# Heterogeneous Real Estate Agents and the Housing Cycle [*]

Sonia Gilbukh[†] Paul Goldsmith-Pinkham[‡]

November 10, 2017

JOB MARKET PAPER

CLICK HERE FOR THE LATEST VERSION

**Abstract**

The real estate market is highly intermediated, with 90% of buyers and sellers hiring an agent to help them transact a house. However, formal training to become an agent is short, and agents primarily learn on the job. Low entry barriers and fixed commission rates result in a market where inexperienced intermediaries have a large market share, especially during and after boom periods. Using rich micro-level data on 10.4 million listings, we first show that seller agents' experience is an important determinant of client outcomes, particularly during real estate busts. Houses listed for sale by inexperienced agents spend more time on the market and have a lower probability of selling. We then study the aggregate implications of the experience distribution on liquidity of the real estate market by building a theoretical entry and exit model of real estate agents with aggregate shocks. Several policies that raise the barriers to entry for agents are considered: 1) increased entry costs; 2) lower commission rates; and 3) more informed clients. Across each counterfactual, increasing barriers to entry shift the distribution of agents across experience to the right, improves liquidity, and reduces the amplitude of liquidity cycles in the housing market.

[*]First version: March 15, 2017. This version: November 10, 2017. We thank Luis Cabral, Stijn Van Nieuwerburgh, Petra Moser, David Backus and John Lazarev for their invaluable support and guidance. Hanna Halaburda, Virgiliu Midrigan, Lawrence White, Alessandro Lizzeri, Boyan Jovanovic, Thomas Sargent, Jarda Borovicka, Pau Roldan, Julian Kozlowski, Diego Dariuch, Andrew Haughwout, Gara Afonso and Argia Sbordone provided helpful discussions and suggestions. We also greatly benefited from comments by Joe Tracy, David Luca, Jacob Wallace, Nic Kozeniauskas, Jihye Jeon and Vadim Elenev. In addition, we thank seminar participants at NYU Stern in Macro, IO and CREFR groups for valuable comments and suggestions. Sonia Gilbukh gratefully acknowledges the hospitality of the New York Federal Reserve Bank of New York where she spent the summer of 2017. The views expressed are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of New York or xbother Federal Reserve System.

[†]New York University. Email: sophia.gilbukh@gmail.com

[‡]Federal Reserve Bank of New York. Email: paulgp@gmail.com

# 1 Introduction

The U.S. housing market is subject to strong boom-bust cycles. The Great Recession provides a particularly severe illustration: from 2006 to 2008, house prices dropped by 10 percent[1], while the likelihood of a listed house selling within a year dropped by 20 percent to 41.3 percentage points[2]. Despite a large literature studying the significance of expectations, financial conditions and other frictions in generating and amplifying the house cycle[3], few studies have focused on a prominent feature of this market: real estate agents. This paper studies the effect of entry, experience accumulation, and exit by real estate agents on housing market dynamics. Combining micro-level empirical evidence and a dynamic model of entry and exit, we show that the presence of inexperienced agents led to reduced liquidity.

Using a rich micro-level data set on 10.4 million transactions in 60 different Multiple Listing Service (MLS) platforms over the 2001-2014 period, we present two sets of facts. First, an agent's work experience is highly predictive of how successfully and quickly they are able to sell homes. All else equal, listings with agents in the 10th percentile of experience have 8 percentage point lower probability of sale than those listed by agents in the 90th percentile. The difference goes up to 12 percentage points in the bust.

Second, due to low entry barriers and fixed percentage commission rates, boom years are accompanied by a significant inflow of inexperienced agents attracted by high housing prices. As the economy shifts from boom to bust, inexperienced agents often remain active. Put together, these facts imply cyclical pressures on housing liquidity. Illiquidity pressures were particularly strong at the onset of the bust, as the market swelled with new agents and when experience mattered most for the probability of sale.

In the recent bust, when illiquidity pressures were strongest, foreclosures increased the cost of lower sale probabilities. Since the housing market collapse coincided with a broader economic downturn, many homeowners struggled to pay their mortgages and attempted to sell their home. Those who failed to sell and fell delinquent on their loan payments were forced into foreclosure. Listed homes that failed to sell in this period had a six percent chance of going into foreclosure in the next two years as compared to one percent for sold properties. Moreover, houses that list in 2008 with

[1]Source: Case Shiller house price index.

[2]Source: Core Logic Multiple Listing Service Database.

[3]Favilukis, Ludvigson, and Van Nieuwerburgh (2017) is the first quantitative paper to illustrate the role of relaxing financial constraints on house prices. For more papers, see Davis, Van Nieuwerburgh et al. (2015) and Guerrieri and Uhlig (2016) for literature review on financial frictions and the housing cycle. Among many papers exploring search and information frictions in this market are Hong and Stein (1999), Anenberg (2016), Head, Lloyd-Ellis, and Sun (2014), and Guren (2016).

inexperienced agents are 2 percentage points more likely to subsequently foreclose compared to those listed with experienced ones. Thus, not only did the inexperienced agents affect individual sale outcomes, but through foreclosures, also had negative externalities on the neighboring properties.[4]

A back-of-the-envelope calculation from our regression results estimates that sales volume would increase 11% if all agents were in the top tercile of the experience distribution; in addition, as many as 20% of foreclosures would have been avoided. This counterfactual ignores the fact that experience accumulation is endogenous, and relies on agents' entry and exit decisions and experience accumulation.

To fully incorporate the dynamic decisions of agents and then vary their incentives to enter, exit, and build experience, we build a dynamic entry and exit model of real estate intermediaries in an economy with aggregate shocks. Each period, homogeneous sellers and buyers enter the market and pair with real estate agents. Some clients look for an agent at random, while the rest seek a recommendation. This implies that each agent is approached by a number of clients (sellers and buyers) that is an increasing function of experience. Next, agents attempt to match buyers and sellers. We assume experience matters for agents' matching ability. Once search outcomes are realized, agents earn commissions on successful sales. Finally, at the end of the period, agents draw a continuation cost of operating from a known distribution and decide whether to remain active or exit. Thus, agents' experience will play an important role in two ways: higher experience generates more clients, and increases the probability of a transaction for each client.

We assume free entry of agents. The compensation scheme, entry costs, and overall market competition (i.e. the distribution of agents by experience level) are all pay-off relevant variables on which real estate agents base exit and entry decisions. This induces a large state space. To solve for the optimal policies, we adopt an oblivious equilibrium concept, introduced in Weintraub, Benkard, and Van Roy (2010). In this equilibrium, agents do not perfectly observe the entire distribution of experience, but instead approximate it.

Our setup includes three aggregate states - bust, boom, and medium - corresponding to different levels of price growth. We match probabilities for each experience group in the three aggregate states, average entry rates, average exit rate and the average experience accumulation by each experience level.

We consider several policies to alter the distribution of agent experience in the market: 1) lower

[4]A body of papers have documented the externalities imposed by foreclosures on local housing markets, including xb Lin, Rosenblatt, and Yao (2009), Campbell, Giglio, and Pathak (2011), Mian, Sufi, and Trebbi (2015), and Gupta (2016).

commission rates; 2) increasing entry cost; and 3) informing the clients about agent experience so that they are more likely to sort into highly-qualified intermediaries. We find all three policies result in a rightward shift in distribution of experience, although this takes place through different channels.

Reducing commission rates makes entry less profitable, decreasing overall entry rates. It also lowers profitability of all agents in the market, thus increasing exit rates for all levels of experience. While increased exit leads to undesirable knowledge loss, this loss is compensated by much faster accumulation of knowledge among existing agents as they make up for reduced commission by working with more listings.

Increasing entry costs also has a negative effect on entry rates. Free entry condition implies that to compensate for increased entry costs, new agents have to work with more clients to earn more profit. As a result entrants learn faster and the more experienced agents learn slower, as their experience share is reduced and overall level of experience increasing in the market.

Informing clients about the importance of agent experience makes it harder for new agents to accumulate experience. As entry becomes less profitable, fewer agents enter thus reducing the overall competition effect. On net, this policy results in fewer entrants and less exit.

### 1.1 Related Literature

Our paper most closely relates to Barwick and Pathak (2015). They study similar data from the Boston area for years 1998-2007 and examine the effect of cheap entry on the probability of sale of listed houses. An important distinction is that we model agent learning as an endogenous process, allowing for differences in experience accumulation across aggregate states and for different overall competition levels. By explicitly modeling this channel, we can measure the learning externality that entering agents impose on other intermediaries. In addition, our data extends through 2014, allowing us to explore the recent housing bust. Hsieh and Moretti (2003) and Han and Hong (2011) also study the effect of cheap entry on market efficiency, specifically focusing on the business stealing externality and abstracting from experience all together.

Our paper contributes to a large literature on the value of real estate agents. Hendel, Nevo, and Ortalo-Magné (2009) compare listing outcomes from an FSBO (for sale by owner) platform to those that were facilitated by an agent. They find that agents provide little value added. Levitt and Syverson (2008) find that agents are able to obtain a better price when they are selling their own homes, rather than those of their clients. These papers abstract from agent heterogeneity, which we argue can have a significant impact on value added for a client.

Finally, the paper adds to a large literature on heterogeneous firm dynamics across the business cycle. Among the papers in this literature are Hopenhayn (1992), Lee and Mukoyama (2015) and Clementi and Palazzo (2016).

## 2 Background and Data

### 2.1 Role of Real Estate Agents

The housing market in the U.S. is highly intermediated with 87% of buyers and 89% of sellers[5] choosing to hire an agent to facilitate selling or buying a home. Among many advantages of working with an agent is their access to Multiple Listing Service database which provides buyers with detailed up-to-date information on all the listings available in the area and allows sellers to list their property on this platform. In essence, an agent has access to a more efficient matching technology. Hiring an agent also gives a client representation in a negotiation process in the final stages of the transaction. Arguably the most valuable role of an agent, however, is that of an adviser. Thus, a listing agent would suggest an appropriate list price for seller's property and advise on improvements or "staging" that may make it more attractive to buyers. Likewise, a buyer agent can advise on whether the asking price is a reasonable one to consider and what details of a property to pay attention to in the process.

Given the importance of intermediaries in facilitating perhaps the single most important transaction in their clients' lives, it is perhaps surprising that in some states one could start on the job after as little as 30 hours of classes and a 50$ exam fee.[6] While the courses familiarize agents with essential terminology and state laws, they do not provide any insight into market conditions for a particular area that an agent is operating in. It is natural then that agents have a lot of room to improve upon entry. In addition to learning about particulars of the area, experience also provides agents with an accumulated network of other agents as well as other professionals that a client might need to be in contact with throughout the process - construction workers, mortgage brokers and appraisers. Curiously we do not find any evidence that clients pay lower commissions to inexperienced agents as compensation for their relative disadvantage as compared to experienced agents[7].

[5] Source: 2015 National Association of REALTORS Profile of Home Buyers and Sellers

[6] The requirements vary somewhat across states with class time ranging from 30-50 hours. I am in the process of collecting 2017 entry data for all relevant states.

[7] While we do not observe listing agent commissions in our data, in some areas we see commission rates offered to a buyer agent as part of the listing description. We find that those rates are near uniform. Since seller agents and buyer agents

In addition to accumulated on-the-job expertise through familiarity with the process and building of professional network, experience might also reflect intrinsic ability of an agent to successfully work with a client. Thus, we might think that all agents enter with a different talent for the job and only those who are most talented stay for long enough to gain experience. For now, we abstract from distinguishing learning and ability as measured by experience, but we will come back to discussing the two channels when we think about policy analysis.

### 2.2 Data

For our empirical analysis we use a comperehensive listing level data set on residential properties for sale and rent collected by CoreLogic. The data come from real estate boards, organizations of real estate agents who each operate a Multiple Listing Service (MLS) system: a platform for advertising listings available to member agents only. Each MLS covers a geographical area that is approximately equal to a commuting zone. An observation includes a large number of fields describing the property and the listing status. In particular, we know the day the property entered the market, the associated listing agent (as well as secondary agent in some cases), the original asking price, as well as the last observed asking price, property features (e.g. living area, number of bedrooms and bathrooms, number of parking spaces, age of the structure, etc.). If the listing sells we observe the close date, close price and the associated buyer agent.

The full Corelogic MLS Dataset has information on over 90 MLS boards. However, the history for each MLS begins at different times, due to variation in contracts with each board, with some MLS data beginning as late as 2009. Since we are interested in studying the full boom and bust period starting in 2001, we focus on the subsample of MLS boards whose data begin in 2001. Due to data concerns, we drop several MLS boards who state their data begins in 2001 or earlier, but experience large jumps in the number of listings during the sample period from 2001-2014 (more than 100% growth in the number of listings in a given year). This drops an additional 10, and leaves our sample with 60 MLS boards. We further exclude listings with asking prices below 1000$. This leaves us with 10.4 million observations. Figure 1 shows the coverage map of the sample.

To document the heterogeneity in agents and it's effect on listing outcomes we need a measure of experience. There is no consensus on the right measure, so we explore a few available to us with the data on hand. Our prefered measure if the number of clients an agent had in the previous year.

often split the commission in half (sometimes mandated by the state), this finding suggests that the total commission is also uniform

We proxy the number of clients by the number of listings originated by the agent in that year and the number of buyers represented by this agent in a transaction that closed in that year. Thus, we measure experience in terms of recent output rather than calendar time since entry, and with strong discounting so that any clients served two or more years prior do not count towards current year's experience. In addition, this measure assumes that all clients contribute to the experience level equally, no matter the outcome of the listing, so that both unsold and sold properties count towards the listing agent experience. The obvious alternatives would be to weight listings that did not sell differently from those that sold (or at the extreme to ignore the unsold ones alltogether), or discount the listings in a different way, so that, for example, older clients matter less than the more recent ones. In addition we could consider years since entry for agents that we observe entering in our data. We have experimented with a few of these measures and found that they did not affect our results (see Appendix B for the discussion). Over the sample period we observe 569148 different agents, on average 175458 active in each year. Table 1 presents summary statistics for each year.

**Table 1:** Summary Statistics

| Year | No. Agents | No. Listing | Frac. Sold | Sale Price$ | Mean Exp. | Med. Exp. |
|---|---|---|---|---|---|---|
| 2001 | 130748 | 686712 | 0.724 | 182222 | . | . |
| 2002 | 151178 | 738907 | 0.714 | 199310 | 14.66 | 11 |
| 2003 | 164444 | 714716 | 0.710 | 221748 | 16.12 | 13 |
| 2004 | 185931 | 768608 | 0.707 | 249651 | 15.98 | 12 |
| 2005 | 219225 | 858817 | 0.661 | 268164 | 14.66 | 11 |
| 2006 | 228993 | 927915 | 0.533 | 262682 | 15.27 | 11 |
| 2007 | 222064 | 903346 | 0.463 | 248521 | 15.57 | 12 |
| 2008 | 197850 | 773629 | 0.477 | 216137 | 16.11 | 12 |
| 2009 | 183682 | 684086 | 0.566 | 205673 | 15.04 | 11 |
| 2010 | 175564 | 689557 | 0.544 | 209144 | 16.1 | 12 |
| 2011 | 166234 | 630017 | 0.610 | 207908 | 16.54 | 13 |
| 2012 | 168496 | 639608 | 0.688 | 223798 | 16.26 | 13 |
| 2013 | 177652 | 705007 | 0.704 | 244036 | 16.45 | 13 |

**Note: This table summarizes the main statistics in our data. For each year, shown here are number of distinct active agents, number of listings originating in that year, fraction of those listings that sold within 360 days, the average sale price as well as the mean experience and the average experience of the listing agents.**

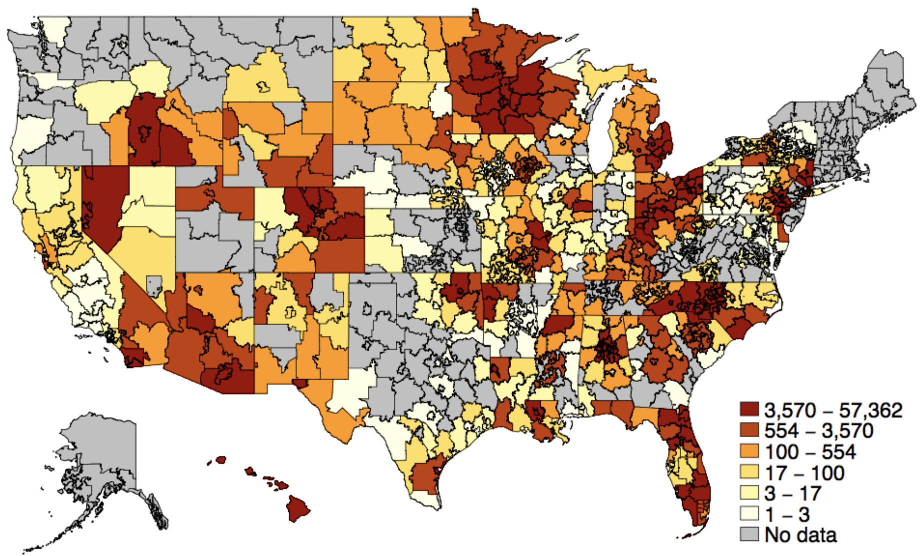
Figure 2(A) shows the distribution of active agents for our preferred measure of experience. The 10th, 25th, 50th, 75th and 90th percentiles correspond to having 0, 0, 3, 9 and 18 clients in the past year. Inexperienced agents are ubiquitous in the market and the industry has been concerned about implications of this phenomenon. In 2015 NAR commissioned a study called "Danger Report" for the purpose of identifying the most threatening challenges for agents, brokers and the market as a whole. For real estate agents the number one concern was found to be "Masses of Marginal Agents Destroy Reputation". Another report, by Inman, one of the more reputable publication in the industry, describes a survey of professionals in the industry. To the question "What are the challenges that the real estate industry is currently facing?", 77% responded "Low-quality agents".[8]

To assess how relevant the experience distribution is for consumers, Figure 2(B) plots the cumulative distribution of listings against our experience measure. Thus, the 25% percentile of listings are handled by agents who had 4 or less clients in the past year, 50% are listed with agents with experience 12 or less, and 75% with experience 24 or less. As we can see, agents with little experience hold a considerable market share and so have a potential for a non-trivial impact on the housing market.

