

sample, 11.1% purchased a house after onsite viewings. Before exiting the search market either through purchasing a house or a search failure, the average and median total number of onsite viewings are 6.502 and 4.000, respectively. These summary statistics are comparable in scale to those in an online report published in 2017 from another Chinese real estate agency.[6] 51.4% of potential buyers are males and a small proportion of buyers make full cash payments when purchasing a house and 66.3% of them leverage on mortgages. To better match buyers' demands, the agency collects the minimum and maximum preferences on housing prices, housing sizes, and the number of bedrooms for each buyer client, and the average (median) buyer has a budget range of RMB 0.853 million (RMB 0.500 million) and an area range of 23.2 square meters (20.0 square meters). Regarding the number of bedrooms, although the average buyer has a range of 0.461 rooms, most of the buyers set a clear target, and do not have any flexibilities in this dimension of needs.

Table 1 Panel B reports the summary statistics of buyers' onsite house viewing activities, buyer characteristics, buyer budget measures, and transaction outcomes using the buyer level subsample for successful buyers only. Among the deal subsample, the average and median total number of onsite house viewings are 9.726 and 8.000, respectively, both are larger than the values in the full sample. We also construct a dummy variable *High_TotalVisit* which equals 1 if a buyer has above median number of onsite viewings. 72.3 % of the buyers in the deal subsample have above median number of onsite viewings. The average (median) houses are sold at RMB 42.35 thousand per square meter (RMB 38.68 thousand per square meter) while the average (median) buyers pay around RMB 3.565 million (RMB 3.000 million) for each house. To address the concern that buyers who intend to buy more expensive houses are more cautious and thus more active in onsite viewings, we further construct some relative price measures. On average 32.6% of the buyers pay beyond the maximum budget total price (*Overpay_Maxprice*). We also compare the per square meter price of the purchasing house with the fair price per square meter, which is constructed as the average per square meter price of the houses sold in the past 6 months within the same business zone.[7] The average (median) fair price (*FP_BusinessZone*) is RMB 39,450 per square meter (RMB 36,380 per square meter), and on average, about 67.1% of the buyers pay prices higher than the calculated fair prices. For robustness check, we also focus on whether a buyer makes an in-budget

[6] The average number of onsite viewings of each buyer client ranges from 2.80 to 5.62 across different months. Retrieved from: https://www.sohu.com/a/210349637_148781.

[7] If the total number of transaction records are less than 6 in the past 6 months, we set it as a missing value. If the total number of months with transaction records are less than 6 but the total number of transaction records are more than 6, we calculate the average price using the actual number of months and the actual transaction records.

purchasing decision. We find that 49.2% of the buyers purchase houses within their budget range (*Inbudget_Price*).

The housing characteristics and transaction outcomes are also observable. An average (median) house is 85.4 square meters (76.8 square meters) large with 2.045 bedrooms (2.000 bedrooms), about on the 7th (5th) floor, 15.960 years (14.000 years) after completion, and charged RMB 1.452 per square meter (RMB 1.370 per square meter) management fees. Most houses are nearby subway stations and schools. Before sold on the market, the average and median total number of visitors of each house are 21.759 and 14.000, respectively. 55.3% of the sellers are male, with the average (median) age of 46.838 (44). 51.3% of the buyers are male, and their average (median) age is 35.55 (33). Buyers are more likely to resort to mortgages as compared to making full cash payments, 62.0% of the deal sample use bank loan financing.

[INSERT TABLE 1 ABOUT HERE]

We present a more intuitive comparison of the distribution of our focal variable, *TotalVisit*, between the full and deal sample. In Figure 1, the blue bars show the distribution of number of onsite visits in the full sample while the yellow ones show that of the deal subsample. For the full sample containing all buyer clients, the mode of the total number of onsite viewings is 1, making up about 15% observations among all records. The frequency exhibits a downward trend as the number of onsite viewings goes up. The pattern suggests that some of the buyers stop onsite viewing more houses after a match failure in the first visit. For the deal subsample containing only buyers who purchased a house, the mode is 5, and the distribution exhibits an inverted U shape, but the smaller numbers represent higher percentages of the deal sample.

[INSERT FIGURE 1 ABOUT HERE]

To compare the impacts of onsite house viewing activities on deal likelihoods and transaction outcomes, we first visualize the relationships in Figure 2 after partitioning the full sample into two groups based on the median number of total onsite viewings among all buyer clients. The yellow bars show the average values of buyers with more onsite viewings while the blue ones show the average values of buyers with fewer onsite viewings.

The variable in the first group is *Deal*, which is a dummy variable that equals 1 if a buyer purchases a house after onsite viewings, and 0 otherwise; *TransacPrc* and *Overpay_MaxPrice* are the measures of transaction prices. We see from Figure 2 that buyers with a greater number of onsite house viewings are more likely to be in the *Deal* group, with 16.72% deal likelihoods, almost three times larger than that for those buyers with fewer onsite viewings (5.86%). Also, we find that buyers who are more active in onsite viewings tend to pay higher transaction prices by 0.38 million, and they have 12.48 percentage points higher chance of making payments over their initial budgets.

[INSERT FIGURE 2 ABOUT HERE]

We present formal univariate test results in Table 2 to check whether the above differences are statistically significant. Panel A compares the buyer search and buyer characteristics between the *Deal* group and *Search Failure* (i.e., *Deal=0*) group. We find that buyers in the *Deal* group have a greater number of onsite house viewings, with the average values equal to 9.726 and 6.100 in the *Deal* and *Search Failure* group, respectively. The difference of 3.626 is statistically significant. Similarly, buyers in the *Deal* group are more likely to have above median onsite viewings. Buyers who rely on bank loans have lower deal likelihoods, with 4.9 percentage points fewer buyers using bank loans to pay for the houses in the *Deal* group. Buyers who have vaguer initial budget settings have lower deal likelihoods as well, where all three differences between the *Deal* and *Search Failure* group are negative and statistically significant. Gender does not have a significant influence on search outcomes.

In Panel B, we equally partition the deal subsample by the median number of onsite house viewings and find consistent relationships as Figure 2 shows. Buyers with more onsite house viewings are more likely to pay higher transaction prices and more likely to make over budget payments. Besides, they also purchase houses with 1,480 RMB higher per square price (*PricePSM*), have a greater chance of purchasing houses at prices above the fair price of the business zone. The probability that they make an in-budget payment is lower.

[INSERT TABLE 2 ABOUT HERE]

Overall, the unconditional univariate tests show that of number of onsite house viewings positively predicts both deal likelihood and deal price, supporting the main hypotheses H1a and H2a over the competing hypotheses H1b and H2b. The combination also suggests that stronger demand is the dominating indication of buyers' onsite viewing activities. We present multivariate regressions evidence in the next section.

5 Empirical Results

5.1 Deal Likelihoods

We start by examining the impact of buyers' onsite viewing activities on deal likelihoods. The dependent variable *Deal* is a dummy variable that equals 1 if the buyer purchases a house after onsite viewings and 0 otherwise. The focal variable *Log_TotalVisit* is the logarithm of the total number of onsite viewings for each buyer. Including fixed effects into binary choice models could lead to inconsistent results, so we mainly use OLS regression in Table 3. We also present the results using Probit regression. Coefficients are reported with standard errors clustered by year-month of the first onsite viewing date.

Specifications (1) and (4) only include our focal variable *Log_TotalVisit*, and buyers' features are controlled in specifications (2) and (5). Specifications (3) and (6) further incorporate year-month fixed effects, which is calculated based on the first onsite viewing time of each buyer. Across all specifications in Table 3, the estimated coefficients of the focal variable *Log_TotalVisit* are positive and statistically significant, indicating that a greater number of onsite house viewings is associated with a higher deal likelihood for each buyer client, supporting hypothesis H1a. Quantitively, 1% more onsite house viewings increases the deal likelihood by 6.9 percentage points as estimated from the OLS regressions or increases the purchasing odds by 47.6%=(exp(0.389)-1) as estimated from the Probit regressions.

Results of other control variables also make intuitive sense. For example, buyers who rely on bank credit have lower chances of purchasing a house. This is possibly because these buyers need to pass additional bank screening and the deployment of bank credit takes time, and the sellers have to wait for longer time before receiving the money. Other things being equal, sellers are less willing to sell their houses to mortgage users. For each potential buyer, the range of his or her budgets, housing sizes and the number of bedrooms measure the accuracy of the housing demand. Thus, negative coefficients of these control variables indicate that aimless buyers with vaguer demands on budgets and housing features have fewer chances of purchasing.

[INSERT TABLE 3 ABOUT HERE]

5.2 Transaction Prices

Next, we investigate how buyers' onsite viewing activities affect transaction prices. The absolute measure of transaction prices *Log_TransacPrc* and a relative measure *Overpay_MaxPrice* are used as dependent variables in odd and even columns. Transaction price measures are only observed in the deal subsample, thus we start from subsample OLS regressions in the first two specifications.[8] While the OLS regression results are intuitive and easy to interpret, they could be contaminated by the sample selection bias, as the results are only estimated using the selected sample of buyers who purchase a house. We address this issue using two-step Heckman Selection models and the report the results in the last two specifications. In the first step, we estimate the deal likelihood using the full specification in Table 3. Then we extract the

[8] In unreported regressions, we use Probit model when the dependent variable is *Overpay_MaxPrice*. Results are consistent.

