Real Estate Agents' Influence on Housing Search*

Seung-Hyun Hong
Department of Economics
University of Illinois, Urbana-Champaign
hyunhong@illinois.edu

July 14, 2021

Abstract

This paper investigates different mechanisms through which real estate agents influence their clients' housing searches. Using residential listing data, I find that sales prices of agent-owned listings are higher than those of client-owned listings, but this difference disappears after controlling for listing prices determined before listing. Moreover, agent-owned listings have higher listing prices than client-owned listings. This suggests a potentially important role of an agent's influence in the listing stage, during which sellers prepare for listing their houses and agents may influence their client-sellers to set lower initial listing prices. To quantify the impact of this mechanism for agents' influence, I develop a new structural model for the home seller's problem particularly in the listing stage, while accounting for other mechanisms. The estimated model shows that agents' influence in the listing stage explains about 60% of the difference in sales price premiums between agent-owned and client-owned listings. Therefore, agents' influence in the listing stage is consequential, because such influence leads to lower listing prices in the listing stage that further affect housing search outcomes even after listing.

Keywords: conflict of interest, agents' influence, housing market, listing price, structural estimation, truncated model

JEL classification: C34, C51, D83, L14, L85, R30

*I am grateful to various seminar participants at Georgetown, Michigan State University, Barcelona GSE Summer Forum, and Stanford Institute for Theoretical Economics for their helpful comments. This paper is based on work supported by (while serving at) the National Science Foundation. Any opinion, findings, and conclusions or recommendations expressed in this paper are those of the author and do not necessarily reflect the views of the National Science Foundation.

1 Introduction

Most consumers rely on real estate agents to buy and sell their homes, since agents are better informed about local housing markets and transactions than consumers who engage in real estate transactions infrequently. In principle, real estate agents have fiduciary duties to their clients in most North American housing markets. However, agents may also seek their own interest at the expense of their clients by influencing housing searches through different channels. In particular, agents may influence sellers to set lower listing prices even in the *listing* stage – during which sellers prepare to list their houses on a Multiple Listing Service (MLS). Agents may also shirk strategically by putting only minimal effort to bring a barely acceptable offer in the *selling* stage – during which houses are listed on a MLS.

This paper investigates these mechanisms and further quantifies the extent of their impacts on housing search. Several studies have already provided important and careful empirical evidence suggesting that sales outcomes might be potentially influenced by real estate agents (e.g., Rutherford, et al., 2005; Levitt and Syverson, 2008a,b; Hendel, et al., 2009; Bernheim and Meer, 2013). However, the aforementioned two channels have not been studied extensively in the literature. Moreover, it is necessary to quantify the extent of specific channels, because separate mechanisms can have different implications on how to discipline real estate agents and minimize their influence.

For example, Levitt and Syverson (2008a) show that agents sell their clients' houses more quickly at lower prices than their own houses. This well-known result tends to suggest a channel in which agents influence sellers to accept a current low offer without waiting longer to receive higher offers. A simple solution for sellers to mitigate this channel is to wait longer than their agents recommend. However, if the initial listing price affects the level of offers more than listing prices revised during the selling stage, agents' influence in the listing stage may result in low offers only, even if sellers wait longer. To avoid such influence, sellers might as well insist on higher listing prices than their agents recommend in the listing stage. Nonetheless, this solution may not work either, if agents make only minimal effort in the selling stage. Moreover, these solutions may even be unnecessary if prices are mainly determined by sellers' motivation to sell, rather than agents' influence.

This paper aims to distinguish between these potential mechanisms and quantify the extent of their impacts. To this end, I consider the home seller's problem in the listing stage separately from her decisions in the selling stage. To examine these decisions, I use the MLS data from downtown areas in a large North American metropolitan statistical area (MSA).[1] The data includes detailed listing information on agent-owned properties, house characteristics, and sales outcomes, including whether listings were sold or withdrawn. In addition, I use the CoreLogic data with all transactions in the MSA. The CoreLogic data allows me to recover each listing's "intrinsic value" based on previous sales prices and local housing price changes which are not related to an agent's influence in a given listing. Because housing prices can reflect various factors unrelated to agents' influence, most of my analyses focus on "price premiums" – ratios of prices to intrinsic values.

Descriptive regressions, controlling for house fixed effects, show that agent-owned listings have higher sales prices than client-owned listings. The same result can be obtained in terms of sales price premiums as well. Once listing prices are included in these regressions, however, sales prices of agent-owned listings are not significantly higher than those of client-owned listings.[2] Moreover, I find that client-owned listings have lower listing prices (or price premiums) than agent-owned listings, which suggests that agents' influence on listing prices during the listing stage might be more important than other channels. However, these results are rather suggestive, and they do not rule out other channels such as sellers' motivation to sell.

Therefore, I focus more on the role of agents' influence in the listing stage, and develop a new structural model that incorporates the home seller's problem in the listing stage, while controlling

[1] Due to a confidentiality agreement, I cannot reveal the identity of the MLS and the MSA.

[2] In this paper, listing prices refer to *initial* listing prices set in the listing stage. Though sellers can also change listing prices in the selling stage, revised listing prices are not observed in my data.

for other mechanisms. The key problem in the listing stage concerns how much to charge for the listing price before listing the house on a MLS. However, a seller and her agent have similar but somewhat different objective functions. Specifically, the seller pays a fixed commission rate, irrespective of the cost of selling, whereas her agent bears the cost that is likely increasing in the listing price premium. As a result, the optimal listing price premium of the seller deviates from that of her agent. Combining these solutions yields an estimable equation for the listing price premium that depends on a seller's motivation to sell, as well as the parameter capturing an agent's influence. The selling stage entails several decisions, but it is difficult to fully model them. Given this paper's focus on the listing stage, I instead consider a reduced form model for key endogenous variables in the selling stage that indirectly reflects other potential channels for agents' influence.

The identification of agents' influence relies on the comparison between agent-owned and client-owned listings, which is the same identification strategy used by Levitt and Syverson (2008a). However, the key departure from the literature is twofold. First, I consider price premiums to control for various factors unrelated to an agent's influence. Second, I jointly examine three main endogenous variables – sales price premiums, listing price premiums, and the dummy variable for whether listings are sold or withdrawn. In particular, I construct a Tobit-type likelihood function that further allows for the error terms of these variables to be correlated. In addition, I use exclusion restrictions to strengthen the identification.

The structural estimation results provide two primary findings. First, even after accounting for various confounding factors and other mechanisms, I find that agents are likely to influence their clients' problem in the listing stage, particularly by influencing their clients to set lower listing price premiums. Second, even controlling for the correlation between unobservables including motivation to sell, I find that listing price premiums also affect sales price premiums and the probability to sell or withdraw. Therefore, agents' influence in the listing stage is consequential, not only because it leads to lower listing price premiums, but also because it has further impacts on housing search outcomes in the selling stage. In fact, I find that this mechanism can explain about 60% of the difference in sales price premiums between agent-owned and client-owned listings.

The estimation results provide three additional findings. First, the correlation coefficients between unobservables are significant, implying that unobserved motivation to sell indeed plays an important role in housing searches. I also find that the estimated model from ignoring these correlations fails to fit the key patterns of the data, which emphasizes the importance of accounting for these correlations. Second, without agents' influence in the listing stage, sales price premiums are not significantly different between agent-owned and client-owned listings, which is inconsistent with the channel in which agents influence sellers in the selling stage to reduce reservation values and accept suboptimal offers. Though this channel may be still important in other housing markets, it is likely insignificant in the market examined in this paper. Third, even after controlling for the role of listing prices in the selling stage, client-owned listings are more likely to be sold than agent-owned listings, which is consistent with strategic shirking, in that agents may make only minimal effort to bring a barely acceptable offer, even though they may not shirk outright.

These results suggest that any policy attempt to weaken real estate agents' influence should focus more on the listing stage as well as agents' incentives. In this regard, my structural model implies that a potential solution to mitigate agents' influence is to reduce the cost of selling borne by agents, and to enable their costs to depend less on listing price premiums. One example is pricing plans that reflect varying costs of offering different levels of services, rather than relying solely on fixed commission rates, regardless of the cost of selling.

These findings contribute to a large empirical literature on conflicts of interest in various other settings,[3] as well as the aforementioned literature on conflicts of interest between home sellers and

[3]See, e.g., Gruber and Owings (1996), Mehran and Stulz (1997), Hubbard (1998), Garmaise and Moskowitz (2004), Woodward and Hall (2012), and Jiang, Nelson, and Vytlacil (2014).

listing agents. Specifically, this paper is the first to provide significant evidence for two mechanisms – agents' influence in the listing stage and strategic shirking in the selling stage – and further use a structural approach to quantify the extent of agents' influence in listing prices. However, which channel is more important may vary across different local housing markets. In particular, this paper does not consider strategic steering and potential collusion (Hatfield, et al., 2020), but two recent empirical studies show the importance of these mechanisms.[4] Therefore, more studies on similar or other specific mechanisms in other housing markets will provide better insights into how to discipline agents' influence. This paper is also related to the literature on the role of listing prices in real estate transactions.[5] Lastly, to the extent that agents' influence can reduce social welfare, this paper is related to market inefficiency in real estate brokerage industry as well (e.g., Hsieh and Moretti, 2003; Han and Hong, 2011; Barwick and Pathak, 2015).[6]

The paper is organized as follows. Section 2 discusses different mechanisms for agents' influence. Section 3 describes my datasets and presents the descriptive results. Sections 4-5 develop my structural model. Section 6 reports the estimation results and discusses the implications of findings. Section 7 concludes the paper.