In addition to CoreLogic dataset, our robustness checks make use of the Zillow's zipcode level house appreciation estimates by house price tiers, as well as the Deeds data of county records on property sales and refinancing.

[8]For more information about the Danger Report refer to their website https://www.dangerreport.com/usa/. The Inman report can be downloaded here: /https://www.inman.com/2015/08/13/special-report-why-and-how-real-estate-needs-to-clean-house/

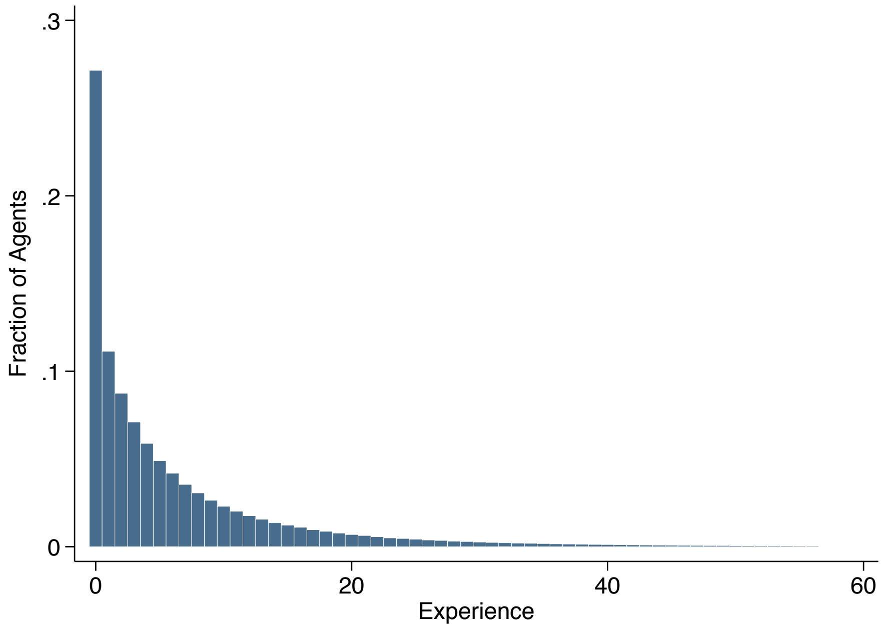
**Figure 1:** Coverage



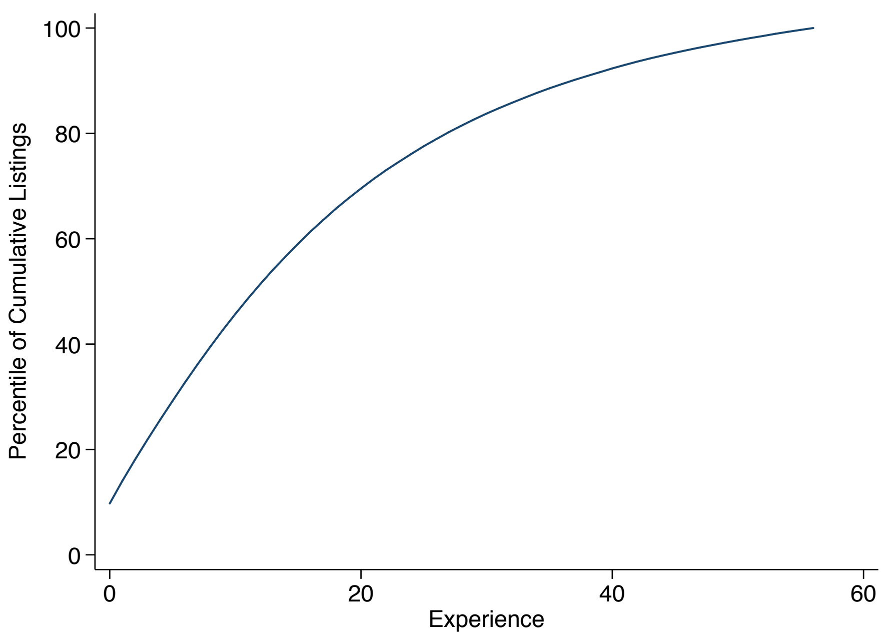

**Note: Above is a heat map representing the spacial coverage of the CoreLogic data sample used in the paper.**

**Figure 2:** Distribution of Experience and Listing Share

(A) Distribution of Experience




(B) Listing Share




**Note: Panel A plots the distribution of experience of active agents in our data for all years. We measure experience as the number of clients an agent had in the previous calendar year. For each experience level, in Panel B, we plot the percentage of listings that were handles by agents of that level of experience or less.**

# 3 Empirical Results

## 3.1 Entry and Exit Patterns

Figure 3(A) plots aggregate entry and exit rates for real estate agents in the U.S., where entry rate is the percentage of currently active agents who were not active in our data set in the previous two years and, similarly, exit rate is the percentage of currently active agents that we do not observe as active in the following two years[9]. The churn in this market is substantial with more than a quarter of all active agents being new in the boom years, and as much as 17% subsequently exiting each year. As the recession hit, the fraction of entrant decreased to around 17%, still a substantial amount. Fraction of exiting agents peaked post 2007 at around 25% and subsequently came down to meet the entry rate around 18% [10]. As expected, agents who are less invested in the profession tend exit with higher probability. In Figure 3(B) we plot average exit rates for each experience level. Entrants and very low experience agents have a particularly high exit rate at above 35%. It come down to around 19% for a median agent and falls to around 5% for agents with high experience levels.

The overall entry and exit rates fluctuate substantially over time. We will next examine how those fluctuations relate to aggregate market conditions. To do that, we assign each agent to a home market (fips code area) in which they have the most activity. We define entry rate in a particular fips as the fraction of agents currently active and assigned to the fips code who we do not observed in our data (including in other fips codes) in the previous two years. Similarly, exit rate is the fraction of agent who are currently active and assigned to the fips code and who we do not observe in following two years. Table 2 summarizes the number of fips codes in the data, as well as the mean and standard deviation of number of active agents, exit rates and entry rates in each fips code. We observe from 663 to 869 distinct fips codes per year that vary substantially in the number of active agents and entry and exit rates.

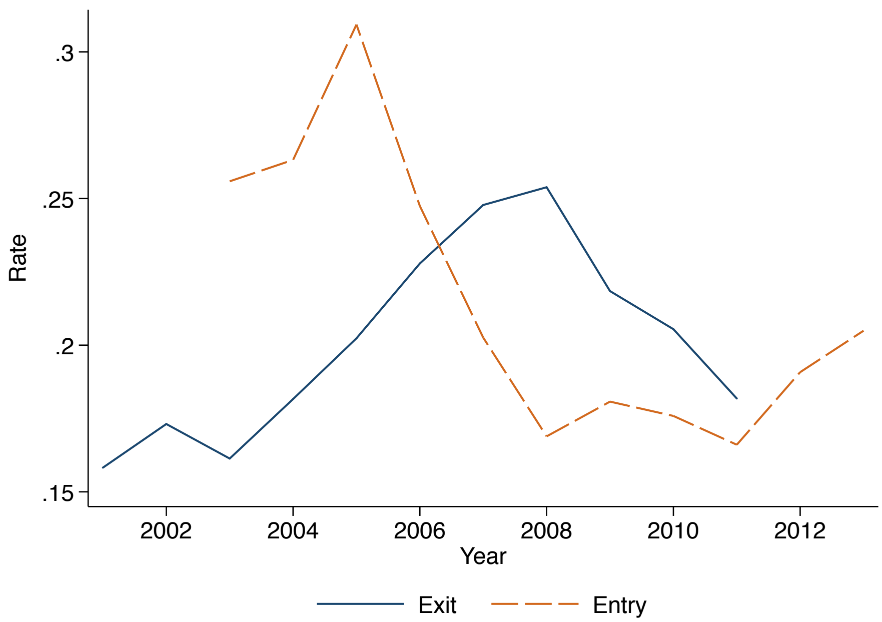
Since real estate agents are compensated with a fixed percentage rate of a sale price, agent earnings are directly related to housing market conditions such as sales volume, house prices, and the ease with which transactions are made. These conditions likely affect agents decision to stay or exit the market. Figure 4 plots the relationship of entry rates to inventory to sale ratio, as well as to percent change in

[9] See appendix A for alternative definitions of entry and exit.

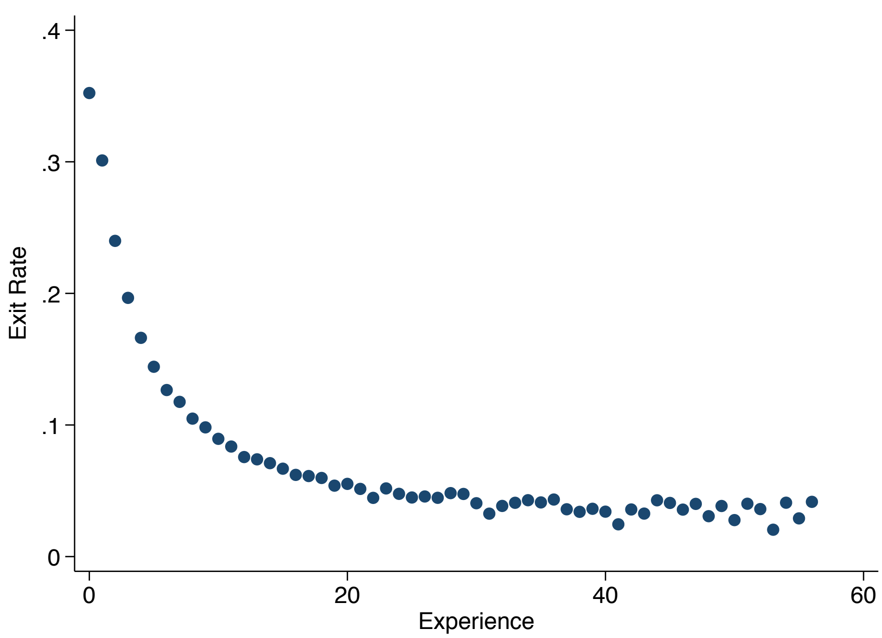
[10] By comparison, both entry and exit rates of the establishments in the U.S.range between 8-12% in the same time period (2000-2015) as reported by the U.S. Census Bureau's Business Dynamics Statistics (BDS), defined by fraction of establishments with positive employment who had/will have 0 employment in the previous/following year. A similar definition for agents (1 year window) delivers an even larger churn than is described in this section (see Appendix A).

**Figure 3:** Entry and Exit Rates

(A) Entry and Exit




(B) Exit by Experience




**Note: Panel A plots entry and exit rates among currently active agents. An active agent is someone who has at least one listing originating in the current year or is marked as a buyer agent for at least one sale in the current year. We define entrant to be agents who are active in the current year, but were not active in the previous two calendar years. Similarly, exiting agents are those we observe active in the current year and inactive in the following two calendar years. Panel B plots average exit rates by each experience level.**

house prices (using sale prices for transactions in that year) and volume (number of listings originated in that year). In the binscatter plots we take out fips code fix effects and control for calendar year. That way we are picking up only variation within a particular fips code that is beyond the aggregate changes in the country. The results are formally summarized in Table 3. Entry rate is very responsive to changes in sales volume and sale price and slightly less responsive to inventory to sale ratio. A 10% increase in sale volume, sale prices and sale to listings ratio increases entry rate by .6, .8 and .4 percentage points respectively.

We do the same exercise for exit rates in Figure 5. We find that exit rates are responsive to all three aggregate variables, but is relatively more sensitive to changes in inventory to sale ratio and the overall volume of sellers. A 10 percentage increase in inventory to sale ratio reduces exit rates by 0.8 percentage points. A 10% increase in sale prices has a small and insignificant effect. In turn, a 10% increase in listing volume reduces exit rate by 0.5 percent.

### 3.2 Outcomes

This section explores the effect of agent experience on listing outcomes observed in the data: sale outcome, duration on the market, and prices. The challenge for this exercise is lack of random assignment between listings and agents. Two types of selection can confound our results - one on property (or listing) characteristics, and another on client characteristics. For example, a more experienced agent might be selecting to work with more motivated clients or easy-to-sell properties. We propose several specifications to address each of these selection channels.

For all outcomes and specifications, we run a version of the following regression:

$$y_{i,t} = \beta_e log(1 + exp) + \sum_{p \in periods} \beta_{e,p} log(1 + exp) \times \mathbf{1}_{t \in p} + \boldsymbol{\delta} \boldsymbol{W}_{i,t} + \boldsymbol{\alpha}_{l(i),t} + \epsilon_{i,t} \tag{1}$$

Here $y_{i,t}$ is the listing outcome of house $i$ at time $t$. Time is defined as list month for most outcomes, except for sale price, where $t$ is the sale month. To account for non-linear relationship between outcomes and experience, we transform the explanatory variable to be exponent of one plus experience. We allow the effect of experience to vary in different time periods. Included in the regression are property specific control variables $\boldsymbol{W}_{i,t}$ and the zip code by month fixed effects $\boldsymbol{\alpha}_{l(i),t}$.

Figure 6 illustrating this regression for sale probability within 365 days with constant experience effect across time. Each dot represents an average value of the outcome variable corresponding to a 5 percent of all observations, taking out zip code by month fixed effects and controlling for house

**Table 2:** Summary Statistics

| | Fips Codes | Agents | Exit Rates | Entry Rates |
|---|---|---|---|---|
| 2002 | 663 | 225 (656) | .18 (.22) | . |
| 2003 | 713 | 228 (692) | .17 (.20) | .31 (.28) |
| 2004 | 747 | 246 (762) | .18 (.22) | .32 (.28) |
| 2005 | 808 | 266 (845) | .20 (.23) | .35 (.28) |
| 2006 | 851 | 263 (832) | .24 (.26) | .30 (.27) |
| 2007 | 853 | 254 (772) | .26 (.25) | .27 (.27) |
| 2008 | 857 | 225 (683) | .26 (.25) | .20 (.24) |
| 2009 | 858 | 209 (656) | .23 (.25) | .19 (.23) |
| 2010 | 851 | 201 (637) | .23 (.25) | .20 (.25) |
| 2011 | 869 | 186 (611) | .21 (0.24) | .20 (.25) |
| 2012 | 861 | 191 (632) | . | .21 (.26) |