Inverse Mill's Ratio (*IMR*) and further include the estimated *IMR* into the second step regression. The focal variable is still *Log_TotalVisit*, and the control variables include both housing features and buyer characteristics. Year-month fixed effects and business zone fixed effects of the purchased properties are controlled in all specifications but not reported. All coefficients are reported with standard errors clustered by business zones.

In the first two columns, we find that a greater number of onsite house viewings predicts a higher total transaction price and a higher chance that a buyer makes an over budget payment. Given the average transaction price being 3,565,000, the coefficient in the first specification translates into RMB 35,650 (=0.010*3,565,000) additional payment for 1% more onsite house viewing. In specification (2), a 1% increase in the total number of onsite house viewings leads to a 7.6 percentage points increase in over budget payments.

In the last two specifications, the significances of the focal variable *Log_TotalVisit* survived the controlling of *IMR* that addresses the sample selection bias. The coeffects are still highly significant at the 1% confidence level, and the economic impacts are stronger. A 1% more onsite viewings will raise the total transaction price by 2.6% (i.e., RMB 92,690) and increase the over budget payment odds by 16.6 percentage points. Results in Table 4 are in favor of hypothesis H2a as opposed to H2b. Combining the findings of Tables 3 and 4 together, we reveal that onsite house viewings positively predict both the transaction probability and sale price, thereby demonstrating that onsite viewings mainly reflect buyers' demand.

For control variables, positive housing attributes such as a larger area, a greater number of bedrooms, nearby subway stations and schools predict higher transaction prices. On the contrary, older houses are more likely to be sold at lower prices. Besides, buyers with vaguer budget ranges, on average, pay higher prices, and it is mechanical that these buyers are less likely to make over budget payments.

[INSERT TABLE 4 ABOUT HERE]

5.3 Identification

Buyers onsite view houses before making purchasing decisions. The sequential setting safely dismisses the possibility of reversed causality. However, the estimated coefficients could be subject to the endogeneities associated with omitted variables. In particular, our data do not have a full set of information of buyer characteristics as in all other studies. Both the total number of onsite house viewings and transaction outcomes could be affected by omitted buyer characteristics, such as buyers' wealth level, income level, education level, family structure, etc.

Motivated by Hu and Lee (2020) that use Melbourne Cup horse races as an exogenous shock on real estate transactions, we use the Chinese Basketball Association (CBA) playoff seasons as the IV for the focal variable *Log_TotalVisit*.

Basketball is very popular in China and the CBA series is one of the most eye-catching national sport events. The 2018-2019 CBA season receives 1.02 trillion TV and 400 million internet viewings, adding up to 1.42 trillion total viewings. The number for China Soccer Leagues series during the same period is only 770 million. For the city Beijing that we collect our data from, its basketball team is one of the 12 founding clubs when CBA league was first established in 1995. In 2011, it welcomed the former NBA all star player Stephon Marbury, the most famous player that has ever joined a Chinese basketball club, making it one of the most popular clubs nationwide. In the following year, it won the first national championship ever in the city's history. During our sample period, Beijing's basketball team won 2 national titles. The joining of superstar players and the championships make the Chinese Basketball Association (CBA) series one of the most attention grabbing sport events in Beijing.

Similar to the NBA series in the U.S., all participating teams in CBA first play a round robin tournament in the regular season, and all teams play games against each other. The top 8 ranking teams in the regular season qualify for playoff seasons, where they play knockout matches until the final champion is determined. Comparing the regular seasons, playoff matches are more exciting and attract more attention. Therefore, we focus on the playoff games instead of the whole season. The choice of playoff seasons as IV is similar to Agarwal, Duchin, and Sosyura (2012) that use the finals of football, basketball, baseball, and hockey seasons in US.

The instrument variable, *Overlapping1*, equals 1 if a buyer's onsite house viewing period overlaps with the CBA playoff seasons as shown in the first two columns in Appendix 2, and 0 otherwise. The CBA playoff seasons increase the opportunities cost of searching, during which potential buyers spend more time watching basketball games as opposed to onsite house viewings. For buyers with searching valuations that are marginally above searching cost, they may choose to postpone or even cease onsite house viewings during the playoff seasons, leading to a lowered total number of visits. Therefore, the IV negatively affects our focal variable. Besides, CBA playoff seasons are exogeneous thus irrelevant to possibly omitted buyer characteristics, and there is no evidence that sport events are related to purchasing likelihoods and transaction prices in the housing market (Giacoletti, Ramcharan, and Yu, 2021), which satisfies the exclusion restriction of the instrumental variable approach.

In Table 5 Panel A, the odd columns report the first stage regression results, where the instrument variable *Overlapping1* exhibits significantly negative coefficients as expected in all specifications, with F statistics higher than 10. The results prove the relevance between CBA playoff seasons and the number of onsite viewings.

The even columns present the estimated coefficients of the second stage of the 2SLS regressions. Specification (2) investigates the impacts of onsite viewing on deal likelihood, and the coefficient of the instrumented focal variable is 0.040 and statistically significant, which confirms the previous baseline results that more onsite house viewings may improve deal likelihoods. The quantitative impacts are weaker than the baseline results in Table 3. 1% more onsite viewings improve deal probability by 4 percentage points. Specifications (4) and (6) focus on the transaction price, and the dependent variables are *Log_TransacPrc* and *Overpay_MaxPrice*, respectively. We find consistent results that buyers who are more active in onsite house viewings are more likely to pay higher total transaction prices, and also more likely to make over budget payments. Quantitatively after addressing the endogeneities issues, a 1% increase in the total number of onsite house viewings, on average, will increase the total price by 3.8% and the probability of spending above the buyers' budget by 17.8 percentage points, which are stronger than the estimated price influence in Table 4. We also reported Cragg-Donald F statistics in the second stage regressions, and the values are much larger than the Stock and Yogo (2005) thresholds.[9] The statistical tests support the validity of the IV used.

CBA playoff seasons are always in February and March across all years, which gives rise to the concern of seasonality. In other words, instead of reflecting the impact of CBA playoff seasons, it could be capturing other factors that may affect housing transactions, such as Spring Festival, weather (Cortés, Duchin, and Sosyura, 2016), etc. To provide a clearer cut identification, we utilize the overlapping between a buyer's onsite house viewing period and the specific game period of team Beijing (*Overlapping2*). Although the full payoff season covers a similar time period in each year, team Beijing's game period varies. In years that Beijing makes it to the final round, its game period covers the entire playoff season (e.g., 2014 and 2015). However, in years that team Beijing is eliminated in early rounds (e.g., 2013 and 2016) or the year that Beijing does not quality the playoffs (e.g., 2017), team Beijing's game period is much shorter than the playoff season. The last two columns of Appendix 2 summarize the

[9] Other weak instrument variable tests also support the validity of IV used, for example, the CLR tests, AR tests, etc.

game periods of team Beijing in our sample period. In the regressions using *Overlapping2* as IV, we restrict the sample to buyers whose house viewing period overlaps with the CBA playoff seasons (i.e., *Overlapping1*=1), which allows us to exploit the variation in onsite viewings within the playoff seasons.

Table 5 Panel B reports the 2SLS regression results. Similar to Panel A, all of the coefficients of *Overlapping2* in odd columns are still negative and statistically significant, showing that even within the playoff seasons, having team Beijing's match leads to a lowered number of onsite house viewings. The coefficients of instrumented *Log_TotalVisit* are significantly positive in the second stage regressions reported in even columns. These results further document that the CBA playoff games offer a clearer cut identification and prove the robustness of the previous findings that more active onsite house viewings increases both deal likelihoods and deal prices.

[INSERT TABLE 5 ABOUT HERE]

5.4 Purchased Housing Features

In our baseline analysis, we mainly investigate the transaction price proxied by total transaction price and an over budget payment dummy. We provide collaborative evidence in this section, examining the impacts of onsite house viewing on the features of purchased housed. As introduced in the institutional background section, a real estate agent asks a buyer client about the ranges of total housing areas and the number of bedrooms that he or she is interested in. The information is collected before the agent makes any recommendations and before the buyers onsite view any houses for sale, thus the budget ranges on the total housing areas and the number of bedrooms reflect the initial house purchasing plans of each buyer client.

In Table 6 Panel A, the dependent variables are the logarithm of house area (*Log_Area*), and a dummy variable indicating if a buyer purchases a house that is larger than his or her planned maximum room area (*Over_MaxArea*). Panel B focus on the number of bedrooms. The dependent variables are the logarithm of number of bedrooms (*Log_NumBedroom*), and a dummy variable indicating if a buyer purchases a house that has more bedrooms than his or her planned maximum number of bedrooms (*Over_MaxBedroom*). Similar to Table 4, both the OLS models and Heckman Selection models are employed in the first and last two specifications of each panel, and the model settings are identical to those in Table 4.

Across both panels in Table 6, we find consistent results that buyers who are more active in onsite viewings are more likely to buy houses with a larger area and a greater number of bedrooms, and they also purchase for additional size and number of bedrooms beyond their initial plans. Empirically, a 1%

increase in the number of onsite house viewings will raise the total hosing areas by 2.9% and the total number of bedrooms by 1.1%. The probability that the buyers purchase houses that are beyond their budget ranges in terms of total housing areas and the number of bedrooms is 6.3 percentage points and 5.0 percentage points higher, respectively.