2 Agents' Influence on Housing Search

The home seller's search problem includes the listing problem and the selling problem. The listing problem occurs during the listing stage, in which sellers prepare to list their houses on a MLS and decide how to list houses – particularly about how much to charge for listing prices. The selling problem concerns two closely related decisions during the selling stage when houses are listed on a

[4]Barwick, et al. (2016) provide careful and exhaustive analysis of agents' steering away from low commission listings. Han and Hong (2016) provide descriptive and structural evidence for strategic steering in in-house transactions.

[5]See, e.g., Bucchianeri and Minson (2013) and Han and Strange (2015, 2016). This paper complements the literature by focusing more on the role of agents' influence, and comparing agent-owned vs. client-owned listings.

[6]In particular, Barwick and Pathak (2015) develop a structural dynamic model for real estate agents' entry/exit. However, given their focus on entry, they use a reduced-form probability model for transactions of individual listings, and do not consider conflicts of interest. Hence, this paper also complements Barwick and Pathak (2015) by modeling listing-level decisions and examining a different type of inefficiency due to conflicts of interest.

MLS – how long to search for buyers' offers, and whether to sell or withdraw. This section discusses three potential channels through which agents can influence home sellers in these decisions.

The first channel occurs even before the house is listed on a MLS. The key decision during the listing stage is how much to charge for listing prices, which can affect potential offers and sales outcomes in the selling stage. However, many sellers might be less informed about local housing markets than real estate agents. As a result, they may rely more on their agents' advice about listing prices. Accordingly, listing agents can use their informational advantage and persuasion skills to influence sellers' decisions on listing prices. For example, agents could selectively pick recently sold listings to show that a certain range of listing prices would be better for their clients' listings, thus inducing sellers to choose lower listing prices than those that would have been chosen by agents themselves for their own houses.

In the second channel, agents may convince sellers to reduce their reservation values during the selling stage, or impose additional costs associated with waiting for more offers. For example, agents may provide misleading information about local housing markets. As a result, sellers might follow agents' recommendations to accept current offers without waiting for more offers. This mechanism is also closely related to the main description of Levitt and Syverson (2008a) on how agents' distortions occur.[7] Another related example is agents' influence in revising listing prices during the selling stage, which appears to be similar to influencing listing prices during the listing stage. However, they need to be distinguished for two reasons. First, sellers are likely more susceptible to an agent's influence in the listing stage than in the selling stage, because sellers may learn more over time. Second, potential buyers can observe the initial listing price in the listing stage as well as listing prices revised in the selling stage. Hence, buyers can use both listing prices to infer the

[7] Levitt and Syverson (2008a) conclude that their evidence is consistent with an agent information distortion explanation. As a hypothetical example for this explanation, they mention that agents would be willing to forgo about $100 to avoid waiting several days, and convince sellers to sell too cheaply and too quickly, which is similar to the second channel above.

quality and potential market for the house, suggesting that agents' influence in initial listing prices may be more consequential than their influence in revising listing prices.

The third channel is related to agent shirking, in that agents may not put enough effort to bring more or higher offers. For example, agents might be shirking outright by ignoring buyers' requests for showings, or not conducting open houses. Levitt and Syverson (2008a) also discuss this type of shirking, but argue that it is inconsistent with the data. Nevertheless, outright shirking is not the only possibility. Agents can also shirk strategically by making efforts only at the minimum level to deceive sellers into thinking that they are working hard to sell houses. In contrast to outright shirking that may lead to no offer, which is not in the best interest of agents, strategic shirking can benefit agents by ensuring that there is at least one offer which could be accepted by their client-sellers. However, if agents sell their own houses and receive only a barely acceptable offer, they would withdraw from the market, rather than accepting such an offer. Therefore, strategic shirking implies that client-owned properties are more likely to be sold than agent-owned properties.

In addition to the three mechanisms above, agents may manipulate sellers in other ways as well. I do not attempt to separately identify these other mechanisms, since identifying them requires more information and different approaches (see, e.g., Barwick, et al, 2016; Han and Hong, 2016). Some of mechanisms above have been discussed in the literature, but an agent's influence during the listing stage and strategic shirking are not extensively examined in the literature. Hence, I focus more on these two mechanisms. Sections 4-5 develop an approach to quantify their impacts on housing search outcomes, while controlling for other mechanisms. The next section describes my data, and provides descriptive results to explore some of mechanisms discussed in this section.

3 Data and Descriptive Evidence

3.1 Data Description

This paper uses two sources of data. The main source is the Multiple Listing Service (MLS) in a large North American metropolitan statistical area. The MLS data contains various information about each listing. In particular, it includes an indicator variable for agent-owned, which is equal to 1 if the property of the listing is owned by an agent. In addition, the data contains rich information on listings, including sales prices for sold listings, listing prices for all listings (sold or withdrawn), and detailed housing characteristics, such as the number of bedrooms, the number of bathrooms, and house age. The final sample excludes outliers in terms of listing prices and days on market.[8]

The MLS data I obtained consists of the residential listing data for downtown areas and a small part of suburban areas in this MSA from 1996 to 2005. This paper presents the results from downtown areas only. To shorten the paper, the corresponding results from suburban areas are reported in the Online Appendix.[9] The MLS data from downtown areas encompasses about 80,000 listings: about 65% of listings were sold, whereas 35% were withdrawn, where I define that an unsold listing is "withdrawn" from the market if it is not re-listed on the MLS within one year after it was de-listed. Table 1 reports summary statistics for the key variables. Column 1 uses only listings that were sold, whereas column 2 uses all listings (sold or withdrawn). The table shows that sold listings include slightly smaller and older houses than all listings, and they tend to have lower listing prices and shorter days on market.

The second source of my data is the CoreLogic data that covers all transactions in the MSA in the 1990s and 2000s. The CoreLogic data includes detailed transaction-level information, as well as detailed information on house characteristics. The information in the CoreLogic data is

[8] Specifically, I drop listings if listing prices are in the top or bottom 1%, or cumulative days on markets are in the top 1%. In addition, I drop observations with missing information or mistakes such as negative days on market.

[9] For example, Table 1 in this section is equivalent to Table D1 in the Online Appendix, except that Table 1 uses the downtown sample, whereas Table D1 reports the same summary statistics for the suburban sample. Note also that the main results for both areas are very similar, despite the differences between downtown and suburban areas.

useful in recovering intrinsic house values to control for various factors including unobserved house characteristics. The next section provides more discussion on intrinsic house values.

3.2 Estimation of Intrinsic Values

The key endogenous variables in this paper include housing prices – specifically, the listing price, $p^L_{j,t}$, and the sales price, $p^S_{j,t}$, for house j at period t. Housing prices are determined by various factors that are not related to real estate agents. To focus on the role of agents' influence, while controlling for these other factors, I posit that housing prices can be decomposed into "price premium", $r_{j,t}$, and "intrinsic value", $h_{j,t}$, capturing other factors. As a result, both $p^L_{j,t}$ and $p^S_{j,t}$ can be written as

$$p^L_{j,t} = h_{j,t} r^L_{j,t} \quad \text{and} \quad p^S_{j,t} = h_{j,t} r^S_{j,t} \tag{1}$$

where $r^L_{j,t}$ and $r^S_{j,t}$ denote house j's listing price premium and sales price premium, respectively. Intrinsic values can change over time, but they are unlikely to vary in the short run. Thus, I define period t to be a year. Hence $h_{j,t}$ changes across different years, but it is fixed during the same year.

The main issue with using $h_{j,t}$ is that it is not observed directly by the econometrician. To address this issue, I exploit the CoreLogic data that includes several years before the period covered by the MLS data. In addition, while my MLS data includes only a part of the MSA, this additional dataset contains all transactions in the entire MSA. Hence, I can recover $h_{j,t}$, using previous sales prices as well as sales prices in nearby areas in all years between house j's previous transaction and current transaction. Specifically, $h_{j,t}$ can be recovered from

$$\ln h_{j,t} = \ln p^S_{j,t'} + \sum_{k=t'+1}^{t-1} P_{l(j),k} \tag{2}$$

where $p^S_{j,t'}$ is the sales price from the previous transaction in period t' that occurred before period t, and $P_{l(j),k}$ is the housing price appreciation in period k in local market $l(j)$ in which house j is located. The Online Appendix provides more details on how $h_{j,t}$ is computed.

The advantage of constructing $h_{j,t}$ from (2) is that it reflects both observed and unobserved house specific characteristics, as well as various housing market factors, thus likely capturing intrinsic house values. Note that $h_{j,t}$ in (2) is computed by excluding $P_{l(j),t}$, which ensures that $p^S_{j,t}$ is not included in the computation of $h_{j,t}$. Moreover, $p^S_{j,t'}$ reflects unobserved house characteristics, but $p^S_{j,t'}$ or $P_{l(j),k}$ $(k < t)$ is unlikely to be related to an agent's influence in period t. For this reason, I treat $h_{j,t}$ as given, and focus more on $r_{j,t}$ instead of $p_{j,t}$.