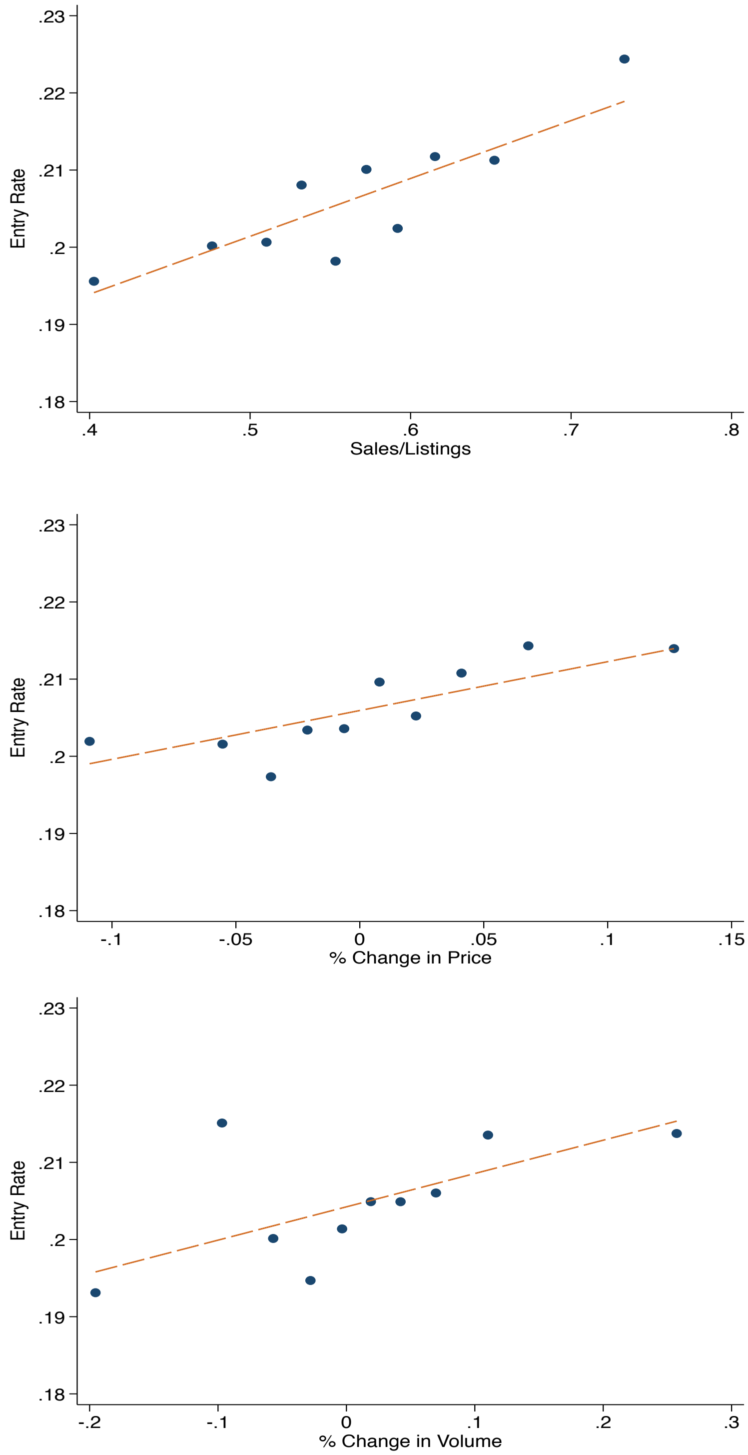
**Note: Represented here are summary statistics for our data at fips code level. For each year the first columns counts the number of distinct fips codes observed in our data, next columns report the mean and standard deviation of number of agents active, exit rates and entry rates.**

**Table 3:** Turnover Rates and Market Conditions

| | Entry | Exit |
|---|---|---|
| Δ Sales Volume | 0.0598*** | -0.0528*** |
| | (0.0157) | (0.0140) |
| Δ Sales Price | 0.0795*** | -0.0303 |
| | (0.0242) | (0.0206) |
| Sales/Listings | 0.0372* | -0.0845*** |
| | (0.0221) | (0.0167) |
| $R^2$ | 0.6885 | 0.7327 |
| Fips Effect | Yes | Yes |
| Year Effect | Yes | Yes |
| $N$ | 4904 | 4790 |

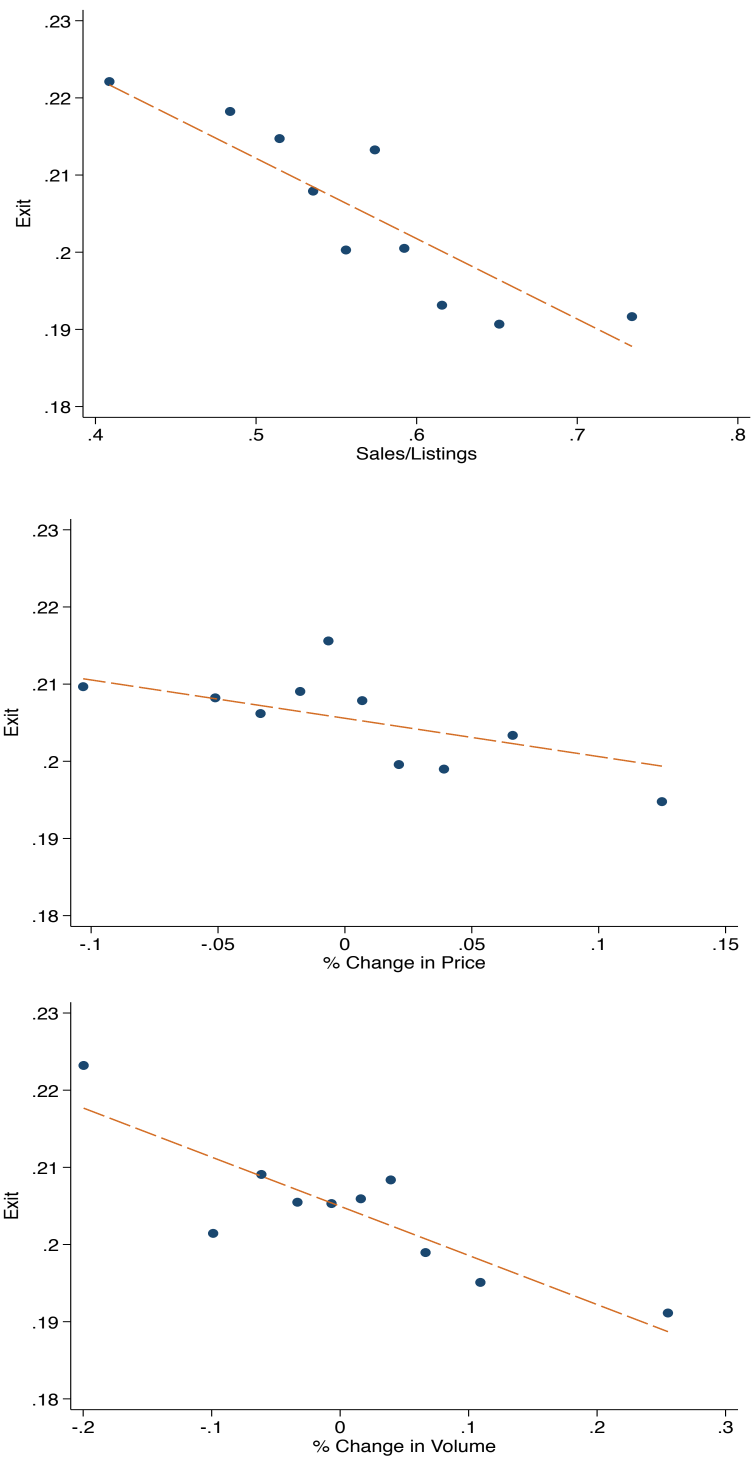
**Note: In this table we explore how turnover rates of real estate agents relates to housing market conditions. We first assign each active agent in the data to a fips code in which they have the most activity. For each fips code we then compute the fraction of those agents who are entrants and a fraction that subsequently exits. We regress entry and exit rates (unweighted by fips characteristics) on change in listing volume and close price, as well as on Sales to Listings ratio as a proxy for how hard it is to transact a property. In these two regressions we include fips code fixed effect and control for calendar year.**

**Figure 4:** Entry Rates and Market Aggregates




**Note: The three figures in this graph represent the relationship between real estate agent entry rate and market conditions. We assign each agent a fips code in which they are most active in a particular year. Fips code specific entry rate is the fraction of the currently active agents assigned to a fips code, who do not appear in our data set in the previous two years. The aggregate market variables considered here are sale to inventory ratio (a signal of how easy it is to sell a property), percentage change in house prices from the previous year, and the percentage change in volume of listings originated in the fips code in that year. The three binscatter graphs include fips code fixed effects and controls for calendar year.**

**Figure 5:** Exit Rates and Market Aggregates

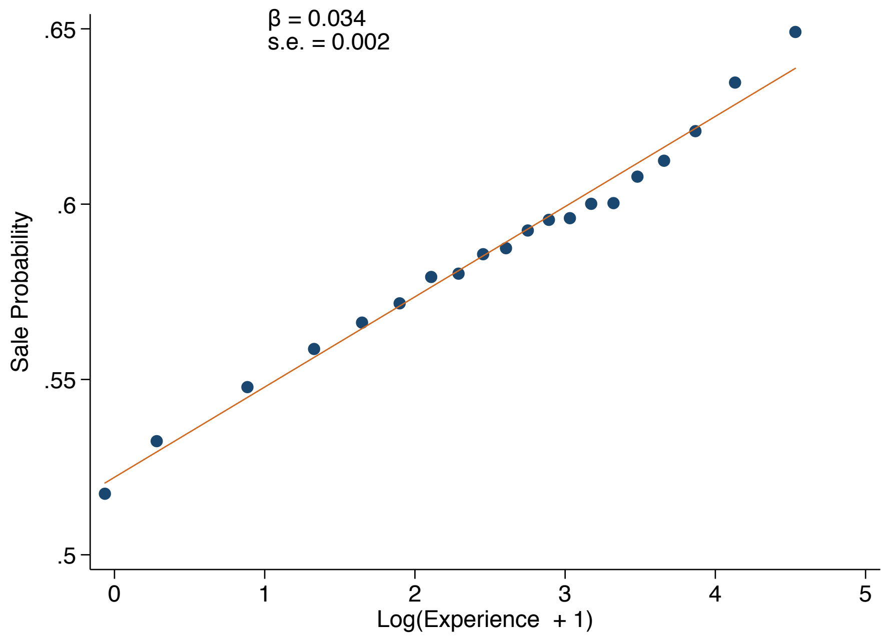



**

**Note: The three figures in this graph represent the relationship between real estate agent exit rate and market conditions. For this exercise we assign each agent a fips code in which they are most active in a particular year. We compute fips code specific exit rates: the fraction of the currently active agents assigned to a fips code, who do not appear in our data set in the following two years. The aggregate market variables considered here are sale to inventory ratio (a signal of how easy it is to sell a property), percentage change in house prices from the previous year, and the percentage change in volume of listings originated in the fips code in that year. The three binscatter graphs include fips code fixed effects and controls for calendar year.**

characteristics. The relationship is strikingly linear with bottom 5th percentile of listings selling almost 15 percentage points less likely compared to the top 95th percentile.

**Figure 6:** Experience Advantage in Sale

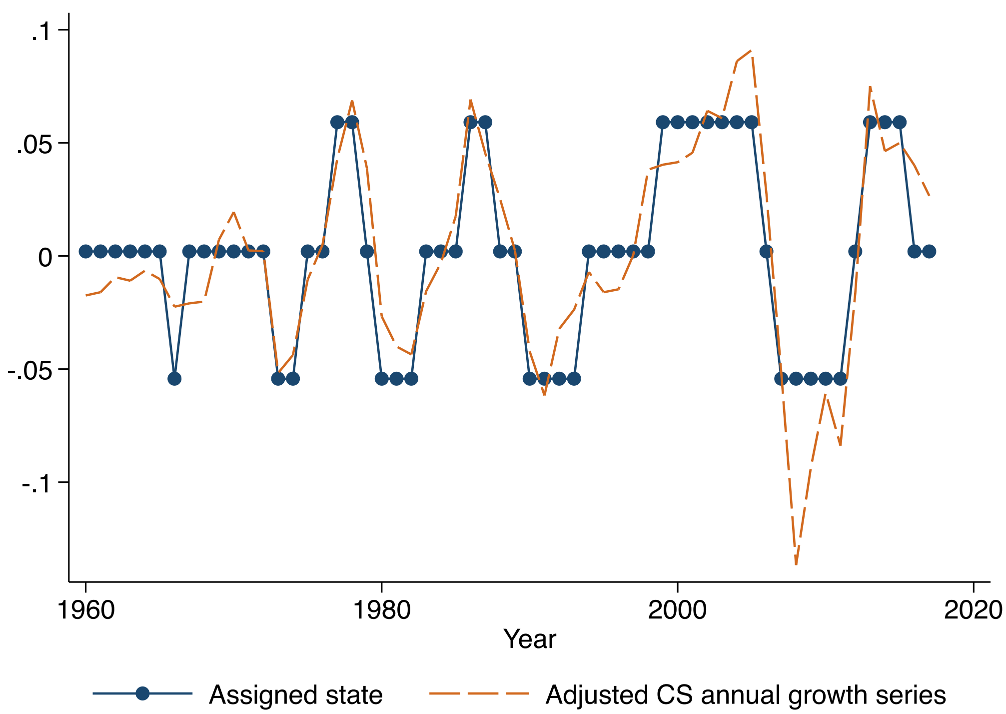



**Note: This figure shows a binscatter plot of the sale outcome against our experience measure log(experience+1). Taken out here are zip code by list month fixed effect as well as controls for housing characteristics.**

In the full empirical analysis the effect of experience is allowed to vary by three different housing market conditions: boom, medium and bust. The assignment of each year to one of the three periods is based on the 12 month house price growth in Case Shiller index deflated by CPI less shelter. Years with average growth rates above 75th percentile are identified as booms, those below 25th percentile are busts, and those in between are assigned to a medium period. Figure 7 illustrates this assignment procedure.

Table 4 presents the results. In the first column we show regression results that do not include controls for house characteristics and where list month and zip code fixed effects are taken separately. Second column shows the same regression, with zip code by list month fixed effects. The third specification adds the following controls for property characteristics: number of bedrooms, bathrooms, garages, living area, type of cooling system, an indicator for whether it is waterfront, and whether is has a view and a fireplace. Coefficients change somewhat from (2) to (3) indicating some selection on housing characteristics. Column 3 is our preferred specification. Doubling listing agent experience increases the probability of sale by 2.8 percentage points. The experience advantage increases

**Figure 7:** Case Shiller Adjusted Series




**Note: This figure plots average yearly 12 month growth rates of the Case Shiller house price index deflated by CPI less shelter. The dots on the plot represent one of the three states assigned to each index value.**

in the bust year, where doubling experience increases probability of sale by as much as 4.1 percentage points. Listings of an agent in the 90th percentile (corresponding to 18 experience) sell with a 8.2 percentage point higher probability than listings of agents in the 10th percentile (corresponding to experience 0). In the bust this gap increases to as much as 12.1 percentage points. For comparison, the base probability of sale in the boom is around 63 and in the bust around 47 percentage points.

**Table 4:** Sale Probability

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Log(Exp+1) | 0.028*** | 0.031*** | 0.028*** | 0.035*** | 0.035*** | 0.035*** |
| | (0.002) | (0.002) | (0.003) | (0.004) | (0.004) | (0.004) |
| Bust X Log(Exp+1) | 0.015*** | 0.010*** | 0.013*** | 0.013*** | 0.013*** | 0.013*** |
| | (0.004) | (0.002) | (0.002) | (0.004) | (0.004) | (0.004) |
| Medium X Log(Exp+1) | 0.009*** | 0.002 | 0.003* | 0.002 | 0.002 | 0.002 |
| | (0.002) | (0.002) | (0.002) | (0.003) | (0.003) | (0.003) |
| Inferred Price | | | | -0.001 | | |
| | | | | (0.013) | | |
| Client Equity | | | | | 0.146*** | |
| | | | | | (0.039) | |
| $R^2$ | 0.0894 | 0.2918 | 0.3276 | 0.3991 | 0.4019 | 0.3991 |
| Time Effect | Yes | - | - | - | - | - |
| Zip Effect | Yes | - | - | - | - | - |
| Time X Zip Effect | No | Yes | Yes | Yes | Yes | Yes |
| House Characteristics | No | No | Yes | Yes | Yes | Yes |
| $N$ | 2584409 | 2584409 | 1999927 | 692677 | 692677 | 692677 |

**Note: This table displays several specification of regression outlined in equation 1. Column one includes fixed effects for zip code and list month, Column 2 instead includes zip code by list month fixed effects, Column 3 adds controls for house characteristics. In Column 4 we consider unobserved heterogeneity by computing inferred price for repeat sales (previous price appreciated using zip code and price tier specific Zillow appreciation rates). Column 5 includes a proxy for client equity (the percent appreciation since last purchase). Column 6 runs specification of Column 3 with the repeat sale sample so that it is comparable to columns 4 and 5.**

To check for selection on unobservables, column four adds a control for inferred price. For listings that we observe selling in the past, we get this value by appreciating the last observed sale price using Zillow zip code- and tier- level house price appreciation indexes. This specification cuts the number of available observations by four and biases the remaining sample towards later years. To evaluate the importance of inferred price control, we use this sub sample to run the preferred specification (3). Column six presents these results. Identical coefficients on the variables of interest indicate that unobserved heterogeneity does not play a significant role in our analysis. In another experiment deferred to the Appendix F, we instead explore whether experience has the same effect in a market where all houses are essentially identical. For this we restrict our analysis to a suburb of San Diego, where price variation in houses is less than 20%. While experience effect is insignificant in the boom, the total effect in the medium and bust periods are 4.9 and 3.5 percentage points, similar to what we find in the full sample.