[INSERT TABLE 6 ABOUT HERE]

5.5 Other Indications

Motivated by prior research, we proposed four possible indications of buyers' active onsite house viewing activities. The baseline results on the impacts of onsite house viewing on deal likelihood and deal price are in favor of the buyer demand indication. In this section, we carefully discuss the rest three possibilities.

The buyer learning by searching affects transaction outcomes through reducing the information disadvantage of buyers. Hollans, Martin, and Munneke (2013) finds that early buyers face stronger information asymmetry when purchasing newly developed proprieties, due to the uncertainties in the maintenance of development standards. A direct implication of learning by searching is that buyers benefit more from searching when they purchase houses in a newly developed complex. To test this indication, we construct a dummy variable *NewComplex*, which equals 1 if a transaction falls into the 6 months after the first transaction record of each complex. We interact *NewComplex* with our focal variable into the models in Table 4, and report the results in Panel A of Table 7.

Consistent with the baseline results, number of onsite house viewings still positively predicts deal price. The negative coefficients *NewComplex* are significant in most specifications, which is consistent with Hollans, Martin, and Munneke (2013) that early buyers of newly constructed houses request for lower price as compensation of the stronger information asymmetry. More importantly, while the coefficients of the interaction term are negative as predicted by the buyer learning indication, none of them are statistically significant. The results indicate that the impacts of buyer learning are trivial.

Another explanation of buyers' onsite house viewing is buyer competition. In our baseline regressions reported in Tables 3 and 4, we control for year-month fixed effects that absorb the impacts of market conditions that change over time, including buyer competition. More formally, we replace our focal variable with the logarithm of the demeaned number of onsite house viewings (*Log_TotalVisit_Demean*), which is the difference between the raw number of onsite house viewings and the monthly average number of onsite house viewings of buyers who started house searching in the same month.

The demeaned number of onsite house viewings only reflects the cross-sectional variations in house searching that are not affected by market conditions. In Panel B Table 7, we estimate the baseline models in Tables 3 and 4 using the demeaned number of onsite house viewings. Results show that after eliminating the possible influence of different market conditions over time, more onsite house viewings still lead to higher chances of purchasing a house and higher transaction price, suggesting that buyer competition is not likely to be the key indication of buyers' onsite house viewings.

The last possible indication that we discuss is bargain hunting, and we use alternative payment measures to examine if bargain hunting is the dominant force of buyers' more active onsite house viewings. In the baseline models in Table 4, we use the raw transaction price and the over-budget payment dummy as proxies for deal price. In Panel C of Table 7, we use three different measures that are closely related to bargain hunting.

A direct implication of bargain hunting is the lower price per square meter of the purchased house. Thus, the first indicator is the logarithm of the average price per square meter in thousand RMB (*Log_PricePSM*). We also compare if the price per square meter of the purchased house is above the fair price per square meter of the business zone (*Overpay_FPBZ*), where the fair price is calculated as the average per square price of houses sold in the same business zone in the past 6 months.[10] For bargain hunters, they are more likely to purchase below fair price houses. However, we see from specifications (1), (2), (4), and (5) that number of onsite house viewings positively predicts price per square meter and the probability of purchasing above fair price houses. The pattern is in line with our baseline result that more active onsite house viewings increase deal prices and rejects the bargain hunting conjecture.

In specifications (3) and (6), we compare a buyer's payment with his or her budget and examine whether a buyer purchases a house whose price falls into the buyer's budget range (*InBudget_Price*). We find evidence against the bargain hunting explanation that buyers onsite viewing more houses have lower chances of making in-budget payments. The results in Panel C decline bargain hunting as the key driver of buyers' onsite house viewings.

[INSERT TABLE 7 ABOUT HERE]

[10] In unreported regressions, we use different time windows to calculate the fair price of each business zone, and the results are consistent.

5.6 Robustness Checks

In this section, we present the robustness tests using alternative sample and model specifications. We see from Figure 1 that the one-time viewers make up the largest percentage of our buyer sample. It is questionable whether a buyer is serious in house purchasing if he or she directly quit searching after a single onsite viewing, and it gives rise to the concern that our findings could be affected by these unserious buyers. We alleviate this concern using an alternative sample that excludes buyers who only have one onsite viewing record. We estimate the models in Tables 3 and 4 using the subsample and report the results in Table 8. The dependent variable in specifications (1) and (4) is the deal dummy, and specifications (2), (5), (3), and (6) examines the impacts of onsite viewings on transaction prices. Results are consistent with our main findings that onsite house viewings improve the probability that a buyer purchases a house, but it also increases the transaction price.

[INSERT TABLE 8 ABOUT HERE]

We use alternative specifications in Table 9. The agency that we collect our data from divides the city Beijing into 225 business zones based on an internal proprietary algorithm. Alternative ways are to allocate all purchased houses into 14 administrative districts or 5,379 complexes. Although the business zone is a finer measure and has clearer economic meaning, the algorithm that partitions business zones are not disclosed to us. Thus we replace business zone fixed effects with district fixed effects in Panel A. In Panel B we control the more granular complex fixed effects. In both panels, the results are similar to our baseline outcomes.

[INSERT TABLE 9 ABOUT HERE]

6 Conclusion

The housing market is far from being perfectly efficient (Case and Shiller, 1990). Due to the heterogeneities of housing attributes, information frictions, the lack of central housing exchange, among other reasons, the housing market can be characterized as a search market, where both buyers and sellers need extra work to look for each other for matches (Wheaton 1990). The majority of the papers look into seller search activities possibly due to data availability, and the research into the buyer side in the market is relatively scarce. However, buyers are much more actively involved in the search process (Baryła and Zumpano, 1995). Thus, empirical research on buyer search is essential for reaching a better understanding of housing markets.

Motivated by prior research, we first propose that buyers' onsite house viewings could be the reflection of buyer demand, buyer learning, buyer competition, and bargain hunting. Using a proprietary dataset covering 4,397,652 onsite viewing records and 621,040 transaction outcomes from a large real estate agency in China, we present empirical evidence that more onsite house viewing increase both deal likelihood and deal price, a pattern that supports buyer demand over the other three indications. The rest three possible indications are carefully ruled out. Our main results survived the instrument variable regression leveraging on the CBA playoff seasons that increase the marginal cost of onsite house viewing, as well as a series of robustness tests using an alternative sample and alternative specifications.

The findings shed light on housing search literature by providing evidence on the important yet insufficiently studied buyer search activities. It also proposes an individual level direct measure of buyers' demand in the housing market. Besides, this paper shows how such kind of big data generated from the PropTech information system can be used to better predict transaction outcomes which will also be interesting and practical to both households and real estate agency practitioners.

References

Agarwal, S., Duchin, R., & Sosyura, D. (2012). In the mood for a loan: The causal effect of sentiment on credit origination. *Available at SSRN 2141030*.

Albrecht, J., Gautier, P. A., & Vroman, S. (2016). Directed search in the housing market. *Review of Economic Dynamics*, 19, 218-231.

Allen, M. T., Cadena, A., Rutherford, J., & Rutherford, R. C. (2015). Effects of real estate brokers' marketing strategies: Public open houses, broker open houses, MLS virtual tours, and MLS photographs. *Journal of Real Estate Research*, 37(3), 343-369.

Anenberg, E. (2016). Information frictions and housing market dynamics. *International Economic Review*, *57*(4), 1449-1479.

Anenberg, E., & Bayer, P. (2020). Endogenous sources of volatility in housing markets: the joint buyer-seller problem. *International Economic Review,* 61(3), 1195-1228.

Anenberg, E., & Kung, E. (2014). Estimates of the size and source of price declines due to nearby foreclosures. *American Economic Review*, *104*(8), 2527-51.

Anglin, P. M. (1997). Determinants of buyer search in a housing market. *Real Estate Economics*, 25(4), 567-589.

Anglin, P. M., Rutherford, R., & Springer, T. M. (2003). The trade-off between the selling price of residential properties and time-on-the-market: The impact of price setting. *The Journal of Real Estate Finance and Economics*, 26(1), 95-111.

Baryla, E., & Zumpano, L. (1995). Buyer search duration in the residential real estate market: The role of the real estate agent. *Journal of Real Estate Research*, 10(1), 1-13.

Baryla, E., Zumpano, L., & Elder, H. (2000). An investigation of buyer search in the residential real estate market under different market conditions. *Journal of Real Estate Research*, 20(1-2), 75-91.

Berkovec, J. A., & Goodman Jr, J. L. (1996). Turnover as a measure of demand for existing homes. *Real Estate Economics*, 24(4), 421-440.

Bian, X., Contat, J. C., Waller, B. D., & Wentland, S. A. (2021). Why Disclose Less Information? Toward Resolving a Disclosure Puzzle in the Housing Market. *The Journal of Real Estate Finance and Economics*, 1-44.

Boehm, T. P., & McKenzie, J. A. (1982). Inflation, taxes, and the demand for housing. *Real Estate Economics*, 10(1), 25-38.

Case, K. E., & Shiller, R. J. (1990). Forecasting prices and excess returns in the housing market. *Real Estate Economics*, 18(3), 253-273.

Chen, Y. and Rosenthal, R. W. (1996a). On the use of celling-price commitments by monopolists. *The RAND Journal of Economics*, 27(2), 207-220.

Chen, Y. and Rosenthal, R. W. (1996b). Asking prices as commitment devices. *International Economic Review*, 37, 129-155.

Cheng, P., Lin, Z., & Liu, Y. (2021). Competing selling strategies in the housing market. *The Journal of Real Estate Finance and Economics*, *63*(3), 394-413.