3.3 Descriptive Results

This section presents descriptive results to explore suggestive evidence on different channels for agents' influence. I consider not only prices but also price premiums defined in the previous section, and additionally examine potential advantages of focusing on price premiums, instead of prices.

I begin with Table 2 which reports summary statistics of $p^L_{j,t}$ and $p^S_{j,t}$, as well as $h_{j,t}$, $r^L_{j,t}$, and $r^S_{j,t}$ for listings whose previous sales prices can be obtained. Column 1 presents the mean of each variable among agent-owned listings, whereas column 2 reports the mean among client-owned listings. In Panel A, the comparison of sales prices in columns 1-2 shows a somewhat puzzling pattern: agent-owned listings have lower sales prices than client-owned listings,[10] which seems to be inconsistent with the findings from the previous literature, including Levitt and Syverson (2008a).

The seemingly puzzling pattern, however, can be explained by intrinsic values, in that agent-owned listings have lower intrinsic values than client-owned listings, suggesting that the pattern is likely to reflect unobserved house characteristics. In fact, if I consider sales price premiums, agent-owned listings have higher r^S than client-owned listings. But this observation alone does not necessarily indicate that agents influence housing searches, for example, by convincing sellers to accept offers with lower price premiums. One more observation from the table is that listing price premiums r^L for agent-owned listings are also higher than those for client-owned listings, which

[10]This is not the case for the suburb sample. Table D2 in the Online Appendix shows that agent-owned listings have higher sales prices than client-owned listings in suburban areas, which is consistent with the literature.

seems to be consistent with a channel in which agents influence sellers to set relatively lower r^L.

Table 3 presents the results from descriptive regressions. I consider two key regressors related to different channels: listing prices and a dummy for agent-owned listings. All regressions attempt to account for unobserved house characteristics either by using house fixed effects or by using price premiums. Columns 1-5 use the sample of sold listings, whereas columns 6-7 use the sample of both sold and withdrawn listings. All estimations, including regressions in this section and the structural estimation, use robust standard errors clustered at the census tract level.

Column 1 of Table 3 closely replicates the results in Levitt and Syverson (2008a). In this column, the dependent variable is the log sales price, and the regression includes house fixed effects. I find that the coefficient estimate on agent-owned listings is about 0.037 in downtown areas.[11] However, column 2 shows that once the log listing price is included in this regression, the coefficient on agent-owned listings becomes very small and statistically insignificant. In addition, if I use the log listing price as the dependent variable in column 6, the agent-owned coefficient is about 0.026 and statistically significant. These results suggest a potentially important role of listing prices in agents' influence, in that agent-sellers may set higher listing prices than client-sellers, which might also explain that agent-owned listings are sold at higher prices than client-owned listings.

Similar results as above can be obtained in columns 3-5 and 7 as well, in which the dependent variable is price premiums, instead of prices. In the regression of r^S, the coefficient estimate on agent-owned listings is significant and positive in columns 3-4,[12] but it becomes very small once r^L is included in column 5. In the regression of r^L in column 7, the coefficient on agent-owned listings is positive and significant. Therefore, using prices or price premiums produces similar results.

[11] The sample size for house fixed effects is much smaller than the full sample for the following two reasons. First, houses in both areas were sold infrequently during the sample period, and so the number of houses sold at least twice is also much smaller. Second, the MLS data contains inaccurate or incomplete addresses and parcel identification numbers for many observations, which prevents me from matching the same house in different listings.

[12] The difference between columns 3 and 4 is that column 3 uses census tract fixed effects, while column 4 includes zip code fixed effects. However, the results are very similar, suggesting that controlling for zip code level unobservables is sufficient. For this reason, I also include zip code fixed effects in my structural estimation, which is helpful in reducing the number of parameters, compared to including census tract fixed effects.

However, I focus more on price premiums, because they provide two additional advantages.

First, the number of observations from using price premiums is larger than that from using house fixed effects. Second, price premiums can control for unobserved house characteristics without using house fixed effects. This is particularly useful in my structural estimation, since it is difficult to include house fixed effects directly in a nonlinear model. By using price premiums, I do not have to use a sophisticated econometric approach to control for unobserved heterogeneity. In contrast, if prices were used instead, I would need house fixed effects or other sophisticated approach.[13]

The results in Table 3 suggest that the role of agents' influence during the listing stage may be more important than other mechanisms for agents' influence. Nevertheless, they are suggestive, and do not necessarily rule out other possibilities. For example, agent-sellers might be less motivated to sell than client-sellers, thus setting higher listing price premiums. In addition, agents' influence in the selling stage may still be important. To quantify the role of agents' influence in the listing stage while accounting for these other possibilities, I develop a structural model in the next section.

4 Structural Model

This section first develops the model for the listing problem, and then considers a reduced form model for the selling problem. As discussed in Section 3.2, I treat listing j's intrinsic value, $h_{j,t}$, as given, and focus on price premiums. In addition, I drop the subscript t to simplify the notation in this section, since all decisions are made in the same period t.

4.1 Listing Problem

This section aims to develop a tractable model that reflects the two key features of the listing problem, in order for the solution from the model to have reasonable and useful economic meanings. The first feature is that the optimal listing price premium r_j^L should balance the essential trade-off

[13] Without house fixed effects, the regression of the log price yields a negative coefficient on the agent-owned dummy in the downtown sample. I do not report this result, because Table 2 shows a similar puzzling pattern.

in the listing problem: lower r_j^L may bring in lower offers, but lower r_j^L may also attract more offers, thus leading to higher offers due to more competition between buyers. The second is that the seller and her agent may have different objective functions, which can allow the model to capture agents' influence in the listing stage and further suggest a potential solution to reduce agents' influence.

To this end, I start with a general model for the listing problem of house j's seller as follows:

$$\max_{r_j^L} \quad (1-\tau)h_j \, E\left[\max_{i\in\{1,2,\ldots,n_j\}} u_{j,i} \middle| I_j \right] - E(C_j|I_j), \tag{3}$$

where τ is commission rates,[14] $u_{j,i}$ denotes i-th potential offer for house j, n_j is the number of potential offers, I_j is information available to seller j during the listing stage, and C_j is the cost incurred during the selling stage. Since the model above cannot be solved generally, I impose the following three assumptions that incorporate the two key features of the listing problem.

The first assumption is that $u_{j,i}$ consists of a common value and a buyer-specific value. The common value is likely proportional to r_j^L, so that it can be written as $\beta_j r_j^L$, where β_j is a random coefficient capturing its relative importance that may vary across different housing market conditions. Denoting the buyer-specific value by $\nu_{j,i}$, I then obtain

$$E\left[\max_{i\in\{1,2,\ldots,n_j\}} u_{j,i} \middle| I_j \right] = \beta_j r_j^L + E\left[\max_{i\in\{1,2,\ldots,n_j\}} \nu_{j,i} \middle| I_j \right], \tag{4}$$

which implies that the seller's expectation of the highest offer is determined by the seller's belief on the distribution of $\nu_{j,i}$ as well as the number of potential offers, n_j.

It is plausible that the second part of (4) is increasing in n_j which is inversely related to r_j^L. The second assumption simply imposes this property by assuming that $E\left[\max_{i\in\{1,2,\ldots,n_j\}} \nu_{j,i} \middle| I_j \right]$ is approximated by $\omega_j - \frac{(r_j^L)^2}{2m_j}$, where ω_j is a parameter that varies across housing markets, and m_j is a positive variable capturing the seller's belief on house j's market. The Online Appendix derives this expression from the exponential distribution on $\nu_{j,i}$ as well as a functional form assumption

[14] If seller j is an agent selling her own house, then τ will not be included in (3).

about n_j and r_j^L. Given the two assumptions above, (4) can be rewritten as

$$E\left[\max_{i\in\{1,2,\ldots,n_j\}} u_{j,i}\middle| I_j\right] = \beta_j r_j^L + \left(\omega_j - \frac{(r_j^L)^2}{2m_j}\right). \tag{5}$$

This expression has the following advantages. First, it allows for the essential trade-off in the listing problem. Second, it results in a simple log-linear equation for the optimal $r_j^{L^*}$, as shown below. Third, this also guarantees that the second order condition to (3) is satisfied. Lastly, (5) incorporates seller's motivation, in that a seller with higher m_j would expect more offers for a given r^L and thus higher prices, in which case the seller is less likely to be motivated to sell easily.