Column five includes a control for client equity. Equity stakes are approximated by house appreciation rate since last purchase. Sellers with low equity stakes are typically more price sensitive, either because of loss aversion, or because they need money for the down payment on the next house they will be purchasing. These sellers might have higher reservation prices and so are more difficult to work with. In line with the prediction, clients with more equity sell with a higher probability. However this regression delivers same coefficients on experience as our preferred specification ran on the identical sample (column six). This indicates that while client equity plays a role in listing outcomes, a selection on equity for different experience agents is unlikely.

To address additional selection on seller motivation we examine a sample of listings that followed a life changing event, such as death or divorce. Specifically, we look at listings that occur within two years after a deeds record of an arms-length transaction there both people have the same last name, but a different first name. Sellers in this sample are likely more motivated in selling the property than an average seller, because they either can not afford maintaining it, or do not have use for it altogether. This sample is relatively small so location is aggregated at a fips code level rather than a zip code and fixed effects are added for fips code and list month. Table 6 presents the results for several outcome variables. For sale probability outcome (column 1), the effect of experience is almost exactly the same as in the full sample.

Table 5 describes the results for days on market and days to sale using the preferred specification. As in the sale probability, other specifications (included in the Appendix C) deliver similar coefficients to the preferred one. In addition to having a lower probability of sale, listings of inexperienced agents spend more time on the market (column 2), even conditional on sale (column 3). All else equal, listings of an agent in the 90th percentile (corresponding to 18 experience) spend 9 more days on the market in the boom than the 10th percentile (0 experience). The difference increases to 12.7 and 17 days in the medium and bust periods respectively. In average, a listing spends around 137, 153 and 179 days in boom, medium and bust states. Conditioning on sale, the 10-90th percentile differences are 6.2, 8.3, and 11.6 days for boom, medium and bust periods on the basis of average values of days to sell of 116,128, and 143 days.

**Table 5:** Experience and outcomes

| | (1)<br>Sale Probability | (2)<br>Days on Market | (3)<br>Days to Sale |
|---|---|---|---|
| Log(Exp+1) | 0.027*** | -3.063*** | -2.108*** |
| | (0.002) | (0.565) | (0.425) |
| Bust X Log(Exp+1) | 0.013*** | -2.700*** | -1.828*** |
| | (0.002) | (0.464) | (0.462) |
| Medium X Log(Exp+1) | 0.003* | -1.252*** | -0.720* |
| | (0.002) | (0.371) | (0.400) |
| $R^2$ | 0.3209 | 0.3283 | 0.3893 |
| Time X Zip Effect | Yes | Yes | Yes |
| House Characteristics | Yes | Yes | Yes |
| $N$ | 1999927 | 1957519 | 1206408 |

**Note: This table displays our preferred specification of regression outcomes in equation 1 for several variables: sale probability, days on market, and days to sale.**

**Table 6:** Experience and outcomes : Motivated sellers

| | (1)<br>Sale Probability | (2)<br>Days on Market | (3)<br>Days to Sale |
|---|---|---|---|
| Log(Exp+1) | 0.027*** | -2.311** | -0.631 |
| | (0.006) | (1.128) | (1.126) |
| Bust X Log(Exp+1) | 0.013 | -3.657** | -2.902* |
| | (0.009) | (1.548) | (1.499) |
| Medium X Log(Exp+1) | 0.019* | -3.684* | -3.763** |
| | (0.011) | (2.120) | (1.793) |
| $R^2$ | 0.136 | 0.135 | 0.150 |
| Fips Effect | Yes | Yes | Yes |
| Time Effect | Yes | Yes | Yes |
| House Characteristics | Yes | Yes | Yes |
| $N$ | 15967 | 15606 | 9590 |

**Note: A set of regressions in this tables are restricted to a sample of motivated sellers: those who have inherited the property or have likely gone through a divorce. Specifically, these listings occur within two years after a deeds record of an arms-length transaction there both people have the same last name, but a different first name. Displayed are our preferred specification of regression outcomes in equation 1 for several variables: sale probability, days on market, and days to sale.**

Next we examine the effect of experience on prices. Tables 8 presents the results. Experienced agents tend to list for lower prices than experienced ones. Comparing 90-10 percentile of listings, the difference in list price is around 2.3 percent in the boom, and as much as 3.5 and 5.9 percent in the medium and bust states respectively. If we look at close prices, the difference in the boom and medium states is not statistically significant, however in the bust more experienced agents get lower sale price on their listings conditional on sale. Comparing again the 10th and 90th percentile of agents, the gap in sale price is around 1 percent in the medium state and around 3.8 percent in the bust.

Perhaps inexperienced agents might be catering to particularly patient sellers. In that case the selection effect should be captures when equity stakes are added to the regression. However the results do not change in this specification (comparing columns 5 and 6 in Table A6). In addition, it is unlikely that the patient sellers would emerge in the bust, when the price differences across experience are largest. We also find that while low experienced agents list at higher prices, they are also more likely to lower the prices over the course of the listings, as described in column 3. Selection on sellers is also inconsistent with out results from the sample of motivated sellers. In Table **??** we find that even among sellers who have experienced death or divorce and thus are likely less patient in selling the property, the price effect of experience persists. We conclude that the selection explanation is not consistent with the data. The inexperienced brokers advise on high prices and, if lucky, sell the property for higher, but most often are forced to lower the price to market levels.

There are a few other plausible reasons why inexperienced agents might choose to list at higher prices. First, there is a trade-off between how much time a listing spends on the market and the list price. While an experienced agent might not find it worthwhile to list a property at a sub-optimally high price, an inexperienced agent has less demands on their time and so might resort to this strategy for a chance of higher commission. Second, an inexperienced agent might have higher uncertainty about what is the optimal price, and chooses to price at higher levels, rather than lower, to "feel" the market. That way they can always lower the price if demand is weak, while listing initially at a lower price makes them responsible for at least the buyer agent fee in case a buyer is found. This information friction is unlikely in more homogeneous markets, where recent sale prices are particularly informative of the current market price for a similar home. Indeed, looking at the homogeneous market of Chula Vista, San Diego, we find that the effect of experience on price is reversed, that is more experienced agents list at higher prices and are able to get better sell prices as well (Table A8). Finally, pricing might be a way for inexperienced agents to attract more clients, or as the industry calls it to"buy a listing". Faced with a choice between two agents, a seller might sign a contract with one that promises to sell their house at a higher price. While this might not be a rational choice, sellers are often susceptible to biases, and often do not have enough information to make the right decision.

**Table 7:** Experience and prices

| | Price (Log) | | |
|---|---|---|---|
| | List | Sale | Frac. Discount |
| Log(Exp+1) | -0.008*** | -0.003 | -0.011*** |
| | (0.002) | (0.002) | (0.002) |
| Bust X Log(Exp+1) | -0.012*** | -0.010*** | 0.001 |
| | (0.002) | (0.002) | (0.002) |
| Medium X Log(Exp+1) | -0.004* | -0.001 | 0.000 |
| | (0.002) | (0.001) | (0.002) |
| | | (0.001) | |
| $R^2$ | 0.870 | 0.891 | 0.344 |
| Time X Zip Effect | Yes | Yes | Yes |
| House Characteristics | Yes | Yes | Yes |
| $N$ | 1956672 | 1211627 | 1202780 |

**Note: This table displays our preferred specification of regression outlines in equation 1 for several variables: sale probability, days on market, and days to sale.**

We now examine the trade-off between prices and probability of sale. To account for heterogeneity in houses, we look at prices relative to our inferred measure of property value: the last sale price recorded for the house, appreciated by zip-code- and price tier- specific appreciation rates taken from Zillow. Figure 8 shows the fraction of houses sold within a year for each relative price point across different experience levels. First, listings with relatively lower prices are more likely to find a buyer than those with high prices. Second, for inexperienced agents, the probability of sale is kinked at a point where list price is equal to inferred price. That is, for prices higher than inferred price, an increase in the markup leads to a larger drop in probability of sale than for prices below the inferred price. Third, for any price point, more experienced agents have higher probability of sale.

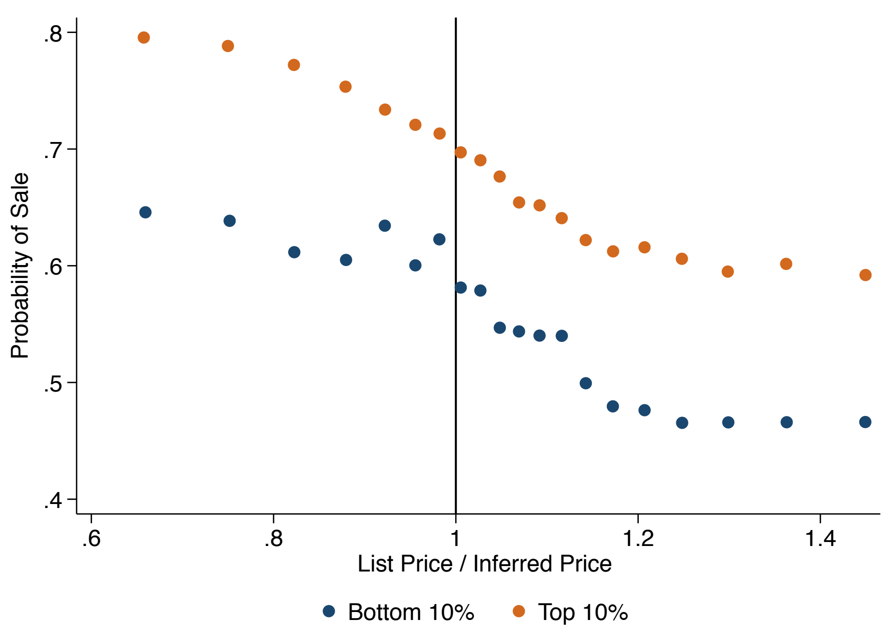
We conclude that while lower experience agents are more likely to list and sell at higher prices, the trade-off is not favorable, unless for extremely patient sellers. In baseline specification of the model, we abstract from price heterogeneity.

**Table 8:** Experience and prices: Motivated Sellers

| | Price (Log) | | |
|---|---|---|---|
| | List | Sale | Frac. Discount |
| Log(Exp+1) | -0.003 | 0.004 | -0.010 |
| | (0.007) | (0.006) | (0.007) |
| Bust X Log(Exp+1) | -0.018 | -0.028** | 0.006 |
| | (0.012) | (0.013) | (0.010) |
| Medium X Log(Exp+1) | -0.021* | -0.029*** | -0.010 |
| | (0.012) | (0.009) | (0.009) |
| $R^2$ | 0.757 | 0.750 | 0.118 |
| Fips Effect | Yes | Yes | Yes |
| Time Effect | Yes | Yes | Yes |
| House Characteristics | Yes | Yes | Yes |
| $N$ | 15689 | 9680 | 9565 |

**A set of regressions in this tables are restricted to a sample of motivated sellers: those who have inherited the property or have likely gone through a divorce. Specifically, these listings occur within two years after a deeds record of an arms-length transaction there both people have the same last name, but a different first name. Displayed are our preferred specification of regression outcomes in equation 1 for several variables: list price, sale price, and the probability of lowering the price.**

**Figure 8:** Pricing and Sale Probability

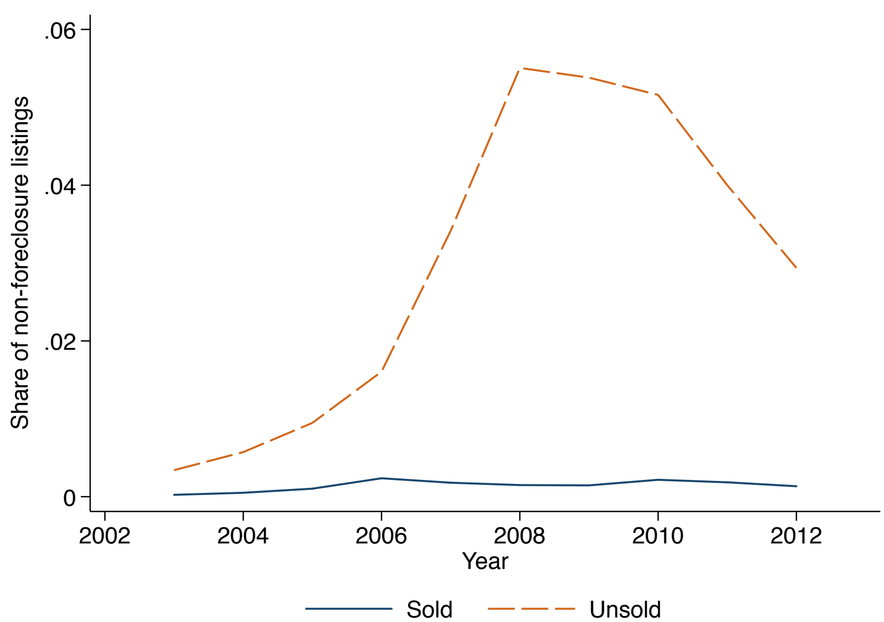



**Note: This graph plots expected sale probability against normalized price: list price relative to our inferred measure of list price. We compute the inferred price as the last price that the property has sold for in the past appreciated to current list date using the Zillow zip code and tier level house price appreciation measure. We plot this relationship for the top and bottom 10 percent of experience distribution.**

### 3.3 Aggregate implications: case of foreclosures

Improved liquidity allows for better allocation of houses as sellers often have to first sell their home, before purchasing a new one. In addition, liquidity in the housing market allows homeowners to reallocate resources across other assets in they portfolio and avoid taking on debt following shocks, such as unexpected expenses or unexpected loss of income. When the recession hit, many households found themselves unable to pay off their mortgages and attempted to sell their properties before falling back on their repayments. Unable to do so, some were forced into default. In Figure 9 plots the fraction of unsold properties that we observe re-listed as foreclosures within the following two years. Relatively low in early 2000s, this fraction spikes to 5.5% in 2007. [11] Similar plot for sold properties suggests that foreclosure could have been avoided if the property had been successfully sold. Through probability of sale alone, real estate intermediaries can play an important role in reducing the number of foreclosures in the housing market. We will come back to this point in the counterfactual exercise in the following section.

**Figure 9:** Probability of Subsequent Foreclosure




**Note: This plot shows the fraction of properties listed (sold and unsold) that we observe going into foreclosure in the next two years.**

Foreclosures result in a substantial financial burden for people who loose their homes. A likely outcome is a substantially lower credit score that limits borrowing ability for years to come. Foreclo-

[11] While we do not have mortgage data, we anticipate that a similar figure for share of unsold properties with subsequently delinquent mortgages also spikes up during the bust.

sures are socially inefficient, because vacant properties tend to depreciate faster, either due to lack of upkeep or through a higher chance of looting and crime. Finally, several studies have documented that foreclosed properties have externalities, putting downward pressure on prices for all houses in the neighboring areas. This was particularly important in the recent bust as lower prices might have caused more homeowners to go underwater and foreclose in their turn. [12]

Note that while substantial, this fraction is likely a lower bound on the actual foreclosure outcome of properties. First, we only observe listings that are marked as foreclosure, meaning that the preceding legal procedures had already been completed. It could very well be that the foreclosure process was initiated within two years, but the property have not been put on the market, so is not counted in our measure. Second, if the lender takes ownership of the property they might not necessarily put it up for sale right away, again excluding a foreclosure scenario from our data.

### 3.4 Partial Equilibrium Counterfactual

How much did real estate agents contribute to the dip in liquidity in the recent housing bust? One way to assess this is to compute what the liquidity would have been if all agents were in a high experience bin. To do that,we use the regression analysis from the empirical section. For each listing, we compute the predicted sale probability for the counterfactual where all variables are fixed except for the experience of the listing agent. Figure 10(A) plots the average yearly probability of sale that we observe in our data and one computed from the predicted counterfactual values. Table 9 summarizes the results. We find that in the trough of the housing bust, as much as 11.5% more listings would have sold under the counterfactual. A similar exercise for our measure of foreclosure probability (illustrated in Figure 10(B)) suggests that 19% of properties that subsequently foreclosed could have avoided foreclosure.