Cortés, K., Duchin, R., & Sosyura, D. (2016). Clouded judgment: The role of sentiment in credit origination. *Journal of Financial Economics*, 121(2), 392-413.

Courant, P. N. (1978). Racial prejudice in a search model of the urban housing market. *Journal of Urban Economics*, 5(3), 329-345.

Cubbin, J. (1974). Price, quality, and selling time in the housing market. *Applied Economics*, 6(3), 171-187.

Diaz, A., & Jerez, B. (2013). House prices, sales, and time on the market: A search-theoretic framework. *International Economic Review*, 54(3), 837-872.

Donald, J., Terry, S., & Daniel, W. (1996). Time on the market: The impact of residential brokerage. *Journal of Real Estate Research*, 12(2), 447-458.

Eaton, J., Eslava, M., Jinkins, D., Krizan, C. J., & Tybout, J. R. (2021). *A search and learning model of export dynamics* (No. w29100). National Bureau of Economic Research.

Elder, H. W., Zumpano, L. V., & Baryla, E. A. (1999). Buyer search intensity and the role of the residential real estate broker. *The Journal of Real Estate Finance and Economics*, 18(3), 351-368.

Fan, Y., Hu, M. R., Wan, W. X., & Wang, Z. (2021). A Tale of Two Cities: Mainland Chinese Buyers in Hong Kong Housing Market.

Ford, J. S., Rutherford, R. C., & Yavas, A. (2005). The effects of the internet on marketing residential real estate. *Journal of Housing Economics*, 14(2), 92-108.

Firoozi, F., Hollas, D., Rutherford, R., & Thomson, T. (2006). Property assessments and information asymmetry in residential real estate. *Journal of Real Estate Research*, *28*(3), 275-292.

Gan, L., Wang, P., & Zhang, Q. (2018). Market thickness and the impact of unemployment on housing market outcomes. *Journal of Monetary Economics*, 98, 27-49.

Gargano, A., Giacoletti, M., & Jarnecic, E. (2020). Attention and Search: Micro-evidence on housing market momentum. *Available at SSRN 3519635*.

Garriga, C., & Hedlund, A. (2020). Mortgage debt, consumption, and illiquid housing markets in the great recession. *American Economic Review*, 110(6), 1603-34.

Genesove, D., & Han, L. (2012). Search and matching in the housing market. *Journal of Urban Economics*, 72(1), 31-45.

Genesove, D., & Mayer, C. J. (1997). Equity and time to sale in the real estate market. *The American Economic Review*, 255-269.

Giacoletti, M., Ramcharan, R., & Yu, E. G. (2021). The FOMC cycle and consumer decision making: Evidence from the residential housing market. *Available at SSRN 3364588*.

Glower, M., Haurin, D. R., & Hendershott, P. H. (1998). Selling time and selling price: The influence of seller motivation. *Real Estate Economics*, 26(4), 719-740.

Guren, A. M., & McQuade, T. J. (2020). How do foreclosures exacerbate housing downturns?. *The Review of Economic Studies*, 87(3), 1331-1364.

Han, L., & Strange, W. C. (2015). The microstructure of housing markets: Search, bargaining, and brokerage. *Handbook of Regional and Urban Economics*, 5, 813-886.

Han, L., & Strange, W. C. (2016). What is the role of the asking price for a house?. *Journal of Urban Economics*, 93, 115-130.

Harding, J. P., Knight, J. R., & Sirmans, C. F. (2003). Estimating bargaining effects in hedonic models: Evidence from the housing market. *Real estate economics*, *31*(4), 601-622.

Harmon, O. R., & Potepan, M. J. (1988). Housing adjustment costs: Their impact on mobility and housing demand elasticities. *Real Estate Economics*, 16(4), 459-478.

Hayunga, D. K., & Munneke, H. J. (2021). Examining both sides of the transaction: Bargaining in the housing market. *Real Estate Economics*, *49*(2), 663-691.

Hayunga, D. K., & Pace, R. K. (2019). The impact of TOM on prices in the US housing market. *The Journal of Real Estate Finance and Economics*, *58*(3), 335-365.

He, X., Lin, Z., Liu, Y., & Seiler, M. J. (2020). Search Benefit in Housing Markets: An Inverted U‐Shaped Price and TOM Relation. *Real Estate Economics*, *48*(3), 772-807.

Head, A., Lloyd-Ellis, H., & Sun, H. (2014). Search, liquidity, and the dynamics of house prices and construction. *American Economic Review*, 104(4), 1172-1210.

Hollans, H., Martin, R. W., & Munneke, H. J. (2013). Measuring price behavior in new residential subdivisions. *The Journal of Real Estate Finance and Economics*, *47*(2), 227-242.

Horowitz, J. L. (1992). The role of the list price in housing markets: Theory and an econometric model. *Journal of Applied Econometrics*, 7(2), 115-129.

Hu, M. R., & Lee, A. D. (2020). Outshine to Outbid: Weather-induced sentiment and the housing market. *Management Science*, 66(3), 1440-1472.

Kang, H. and M. Gardner. (1989). Selling price and marketing time in the residential real estate market. *Journal of Real Estate Research,* 4(1), 21–35.

Knight, J. R. (2002). Listing price, time on market, and ultimate selling price: Causes and effects of listing price changes. *Real Estate Economics*, 30(2), 213-237.

Krainer, J. (2001). A theory of liquidity in residential real estate markets. *Journal of Urban Economics*, 49(1), 32-53.

Krainer, J., & LeRoy, S. F. (2002). Equilibrium valuation of illiquid assets. *Economic Theory*, *19*(2), 223-242.

Levitt, S. D., & Syverson, C. (2008). Market distortions when agents are better informed: The value of information in real estate transactions. *The Review of Economics and Statistics*, *90*(4), 599-611.

Marcin, T. C., & Kokus Jr, J. (1975). Economic and demographic factors affecting housing demand into the 1980s and 1990s. *Real Estate Economics*, 3(3), 81-93.

McGreal, S., Adair, A., Brown, L., & Webb, J. (2009). Pricing and time on the market for residential properties in a major UK city. *Journal of Real Estate Research*, 31(2), 209-234.

Merlo, A., & Ortalo-Magne, F. (2004). Bargaining over residential real estate: Evidence from England. *Journal of Urban Economics*, 56(2), 192-216.

Mitchell, A., & Petersen. (2009). Estimating standard errors in finance panel data sets: comparing approaches. *Review of Financial Studies*, 22(1), 435-480.

Ortalo-Magne, F., & Rady, S. (2006). Housing market dynamics: On the contribution of income shocks and credit constraints. *The Review of Economic Studies*, *73*(2), 459-485.

Piazzesi, M., Schneider, M., & Stroebel, J. (2020). Segmented housing search. *American Economic Review*, 110(3), 720-759.

Pope, J. C. (2008). Buyer information and the hedonic: the impact of a seller disclosure on the implicit price for airport noise. *Journal of Urban Economics*, *63*(2), 498-516.

Qiu, L., & Zhao, D. (2018). Information and housing search: theory and evidence from China market. *Applied Economics*, *50*(46), 4954-4967.

Springer, T. M. (1996). Single-family housing transactions: Seller motivations, price, and marketing time. *The Journal of Real Estate Finance and Economics*, 13(3), 237-254.

Steegmans, J., & Hassink, W. (2017). Financial position and house price determination: An empirical study of income and wealth effects. *Journal of Housing Economics*, 36, 8-24.

Stein, J. C. (1995). Prices and trading volume in the housing market: A model with down-payment effects. *The Quarterly Journal of Economics*, 110(2), 379-406.

Stull, W. J. (1978). The landlord's dilemma: Asking rent strategies in a heterogeneous housing market. *Journal of Urban Economics*, *5*(1), 101-115.

Turnbull, G. K., & Dombrow, J. (2006). Spatial competition and shopping externalities: Evidence from the housing market. *The Journal of Real Estate Finance and Economics*, *32*(4), 391-408.

Turnbull, G. K., Dombrow, J., & Sirmans, C. F. (2006). Big house, little house: relative size and value. *Real Estate Economics*, *34*(3), 439-456.

Turnbull, G. K., & Zahirovic-Herbert, V. (2011). Why do vacant houses sell for less: holding costs, bargaining power or stigma?. *Real Estate Economics*, *39*(1), 19-43.

Van Dijk, D. W., & Francke, M. K. (2018). Internet search behavior, liquidity and prices in the housing market. *Real Estate Economics*, 46(2), 368-403.

Wheaton, W. C. (1990). Vacancy, search, and prices in a housing market matching model. *Journal of Political Economy*, 98(6), 1270-1292.

Wilhelmsson, M. (2008). Evidence of buyer bargaining power in the Stockholm residential real estate market. *Journal of Real Estate Research*, 30(4), 475-500.

Williams, J. T. (2018). Housing markets with endogenous search: Theory and implications. *Journal of Urban Economics*, 105, 107-120.

Wong, S. K., Yiu, C. Y., & Chau, K. W. (2012). Liquidity and information asymmetry in the real estate market. *The Journal of Real Estate Finance and Economics*, *45*(1), 49-62.

Yavas, A., & Yang, S. (1995). The strategic role of listing price in marketing real estate: Theory and evidence. *Real Estate Economics*, 23(3), 347-368.

Ziegert, A. L. (1988). The demand for housing additions: An empirical analysis. *Real Estate Economics*, 16(4), 479-496.