The third assumption is that $E(C_j|I_j)$ in (3) is invariant to r_j^L for seller j, but for seller j's agent, it increases in r_j^L. As Levitt and Syverson (2008a) pointed out, real estate agents typically bear the cost of selling the house (e.g. advertising, hosting open houses, etc.), while sellers do not. Moreover, the agent's cost of selling is likely to increase in r_j^L, because if r_j^L is higher, the house becomes more difficult to sell, thus requiring the agent to incur more efforts to find buyers. Specifically, I assume that $E(C_j|I_j) = \kappa_j h_j$ for seller j, while $E(C_j|I_j) = (\kappa_j^{agent} + \varphi r_j^L)h_j$ for seller j's agent, where κ_j, κ_j^{agent}, and φ are positive parameters for different levels of the expected cost.[15]

The assumptions above imply that seller j's listing problem in (3) can be rewritten as

$$\max_{r_j^L} \quad (1-\tau)h_j\left[\beta_j r_j^L + \left(\omega_j - \frac{(r_j^L)^2}{2m_j}\right)\right] - \kappa_j h_j. \tag{6}$$

Taking the first order condition and solving for the optimal $r_j^{L^*}$ yields

$$r_j^{L^*} = \beta_j m_j. \tag{7}$$

The optimal $r_j^{L^*}$ in (7) suggests a few intuitive implications. First, motivated sellers (i.e. lower m_j) will set lower listing price premiums. Second, if sellers expect that r_j^L is relatively less important than a buyer-specific value in determining the expected offer value (i.e. lower β_j), sellers will set lower listing price premiums, because a decrease in r_j^L would be outweighed by an increase in n_j.

[15]The model allows the seller to have a positive cost $\kappa_j h_j$, although κ_j can be negligible. Note also that I factor out h_j to be consistent with price premiums.

The listing problem for seller j's agent is similar to (6), except that the agent receives a τ fraction of the sales price, and the expected cost is also different. Seller j's agent thus solves

$$\max_{r_j^L} \quad \tau h_j \left[\beta_j r_j^L + \left(\omega - \frac{(r_j^L)^2}{2m_j} \right) \right] - (\kappa_j^{agent} + \varphi r_j^L) h_j,$$

Taking the first order condition and solving for the optimal $r_j^{L,agent^*}$ yields

$$r_j^{L,agent^*} = \left(\beta_j - \frac{\varphi}{\tau} \right) m_j. \tag{8}$$

Comparing (7) and (8) shows that if agents seek their own interest and influence their clients' listing problem, the resulting r_j^L will be lower than (7). Hence, the observed r_j^L can be written as

$$\ln(r_j^L) = \ln(\beta_j + \vartheta d_j) + \ln(m_j), \tag{9}$$

where d_j is the dummy variable for client-owned listings (i.e. $d_j = 0$ for agent-owned listings), and ϑ captures the extent to which agents influence their clients' listing problem.[16]

To estimate the model, I consider the following linear approximation of (9):

$$\ln(r_j^L) = X_j \delta + \theta d_j + \epsilon_{1j}, \tag{10}$$

where X_j is a vector of variables related to β_j, δ is a corresponding vector of parameters, θ is the parameter reflecting ϑ above, and ϵ_{1j} is the error term mainly capturing m_j.[17] The model above provides two useful implications. First, the agent's influence in the listing stage can be measured by θ in (10), and its identification boils down to controlling for X_j and m_j captured by ϵ_{1j}. Second, the agent's optimal r_j^L deviates from the seller's optimal choice, because the agent bears the cost of selling, and this cost increases in r_j^L. This suggests that the agent's influence may be reduced if $\varphi = 0$, for example, by allowing agents to provide only standardized services for fixed fees.

[16]Note that ϑ reflects both $-\varphi/\tau$ and the degree to which client-sellers follow their agents' advice. These two cannot be separated, and so I only consider ϑ.

[17]For three reasons, I use (10) in the estimation, instead of (9). First, though the approximation error from the linearization is also included in ϵ_{1j}, it is likely minor and unlikely to create any potential bias. Second, the degree of an agent's influence in the listing stage can be still captured by θ in (10). Third, the linear model in (10) can easily incorporate various fixed effects to account for local market unobservables.

4.2 Selling Problem

The selling problem occurs once the house is listed on a MLS. This problem entails the seller's decisions on how to search for offers (e.g. holding open houses, revising listing prices), whether to accept an offer if it arrives, or continue to search, or withdraw. These decisions depend on various factors, including the seller's reservation value, the distribution of offers, bargaining between sellers and buyers, as well as agents' influence. They can be also affected by r_j^L determined in the listing stage. Hence, it is difficult to fully model them. To study the role of an agent's influence in the listing stage, however, a reduced form model for the selling stage is sufficient.

Accordingly, I consider a reduced form model for two key endogenous variables in the selling stage. The first is seller j's sales price premium, r_j^S, written as

$$\ln(r_j^S) = \alpha \ln(r_j^L) + W_j \zeta + \xi_j, \tag{11}$$

where W_j and ξ_j are respectively observed and unobserved variables related to sales price premiums, ζ is a vector of coefficients corresponding to W_j, and the coefficient α captures the degree to which r_j^L affects sales price premiums. The other key endogenous variable is the dummy variable for sold listings, denoted by s_j. I assume that s_j is given by

$$s_j = 1\{y_j > 0\}, \text{ and } y_j \equiv \gamma \ln(r_j^L) + Z_j \eta + \upsilon_j, \tag{12}$$

where $1\{\cdot\}$ is the indicator function; y_j is a latent variable determined by r_j^L as well as Z_j and υ_j, denoting observed and unobserved variables related to sale and withdrawal, respectively; γ and η are parameters corresponding to $\ln(r_j^L)$ and Z_j.

One of variables I consider in W_j and Z_j is the dummy variable for client-owned listings, d_j. The coefficients on d_j in (11) and (12) are likely to reflect not only agents' influence but also the difference between agent-sellers and client-sellers in the selling stage. Although I cannot isolate a particular mechanism from the coefficient on d_j, the estimate may still provide suggestive evidence on some mechanisms. For example, agents' influence in the selling stage may reduce sellers' reservation values, implying a negative coefficient on d_j in (11). However, client-sellers may also have lower reservation values than agent-sellers, in which case the coefficient on d_j will be biased downward. Hence, if the estimated coefficient is not significantly negative, I can rule out a mechanism in which agents influence sellers to accept suboptimal offers during the selling stage.

Note that r_j^L is always determined before s_j and r_j^S are determined. As a result, r_j^L enters both (11) and (12), but s_j or r_j^S does not enter the equation for r_j^L in (10). This also reflects the potential role of r_j^L in the selling stage, given that r_j^L is likely to affect the distribution of offers and search outcomes, and may also change the cost of selling, thus affecting the agent's behavior in the selling stage. Though I do not explicitly model how r_j^L affects the selling stage, the effect of r_j^L on s_j and r_j^S can be still reflected in the reduced form model above.

Note also that the combined model for the listing stage and the selling stage can allow for the correlation among m_j, ξ_j, and υ_j. Allowing for this correlation is particularly important, because both ξ_j and υ_j are likely to include various factors in the selling stage that are correlated with the seller's motivation, m_j. In addition, by allowing m_j to be correlated with ξ_j and υ_j, I can address the endogeneity of r_j^L in (11) and (12), given that this endogeneity issue mainly arises from the correlation between these unobservables. Moreover, to the extent that m_j is correlated with r_j^S and s_j through ξ_j and υ_j, jointly estimating r_j^L with r_j^S and s_j controls for ϵ_{1j} (i.e. m_j) in (10) indirectly, which would not be possible if the equation (10) were to be estimated alone. The next section describes how this correlation is incorporated into the model econometrically.

5 Econometric Model

5.1 Likelihood Function

I begin by deriving the likelihood function for the model developed in the previous section. The main endogenous variables observed in the data include the indicator variable for sale or withdrawal, s_j, and listing price premium, r_j^L, as well as sales price premium, r_j^S, which is observed only if $s_j = 1$. The individual likelihood function for listing j, denoted by ℓ_j, can be written as

$$\ell_j = f(r_j^L)\left[\Pr(s_j = 0|r_j^L)\right]^{1-s_j}\left[\Pr(s_j = 1|r_j^L)g(r_j^S|s_j = 1, r_j^L)\right]^{s_j},$$

where $f(r_j^L)$ denotes the probability density function (PDF) for r_j^L, and $g(r_j^S)$ denotes the PDF for r_j^S. To obtain the likelihood function for the estimation, I further rewrite ℓ_j as

$$\ell_j = f(r_j^L)\left[\int_{-\infty}^{0} q(y_j|r_j^L)dy\right]^{1-s_j}\left[\int_{0}^{\infty} q(y_j|r_j^S, r_j^L)dy \times g(r_j^S|r_j^L)\right]^{s_j}, \tag{13}$$

where y_j defined in (12) is used for s_j, and $q(y_j)$ denotes the PDF for y_j.