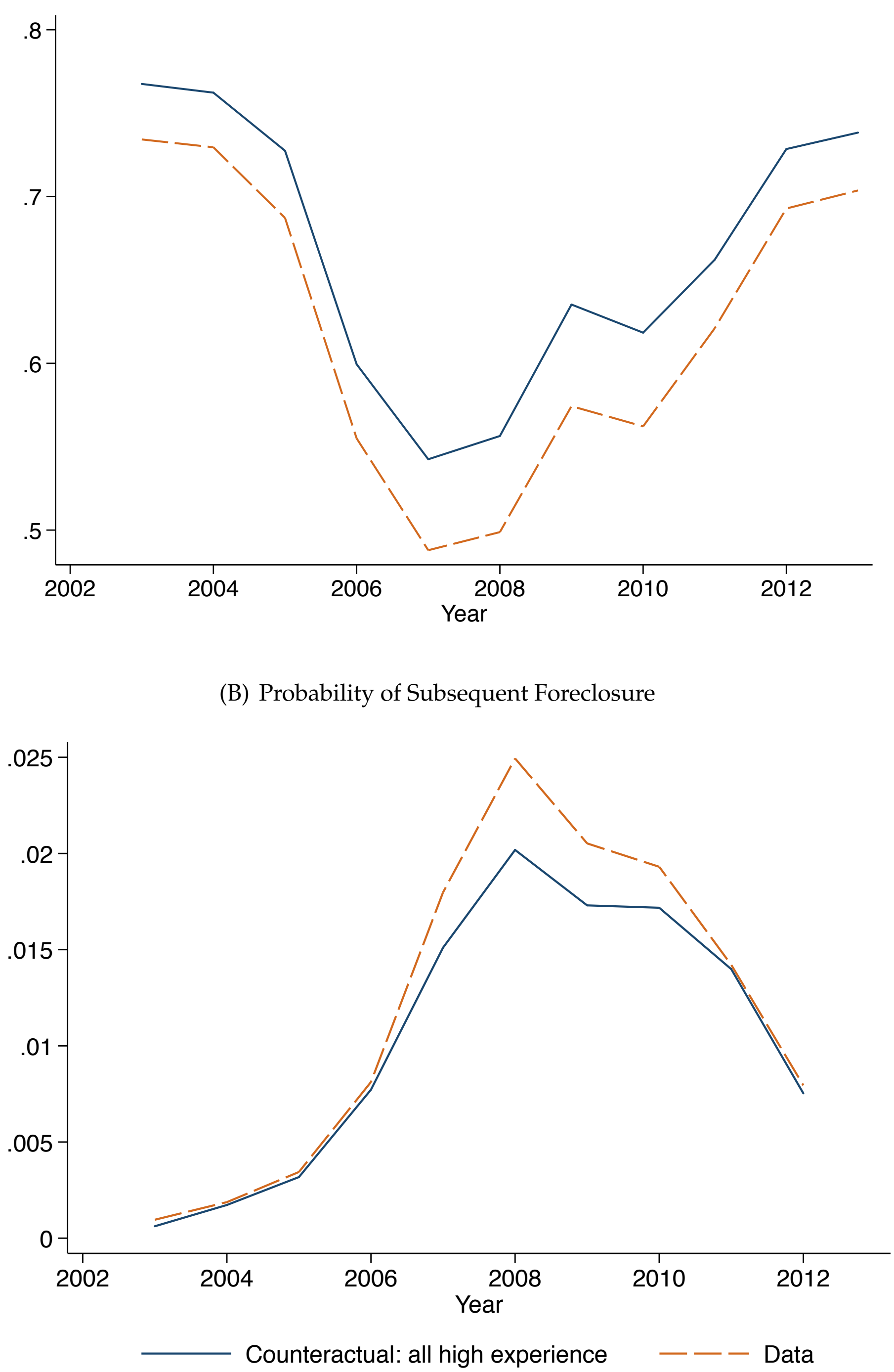
Counterfactual here can not be achieved in practice, because experience is an endogenous variable whose composition depends on entry and exit decisions, as well as learning opportunities that agents have in each time period. It is impossible to directly manipulate experience of agents, but through changing incentives, a policymaker might hope to affect relevant aspects of the market. In this section we proceed by building a structural model of real estate intermediaries and assess the effect of different policies on the distribution of experience as well as aggregate outcomes in the housing market.

[12] Notable papers examining foreclosure externalities include Lin, Rosenblatt, and Yao (2009), Campbell, Giglio, and Pathak (2011), and Mian, Sufi, and Trebbi (2015), Gupta (2016)

**Figure 10:** Partial Equilibrium Counterfactuals

(A) Sale Probability

(B) Probability of Subsequent Foreclosure




**Note: Data plotted in the orange graph is the fraction of listings posted in the corresponding year that result in a sale. We then regress sale probability on the house characteristics, zip code level by calendar month fixed effects, as well as listing agent experience agent interacted with each year. Using the coefficients of this regression, we then predict sale probability for a counterfactual where all agents are in a high experience bin. The blue line then plots the average counterfactual sale probability using the predicted values.**

**Table 9:** Partial Equilibrium Analysis

| | Sale Probability | | | Foreclosure Probability | | |
|---|---|---|---|---|---|---|
| | Data | Counterf. | %Δ | Data | Counterf. | %Δ |
| 2003 | 0.73 | 0.77 | 4.53 | 0.0010 | 0.0006 | -35.33 |
| 2004 | 0.73 | 0.76 | 4.49 | 0.0019 | 0.0017 | -7.88 |
| 2005 | 0.69 | 0.73 | 5.86 | 0.0034 | 0.0032 | -7.69 |
| 2006 | 0.55 | 0.60 | 8.01 | 0.0081 | 0.0077 | -4.93 |
| 2007 | 0.49 | 0.54 | 11.18 | 0.0180 | 0.0151 | -15.99 |
| 2008 | 0.50 | 0.56 | 11.55 | 0.0249 | 0.0202 | -19.07 |
| 2009 | 0.57 | 0.64 | 10.61 | 0.0205 | 0.0173 | -15.69 |
| 2010 | 0.56 | 0.62 | 9.99 | 0.0193 | 0.0172 | -11.03 |
| 2011 | 0.62 | 0.66 | 6.62 | 0.0142 | 0.0140 | -1.57 |
| 2012 | 0.69 | 0.73 | 5.14 | 0.0079 | 0.0075 | -5.21 |
| 2013 | 0.70 | 0.74 | 4.93 | . | . | . |
| 2014 | 0.45 | 0.48 | 6.51 | . | . | . |

**Note: This table shows results from partial equilibrium counterfactual exercise. For each outcome $y$ (sale and identifier of future foreclosure), we run the following regression: $y_{i,t} = \sum_{E\in L,M,H}\left(\beta_E 1_{Exp=E} + \sum_{s=2002}^{2013}\beta_{E,s}1_{Exp=E}\times 1_{yr_t=s}\right) + \delta W_{i,t} + \alpha_{l(i),t} + \epsilon_{i,t}$, where $W_{i,t}$ are detailed property characteristics and $\alpha$ are zip code by month fixed effects. For each observation we then predict the outcome values when listing agent experience is set to be high. Columns labeled "Counterf." show yearly averages for these predicted values. Columns labeled "Data" show yearly averages of the actual outcome values. Finally "%Δ" columns show the percentage difference between the two.**

# 4 Model

## 4.1 Model Setup

There are three types of agents in the model: buyers, sellers and real estate agents. All the houses in the economy are identical and there is no heterogeneity in buyers or sellers, however agents differ by their experience $e$ in the market. We aim to be consistent with our empirical exercise, so experience is defined as the number of past listings that an agent had added to the number of successful transactions they facilitated when representing a buyer. We discount experience at a rate $\delta_e = 0.5$ to mimic the more extreme analogue in the data where only clients accumulated in the prior two years count. We will come back to a formal definition when we describe how experience is updated.

Time is discrete $t \in \mathbf{N}(\mathbf{N} = \{0, 1, 2, ...\})$ and all entering agents are assigned a unique index $i$, so that the experience level of an agent $i$ at time $t$ is $e_{i,t} \in \mathbf{N}$. We define a competition state $n_t^a$ to be a vector over experience levels that specifies the number of all active agents of experience $e$. For a particular agent $i$, the set of competitors can be described as $n^a_{-i,t}$, where $n^a_{-i,t}(e) = n^a_t(e) - 1$ if $e = e_{it}$ and $n^a_{-i,t}(e) = n^a_t(e)$ otherwise. In addition to competition level, each period is also characterized by an industry state $z_t = (n_t^s, v_t)$ that is common across all agents and has two components: a time specific number of sellers that are looking to sell their property $n_t^s$ and the valuation $v_t$ at which the buyers value a purchase of a home. We assume that the industry state evolves according to a Markov process with transition probabilities $P_z$ and takes on three values $z_t \in \{z_1, z_2, z_3\}$ representing bust, average and boom activity in the housing market. Finally, we denote $n_t^b$ to be the total number of buyers (determined endogenously) that will be searching for a house.

In the beginning of each period, the industry state is realized $z_t = (n_t^s, v_t)$ and competition level $n_t^a$ is observed. There is an infinite pool of potential listing agents that have an option to pay an entry cost $c_e$ to get licensed and enter in the current period with experience level $e = 0$.

Following agent entry decisions, an infinite pool of potential buyers decide whether to pay a search cost $c_b$ to enter the search market.

Next, all buyers and sellers are paired with an agent. We assume that a fraction $\phi$ of them contact an agent at random, and the remaining fraction get a referral and match with a particular agent with a probability proportional to the agent's experience share. Formally, the number of sellers an agent with experience $e$ is expected to work with is

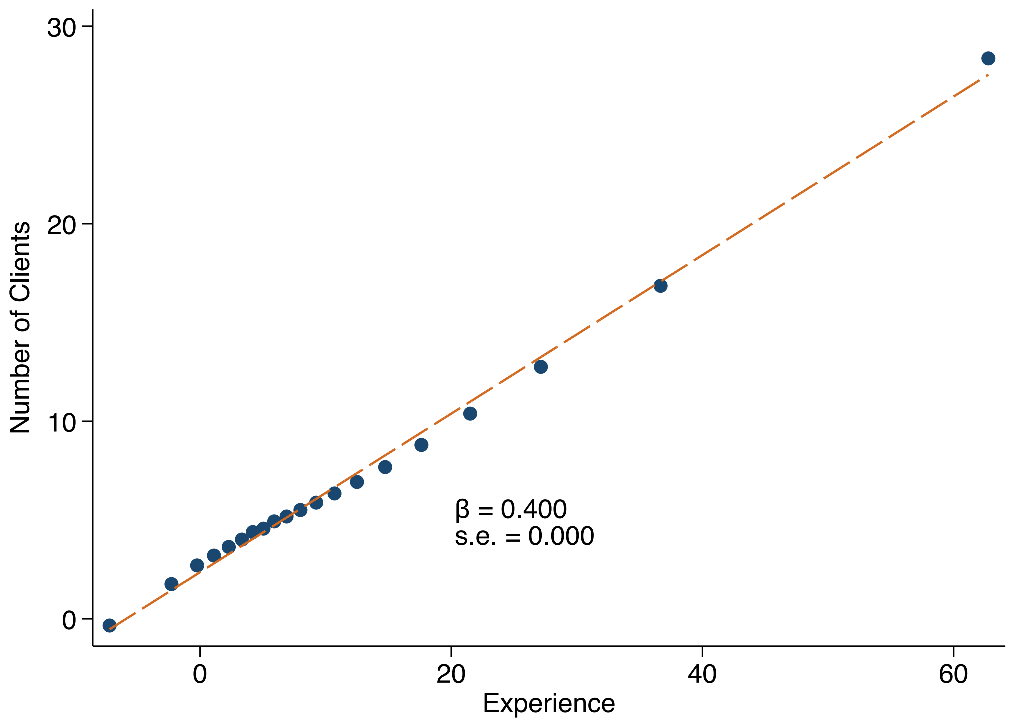
$$s(e, n_t^s; n_t^a) = \phi n_t^s \frac{1}{\sum_{\tilde{e}} n_t^a(\tilde{e})} + (1-\phi) n_t^s \frac{e}{\sum_{\tilde{e}} n_t^a(\tilde{e})\tilde{e}} \tag{2}$$

Similarly, the number of buyers that an agent with experience $e$ is expected to work with is

$$b(e; n^a, n_t^b) = \phi n_t^b \frac{1}{\sum_{\tilde{e}} n_t^a(\tilde{e})} + (1-\phi) n_t^b \frac{e}{\sum_{\tilde{e}} n_t^a(\tilde{e})\tilde{e}} \tag{3}$$

An experienced agent can then expect to have more clients on both seller and buyer side.[13] For non-integer $s(e, n_t^s; n_t^a)$ and $b(e; n^a, n_t^b)$ we assume rationing among agents of experience $e$.

**Figure 11:** Clients and Experience




**Note: This figure is binscatter of number of clients (this includes all listings and successful buyers) that an agent is observed working with on the experience level of the agent in that year. All listings are attributed to the original list year, and all buyers are counted for the close year of the property they bought, thus there is no overlap between clients across different years. Experience is defined as the number of clients that and agent had in the previous two years.**

Clients fully delegate the housing search process to their agents and thus have no further role in

[13]While a linear relationship between experience and number of listings might seem ad hoc, it's a surprisingly accurate representation of what we observe in the data. Figure (11) shows a binscatter plot of number of clients we observe in the data (this includes all listings and successful buyers) against our measure of agent experience. In this plot, we control for cbsa level fixed effects associated with each agent. Table 10 explores the relationship more formally in a regression. We find a strong linear relation with the slope that depends little on the time period.

**Table 10:** Number of Clients

| | (1) | (2) |
|---|---|---|
| Experience | 0.40*** | 0.41*** |
| | (0.00) | (0.00) |
| Bust | | -0.64*** |
| | | (0.01) |
| Recovery | | -0.66*** |
| | | (0.02) |
| Bust X Experience | | -0.03*** |
| | | (0.00) |
| Recovery X Experience | | 0.04*** |
| | | (0.00) |
| $R^2$ | 0.4936 | 0.4971 |
| Fips Effect | Yes | Yes |
| $N$ | 2021861 | 2021861 |

**Note: This table shows a regression of number of clients we observe in the data (this includes all listings and successful buyers) against experience of the agent. Experience here is measured as the number of clients that the agent had in the previous two years. All listings are attributed to the original list year, and all buyers are counted for the close year of the property they bought, thus there is no overlap between clients across different years. To exclude the outliers with unreasonable number of clients, the sample truncates the top 1% of agent by year observations. The first specification controls only for location, where cbsa used for each observation is one where an agent has the most number of clients in a particular year. The second specification includes three time periods for boom bust and recovery and their interaction with the experience measure. We find that while the total number of clients drops in the bust and recovery, there is very little effect on the slope.**

the model. If a transaction is made, a buyer has to pay the negotiated price, 6% of which is equally split between the buyer agent and the seller agent and the remaining 94% is received by the seller. We further assume that all client - agent pairs can be treated as independent of other links that the two parties might have. That is, an agent who is working with both a seller and a buyer is not able to pair the two clients for a transaction. Instead, the search market operates as if each client was represented by their own individual agent. We now describe the search market in more detail.

We model the housing market using the directed search framework, a standard setting in labor, finance and Industrial Organization literature.[14] In this setting buyer agents can direct their search towards houses whose listings agents are of a particular experience. This effectively creates different sub-markets that are indexed by the experience of selling agents operating in that market.

In each sub-market $j$ that has $s$ seller agents and $b$ buyer agents, $s(1-e^{-b\lambda(e_j)/s})$ matches are realized, where $e_j$ is experience level of listing agents in that market.[15] Function $\lambda(e)$ captures the experience advantage of attracting clients to a property. There are a few channels through which this happens in practice. First, agents with more experience tend to be more connected to other agents and also former clients. Thus, they are more likely to attract a match either through directly reaching out to potential buyers or through contacting other agents and tapping into their network of clients. Second, a more experience agent can more effectively market a property to attract viewings and increase desirability for buyers who view the house. They might be better at drawing more interest from buyers by using a professional photographer for listings to present the house in a better light and accentuate it's positive characteristics. In addition, an experienced agent might know how to stage the house in a way that helps a buyer envision the space as their own (this typically involves painting the walls white and advising the current owner to remove all personal items from the space). Finally, an experienced agent might appear more knowledgeable and trustworthy to the buyers who are then more likely to go through with the purchase. We impose $\lambda$ to have the following functional form $\lambda(e) = \lambda_1 e^{\lambda_2}$. Power functions are useful in this setting as they allow for a decreasing returns to

[14]While the set up and the solution method of our model echoes the standard directed search model (see Moen (1997) and Shimer (1996)), it differs in a significant way. Standard directed search involves both optimal price setting on one side and the ability to direct search to particular prices on the other (each market only differing in prices). Instead, markets in our model differ in matching function, so home buyers direct their search to a particular technology, while the prices are determined upon meeting. The ability for buyers to select into different technologies combined with certain class of matching functions makes the equilibrium block-recursive, one of the main appeals of the directed search framework.

[15]This matching function is an approximation of an urn-ball matching function for a large number of agents. The formulation is convenient because it restricts the probability of match to be between 0 and 1. In addition, match probabilities for each side exhibit constant return to scale which allows to keep track of the market tightness only, rather than the number of counterparties on each side of the market. For a more detailed discussion refer to Rogerson, Shimer, and Wright (2005)

scale, meaning faster "learning" by inexperienced agents observed in the data.[16]

Match probability for a buyer is then a function of listing agents experience $e$ and the market tightness $\theta = b/s$: $\eta(e,\theta) = \frac{1}{\theta}(1 - e^{-\lambda(e)\theta})$. Similarly, the probability of match for a seller is $\mu(e,\theta) = 1 - e^{-\lambda(e)\theta} = \theta\eta(e,\theta)$.