Zuehlke, T. W. (1987). Duration dependence in the housing market. *The Review of Economics and Statistics*, 701-709.

Figure 1: Distribution of Total Number of Onsite Viewings

This figure plots the distribution of the total number of onsite house viewings using the full sample containing all buyer clients and the deal subsample containing buyers who purchased a house in Beijing from January 2013 to December 2017, respectively.

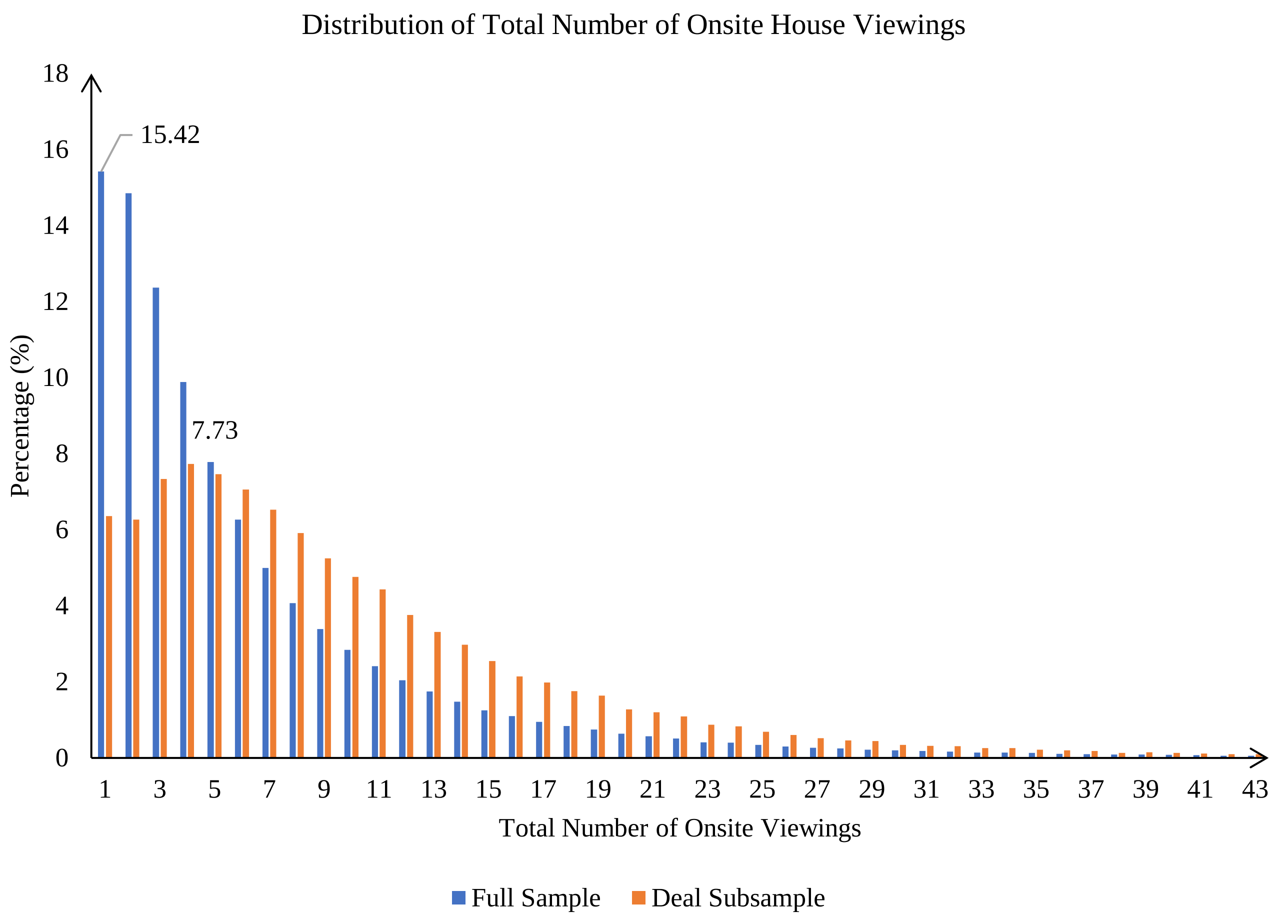

Figure 2: Onsite House Viewings and Transaction Outcomes

This figure shows the relationship between the onsite house viewings and transaction outcomes. We partition all buyers into two groups by the median total number of onsite viewings. *Deal* is a dummy variable that equals 1 if a buyer purchases a house after onsite viewings some delegated properties, and 0 otherwise. *TransacPrc* is the raw total transaction price in million RMB. *Overpay_MaxPrice* indicates if the transaction price is above the buyer's maximum budget. The definition of all variables is presented in Appendix 1.

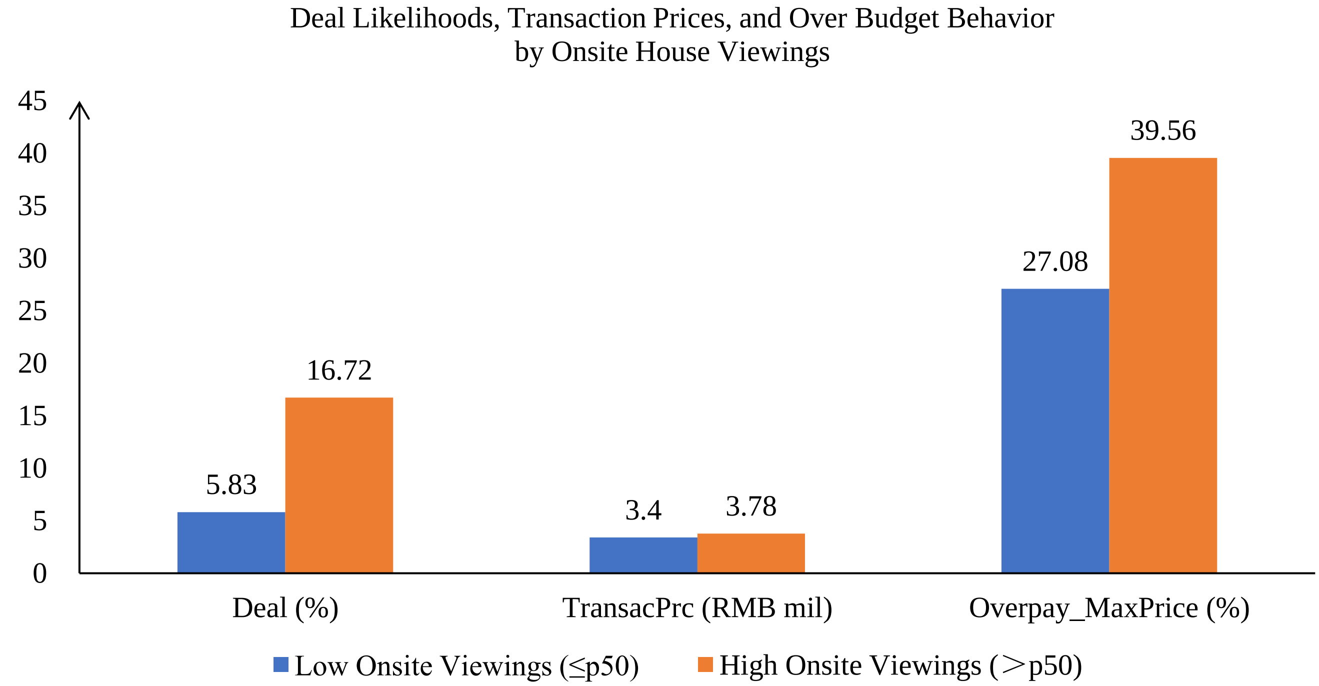

Table 1: Summary Statistics

This table reports the summary statistics of the key variables. Panel A uses the buyer level full sample covering all onsite viewing and transaction records from January 2013 to December 2017 in Beijing. Panel B uses the deal subsample containing only buyers who purchased a house after onsite viewings. The definition of all variables is presented in Appendix 1.

Panel A: Full Sample

Variable	N	mean	sd	p50	min	max
Deal (0-1)	614,198	0.111	0.314	0.000	0.000	1.000
TotalVisit	614,198	6.502	6.610	4.000	1.000	43.000
High_TotalVisit	614,198	0.475	0.499	0.000	0.000	1.000
MaleBuyer	614,083	0.514	0.500	1.000	0.000	1.000
BankLoan	614,198	0.663	0.473	1.000	0.000	1.000
Aimless_Price (mil)	608,022	0.853	0.907	0.500	0.000	6.500
Aimless_Area (100 m^2)	608,642	0.232	0.187	0.200	0.000	1.200
Aimless_NumBedRoom	614,083	0.461	0.566	0.000	0.000	5.000