Note that (13) contains three conditional PDFs. To obtain tractable expressions for these conditional PDFs, while allowing for the correlation between ξ_j, υ_j, and m_j, I first rewrite the model in (11) and (12) by replacing $\ln(r_j^L)$ with (10) as follows:

$$\begin{aligned} \ln(r_j^S) &= \alpha(X_j\delta + \theta d_j) + W_j\zeta + \epsilon_{2j} \\ y_j &= \gamma(X_j\delta + \theta d_j) + Z_j\eta + \epsilon_{3j} \end{aligned} \tag{14}$$

where $\epsilon_{2j} = \xi_j + \alpha\epsilon_{1j}$ and $\epsilon_{3j} = \upsilon_j + \gamma\epsilon_{1j}$. I then assume that ϵ_{1j}, ϵ_{2j}, and ϵ_{3j} follow the multivariate normal distribution with zero means, which implies that $\boldsymbol{Y} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{Y} = \begin{pmatrix} \ln(r_j^L) \\ \ln(r_j^S) \\ y_j \end{pmatrix}, \ \boldsymbol{\mu} = \begin{pmatrix} X_j\delta + \theta d_j \\ \alpha(X_j\delta + \theta d_j) + W_j\zeta \\ \gamma(X_j\delta + \theta d_j) + Z_j\eta \end{pmatrix}, \ \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho_{1,2}\sigma_1\sigma_2 & \rho_{1,3}\sigma_1\sigma_3 \\ \rho_{1,2}\sigma_1\sigma_2 & \sigma_2^2 & \rho_{2,3}\sigma_2\sigma_3 \\ \rho_{1,3}\sigma_1\sigma_3 & \rho_{2,3}\sigma_2\sigma_3 & \sigma_3^2 \end{pmatrix} \tag{15}$$

and σ_1^2, σ_2^2, and σ_3^2 denote the variance of $\ln(r_j^L)$, $\ln(r_j^S)$, and y_j, respectively; the correlation coefficients between these three variables are denoted by $\rho_{1,2}$, $\rho_{1,3}$, and $\rho_{2,3}$. Note that using the log of r_j^L and r_j^S in (15) indicates that r_j^L and r_j^S are assumed to follow the log-normal distribution.

The key advantage from the normal distribution assumption above is that its conditional distribution is also normal with known expressions for the conditional mean and covariance. Specifically, the conditional distribution of y_j conditional on $\ln(r_j^L)$ is given by

$$N\left(\gamma(X_j\delta+\theta d_j)+Z_j\eta+\frac{\rho_{1,3}}{\sigma_1}(\ln(r_j^L)-(X_j\delta+\theta d_j)),\ (1-\rho_{1,3}^2)\right), \tag{16}$$

where σ_3^2 is normalized to 1. Similarly, the conditional distribution of $\ln(r_j^S)$ given $\ln(r_j^L)$ follows

$$N\left(\alpha(X_j\delta+\theta d_j)+W_j\zeta+\frac{\rho_{2,3}\sigma_2}{\sigma_1}(\ln(r_j^L)-(X_j\delta+\theta d_j)),\ (1-\rho_{2,3}^2)\sigma_2^2\right). \tag{17}$$

Finally, the conditional distribution of y_j, conditional on $\ln(r_j^S)$ and $\ln(r_j^L)$ is given by

$$N(\boldsymbol{\mu}_2+\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\boldsymbol{Y}_1-\boldsymbol{\mu}_1),\ \boldsymbol{\Sigma}_{22}-\boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}), \tag{18}$$

where $\boldsymbol{Y}_1=(\ln(r_j^L),\ln(r_j^S))'$, $\boldsymbol{\mu}_1=(X_j\delta+\theta d_j,\alpha(X_j\delta+\theta d_j)+W_j\zeta)'$, $\boldsymbol{\mu}_2=\gamma(X_j\delta+\theta d_j)+Z_j\eta$, and $\boldsymbol{\Sigma}$ in (15) is partitioned into four blocks with (2×2) matrix $\boldsymbol{\Sigma}_{11}$, (1×2) matrix $\boldsymbol{\Sigma}_{12}$, (2×1) matrix $\boldsymbol{\Sigma}_{21}$, and $\boldsymbol{\Sigma}_{22}=\sigma_3^2=1$. The likelihood function (13) can be rewritten, using conditional PDFs from (16), (17), and (18). These conditional PDFs account for truncated or non-randomly selected observations.[18] The resulting likelihood function is therefore related to a Tobit model. The predicted values of the dependent variables can be computed by using the inverse Mills ratio, as shown in Section 6.2. The model is then estimated by the maximum likelihood estimation.

5.2 Identification

The advantages of using price premiums are discussed in Sections 3.2-3.3. In particular, listing j's intrinsic value can control for unobserved house characteristics, and help isolating an agent's influence that might arise during the listing stage and the selling stage of listing j, because it is recovered from its previous transaction price and sales prices of other houses in the same neighborhood before listing j occurred. However, there are two more identification issues to examine.

The first is that r_j^L is likely to be correlated with ξ_j in (11) and υ_j in (12). To address this issue, I allow for the correlation between m_j, ξ_j, and υ_j, which is modeled in the previous section.

[18] Note that r_j^S is observed only if $y_j>0$. In addition, the realized values of s_j and r_j^S depend on r_j^L. Such truncated or non-randomly selected observations are modeled parametrically in these conditional PDFs.

Nevertheless, the issue may still arise if the same variables in X_j and d_j are also included in W_j or Z_j in (14). To avoid this issue, I need variables that enter only X_j, but not W_j or Z_j. These variables are essentially instruments for r_j^L in (11) and (12). However, such variables are not readily available, since most housing variables are likely to enter not only X_j, but also W_j and Z_j.

To find such instruments, I use the following two approaches. First, I include detailed house characteristics,[19] as well as zip code fixed effects and year×month fixed effects in X_j, W_j, and Z_j. Second, as instrument variables, I use log h_j, and tract-level yearly average log listing prices that are computed by excluding house j. Both instruments are likely correlated with r_j^L, because r_j^L is determined relative to h_j, and sellers tend to consider neighbors' listing prices when they choose their listing prices. Of course, they can be also correlated with r_j^S and s_j. However, the key identifying assumption is that conditional on observed house characteristics, as well as local market and time-varying unobservables, these instruments should affect r_j^S and s_j only through r_j^L.

One concern related to the second instrument is that listing prices of those adjacent to house j may also affect ξ_j and υ_j, in addition to r_j^L. For example, house j's neighbors may be more motivated to sell, thus setting lower listing prices. This is likely to put pressure on r_j^L, but the concern is that this may also alter the distribution of buyers' offers. To the extent that W_j and Z_j can control for the offer distribution, this concern can be alleviated. To further address this concern, I modify the instrument by excluding nearby houses as well. However, these modified instruments do not change the results significantly.[20] For this reason, I use the instruments proposed above.

Lastly, I consider standard tests for the validity of instruments that use the first-stage F-statistic and Sargan's J-test. The results from these tests are reported in Table 4, where I use the two approaches described above. In column 1, the dependent variable is r_j^S, and only sold listings

[19] House characteristics include #bedrooms, #rooms, #bathrooms, #garages, and various dummy variables for property types, basement types, and house ages.

[20] Note that these instruments are also similar to the instruments used in Bayer, et al. (2007). In the Online Appendix, the results from these modified instruments are reported in Table C1 that can be compared with Table 4.

are used. In column 2, the dependent variable is the dummy variable for sold listings, and the sample includes both sold and withdrawn listings. The table shows that instruments are indeed important in determining r_j^L, but are unlikely to be correlated with the error terms in r_j^S and s_j. This result also supports the key identifying assumption.

The second identification issue is that sellers' motivation reflected by m_j needs to be distinguished from agents' influence in the listing stage captured by the coefficient on d_j in (10). If client-sellers have lower m_j (i.e. more motivated to sell) than agent-sellers, r_j^L will be lower for client-sellers than agent-sellers even in the absence of an agent's influence. In this case, d_j will be correlated with ϵ_{1j} in (10). This identification issue is addressed as follows. First, X_j includes the instruments and other control variables described above. This means that d_j needs to be uncorrelated with ϵ_{1j}, conditional on X_j, which is more plausible than the case without these control variables. Note also that X_j includes year$\times$month fixed effects, thus adjusting for seasonality.[21] Second, m_j is allowed to be correlated with the error terms in r_j^S and s_j. To the extent that the difference in m_j between agent-sellers and client-sellers is reflected in this correlation, it can also be captured by the difference in r_j^S and s_j between different sellers. Hence, jointly estimating r_j^L, r_j^S, and s_j helps to address the identification issue. Third, I allow the variance of r_j^L and r_j^S to depend on d_j, which further controls for the difference in m_j between agent-sellers and client-sellers.

6 Results

This section first discusses the model estimates, and also assesses the fit of the model. I then examine the key findings from the full model. I further discuss their implications on how to discipline agents. I propose a potential solution to reduce agents' influence, which also motivates a

[21]One may be also concerned that agent-sellers might be more flexible than client-sellers in terms of when to list their houses. To explore this concern, I regress the dummy for an agent-owned listing on the year and month dummies for when the house was listed, controlling for house fixed effects. I find that the coefficients on all time dummies are small and statistically insignificant, suggesting that when to list the house is not necessarily different between agent-sellers and client-sellers.

counterfactual exercise to quantify the impacts of agents' influence in the listing stage.

6.1 Estimation Results

The estimation results are presented in Table 5. Each panel in the table reports the key estimates of each model component: Panel A includes the equation for $\ln r_j^L$ in (10), and reports θ, the coefficient on d_j, the dummy for client-owned listings; Panel B shows the equation for $\ln r_j^S$, including α, the coefficient on $\ln r_j^L$, and the coefficient on d_j in (11); the equation for s_j is in Panel C which contains γ, the coefficient on $\ln r_j^L$, and the coefficient on d_j in (12); Panels D-F present the estimates on the variance-covariance matrix Σ in (15).