Once a meeting occurs, prices are determined via Nash bargaining with bargaining parameter $\gamma$ for the buyer. We assume that a seller of an unsold house and a buyer who purchases a home value the changes in future resale value the same, in which case the total surplus of a transaction will not be affected by the continuation value of holding on to the property and will be simply $v_t$. The prices will then be the same in each market and will be equal to

$$p(v_t) = \gamma v_t \tag{4}$$

Buyer agents maximize buyer valuation and solve

$$V^B = -c_b + \max_j \eta(\theta_{j,t})(v_t - p_t) \tag{5}$$

Since prices do not differ by markets, it must be that $\eta(\theta_{j,t})$ is also constant. Otherwise only markets with highest $\eta(\theta_{j,t})$ will attract buyers. Intuitively, this means that while some markets have a better technology, they also attract longer lines equalizing the overall probability of match for each buyer. The buyer free entry condition implies that buyers will enter to the point that $V^B = 0$. The free entry condition combined with the equilibrium result of equal match rates determines the technology - queue trade-off for the buyers:

$$\eta(e,\theta_{j,t}) = \frac{c_b}{(1-\gamma)v_t} = \frac{1}{\theta_{j,t}}(1 - e^{-\lambda(e)\theta_{j,t}}) \tag{6}$$

The right hand side is decreasing in $\theta$, while the left hand side is constant. Thus there is a unique $\theta_{j,t}$ for each market that satisfies the equilibrium conditions for free entry and market indifference. Solving for $\theta_{j,t} = \theta(e, v_t)$ allows us to compute the match probabilities for the seller side $\mu(e,v_t) = 1 - e^{-\lambda(e)\theta(e,v_t)}$. After the matches are realized, buyers pay $p_t$, of which 3% goes to the buyer agent, 3% to the seller agent and the remaining 94% is taken by the seller.

While in equilibrium $\eta(v_t) = \frac{c_b}{(1-\gamma)v_t}$ is constant across market, $\mu(e,v_t)$ is increasing in the experi-

[16] Some recent papers that use power functions to describe experience effect on production include Benkard (2000), Kellogg (2011) and Levitt, List, and Syverson (2013)

ence of a listing agent operating in each market $j$ through the $\lambda$ function. Thus, the experience of an agent only affects outcomes of sellers and does not improve outcomes for the buyers. This is a simplifying assumption that allows us abstract from heterogeneity on both sides of the search market, but we think it is somewhat realistic. While the marketing effort and expertise is often crucial in whether a house find a buyer, the buyer agent mainly engages in scheduling viewings for existing homes up for sale, which arguably requires less know-how.

For a particular distribution $n_t^a$ across agents, we can now compute the total number of buyers $n_t^b$:

$$n_t^b = \sum_e n_t^a(e) s(e, n_t^s; n_t^a) \theta(e, v_t) \tag{7}$$

The equation sums up buyers that are present in each market, which is the equilibrium market tightness multiplied by the number of listings allocated to the corresponding experience group.

We can now compute the per-period expected profit function for each agent of experience $e$.

$$E[\pi(e)|z_t, n_t^a, n_t^b] = s(e, n_t^s; n_t^a)\mu(e, v_t)\psi p(v_t) + b(e; n_t^b, n_t^a)\eta(v_t)\psi\gamma v_t, \tag{8}$$

Where $\psi = 0.03$ is the commission rate that each agents earns on a successful transaction. They expect to get $s(e, n_t^s; n_t^a)$ listings that will sell with probability $\mu(e, v_t)$ as well as $b(e; n_t^b, n_t^a)$ buyers who buy with probability $\eta(v_t)$ . All transacted properties will earn the agent 3% of the sale price $p(v_t)$.

At the end of the period, experience of all agents is updated. We assume that only successful buyers count towards experience, while all listings contribute to experience equally no matter if they are sold. In addition, all previous experience is discounted at rate $\delta$. Then the expected experience level of an agent entering time $t$ with experience $e_t$ is

$$E[e_{t+1}|e_t, z_t; n_t^b, n_t^a] = \delta e_t + s(e, n_t^s; n_t^a) + b(e; n_t^b, n_t^a)\eta(v_t) \tag{9}$$

We calibrate $\delta = 1/2$ so that constant flow of clients leads to a constant level of experience, as it does in the empirical counterpart, where the experience is taken to be the number of clients in the previous two years.

At the end of the period, but before the next aggregate state is realized, all agents draw an idiosyncratic cost of operating $c_{i,t}$ from a log normal distribution, so that $log(c_{i,t}) \sim N(\mu_{fc}, \sigma_{fc})$. If the cost drawn exceeds their expected value of staying in the business, they choose to exit the market.

The expected value of an agent $i$ of experience $e$ entering time $t$ is then

$$V_t(e_{i,t}, z_t; n_t^b, n_t^a) = E[\pi(e_{i,t})|z_t, n_t^a, n_t^b] + \beta E_t[\max\{0, -c_{i,t} + V_{t+1}(e_{i,t+1}, z_{t+1}; n_{t+1}^b, n_{t+1}^a)\}] \tag{10}$$

A value of an entrant entering time $t$ is similarly

$$V_t(0, z_t; n_t^b, n_t^a) = -c_e + E[\pi(0)|z_t, n_t^a, n_t^b] + \beta E_t[\max\{0, -c_{i,t} + V_{t+1}(e_{i,t+1}, z_{t+1}; n_{t+1}^b, n_{t+1}^a)\}] \tag{11}$$

Since both number of clients and the probability of sale is increasing with experience, $V$ is strictly increasing with experience as well. Then the optimal exit strategy $\rho_t(e_{i,t+1}, c_{i,t}))$ follows a cut-off rule:

$$\rho_t(e_{i,t+1}, c_{i,t})) = \begin{cases} 1 & \text{if } c_{i,t} > E_t[V_t(e_{i,t+1}, z_{t+1}; n_{t+1}^b, n_{t+1}^a)] \\ 0 & \text{otherwise} \end{cases}. \tag{12}$$

The free entry condition for real estate agents implies that if any agents find it profitable to enter, the value of entry will be driven down to 0. If, however, no entry happens, then the value of entry must be negative. Formally if $\lambda_t$ is the entry rate at time $t$, then $\lambda_t V_t(0, z_t; n_t^b, n_t^a) = 0$. [17]

### 4.2 Model Equilibrium

We allow the exogenous aggregate state $z_t = (n_t^s, v_t)$ to take on three different pairs of values corresponding to boom, bust and recovery. The endogenous measure of buyers $n_t^b$ is a direct function of $v_t$, $n_t^s$ and $n_t^a$, as described in equation 7, so does not need to be monitored it separately. However allowing agents to keep track of the entire distribution $n_t^a$ makes the state space essentially infinite (since each $n_t^a(e)$ is a state variable). While in a static setting, this distribution might reduce to one profit-relevant value that affects competition (such as the overall experience level in the market), in a dynamic setting the entire distribution is needed to project how competition will evolve over time. To simplify the problem, we adopt the Extended Oblivious Equilibrium concept described in Weintraub, Benkard, and Van Roy (2010). In this equilibrium agents approximate the distribution $n_t^a$ using a long run average corresponding to a recent history of aggregate states $z_t$. Adopting the notation in the original paper, let $\{w_t = (z_t, z_{t-1})\}$ be a Markov chain adopted to the filtration generated by $\{z_t : t \geq 0\}$. Let $\lambda(w_t)$ be the entry rate and $\rho(e, w_t)$ be the exit policy at state $w_t$. We define $\tilde{n}^a_{\lambda,\rho}(w_t)$ to be the predicted distribution of agents at state $w_t$, which corresponds to the long run average distribution

[17] While we match the aggregate state $n_t^s$ (number of sellers) to the actual number of listings we observe in the data,we abstract from issues of discreteness for other measures and allow for non-integer values of $n_t^b, n_t^a$ and the entry rate $\lambda_t$.

under entry rates $\lambda$ and policy function $\rho$. We now define agent's value function:

$$\tilde{V}(e,w|\rho',\rho,\lambda) = E[\pi(e,w)] + \beta E[\max\{0, -c + \tilde{V}(e',w')|e,w,\rho',\rho,\lambda] \tag{13}$$

Similarly, an entrant's value is

$$\tilde{V}(0,w|\rho',\rho,\lambda) = -c_e + E[\pi(0,w)] + \beta E[\max\{0, -c + \tilde{V}(e',w')|e,w,\rho',\rho,\lambda] \tag{14}$$

**Definition** An *extended oblivious equilibrium* consists of

1. an exit strategy $\rho(e,w)$, and entry rate $\lambda(w)$ that satisfy the following conditions

   (a) Agents optimize the extended oblivious value function:

   $$\sup_{\rho'} \tilde{V}(e,w|\rho',\rho,\lambda) = \tilde{V}(e,w|\rho,\rho,\lambda)$$

   (b) Either the oblivious expected value of an entering agent is zero or the optimal entry rate is zero (or both):

   $$\begin{aligned} \lambda(w)\tilde{V}(0,w|\rho',\rho,\lambda) &= 0, \\ \tilde{V}(0,w|\rho',\rho,\lambda) &\leq 0, \\ \lambda(w) &\geq 0, \forall w \in Z \times Z \end{aligned}$$

2. $n^b$, entry rate of buyers such that the value of entry is zero (there are always some entrants as long as $v_t \gg c_b$)

3. A belief $\tilde{n}^a(w)$ over the distribution of agents that corresponds to the long run average distribution of agents across experience.

We adopt the solution method described in Weintraub, Benkard, and Van Roy (2010) with slight modification. The full algorithm is described in detail in Appendix E.

### 4.3 Calibration

Fitting the model to the data involves three nested steps. First, we define the stochastic behavior of $z_t$ and fit the behavior of the common industry states for each $z_t$: $v_t, n_t^s$, corresponding to the housing valuation and the number of sellers look to to sell their property. Next, for a given state $z_t$, we calibrate the directed search model to the sale probabilities for each experience group. Finally, given the parameters from the previous two steps, we fit the dynamic entry and exit model to the observed entry and exit rates for every state and experience.

First, we define three states for $z_t$ using the historical series of the Case Shiller house price index for years 1940-2017. The evolution of states in this dataset will allow us to compute a Markov transition probability matrix $P$ for the aggregate states. To estimate $P$, we first deflate the index by Consumer Price Index less shelter compute yearly average of the 12 month growth rate. We define years with growth rates in the lower and higher quartile of the data to be bust and boom years, respectively. The rest of the years correspond to the medium state. Figure (7) plots the adjusted growth rates together with our approximation for the state process. This identifies the time periods corresponding to each aggregate state in our dataset, as well as $P$.

Given these three states, we use the data to compute the observed number of sellers, $n^{s,obs}(z_t)$, and the observed average price levels, $p^{obs}(z_t)$, in each state. For a given price, the parameters of interest, $(v(z_t), \gamma)$ are not separately identified, as they always enter in our model as multiples of each other. Hence, we normalize $\gamma = 0.5$ and fit $v(z_t)$ to match the observed average prices: $p^{obs}(z_t) = \gamma v(z_t)$.

Next, we use the observed sale probabilities for each experience group and aggregate state to calibrate the parameters of the housing search markets. Since the probability of sale does not depend on the distribution of experience, we can calibrate the search parameters without computing for the equilibrium of the model. We match the probability of sale for each experience value, $e$, in different aggregate states, $z_t \in (bust, medium, boom)$, to their counterparts in the model $\mu(e, z_t) = 1 - e^{-\lambda_1(z_t)e^{\lambda_2}\theta(e,z_t)}$. In equilibrium, $\theta(e, z_t)$ is a function of $c_b, v(z_t)$ and $\gamma$ due to free entry of the buyers (see equation (6)). Since the cost of entry for the buyer, $c_b$, identifies the overall level of sale probabilities across all states, we normalize $\lambda_1(bust) = 1$, such that $\lambda_1(recovery)$ and $\lambda_1(boom)$ measure the differences in sale probabilities across aggregate states. Finally, $\lambda_2$ matches the differences in sale probability across experience levels. Formally, let $\Theta_1 = (c_b, \lambda_1(medium), \lambda_1(bust), \lambda_2)$ be the parameters of interest, while the set of moments are $g(e, z, \Theta) = (\tilde{\mu}^{obs}(e, z_t) - \mu_{model}(e, z_t, \Theta))$. The chosen parameters $\hat{\Theta}_1$ are then

$$\hat{\Theta}_1 = argmin_{\Theta_1} \sum_{e,z} g(e, z, \Theta_1)^2. \tag{15}$$

Finally, we estimate the remaining parameters, $c_e$, $\mu_{fc}$, and $\sigma_{fc}$ governing the entry and exit rates of real estate agents. Computing the entry and exit rates implied by these parameters involves a computation of the equilibrium decision, using the distribution of $z_t$, $P$, the values of $(n^s(z_t), v(z_t)$ in each aggregate state, and the parameters from the previous step, $\hat{\Theta}_1$. We choose $c_e$, $\mu_{fc}$, and $\sigma_{fc}$ to minimize the difference between the observed exit rates, $\rho^{obs}(e, \omega_t = (z_t, z_{t-1}))$ and entry rates, $\Lambda($, $\rho(e, \omega_t = (z_t, z_{t-1}))$, and their counterparts in the model. Formally, let $\Theta_2 = (c_e, \mu_{fc}, \sigma_{fc})$, $g_1(e, \omega\Theta) = (\tilde{\rho}^{obs}(e, \omega_t) - \rho_{model}(e, \omega_t, \Theta))$ and $g_2(\omega, \Theta) = (\tilde{\Lambda}^{obs}(\omega_t) - \rho_{model}(\omega_t, \Theta))$. Hence,

$$\hat{\Theta}_2 = argmin_{\Theta_2} \sum_{e,\omega} (g_1(e, \omega, \Theta)^2 + g_2(\omega, \Theta)^2. \tag{16}$$

We summarize the parameter values and the calibration strategy in Table 11.

**Table 11:** Model Calibration

| Parameter | Value | Matching Moment |
|---|---|---|
| $n^s(z)$ | [221193 195023 240191] | number of listings |
| $v$ | [342540 367810 381710] | price level |
| $\gamma$ | 0.5 | - |
| $\lambda_1(z)$ | (1,1.19,1.03) | norm / average sale probability by state |
| $\lambda_2$ | 0.024 | sale probability by exp. |
| $c_b$ | 35918 | overall sale probability |
| $c_e$ | 10000 | entry rates |
| $\mu_c$ | 8 | exit rates across |
| $\sigma_c$ | 2.75 | experience groups |
| Transition Probabilities | | historical price data |

**Note: This table shows parameter values used in the data together with the description of the calibration approach.**

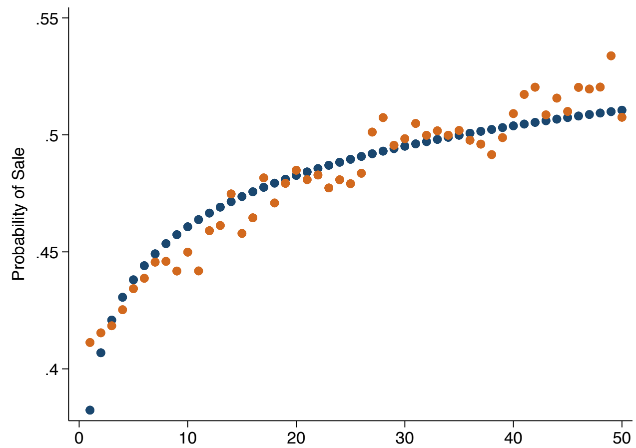
### 4.4 Model Fit

To evaluate how well the model captures key aspects of the real estate intermediation industry, we compare several moments in the model, both explicitly targeted in the calibration exercise and those not targeted, to their counterparts in the data. Figure 12 plots the probability of sale for each state $z_t$ as predicted by the model together with the equivalent values in the data. The model captures these perfectly. Figure 13 plots average entry and exit rates for each experience level. The model predict a more rapid falloff in exit rates for low experienced agents as compared to the data. In addition, the entry rates are slightly lower than observed in the data.