Panel B: Subsample for Successful Transactions only

Variable	N	mean	Sd	p50	min	max
Buyer Onsite Viewing Records						
TotalVisit	67,994	9.726	7.628	8.000	1.000	43.000
High_TotalVisit	67,994	0.723	0.447	1.000	0.000	1.000
Price and Budget						
TransacPrc (mil)	67,994	3.565	2.224	3.000	0.222	62.000
PricePSM (10k/m^2)	66,181	4.235	1.693	3.868	1.607	10.175
Overpay_Maxprice	67,215	0.326	0.469	0.000	0.000	1.000
InBudget_Price	66,793	0.492	0.500	0.000	0.000	1.000
FP_BusinessZone (10k/m^2)	63,549	3.945	1.419	3.638	1.714	9.610
Overpay_FPBZ	63,549	0.671	0.470	1.000	0.000	1.000
ListingPrc (mil)	67,535	3.636	2.290	3.050	0.250	69.800
Transaction Characteristics						
Area (100 m^2)	67,535	0.854	0.384	0.768	0.070	6.400
NumBedroom	67,535	2.045	0.776	2.000	0.000	9.000
MGMTFee	66,861	1.452	1.038	1.370	0.000	5.800
Floor	66,536	7.279	5.779	5.000	-3.000	39.000
HouseAge (y)	62,440	15.960	8.505	14.000	0.000	66.000
Subway	67,535	0.701	0.458	1.000	0.000	1.000
School	67,535	0.766	0.423	1.000	0.000	1.000
NumVisitor	66,875	21.759	23.114	14.000	1.000	137.000
SellerAge	67,175	46.838	14.077	44.000	18.000	115.000
MaleSeller	67,415	0.553	0.497	1.000	0.000	1.000
BuyerAge	67,802	35.551	9.499	33.000	18.000	116.000
MaleBuyer	67,994	0.513	0.500	1.000	0.000	1.000
BankLoan	67,994	0.620	0.485	1.000	0.000	1.000
Aimless_Price (mil)	67,749	0.683	0.738	0.500	0.000	6.500
Aimless_Area (100 m^2)	67,726	0.204	0.160	0.200	0.000	1.200
Aimless_NumBedroom	67,994	0.403	0.532	0.000	0.000	5.000

Table 2: Univariate Tests

Panel A compares the search activities and buyer characteristics between deal and search failure subsample, where the full sample is partitioned by whether a buyer purchases a house after onsite viewings or not. Panel B compares the transaction prices and house characteristics between buyers of above and below median onsite house viewings. Only the deal subsample is used in Panel B. The number of observations, the sample means, standard deviations, the differences in means, and the t-test significance levels are presented here. ***, **, and * denote significance at the 1%, 5%, and 10% level, respectively. The definition of all variables is presented in Appendix 1.

Panel A: by House Purchasing (Full Sample)

	Deal=1			Deal=0			Diff.
	N	Mean	Sd	N	Mean	Sd	in mean
TotalVisit	67,994	9.726	7.628	546,204	6.100	6.358	3.626***
High_TotalVisit	67,994	0.723	0.723	546,204	0.444	0.444	0.279***
MaleBuyer	67,994	0.513	0.500	546,089	0.515	0.500	-0.002
BankLoan	67,994	0.620	0.485	546,204	0.669	0.471	-0.049***
Aimless_Price	67,749	0.683	0.738	540,273	0.875	0.924	-0.191***
Aimless_Area	67,726	0.204	0.160	540,916	0.235	0.190	-0.031***
Aimless_NumBedroom	67,994	0.403	0.532	546,089	0.469	0.570	-0.066***

Panel B: by Total Number of Onsite Viewings (Subsample for Successful Transaction only)

	High_TotalVisit=1			High_TotalVisit=0			Diff.
	N	Mean	Sd	N	Mean	Sd	in mean
TransacPrc (mil)	30,837	3.769	2.221	37,157	3.396	2.212	0.373***
Overpay_Maxprice	30,577	0.393	0.488	36,638	0.271	0.444	0.122***
InBudget_Price	30,412	0.421	0.494	36,381	0.552	0.497	-0.131***
PricePSM (10k/m^2)	30,112	4.316	1.708	36,069	4.168	1.677	0.148***
FP_BusinessZone (10k/m^2)	28,910	4.004	1.433	34,639	3.895	1.406	0.109***
Overpay_FPBZ	28,910	0.684	0.465	34,639	0.659	0.474	0.024***

Table 3: Deal Likelihoods and Onsite House Viewings

This table presents the estimation results on the relationship between the total number of onsite house viewings and the house deal likelihoods. The OLS models and Probit models are used. The full sample covering all buyers in Beijing from January 2013 to December 2017 is used. The dependent variable is *Deal*, a dummy equal to 1 if the buyer purchases a house after onsite viewings and 0 otherwise. The focal variable, *Log_TotalVisit*, is logarithmic of the total number of onsite house viewings of each buyer. Buyer characteristics and year-month fixed effects of the first onsite viewing activity for each buyer are added in specifications (2), (5) and (3), (6) gradually. Coefficients are reported in all specifications with standard errors clustered by year-month in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% level, respectively. The definition of all variables is presented in Appendix 1.

Dependent Variable: Deal	(1)	(2)	(3)	(4)	(5)	(6)
		OLS			Probit	
Log_TotalVisit	0.064***	0.067***	0.069***	0.355***	0.372***	0.389***
	(0.002)	(0.001)	(0.002)	(0.008)	(0.008)	(0.007)
MaleBuyer		-0.000	-0.001		-0.001	-0.004
		(0.001)	(0.001)		(0.004)	(0.004)
BankLoan		-0.021***	-0.022***		-0.119***	-0.123***
		(0.001)	(0.001)		(0.007)	(0.006)
Log_Aimless_Price		-0.003***	-0.003***		-0.013***	-0.013***
		(0.000)	(0.000)		(0.001)	(0.001)
Aimless_Area		-0.016***	-0.016***		-0.091***	-0.091***
		(0.001)	(0.001)		(0.004)	(0.004)
Aimless_NumBedroom		-0.024***	-0.024***		-0.137***	-0.139***
		(0.002)	(0.002)		(0.010)	(0.010)
Constant	0.017***	0.116***	0.060***	-1.803***	-1.274***	-4.462***
	(0.003)	(0.003)	(0.002)	(0.025)	(0.020)	(0.228)
Year-Month FE	NO	NO	YES	NO	NO	YES
Clustered SE	Year-Month	Year-Month	Year-Month	Year-Month	Year-Month	Year-Month
Number of Obs.	614,198	604,420	604,420	614,198	604,420	604,331
Adjusted R-square	0.037	0.042	0.044			
Pseudo R-square				0.053	0.062	0.065

Table 4: Transaction Prices, Over Budget Payments, and Onsite House Viewings

This table presents the impacts of buyers' onsite house viewings on transaction prices and over budget payments. Columns (1)-(2) and Columns (3)-(4) report the OLS and Heckman Selection regression results. The dependent variable in odd columns is the logarithm of the total transaction price in million RMB (*Log_TransacPrc*) and is a dummy variable that indicates if the total transaction price is above the buyer's maximum budgets (*Overpay_MaxPrice*) in even columns. The focal variable *Log_TotalVisit*, is the logarithm of the total number of onsite house viewings of each buyer. *IMR* is constructed from estimating the deal likelihood using the full specification in Table 3. Year-month fixed effects and business zone fixed effects are added in all specifications. Coefficients are reported in all specifications with standard errors clustered by business zone in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% level, respectively. The definition of all variables is presented in Appendix 1.

	(1)	(2)	(3)	(4)
Dependent Variable	Log_TransacPrc	Overpay_MaxPrice	Log_TransacPrc	Overpay_MaxPrice
	OLS		Heckman Twostep	
Log_TotalVisit	0.010***	0.076***	0.026***	0.166***
	(0.001)	(0.002)	(0.004)	(0.012)
Log_Area	0.801***	0.251***	0.801***	0.250***
	(0.019)	(0.013)	(0.019)	(0.013)
Log_NumBedroom	0.158***	0.043***	0.157***	0.042***
	(0.019)	(0.015)	(0.019)	(0.014)
Log_MgmtFee	0.085***	0.028***	0.085***	0.028***
	(0.011)	(0.010)	(0.011)	(0.010)
Log_Floor	-0.032***	-0.014***	-0.032***	-0.014***
	(0.004)	(0.004)	(0.004)	(0.004)
Log_HouseAge	-0.090***	0.001	-0.090***	0.001
	(0.017)	(0.010)	(0.017)	(0.010)
Subway	0.047***	0.018***	0.047***	0.018***
	(0.012)	(0.007)	(0.012)	(0.007)
School	0.055***	0.015***	0.055***	0.015***
	(0.009)	(0.005)	(0.009)	(0.005)
Log_NumVisitor	-0.002	-0.008***	-0.002	-0.008***
	(0.001)	(0.002)	(0.001)	(0.002)
Log_SellerAge	-0.013**	-0.000	-0.013**	-0.001
	(0.006)	(0.008)	(0.006)	(0.008)
MaleSeller	0.000	0.003	0.000	0.003
	(0.002)	(0.004)	(0.002)	(0.004)
Log_BuyerAge	0.014***	-0.040***	0.014***	-0.042***
	(0.005)	(0.010)	(0.005)	(0.010)
MaleBuyer	-0.005***	-0.008**	-0.005***	-0.008**
	(0.001)	(0.004)	(0.001)	(0.004)
BankLoan	0.033***	-0.016***	0.028***	-0.045***
	(0.003)	(0.004)	(0.003)	(0.005)
Log_Aimless_Price	0.004***	-0.025***	0.004***	-0.028***
	(0.000)	(0.001)	(0.000)	(0.001)
Log_Aimless_Area	0.005***	-0.061***	0.001	-0.083***
	(0.002)	(0.003)	(0.002)	(0.004)
Log_Aimless_NumBedroom	0.008**	-0.042***	0.002	-0.075***
	(0.003)	(0.006)	(0.003)	(0.007)
IMR			0.051***	0.295***
			(0.014)	(0.040)
Constant	1.001***	0.295***	0.910***	-0.232**
	(0.049)	(0.057)	(0.056)	(0.093)
Year-Month FE	YES	YES	YES	YES
BusinessZone FE	YES	YES	YES	YES
Clustered SE	BusinessZone	BusinessZone	BusinessZone	BusinessZone
Number of Obs.	59,523	59,024	604,331	604,331
Number of Selected			59,523	59,024
Adjusted R-square	0.897	0.101	0.897	0.102

Table 5: Instrument Variable Regressions

This table presents the first stage and second stage regression results of the 2SLS models. In Panel A, the instrument variable is *Overlapping1*, a dummy variable that equals 1 if a buyer's onsite house viewing period overlaps with the CBA playoff seasons, and 0 otherwise. The full sample is used. In Panel B, the instrument variable is *Overlapping2*, a dummy variable that equals 1 if a buyer's onsite house viewing period overlaps with the Beijing basketball team's playoff game periods, and 0 otherwise. Panel B only includes the subsample of buyers whose housing viewing periods fall into the CBA playoff seasons. The dependent variables and other controls are the same as in Tables 3 and 4. Both Cragg-Donald F statistics and Stock and Yogo (2005) threshold values are presented. Coefficients are reported in all specifications with standard errors clustered by business zone in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% level, respectively. The definition of all variables is presented in Appendix 1.