In the table, column 4 uses all approaches described in Section 5, whereas columns 1-3 use simpler approaches for comparison. In column 1, the model includes all three endogenous variables. However, it does not allow the error terms to be correlated, though their variances, except for σ_3^2 which is normalized to be 1, are allowed to vary between client-sellers and agent-sellers. All control variables – house characteristics, zip code fixed effects, and year×month fixed effects – are included in all columns, but column 1 does not use instruments for $\ln r_j^L$, and also excludes $\ln r_j^L$ in the equations for $\ln r_j^S$ and s_j. In contrast, column 2 adds $\ln r_j^L$ to Panels B and C to reflect the impact of r_j^L on the selling stage. Column 3 further includes instruments in the equation for $\ln r_j^L$. Lastly, only column 4 allows the error terms to be correlated with each other.

The comparison of columns 1-4 provides two key observations of the model estimates. First, simpler approaches overestimate the coefficient estimates on d_j in Panels A-C. In column 1, these estimates are all significant, which appears to suggest the importance of agents' influence in both the listing stage and the selling stage. However, as shown in column 2, adding $\ln r_j^L$ to the equation for $\ln r_j^S$ significantly reduces the magnitude of the coefficient estimates on d_j in Panel B, which is also consistent with the regression results reported in columns 4-5 of Table 3. In column 4, the estimate in Panel B becomes even positive, and statistically insignificant. In Panel A, the magnitude of the estimate is also reduced by about one third from column 1 to column 4, but it is still significant in column 4. In Panel C, the coefficient on d_j is significant in all columns.

Second, the estimated correlation coefficients between the three error terms are all statistically significant in column 4. In particular, the correlation coefficient between the error terms in $\ln r_j^L$ and $\ln r_j^S$ is 0.937, which implies that sellers' motivation captured by m_j is strongly correlated with the error term in $\ln r_j^S$. For example, sellers with lower m_j will set lower r_j^L even without agents' influence, and are also willing to accept lower offers, thus resulting in lower r_j^S. Therefore, ignoring these correlations is likely to bias the estimates, which explains the difference in the coefficients on d_j and $\ln r_j^L$ between column 4 and the other columns.

6.2 Model Fit

In this section, I assess the fit of the main model in column 4 of Table 5, relative to simpler models in columns 1-3. To this end, I compute the predicted values of price premiums for different cases. In particular, I calculate the expected value of $\ln r_j^L$ and $\ln r_j^S$, conditional on d_j and s_j. The results are reported in Table 6. Columns 1-4 of this table report the predicted values from the estimated models, where each column corresponds to the same column in Table 5. In addition, column 5 of Table 6 presents the conditional mean values of observed price premiums.

Note that I consider log price premiums, instead of price premiums, because $\ln r_j^L$ and $\ln r_j^S$ are assumed to follow the normal distribution, and their conditional expectation can be computed by using the property of the bivariate normal distribution and the inverse Mills ratio.[22] However, Table 6 presents $\exp\left[E(\ln r_j^L|d_j, s_j)\right]$ and $\exp\left[E(\ln r_j^S|d_j, s_j)\right]$, instead of $E(\ln r_j^L|d_j, s_j)$ and $E(\ln r_j^S|d_j, s_j)$, because the former take values close to r_j^L and r_j^S, and are easier to interpret than

[22] Since $s_j = 1$ if $y_j > 0$, $E(\ln r_j^L|d_j, s_j = 1)$ is given by $X_j\delta + \theta d_j + \frac{\rho_{1,3}\sigma_1\phi\left[-\gamma(X_j\delta+\theta d_j)-Z_j\eta\right]}{1-\Phi\left[-\gamma(X_j\delta+\theta d_j)-Z_j\eta\right]}$. A similar expression for $E(\ln r_j^L|d_j, s_j = 0)$ can be obtained by modifying the inverse Mills ratio. Likewise, the conditional expectation for $\ln r_j^S$ is given by $E(\ln r_j^S|d_j, s_j = 1) = \alpha(X_j\delta + \theta d_j) + W_j\zeta + \frac{\rho_{2,3}\sigma_2\phi\left[-\gamma(X_j\delta+\theta d_j)-Z_j\eta\right]}{1-\Phi\left[-\gamma(X_j\delta+\theta d_j)-Z_j\eta\right]}$. Note that these expressions are used for the predicted values from the main model in column 4. For simpler models in columns 1-3, the conditional expectation can be obtained by excluding the inverse Mills ratio in these expressions.

the latter. To be comparable with those in columns 1-4, the values in column 5 are also the exponential function of the conditional mean values of log price premiums.

Table 6 reveals that the main model fits the data better than simpler models. The values in column 4 are clearly the closest to those in column 5. Moreover, those in column 4 show the following three patterns from the observed values in column 5: first, price premiums of agent-owned listings are higher than those of client-owned listings; second, sales price premiums are lower than listing price premiums; third, listing price premiums of sold listings are lower than those of withdrawn listings. In contrast, the values from simpler models exhibit only the first pattern, whereas the second and third patterns are either absent or only weakly shown in columns 1-3. Therefore, if the correlation between the three error terms of endogenous variables is ignored, the estimated model fails to fit the key patterns of the data.

6.3 Main Findings

Given the results on the model fit, I now focus on the full model in column 4 of Table 5. Five findings emerge from column 4. First, even after controlling for various confounding factors as well as the significant correlations between the error terms, I find that listing price premiums of client-owned listings are lower than those of agent-owned listings by 2.84%. This supports the first channel of an agent's influence discussed in Section 2.

Second, the coefficient estimates on $\ln r_j^L$ are mostly large and significant in Panels B and C, thus indicating that listing price premiums play an important role in the selling stage. Hence, agents' influence in the listing stage can further affect the selling stage through listing prices. Third, the coefficient estimate on $\ln r_j^L$ in Panel C is 0.246, whereas $\rho_{1,3}$, the correlation coefficient between the error terms in $\ln r_j^L$ and s_j, is -0.231. Therefore, a high listing price premium itself may improve the chance to sell, for example, by attracting offers that satisfy the seller, but it also reflects the seller's lower motivation to sell (i.e. higher m_j) which reduces the probability to sell. This result also implies that if the correlation between the error terms is ignored (as in columns 2-3), the coefficient estimate on $\ln r_j^L$ in Panel C is biased due to unobserved seller motivation.

Fourth, the coefficient estimate on d_j in Panel B is small and insignificant. This result is inconsistent with the second mechanism discussed in Section 2, in which agents might influence sellers in the selling stage to reduce reservation values, and accept suboptimal offers. To the extent that this estimate can be biased downward if client-sellers have lower reservation values than agent-sellers, the second mechanism is more likely ruled out.[23]

Fifth, the coefficient estimate on d_j in Panel C is positive and significant. Hence, even after controlling for the role of listing price premiums, client-owned properties are more likely sold than agent-owned properties, which is inconsistent with outright shirking. However, it is consistent with strategic shirking, in that agents will ensure that their clients accept a barely sufficient offer that they would reject for their own houses. In addition, strategic shirking implies that once controlling for listing price premiums, sales price premiums of client-owned listings will not be significantly different from those of agent-owned listings, given that agents still make efforts to bring a minimally acceptable offer to their clients, but no extra effort to bring higher or more offers. Accordingly, the insignificant coefficient on d_j in Panel B is also consistent with strategic shirking.

6.4 Discussion

Given the findings above, what are the implications on how to discipline agents and avoid agents' influence? In the market examined in this paper, the second mechanism, in which agents influence their clients' reservation values or search costs during the selling stage, is unlikely important. As a result, sellers may not be better off by rejecting their agents' advice during the selling stage and waiting longer for higher offers, because simply waiting longer will not result in higher offers. Of course, sellers can additionally increase their listing prices during the selling stage, in order to attract

[23] An alternative interpretation is that agents' influence in the listing stage might be already sufficient, so that further manipulations during the selling stage may be unnecessary.

higher offers. However, unless sellers can delete the history of previous listing prices completely, most potential buyers will take into account revised listing prices as well as initial listing prices. Moreover, as the results above have shown, the role of initial listing prices is consequential in the selling stage, suggesting that revising listing prices alone may not necessarily improve the outcomes.

Another key finding above is that agents' influence in the listing stage is important, not only because they influence sellers to set lower r^L, but also because lower r^L further affects housing search outcomes in the selling stage. In this regard, one useful way for sellers to avoid agents' influence is to discount their agents' advice in the listing stage and set r^L higher than agents recommend. If this does not alter agents' behavior, sellers may be able to sell at higher prices. However, there are three potential issues. First, sellers may not know how much to discount agents' advice in the listing stage. Second, if agents anticipate that sellers might discount their advice on listing prices, they might recommend even lower r^L to ensure that sellers still choose lower r^L. Third, if higher r^L also increases the minimum level of effort to bring any offer, sellers' attempts to avoid agents' influence by increasing r^L would discourage agents from making even minimal effort to bring a barely acceptable offer for sellers, which could create another distortion.