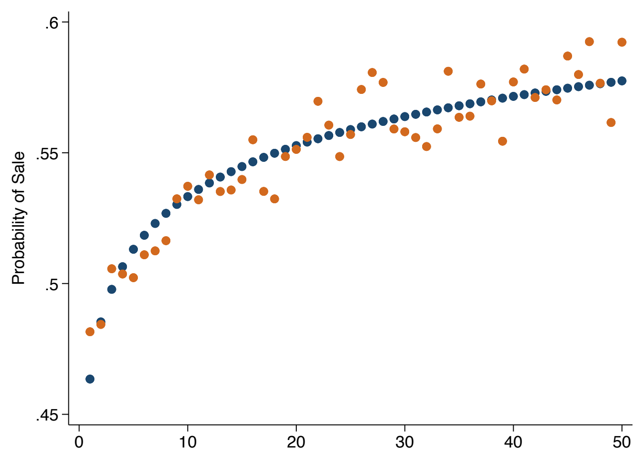
In addition to moments directly used by the estimation, the model delivers entry and exit rates

**Figure 12:** Sale Probability: baseline vs. data

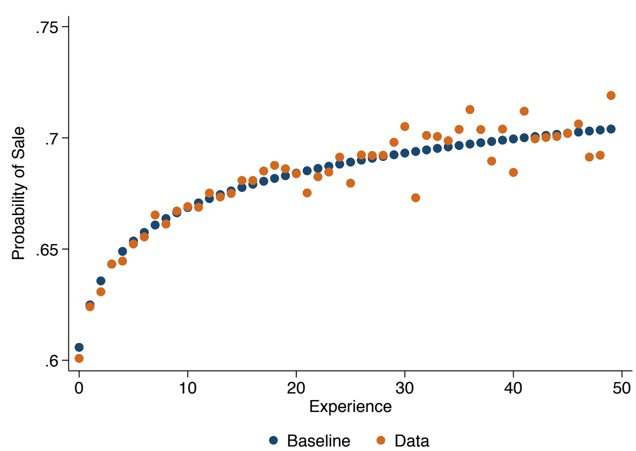
(A) Bust




(B) Median



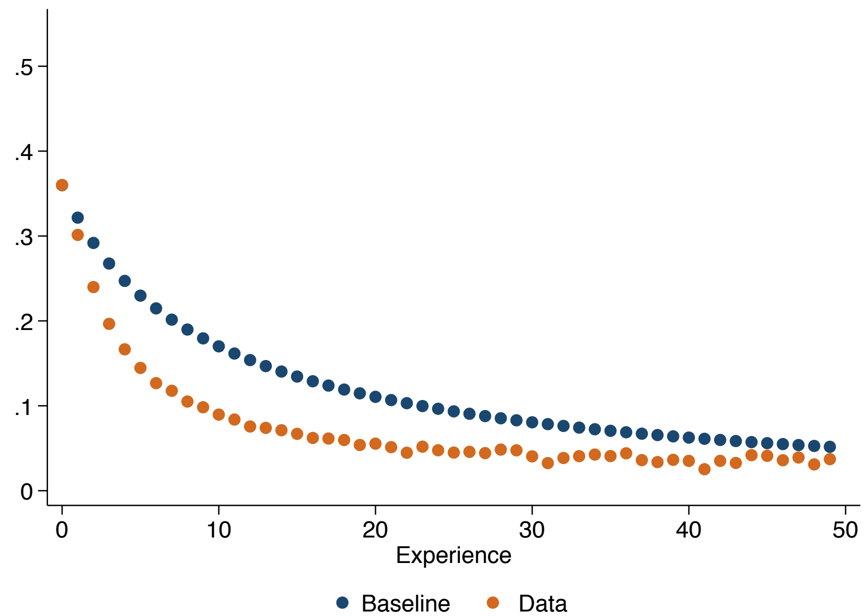

(C) Boom




**Note: We plot here sale probability by each experience level of an agent together with the data counterpart. Since these values do not depend on the distribution of experience, they vary only by aggregate state $z$ which takes three values corresponding to housing boom, a median state and the housing bust. For the data counterpart, we plot the coefficients on experience levels from the following regression: we regress the sale outcome variable on house characteristics and an identifier for each experience level of the listing agent, adding fixed effects for zip code by list month.**

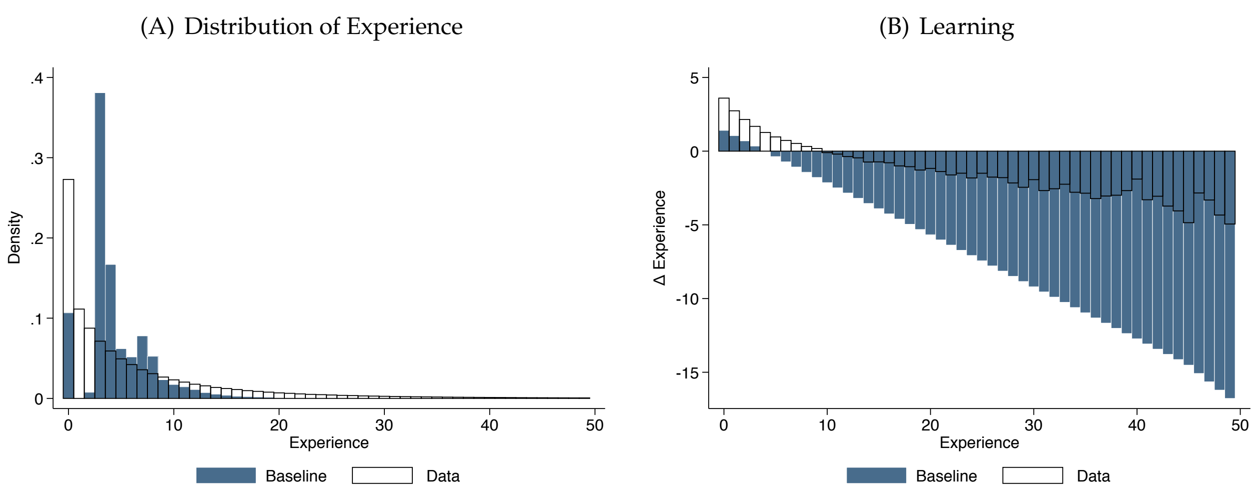
within each of the aggregate states, as well as the learning accumulation, and the resulting distribution of agents across experience groups for each state. The model predicts no entry in certain periods $w$ that follow big spikes in entry in the previous year. The model is, however, able to match exit rates fairly well. To capture the distribution of agents we compute the 25th, 50th and 75th percentile of agent experience. To capture how fast agents accumulate experience, we compute the change in experience of agents conditional on staying in the market and present the experience change for difference experience points. Table 12 summarizes the fit across several states we observe in the data. We plot the aggregate distribution and learning accumulation in Figure 14.

**Figure 13:** Aggregate Turnover: Baseline vs. Data




**Note: Presented here are aggregate entry and exit rates across different experience bins in the equilibrium of the model and as observed in the data.**

**Figure 14:** Distribution and Learning: Baseline vs. Data




**Note: Panel A plots an average distribution of agents by each experience level with the data counterpart. Panel B plots average experience accumulation conditional on staying in the market the following year.**

**Table 12:** Model Fit

| | Exit Rates | | | | | | Entry Rates | |
|---|---|---|---|---|---|---|---|---|
| | Experience 0 | | Experience 10 | | Experience 40 | | | |
| | Baseline | Data | Baseline | Data | Baseline | Data | Baseline | Data |
| Bust Bust | 0.38 | 0.39 | 0.26 | 0.10 | 0.14 | 0.04 | 0.00 | 0.17 |
| Bust Medium | 0.33 | . | 0.14 | . | 0.05 | . | 0.59 | 0.19 |
| Medium Bust | 0.33 | 0.41 | 0.16 | 0.11 | 0.06 | 0.04 | 0.33 | 0.20 |
| Medium Boom | 0.33 | . | 0.13 | . | 0.04 | . | 0.59 | 0.20 |
| Boom Medium | 0.33 | 0.38 | 0.19 | 0.08 | 0.08 | 0.06 | 0.15 | 0.25 |
| Boom Boom | 0.37 | 0.30 | 0.25 | 0.07 | 0.12 | 0.02 | 0.00 | 0.19 |

| | Distribution | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 25th Percentile | | 50th Percentile | | 75th Percentile | | 95th Percentile | |
| | Baseline | Data | Baseline | Data | Baseline | Data | Baseline | Data |
| Bust Bust | 4 | 1 | 4 | 3 | 7 | 8 | 12 | 24 |
| Bust Medium | 0 | 0 | 0 | 3 | 4 | 8 | 8 | 24 |
| Medium Bust | 0 | 0 | 3 | 3 | 5 | 8 | 9 | 23 |
| Medium Boom | 0 | 0 | 0 | 3 | 4 | 8 | 7 | 24 |
| Boom Medium | 3 | 0 | 3 | 3 | 5 | 8 | 11 | 23 |
| Boom Boom | 3 | 0 | 3 | 3 | 6 | 8 | 11 | 23 |

| | Learning | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Experience 0 | | Experience 5 | | Experience 10 | | Experience 40 | |
| | Baseline | Data | Baseline | Data | Baseline | Data | Baseline | Data |
| Bust Bust | 2.2 | 3.4 | -1.1 | 0.7 | -4.5 | -0.4 | -24.7 | -4.1 |
| Bust Medium | 2.9 | 3.4 | 3.2 | 1.0 | 3.5 | 0.3 | 5.2 | -1.3 |
| Medium Bust | 3.6 | 3.6 | 3.2 | 0.8 | 2.8 | -0.6 | 0.3 | -3.0 |
| Medium Boom | 3.0 | . | 4.1 | . | 5.3 | . | 10.0 | . |
| Boom Medium | 2.9 | 3.6 | 0.9 | 0.9 | -1.1 | -0.3 | -13.1 | 0.3 |
| Boom Boom | 2.0 | 4.0 | -1.3 | 1.4 | -4.7 | 0.4 | -24.5 | -1.4 |

## 5 Results

The model allows for evaluation of various policy interventions. We consider three policies corresponding to the following counterfactuals. With the rest of the structural parameters fixed, the equilibrium of the model is recomputed with 1) lower commission rates 2) more informed clients, meaning a lower fraction of buyers and sellers who go to a random agent; 3) increased entry costs.

We are interested in how those policies change the composition of experience overall and specifically in the bust state that is followed by the boom, the period non-sale outcomes are most costly. This shift in equilibrium experience distribution comes from three different channels - entry, exit, and learning. We estimate how each channel is affected by different policies, and how does the overall change in the distribution translate into aggregate probability of sale. In addition to liquidity, the model also speaks to welfare consequences for sellers in this market (buyers and agents have free entry, so have 0 value independent of parameter values). While sellers are not modeled as dynamic agents, we can assume that sellers who do not sell their home return to the market the next period and repeat the effort to sell. Their ex-ante value is then computed as follows:

$$V^s(w) = \sum_{\tilde{e}} \left( \phi \frac{n^a(w,\tilde{e})}{\sum_e n^a(w,e)} + (1-\phi) \frac{n^a(w,\tilde{e})}{\sum_e n^a(w,e)e} \right) \left( \mu(\tilde{e}, v(w))(1-\psi)p(v(w)) + \beta(1-\mu(\tilde{e}, v(w)))E[V^S(w')|w] \right) \tag{17}$$

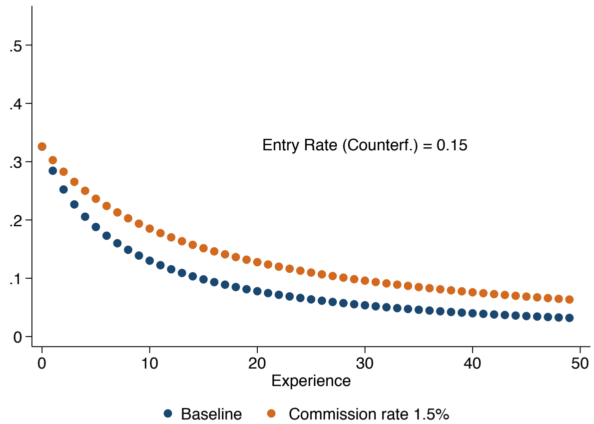
It is a sum over each experience level, of a chance of being matched to an agent in that experience group, multiplied by the probability of sale at a current price, less the commission rate and the complementary probability of moving into the next period with an unsold house.

### 5.1 Low Commission Rates

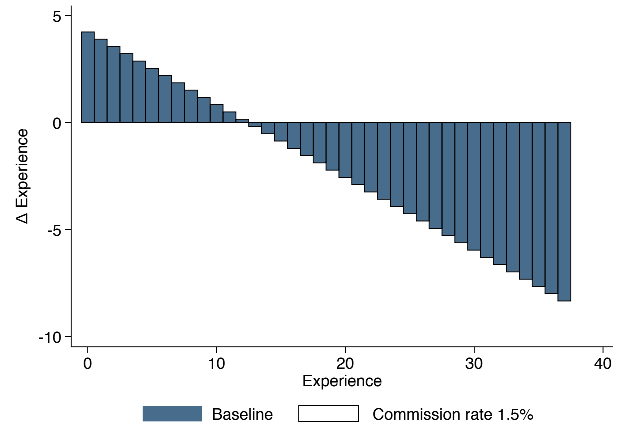
The first counterfactual exercise is to vary commission rates. Qualitatively, reduced commission rates, make entry less profitable, reducing the overall entry rates. It also lowers profitability of all agents in the market, thus increasing exit rates for all levels of experience. In general increased exit rates is not desirable, as exit leads to loss of knowledge in the market. However this loss is compensated by much faster accumulation of knowledge among existing agents. This is because in order to compensate for fixed and entry costs, agents have to make up for reduced commission by working with more listings. Figure 15 illustrates the effect, by contrasting the baseline equilibrium with one where the commission rate is cut by half to 1.5%. Panel A shows that while entry rate decreases, exit rates go up for all experience levels. In Panel B, expected change in experience conditional on remaining active is higher for all experience agents. Finally Panel C plots the overall effect on the distribution.

**Figure 15:** Baseline vs. 1.5% commission rates

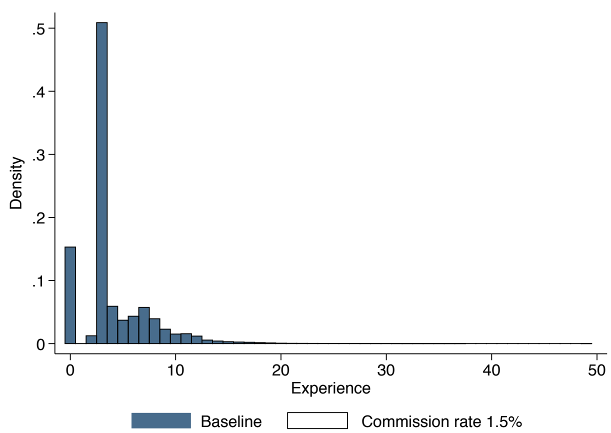
(A) Exit Rates by Experience




(B) Experience Accumulation




(C) Distribution of Experience



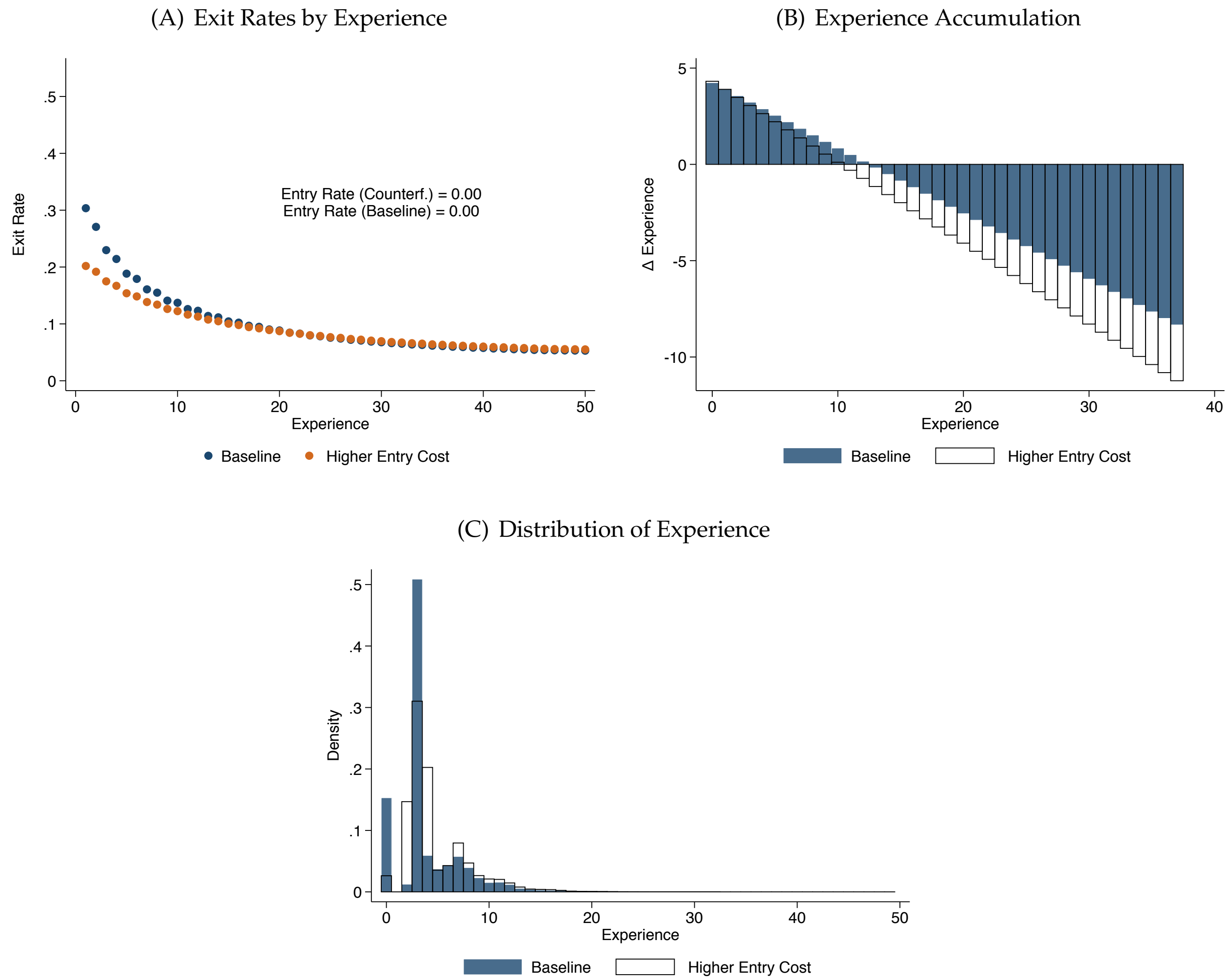

**Note: Contrasted here are the baseline specification of the model and the counterfactual of reduced commission rates. Panel A plots exit rates by different experience groups. Panel B captures learning - the expected change in experience level conditional on staying in the following period. Panel C plots agent distribution across experience groups**

To quantitatively evaluate this policy we consider several levels of commission rates. For each, we compute aggregate probability of sale and look at how it compares to the baseline specification.