Panel A: Playoff Seasons as IV (Full Sample)

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	Log_ TotalVisit	Deal	Log_ TotalVisit	Log_ TransacPrc	Log_ TotalVisit	Overpay_ MaxPrice
Overlapping1	-0.785***		-0.596***		-0.596***	
	(0.031)		(0.018)		(0.018)	
Log_TotalVisit		0.040***		0.038**		0.178***
		(0.002)		(0.007)		(0.014)
Other Controls	YES	YES	YES	YES	YES	YES
Year-Month FE	YES	YES	YES	YES	YES	YES
BusinessZone FE	NO	YES	YES	YES	YES	YES
Clustered SE	BusinessZone	BusinessZone	BusinessZone	BusinessZone	BusinessZone	BusinessZone
Number of Obs.	604,420	604,420	59,523	59,523	59,024	59,024
First Stage F Stats	638.92		101.58		1,129.27	
Adjusted R-square	0.037		0.895		0.069	
C-D Wald F Statistic		18,367.010		1,449.160		1,435.957
Stock Yolo 10%		16.38		16.38		16.38

Panel B: Beijing's Playoff Game Periods as IV (Playoff Season Subsample)

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	Log_ TotalVisit	Deal	Log_ TotalVisit	Log_ TransacPrc	Log_ TotalVisit	Overpay_ MaxPrice
Overlapping2	-0.359***		-0.187***		-0.184***	
	(0.001)		(0.034)		(0.035)	
Log_TotalVisit		0.201***		0.073*		0.153*
		(0.001)		(0.039)		(0.091)
Other Controls	YES	YES	YES	YES	YES	YES
Year-Month FE	YES	YES	YES	YES	YES	YES
BusinessZone FE	NO	NO	YES	YES	YES	YES
Clustered SE	BusinessZone	BusinessZone	BusinessZone	BusinessZone	BusinessZone	BusinessZone
Number of Obs.	96,379	96,379	10,002	10,002	9,930	9,930
First Stage F Stats	1.1e+05		30.15		28.41	
Adjusted R-square	-0.008		0.876		0.062	
C-D Wald F Statistic		1,919.500		33.404		32.336
Stock Yolo 10%		16.38		16.38		16.38

Table 6: Housing Features, Over Budget Attributes, and Onsite House Viewings

This table presents the impacts of buyers' onsite house viewings on purchased housing features in terms of the total areas in Panel A and the number of bedrooms in Panel B, respectively. Within each panel, columns (1)-(2) and columns (3)-(4) report the OLS and Heckman Selection regression results. In Panel A, the dependent variables are the logarithm of the total area in square meters (*Log_Area*), and a dummy variable indicates if the area of the purchased house is above the maximum budget area (*Over_MaxArea*). In Panel B, the dependent variables are the logarithm of the number of bedrooms of the purchased house (*Log_NumBedroom*), and a dummy variable indicates if the number of bedrooms of the purchased house is above the maximum budget number of bedrooms (*Over_MaxBedroom*). The focal variable *Log_TotalVisit*, is the logarithm of the total number of onsite house viewings of each buyer. *IMR* is constructed from estimating the deal likelihood using the full specification in Table 3. Year-month fixed effects and business zone fixed effects are added in all specifications. Coefficients are reported in all specifications with standard errors clustered by business zone in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% level, respectively. The definition of all variables is presented in Appendix 1.

Panel A: Housing Areas and Onsite House Viewings

	(1)	(2)	(3)	(4)
Dependent Variable	Log_Area	Over_MaxArea	Log_Area	Over_MaxArea
	OLS		Heckman Twostep	
Log_TotalVisit	0.008***	0.039***	0.029***	0.063***
	(0.001)	(0.002)	(0.005)	(0.011)
IMR			0.068***	0.079**
			(0.017)	(0.035)
Other Controls	YES	YES	YES	YES
Year-Month FE	YES	YES	YES	YES
BusinessZone FE	YES	YES	YES	YES
Clustered SE	BusinessZone	BusinessZone	BusinessZone	BusinessZone
Number of Obs.	59,522	59,026	604,331	604,331
Number of Selected			59,522	59,026
Adjusted R-square	0.727	0.187	0.727	0.191

Panel B: Number of Bedrooms and Onsite House Viewings

	(1)	(2)	(3)	(4)
Dependent Variable	Log_NumBedroom	Over_MaxBedroom	Log_NumBedroom	Over_MaxBedroom
	OLS		Heckman Twostep	
Log_TotalVisit	0.005***	0.021***	0.011***	0.050***
	(0.001)	(0.001)	(0.004)	(0.007)
IMR			0.021*	0.096***
			(0.012)	(0.022)
Other Controls	YES	YES	YES	YES
Year-Month FE	YES	YES	YES	YES
BusinessZone FE	YES	YES	YES	YES
Clustered SE	BusinessZone	BusinessZone	BusinessZone	BusinessZone
Number of Obs.	59,522	59,523	604,331	604,331
Number of Selected			59,522	59,523
Adjusted R-square	0.658	0.155	0.656	0.155

Table 7: Other Indications

This table examines the other possible indications of buyers' onsite house viewings. Panel A tests the buyer learning indication, by examining the heterogeneities of the impacts of buyers' onsite house viewings on transaction prices between newly built complex and old complex. *NewComplex* is a dummy variable equals 1 if the purchase time falls into the 6 months after the first transaction of each complex. Panel B tests buyer competition indication. We replace the focal variable with the logarithm of the demeaned number of onsite house viewings of each buyer, *Log_TotalVisit_Demean,* which is the difference between the raw number of onsite house viewings and the monthly average number of onsite house viewings of buyers who started house searching in the same month. Panel C tests bargain hunting indication, which uses alternative measures of the deal price. The dependent variables are the logarithm of the average per square price in thousand RMB (*Log_PricePSM*), and a dummy variable indicates if the average transaction price is above the fair price per square meter within the same business zone (*Overpay_FPBZ*), and a dummy variable indicates if the total transaction price is within a buyer's budget range (*InBudget_Price*), respectively. The last two specifications in Panel A and Panel B, and the last three specifications in Panel C use two-step Heckman regression, where the *IMR* is constructed from estimating the deal likelihood using the full specification in Table 3. All other settings are the same as in Tables 3 and 4. Coefficients are reported in all specifications with standard errors clustered by year-month for specifications (1) and (4), and by BusinessZone for other specifications in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% level, respectively. The definition of all variables is presented in Appendix 1.

Panel A: Buyer Learning and Local Buyer

	(1)	(2)	(3)	(4)
Dependent Variable	Log_TransacPrc	Overpay_MaxPrice	Log_TransacPrc	Overpay_MaxPrice
	OLS		Heckman Twostep	
NewComplex*Log_TotalVisit	-0.002	-0.002	-0.002	-0.006
	(0.002)	(0.004)	(0.002)	(0.004)
Log_TotalVisit	0.010***	0.076***	0.028***	0.172***
	(0.001)	(0.003)	(0.004)	(0.013)
NewComplex	-0.018***	-0.018*	-0.017**	-0.012
	(0.007)	(0.010)	(0.007)	(0.010)
IMR			0.056***	0.306***
			(0.135)	(0.040)
Constant	1.030***	0.321***	0.929***	-0.230**
	(0.049)	(0.058)	(0.056)	(0.093)
Other Controls	YES	YES	YES	YES
Year-Month FE	YES	YES	YES	YES
BusinessZone FE	YES	YES	YES	YES
Clustered SE	BusinessZone	BusinessZone	BusinessZone	BusinessZone
Number of Obs.	60,089	59,586	604,331	604,331
Number of Selected			60,089	59,586
Adjusted R-square	0.898	0.105	0.898	0.106