From the agent's perspective, it might be more effective and efficient to influence their clients in the listing stage than in the selling stage for two reasons. First, many sellers might be less informed in the listing stage than in the selling stage, so that they might be more easily deceived in the listing stage. Second, lower r^L might imply that the minimum level of effort to bring any acceptable offer is also likely lower. This suggests that removing agents' distortion in the listing stage may not be easy, and an attempt to do so needs to take into account agents' incentives and potential changes in their behavior.

In this regard, the simple model of the agent's problem in (8) discussed in Section 4.1 suggests a potential solution. In this model, the agent's optimal listing price premium deviates from the seller's optimal choice, because agents bear the cost of selling, and more importantly, this cost is increasing in listing price premiums. This suggests that the agent's deviation from the seller's optimal choice could be reduced if agents bear the cost less and their costs depend less on listing price premiums. For example, agents could provide standardized services for fixed fees, but they could also offer upgraded services for additional fees.

6.5 Counterfactual Results

In this section, I use the estimates from the main model in column 4 of Table 5, and quantify the impacts of the first channel for an agent's influence in terms of changes in price premiums. This exercise is also motivated by the potential solution proposed at the end of the previous section which can eliminate the agent's influence in the listing stage by requiring the agent's cost not to depend on the listing price premium. I consider a counterfactual scenario where I remove an agent's influence in the listing stage by setting $\theta = 0$. The results are presented in Table 7 in which the conditional expectation of log price premiums is computed similarly as Table 6. In fact, column 1 in Table 7 is the same as column 4 of Table 6. The difference is that Table 7 also reports the predicted values under the counterfactual scenario above, and it does not report the values for agent-owned listings, because only client-owned listings are affected under the counterfactual scenario.

Column 2 reports the predicted values under $\theta = 0$, in which agents' influence in the listing stage is removed, while listing price premiums continue to affect the selling stage. The result shows that both sales price premiums (Panel A) and listing price premiums (Panel B) for client-owned listings increase, which illustrates that agents' influence in the listing stage can be consequential not only by lowering listing price premiums during the listing stage, but also by lowering sales price premiums during the selling stage. To further quantify the impact of agents' influence, I use the counterfactual price premiums in column 2, and compare them with the predicted price premiums in column 1. This difference can be attributable to agents' influence in the listing stage. I then compute the percentage of this difference, relative to the difference in price premiums between agent-owned, vs. client-owned listings in column 4 of Table 6.

Column 3 of Table 7 shows that agents' influence in the listing stage can explain about 61% of the difference in sales price premiums between agent-owned and client-owned listings. In other words, the proposed solution at the end of the previous section can remove about 61% of the difference in sales price premiums between agent-owned and client-owned listings. Column 3 also reports that the impact of agents' influence in the listing stage on the listing price premium is consistently significant, explaining about 69% (sold listings) or 67% (withdrawn listings) of the difference in listing price premiums between agent-owned vs. client-owned listings.

7 Conclusion

This paper examines different ways for real estate agents to influence a housing search, particularly focusing on two mechanisms that have not been studied extensively in the literature – agents' influence in the listing stage and strategic shirking. Using the MLS data, I first find that agent-owned listings have higher sales prices than client-owned listings, but this well-known result is not robust when including listing prices set before listing. Hence, I further investigate agents' influence in the listing stage by developing a structural model for the home seller's problem in the listing stage, while accounting for other potential mechanisms for agents' influence in the housing search. One key feature of the structural model is to allow for the correlation between unobservables including motivation to sell. The estimation results show that this feature is important, because the key patterns of the data do not fit if such correlation is ignored, and unobserved motivation to sell also plays a significant role in the housing search.

The primary findings from the structural estimation indicate that agents are likely to influence sellers to set lower listing price premiums in the listing stage, and these listing price premiums also affect sales price premiums in the selling stage, even after controlling for the correlation between unobservables. I find that this mechanism explains 61% of the difference in sales price premiums between agent-owned and client-owned listings. In addition, the estimation results suggest that agents may make only minimal effort to bring a barely acceptable offer in the selling stage. Therefore, real estate agents indeed seek their own interest by taking advantage of sellers' insufficient knowledge and experience particularly in the listing stage, and by strategically minimizing their costs at the expense of their client-sellers.

These results suggest that any policy attempt to mitigate an agent's influence needs to focus more on the listing stage as well as agents' incentives and possible changes in their behavior. In this respect, a potential solution implied by my structural model is to reduce the cost of selling borne by agents, and to enable their costs to depend less on listing price premiums. For example, this can be achieved by allowing agents to be compensated for different levels of services, instead of using only the current pricing that typically pays agents fixed commission rates, regardless of their costs of selling. Therefore, an avenue for future research is to investigate similar solutions or other alternatives, together with real estate agents' incentives.

References

Barwick, P. J. and Pathak, P. (2015), "The Costs of Free Entry: An Empirical Study of Real Estate Agents in Greater Boston," *RAND Journal of Economics*, *46(1)*, pp. 103-145.

Barwick, P. J., Pathak, P., and Wong, M. (2017), "Conflicts of Interest and Steering in Residential Brokerage," *American Economic Journal: Applied Microeconomics*, *9(3)*, pp. 191-222.

Bayer, P., Ferreira, F., and McMillan, R. (2007), "A Unified Framework for Measuring Preferences for Schools and Neighborhoods," *Journal of Political Economy*, *115(4)*, pp. 588-638.

Bernheim, D. and Meer, J. (2012), "Do Real Estate Brokers Add Value When Listing Services are Unbundled?," *Economic Inquiry*, *51*, pp. 1166-1182.

Bucchianeri, G. and Minson, J. (2013), "A Homeowner's Dilemma: Anchoring in Residential Real Estate Transactions," *Journal of Economic Behavior & Organization*, *89*, pp. 76-92.

Garmaise, M. and Moskowitz, T. (2004), "Confronting Information Asymmetries: Evidence from Real Estate Markets," *Review of Financial Studies*, *17(2)*, pp. 405-37.

Gruber, J. and Owings, M. (1996), "Physician Financial Incentives and Cesarean Section Delivery," *RAND Journal of Economics*, *27(1)*, pp. 99-123.

Han, L. and Hong, S.-H. (2011), "Testing Cost Inefficiency Under Free Entry in the Real Estate Brokerage Industry," *Journal of Business and Economic Statistics*, *29(4)*, pp. 564-578.

Han, L. and Hong, S.-H. (2016), "Understanding In-House Transactions in the Real Estate Brokerage Industry," *RAND Journal of Economics*, *47(4)*, pp. 1057-1086.

Han, L. and Strange, W. (2015), "The Microstructure of Housing Markets: Search, Bargaining, and Brokerage," *Handbook of Regional and Urban Economics*, *Vol. 5*, pp. 813-886.

Han, L. and Strange, W. (2016), "What is the Role of the Asking Price for a House?" *Journal of Urban Economics*, *93*, pp. 115-130.

Hatfield, J., Kominers, S., and Lowery, R. (2020), "Collusion in Brokered Markets," Working paper, Harvard University.

Hendel, I., Nevo, A., and Ortalo-Magne, F. (2009), "The Relative Performance of Real Estate Marketing Platforms: MLS versus FSBOMadison.com." *American Economic Review*, *99*, pp.1878-98.

Hsieh, C. and Moretti, E. (2003), "Can Free Entry Be Inefficient? Fixed Commissions and Social Waste in the Real Estate industry," *Journal of Political Economy*, *111(5)*, pp. 1076-1122.

Hubbard, T. (1998), "An Empirical Examination of Moral Hazard in the Vehicle Inspection Market," *RAND Journal of Economics*, *29*, pp. 406-426.

Jiang, W., Nelson, A., and Vytlacil, E. (2014), "Liar's Loan? Effects of Origination Channel and Information Falsification on Mortgage Deliquency," *Review of Economics and Statistics*, *96(1)*, pp. 1-18.

Levitt, S., and Syverson, C. (2008a), "Market Distortions When Agents Are Better Informed: The Value of Information in Real Estate," *Review of Economics and Statistics*, *90(4)*, 599-611.

Levitt, S., and Syverson, C. (2008b), "Antitrust Implications of Home Seller Outcomes when using Flat-Fee Real Estate Agents" *Brookings-Wharton Papers on Urban Affairs.*

Mehran, H. and Stulz R. (2007), "The Economics of Conflicts of Interest in Financial Institutions," *Journal of Financial Economics*, 85 (2), 267-96.

Rutherford, R., Springer, T., and Yavas, A. (2005), "Conflicts between Principals and Agents: Evidence from Residential Brokerage," *Journal of Financial Economics*, 76 (3), 627-65.

Woodward, S. and Hall, B. (2012), "Diagnosing Consumer Confusion and Sub-Optimal Shopping Effort: Theory and Mortgage-Market Evidence," *American Economic Review*, *102(7)*, pp. 3249-3276.

Table 1: Summary Statistics[a]

	Sold listings only	All listings (sold or withdrawn)
	(1)	(2)
listing price	420786.12	450918.77
sales price	417297.49	
sold	1.00	0.65
cumulative days on market	79.59	102.15
number of bedrooms	1.72	1.78
number of baths	1.61	1.65
house age (years)	19.10	18.46
condo	0.91	0.90
agent-owned	0.07	0.09
observations	53913	82748

[a]The table reports the mean of each variable. Column 1 uses only sold listings, while column 2 uses all sample that includes both sold listings and withdrawn listings. All prices (listing and sale) are in 2010 dollar, deflated by the Consumer Price Index. Condo is the indicator variable for whether housing tenure is condo. Agent-owned is the indicator for whether the listing was owned by an agent-seller. Sold is the dummy for whether the listing was sold.

Table 2: Price and Intrinsic Value: Agent- vs. Client-owned[a]

	Agent-owned	Client-owned
	(1)	(2)
	A. Sold listings only	
p^S (sales price)	386876.53	408647.61
intrinsic value	382468.65	426373.00
r^S (sales price premium)	1.036	0.985
observations	1851	23194
	B. All listings (sold or withdrawn)	
p^L (listing price)	418002.76	433204.71
intrinsic value	398480.93	435963.79
r^L (listing price premium)	1.065	1.014
observations	2892	31469

[a]The table reports the mean of each variable among the sample for which intrinsic values can be computed. Panel A uses only sold listings. Panel B uses all sample that includes both sold listings and withdrawn listings. All prices and intrinsic values are in 2010 dollar, deflated by the Consumer Price Index. Listing price premium, r^L, is equal to listing price/intrinsic value, and sales price premium, r^S, is equal to sales price/intrinsic value.

Table 3: Regression Results[a]

dependent variable	$\ln(p^S)$	$\ln(p^S)$	r^S	r^S	r^S	$\ln(p^L)$	r^L
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
agent-owned	0.0370*	0.0073	0.0509**	0.0520**	0.0042*	0.0263**	0.0518**
	(0.0141)	(0.0047)	(0.0093)	(0.0097)	(0.0020)	(0.0090)	(0.0096)
ln(listing price)		0.8594**					
		(0.0351)					
r^L					0.9202**		
					(0.0045)		
year×month fixed effects	yes	yes	yes	yes	yes	yes	yes
house fixed effects	yes	yes	no	no	no	yes	no
house characteristics	no	no	yes	yes	yes	no	yes
census track fixed effects	no	no	yes	no	no	no	no
zip code fixed effects	no	no	no	yes	yes	no	yes
observations	14433	14433	25045	25045	25045	29761	34361
R^2	0.984	0.996	0.088	0.084	0.932	0.981	0.088

[a]The table reports the key coefficient estimates from regressions of different price variables. The dependent variable is the log of sales price, $\ln(p^S)$, in columns 1-2; the sales price premium, r^S, in columns 3-5; the log of listing price, $\ln(p^L)$, in column 6; and the listing price premium, r^L, in column 7. House characteristics include #bedrooms, #rooms, #bathrooms, #garages, and various dummy variables for property types, basement types, and house ages. Robust standard errors are in parentheses and clustered at the census tract level. + denotes significance at a 10% level, * denotes significance at a 5% level, and ** denotes significance at 1% level.

Table 4: Instrumental Variable Regression Results[a]

dependent variable	sales price premium, r^S	dummy for sold listing
	(1)	(2)
agent-owned	0.0011	-0.0937**
	(0.0024)	(0.0086)
r^L	0.9869**	0.1308**
	(0.0088)	(0.0283)
year×month fixed effects	yes	yes
house characteristics	yes	yes
zip code fixed effects	yes	yes
1st stage F-stat for instruments	64.49	77.78
Sargan's J-test (p-value)	0.377	0.222
observations	25045	34361
R^2	0.927	0.051

[a]The table reports the results from instrumental variable regressions of two key dependent variables. The instruments for r^L include each house's intrinsic value and tract-level yearly average log listing price computed, excluding the house. Robust standard errors are in parentheses and clustered at the census tract level. + denotes significance at a 10% level, * denotes significance at a 5% level, and ** denotes significance at 1% level.

Table 5: Structural Estimation Results[a]

	(1)	(2)	(3)	(4)
	A. $\ln r^L$			
client-owned	-0.0464**	-0.0464**	-0.0284**	-0.0284**
	(0.0092)	(0.0092)	(0.0078)	(0.0078)
	B. $\ln r^S$			
client-owned	-0.0471**	-0.0046*	-0.0046*	0.0051
	(0.0092)	(0.0022)	(0.0022)	(0.0032)
$\ln r^L$		0.9408**	0.9408**	1.0285**
		(0.0080)	(0.0080)	(0.0092)
	C. sold dummy			
client-owned	0.2611**	0.2280**	0.2280**	0.2686**
	(0.0359)	(0.0346)	(0.0346)	(0.0381)
$\ln r^L$		-0.7407**	-0.7407**	0.2462+
		(0.0769)	(0.0769)	(0.1463)
	D. σ_1 for $\ln r^L$			
client-owned	-0.0058	-0.0058	-0.0035	-0.0066
	(0.0047)	(0.0047)	(0.0039)	(0.0047)
constant	0.1841**	0.1841**	0.1635**	0.1677**
	(0.0077)	(0.0077)	(0.0061)	(0.0070)
	E. σ_2 for $\ln r^S$			
client-owned	-0.0045	-0.0016	-0.0016	-0.0048
	(0.0062)	(0.0029)	(0.0029)	(0.0053)
constant	0.1756**	0.0505**	0.0505**	0.1593**
	(0.0082)	(0.0027)	(0.0027)	(0.0077)
	F. correlation coefficients: ρ			
$\rho_{1,2}$				0.9366**
				(0.0092)
$\rho_{1,3}$				-0.2311**
				(0.0233)
$\rho_{2,3}$				0.0254*
				(0.0108)
year×month FE in equations A-C	yes	yes	yes	yes
house characteristics in equations A-C	yes	yes	yes	yes
zip code FE in equations A-C	yes	yes	yes	yes
instruments for $\ln r^L$ in equations B and C	no	no	yes	yes
observations	34361	34361	34361	34361

[a]The table reports the key coefficient estimates from the structural estimation, using the same observations in column 7 of Table 3. Column 1 excludes $\ln r^L$ in the equations for $\ln r^S$ (Panel B) and s_j (Panel C), and both columns 1-2 do not use instruments for $\ln r^L$. Columns 3-4 include the same instruments used in Table 4. Only column 4 allows the correlation between the error terms of the three endogenous variables. Robust standard errors are in parentheses and clustered at the census tract level. + denotes significance at a 10% level, * denotes significance at a 5% level, and ** denotes significance at 1% level.

Table 6: Model Fit[a]

	Predicted				Observed
	(1)	(2)	(3)	(4)	(5)
	A. $\exp\left[E(\ln r^S)\right]$				
Client-owned and Sold listings only	1.0410	0.9888	0.9888	0.9652	0.9690
Agent-owned and Sold listings only	1.1226	1.0476	1.0476	1.0120	1.0191
	B. $\exp\left[E(\ln r^L)\right]$				
Client-owned and Sold listings only	0.9948	0.9948	0.9959	0.9802	0.9848
Agent-owned and Sold listings only	1.0419	1.0419	1.0434	1.0212	1.0294
Client-owned and Withdrawn listings only	1.0000	1.0000	0.9968	1.0409	1.0285
Agent-owned and Withdrawn listings only	1.0488	1.0488	1.0460	1.0860	1.0715

[a]The table reports $\exp\left[E(\ln r_j^L|d_j, s_j)\right]$ and $\exp\left[E(\ln r_j^S|d_j, s_j)\right]$, where d_j is the dummy for client-owned listings, and s_j is the dummy for sold listings. Columns 1-4 report the predicted values from the estimated models, where each column in this table corresponds to the same column in Table 5. The conditional expectation of log price premiums is computed by using the property of the bivariate normal distribution and the inverse Mills ratio. For comparison, column 5 presents the conditional mean values of observed price premiums. To be comparable with those in columns 1-4, I compute the conditional mean values of log price premiums, and then report their exponential values in column 5.

Table 7: Counterfactual Expected Values[a]

	Predicted	Counterfactual $\theta = 0$	Difference in price premium b/w agent- vs. client-owned due to agents' influence in r^L
	(1)	(2)	(3)
	A. $\exp\left[E(\ln r^S)\right]$		
Client-owned and Sold	0.9652	0.9938	61.11%
	B. $\exp\left[E(\ln r^L)\right]$		
Client-owned and Sold	0.9802	1.0085	69.02%
Client-owned and Withdrawn	1.0409	1.0710	66.74%

[a]The table reports $\exp\left[E(\ln r_j^L|d_j = 1, s_j)\right]$ and $\exp\left[E(\ln r_j^S|d_j = 1, s_j)\right]$, where d_j is the dummy for client-owned listings, and s_j is the dummy for sold listings. Column 1 is the same as column 4 of Table 6. Column 2 present the results under a counterfactual scenario where agents' influence in listing prices is removed by setting $\theta = 0$ in (10). Column 3 reports the difference between the counterfactual price premium in column 2 and the predicted price premium in column 1, relative to the difference in price premiums between agent-owned, vs. client-owned listings in column 4 of Table 6.