Panel B: Buyer Competition and Demeaned Number of Onsite Viewings

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	Deal	Log_ TransacPrc	Overpay_ MaPrice	Deal	Log_ TransacPrc	Overpay_ MaxPrice
		OLS		Probit	Heckman Twostep	
Log_TotalVisit_Demean	0.069***	0.010***	0.075***	0.389***	0.008***	0.075***
	(0.002)	(0.001)	(0.003)	(0.007)	(0.001)	(0.003)
IMR					0.024**	0.016
					(0.009)	(0.033)
Other Controls	YES	YES	YES	YES	YES	YES
Year-Month FE	NO	YES	YES	NO	YES	YES
BusinessZone FE	NO	YES	YES	NO	YES	YES
Clustered SE	Year-Month	BusinessZone	BusinessZone	Year-Month	BusinessZone	BusinessZone
Number of Obs.	604,420	59,523	59,024	604,331	604,331	604,331
Number of Selected					59,523	59,024
Adjusted R-square	0.044	0.898	0.101		0.897	0.106
Pseudo R-square				0.065		

Panel C: Bargain Hunting and Alternative Payment Measures

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	Log_ PricePSM	Overpay_ FPBZ	Inbudget_ Price	Log_ PricePSM	Overpay_ FPBZ	Inbudget_ Price
		OLS			Heckman Twostep	
Log_TotalVisit	0.009***	0.020***	-0.099***	0.024***	0.051***	-0.192***
	(0.001)	(0.002)	(0.002)	(0.004)	(0.011)	(0.012)
IMR				0.051***	0.101***	-0.302***
				(0.013)	(0.037)	(0.038)
Other Controls	YES	YES	YES	YES	YES	YES
Year-Month FE	YES	YES	YES	YES	YES	YES
BusinessZone FE	YES	YES	YES	YES	YES	YES
Clustered SE	BusinessZone	BusinessZone	BusinessZone	BusinessZone	BusinessZone	BusinessZone
Number of Obs.	58,434	56,035	58,717	604,331	604,331	604,331
Number of Selected				58,434	56,035	58,717
Adjusted R-square	0.848	0.161	0.097	0.848	0.161	0.098

Table 8: Alternative Sample: Serious Buyers Only

This table presents the robustness checks for the impacts of buyers' onsite house viewings on deal likelihoods and transaction prices using the subsample for serious buyers, which excludes the buyers with only one onsite viewing record. The focal variable *Log_TotalVisit*, is the logarithm of the total number of onsite house viewings of each buyer. *IMR* is constructed from estimating the deal likelihood using the full specification in Table 3. All other settings are the same as in Tables 3 and 4. Coefficients are reported in all specifications with standard errors clustered by year-month for specifications (1) and (4), and by BusinessZone for other specifications in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% level, respectively. The definition of all variables is presented in Appendix 1.

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	Deal	Log_ TransacPrc	Overpay_ MaPrice	Deal	Log_ TransacPrc	Overpay_ MaxPrice
		OLS		Probit	Heckman Twostep	
Log_TotalVisit	0.063***	0.010***	0.076***	0.283***	0.020***	0.128***
	(0.002)	(0.001)	(0.002)	(0.011)	(0.004)	(0.010)
IMR					0.047***	0.238***
					(0.015)	(0.043)
Other Controls	YES	YES	YES	YES	YES	YES
Year-Month FE	NO	YES	YES	NO	YES	YES
BusinessZone FE	NO	YES	YES	NO	YES	YES
Clustered SE	Year-Month	BusinessZone	BusinessZone	Year-Month	BusinessZone	BusinessZone
Number of Obs.	515,670	59,523	59,024	515,613	515,613	515,613
Number of Selected					59,523	59,024
Adjusted R-square	0.025	0.897	0.101		0.897	0.102
Pseudo R-square				0.032		

Table 9: Alternative Specifications: Administrative District Fixed Effects and Complex Fixed Effects

This table presents the impacts of buyers' onsite house viewings on transaction prices using alternative location fixed effects. The focal variable *Log_TotalVisit,* is the logarithm of the total number of onsite house viewings of each buyer. *IMR* is constructed from estimating the deal likelihood using the full specification in Table 3. Year-month fixed effects and administrative district fixed effects are added in all specifications in Panel A, while year-month fixed effects and complex fixed effects are added in all specifications in Panel B. All other settings are the same as in Tables 4. Coefficients are reported in specifications with standard errors clustered by the administrative district for other specifications in Panel A and the complex for other specifications in Panel B in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% level, respectively. The definition of all variables is presented in Appendix 1.

Panel A: Administrative District Fixed Effects

	(1)	(2)	(3)	(4)
Dependent Variable	Log_TransacPrc	Overpay_MaxPrice	Log_TransacPrc	Overpay_MaxPrice
	OLS		Heckman Twostep	
Log_TotalVisit	0.010**	0.076***	0.029***	0.169***
	(0.004)	(0.003)	(0.006)	(0.006)
IMR			0.063**	0.302***
			(0.026)	(0.021)
Other Controls	YES	YES	YES	YES
Year-Month FE	YES	YES	YES	YES
District FE	YES	YES	YES	YES
Clustered SE	District	District	District	District
Number of Obs.	59,523	59,024	604,331	604,331
Number of Selected			59,523	59,024
Adjusted R-square	0.821	0.098	0.821	0.099

Panel B: Complex Fixed Effects

	(1)	(2)	(3)	(4)
Dependent Variable	Log_TransacPrc	Overpay_MaxPrice	Log_TransacPrc	Overpay_MaxPrice
	OLS		Heckman Twostep	
Log_TotalVisit	0.007***	0.073***	0.010***	0.161***
	(0.001)	(0.002)	(0.003)	(0.012)
IMR			0.010	0.286***
			(0.009)	(0.040)
Other Controls	YES	YES	YES	YES
Year-Month FE	YES	YES	YES	YES
Complex FE	YES	YES	YES	YES
Clustered SE	Complex	Complex	Complex	Complex
Number of Obs.	58,803	58,311	604,331	604,331
Number of Selected			58,803	58,311
Adjusted R-square	0.956	0.113	0.687	0.184

Appendix 1: Variable Definition

Variable	Definition
Transaction Characteristics	
Deal (0-1)	A dummy variable that equals 1 if the buyer purchases a house after onsite viewings, and 0 otherwise.
TransacPrc (mil)	The total transaction price of the purchased house in million RMB.
LisringPrc (mil)	The listing price of the purchased house in million RMB.
PricePSM (10k/m^2)	The average per square meter price of the purchased house in ten thousand RMB.
InBudget_Price	A dummy variable that equals 1 if the total transaction price is within the minimum and the maximum values of a buyer's budget, and 0 otherwise.
Overpay_MaxPrice	A dummy variable that equals 1 if the total transaction price is above a buyer's maximum budget, and 0 otherwise.
Overpay_FPBZ	A dummy variable that equals 1 if the average per square price of the purchased house is above the average per square price of the houses in the same business zone over the last six months, and 0 otherwise. Missing if the total number of deals over the last six months within the same business zone is less than 6.
NumVisitor	The total number of visitors of each delegated house.
NumBedroom	The total number of bedrooms of each house.
MgmtFee	The per square meter management fee of each house in RMB.
Area (100 m^2)	The total area of each house in 100 square meters.
Floor	The floor number of the housing unit.
Subway	A dummy variable that equals 1 if the house is nearby a subway station, and 0 otherwise.
School	A dummy variable that equals 1 if the house is nearby a school, and 0 otherwise.
HouseAge (y)	The number of years after the house is built.
SellerAge	The age of each seller.
MaleSeller	A dummy variable that equals 1 if the seller is a male, and 0 otherwise.
BuyerAge	The age of each buyer.
BusinessZone	The business zone where each house locates.
District	The administrative district where each house locates.
Complex	The complex where each house locates.
Buyer Characteristics	
MaleBuyer	A dummy variable that equals 1 if the buyer is a male, and 0 otherwise.
BankLoan	A dummy variable that equals 1 if the buyer uses a bank loan to finance house purchasing.
Aimless_NumBedroom	The difference between the maximum and the minimum total number of bedrooms that a buyer intends to buy.
Aimless_Price (mil)	The difference between the maximum and the minimum total transaction price in million RMB that a buyer intends to pay.
Aimless_Area (100 m^2)	The difference in 100 square meters between the maximum and the minimum house area in square meters that a buyer intends to buy.
TotalVisit	The total number of onsite house viewings of each buyer.
TotalVisit_Demean	The total number of onsite house viewings of each buyer minus the average number of onsite house viewing for buyer started visit houses in the same month.
High_TotalVisit	A dummy variable that equals 1 if a buyer has above median number of onsite house viewings, and 0 otherwise.
Instrument Variables	
Overlapping1	A dummy variable that equals 1 if a buyer's onsite house viewing period overlaps with the whole CBA playoff seasons, and 0 otherwise.
Overlapping2	A dummy variable that equals 1 if a buyer's onsite house viewing period overlaps with team Beijing's playoff games period, and 0 otherwise.

Appendix 2: CBA Playoff Seasons and Beijing Team Game Periods

This table presents the first CBA playoff game date, the last CBA playoffs date, the first Beijing playoff game date, and the last Beijing playoff game date from the year 2013 to 2017, respectively.

Year	First Playoff Game	Final Playoff Game	First Beijing Playoff Game	Last Beijing Playoff Game
2013	2013/2/27	2013/3/29	2013/2/27	2013/3/15
2014	2014/2/18	2014/3/30	2014/2/19	2014/3/30
2015	2015/2/6	2015/3/22	2015/2/6	2015/3/22
2016	2016/2/15	2016/3/20	2016/2/15	2016/2/21
2017	2017/2/24	2017/4/7	Did not qualify playoffs	Did not qualify playoffs