## 5.2 Increasing Entry Costs

Next counterfactual examines the effect of changing entry costs directly. This policy is perhaps the most straightforward to implement as states can simply raise the licensing costs of real estate agents. Increasing entry costs has a negative effect on entry rates. Free entry condition implies that to compensate for increased entry costs, new agents would have to work with more agents to earn more profit. As a result entrants learn faster while the more experienced agents learn slower, as their experience share is reduced with overall level of experience increasing in the market. Figure 16 illustrates these channels for an increased entry cost of 30000$.

**Figure 16:** Baseline vs. entry cost set to 30000$



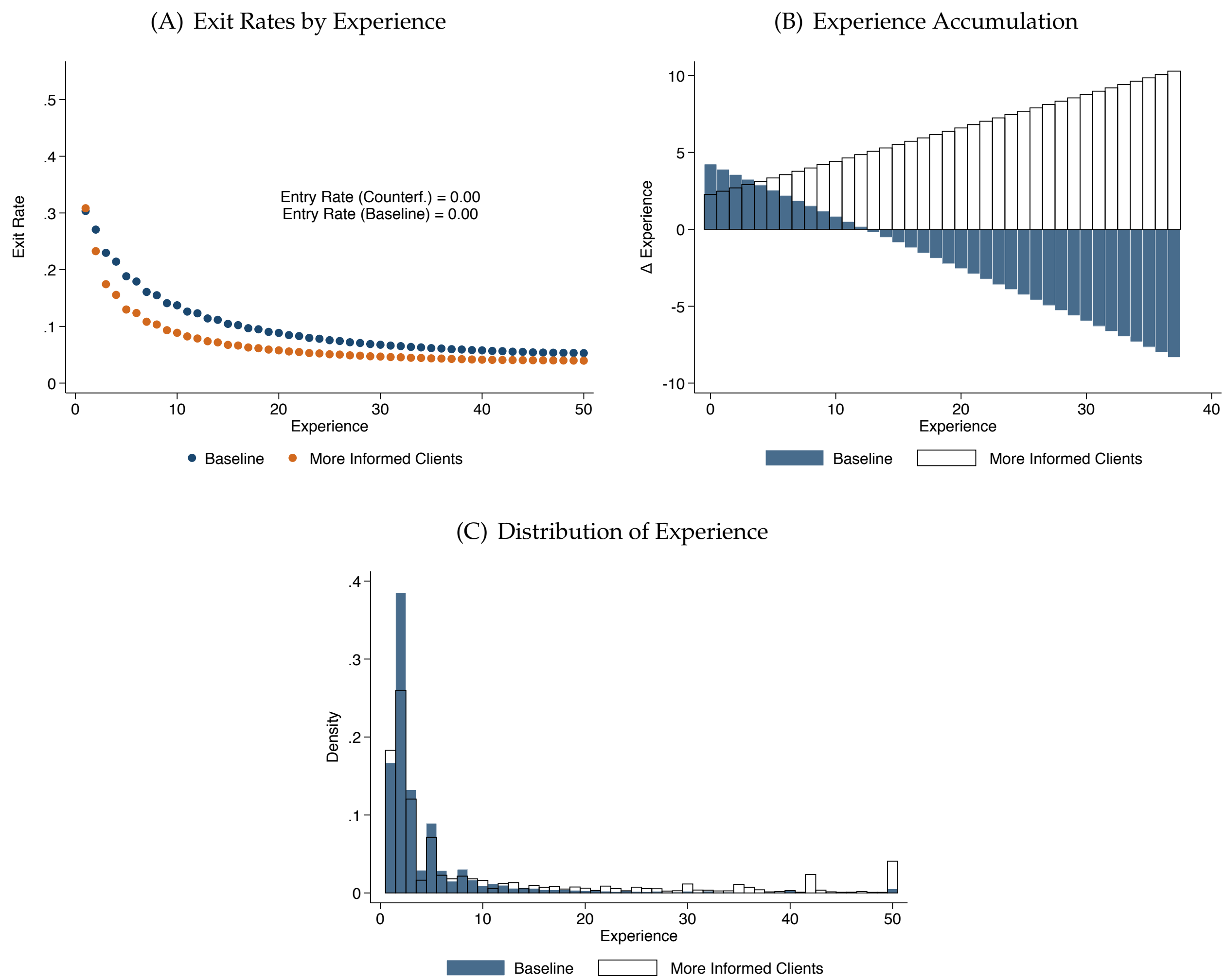

**Note: This plot compares exit rates and distribution od experience for the baseline equilibrium to those in a counterfactual equilibrium where entry costs are set to 30000$.**

## 5.3 Informing Clients on Importance of Experience

Last counterfactual speaks to policies that improve client awareness on the importance of experience. If sellers knew the extent to which the outcome of their listing depends on the agent they choose, they would seek references or evidence of past experience when hiring an intermediary. In the model, this policy would reduce the fraction of clients $\phi$ that look for an agent at random and increase the complementary fraction that match with agents through referrals.

This policy essentially shift the industry profits from low experienced agents towards more experienced ones. This greatly reduces the incentives to enter the market and results in much lower entry rates than those we see in the baseline model. With fewer agents remaining and higher expected returns to experience, exit rates in this counterfactual fall for most experience groups allowing for more knowledge to remain in the market. However knowledge accumulation is slow for the entrants which can actually increase exit rates for the lowest experience groups. Figure 17 illustrates these channels for an increased entry cost of 30000$.

**Figure 17:** Baseline vs. more informed clients




**Note: This plot compares exit rates and the distribution of experience for the baseline equilibrium compared to those in a counterfactual equilibrium where clients are more informed, that is only 10% of all clients seek out a random agent, while the remaining 90% ask for a referral and are assigned to agents with probability proportional to agent experience.**

## 5.4 Employment

Policymakers might also find it valuable to compute the impact of the policies we considered on employment. In Table 15 we compute the total number of agents that operate in each state under

**Table 13:** Aggregate Liquidity

| | Data | Basel. | 1.5% $\phi$ | | $\psi = 0.1$ | | c = 30k | |
|---|---|---|---|---|---|---|---|---|
| | Mean | Mean | Mean | %Δ | Mean | %Δ | Mean | %Δ |
| Bust Bust | 0.466 | 0.457 | 0.458 | 0.3 | 0.467 | 2.2 | 0.456 | -0.1 |
| Bust Med | 0.538 | 0.520 | 0.617 | 18.6 | 0.624 | 19.9 | 0.616 | 18.4 |
| Bust Boom | . | 0.646 | 0.606 | -6.2 | 0.612 | -5.2 | 0.604 | -6.4 |
| Med Bust | 0.463 | 0.453 | 0.456 | 0.6 | 0.467 | 3.0 | 0.455 | 0.4 |
| Med Med | . | 0.527 | 0.616 | 16.9 | 0.625 | 18.5 | 0.615 | 16.7 |
| Med Boom | 0.657 | 0.645 | 0.605 | -6.3 | 0.613 | -5.0 | 0.604 | -6.4 |
| Boom Bust | . | 0.447 | 0.456 | 2.0 | 0.467 | 4.5 | 0.456 | 2.0 |
| Boom Med | 0.533 | 0.526 | 0.616 | 17.0 | 0.624 | 18.6 | 0.616 | 17.0 |
| Boom Boom | 0.653 | 0.651 | 0.605 | -7.1 | 0.614 | -5.7 | 0.604 | -7.1 |

**Note: This table shows average probability of sale in each of the nine states. Column one reports mean sale probability observed in the data. The following column shows the values for our baseline calibration. The next six columns correspond to the counterfactual equilibria and the percentage difference of those values from the baseline. The three policies shown here are 1) lowering commission rates to 1.5% of the transaction price; 2) Having more informed clients, meaning lower percentage of both buyers and sellers contact a seller at random and thus higher chance of referrals; 3) Lastly we consider raising the entry costs directly to 30k.**

**Table 14:** Seller Valuation

| | Baseline | 1.5% Commission | | Informed Clients | | Entry Cost 30k | |
|---|---|---|---|---|---|---|---|
| | Mean | Mean | %Δ | Mean | %Δ | Mean | %Δ |
| Bust Bust | 148,690 | 158,680 | 6.72 | 154,200 | 3.71 | 153,690 | 3.36 |
| Bust Medium | 158,070 | 160,930 | 1.81 | 156,220 | -1.17 | 155,910 | -1.37 |
| Bust Boom | 167,840 | 177,650 | 5.84 | 172,570 | 2.82 | 172,110 | 2.54 |
| Medium Bust | 148,610 | 158,630 | 6.74 | 154,200 | 3.76 | 153,660 | 3.40 |
| Medium Medium | 158,280 | 160,910 | 1.66 | 156,250 | -1.28 | 155,900 | -1.50 |
| Medium Boom | 167,810 | 177,610 | 5.84 | 172,590 | 2.85 | 172,080 | 2.54 |
| Boom Bust | 148,460 | 158,620 | 6.84 | 154,200 | 3.87 | 153,680 | 3.52 |
| Boom Medium | 158,260 | 160,900 | 1.67 | 156,230 | -1.28 | 155,910 | -1.48 |
| Boom Boom | 168,010 | 177,620 | 5.72 | 172,630 | 2.75 | 172,110 | 2.44 |

**Note: This table shows seller valuation in each of the nine states. Column one reports the values for our baseline calibration. The next six columns correspond to the counterfactual equilibria and the percentage difference of those values from the baseline. The three policies shown here are 1) lowering commission rates to 1.5% of the transaction price; 2) Having more informed clients, meaning lower percentage of both buyers and sellers contact a seller at random and thus higher chance of referrals; 3) Lastly we consider raising the entry costs directly to 30k.**

different policies.

**Table 15:** Employment

| | Baseline | 1.5% Commission | | Informed Clients | | Entry Cost 30k | |
|---|---|---|---|---|---|---|---|
| | Mean | Mean | %Δ | Mean | %Δ | Mean | %Δ |
| Bust Bust | 147,826 | 58,416 | -60.48 | 53,358 | -63.90 | 68,386 | -53.74 |
| Bust Medium | 102,851 | 42,333 | -58.84 | 32,869 | -68.04 | 32,740 | -68.17 |
| Bust Boom | 121,713 | 46,192 | -62.05 | 37,075 | -69.54 | 36,698 | -69.85 |
| Medium Bust | 62,876 | 25,579 | -59.32 | 22,141 | -64.79 | 26,825 | -57.34 |
| Medium Medium | 145,257 | 59,992 | -58.70 | 47,586 | -67.24 | 64,632 | -55.51 |
| Medium Boom | 121,674 | 46,184 | -62.04 | 37,087 | -69.52 | 36,689 | -69.85 |
| Boom Bust | 87,616 | 32,374 | -63.05 | 29,245 | -66.62 | 27,831 | -68.23 |
| Boom Medium | 87,293 | 34,107 | -60.93 | 28,575 | -67.27 | 33,981 | -61.07 |
| Boom Boom | 202,571 | 79,795 | -60.61 | 58,866 | -70.94 | 81,807 | -59.62 |

**Note: This table shows employment nine states. Column one reports the number of agents in our baseline calibration. The next six columns correspond to the counterfactual equilibria and the percentage difference of those values from the baseline. The three policies shown here are 1) lowering commission rates to 1.5% of the transaction price; 2) Having more informed clients, meaning lower percentage of both buyers and sellers contact a seller at random and thus higher chance of referrals; 3) Lastly we consider raising the entry costs directly to 30k.**

# 6 Conclusion

Agents have heterogeneous abilities in alleviating search frictions in the housing market and distribution of agents across experience has important implications for aggregate housing market outcomes. Easy entry and fixed commission results in a significant inflow of inexperienced agents during and following periods of house price appreciation.

A back-of-the-envelope calculation from the regression results estimates that in the recent bust sales volume would increase 11% if all agents were in the top tercile of the experience distribution; moreover, as many as 20% of foreclosures would have been avoided.

Using A structural entry and exit model we estimate counterfactuals that incorporates the dynamic decisions of the real estate agents. Several policies are considered: 1) increased entry costs; 2) lower commission rates; and 3) more informed clients. The counterfactual estimations imply that all policies, although through different channels, lead to lower entry rates and a rightward shift of the distribution of experience.

## References


Anenberg, Elliot. 2016. "Information frictions and housing market dynamics." *International Economic Review* 57 (4):1449–1479.

Barwick, Panle Jia and Parag A Pathak. 2015. "The costs of free entry: an empirical study of real estate agents in Greater Boston." *The RAND Journal of Economics* 46 (1):103–145.

Benkard, C Lanier. 2000. "Learning and Forgetting: The Dynamics of Aircraft Production." *The American Economic Review* 90 (4):1034–1054.

Campbell, John Y, Stefano Giglio, and Parag Pathak. 2011. "Forced sales and house prices." *The American Economic Review* 101 (5):2108–2131.

Clementi, Gian Luca and Berardino Palazzo. 2016. "Entry, exit, firm dynamics, and aggregate fluctuations." *American Economic Journal: Macroeconomics* 8 (3):1–41.

Davis, Morris A, Stijn Van Nieuwerburgh et al. 2015. "Housing, Finance, and the Macroeconomy." *Handbook of Regional and Urban Economics* 5:753–811.

Favilukis, Jack, Sydney C Ludvigson, and Stijn Van Nieuwerburgh. 2017. "The macroeconomic effects of housing wealth, housing finance, and limited risk sharing in general equilibrium." *Journal of Political Economy* 125 (1):140–223.

Guerrieri, V and H Uhlig. 2016. "Housing and credit markets: booms and busts." *Handbook of Macroeconomics* 2:1427–1496.

Gupta, Arpit. 2016. "Foreclosure contagion and the neighborhood spillover effects of mortgage defaults." .

Guren, A. 2016. "House price momentum and strategic complementarity. Forthcoming." *Journal of Political Economy* .

Han, Lu and Seung-Hyun Hong. 2011. "Testing cost inefficiency under free entry in the real estate brokerage industry." *Journal of Business & Economic Statistics* 29 (4):564–578.

Head, Allen, Huw Lloyd-Ellis, and Hongfei Sun. 2014. "Search, liquidity, and the dynamics of house prices and construction." *The American Economic Review* 104 (4):1172–1210.

Hendel, Igal, Aviv Nevo, and François Ortalo-Magné. 2009. "The Relative Performance of Real Estate Marketing Platforms: MLS versus FSBOMadison. com." *American Economic Review* 99 (5):1878–1898.

Hong, Harrison and Jeremy C Stein. 1999. "A unified theory of underreaction, momentum trading, and overreaction in asset markets." *The Journal of finance* 54 (6):2143–2184.

Hopenhayn, Hugo A. 1992. "Entry, exit, and firm dynamics in long run equilibrium." *Econometrica: Journal of the Econometric Society* :1127–1150.

Hsieh, Chang-Tai and Enrico Moretti. 2003. "Can free entry be inefficient? Fixed commissions and social waste in the real estate industry." *Journal of Political Economy* 111 (5):1076–1122.

Kellogg, Ryan. 2011. "Learning by Drilling: Interfirm Learning and Relationship Persistence in the Texas Oilpatch." *The Quarterly Journal of Economics* 126 (4):1961–2004. URL https://ideas.repec.org/a/oup/qjecon/v126y2011i4p1961-2004.html.

Lee, Yoonsoo and Toshihiko Mukoyama. 2015. "Entry and exit of manufacturing plants over the business cycle." *European Economic Review* 77:20–27.

Levitt, Steven D., John A. List, and Chad Syverson. 2013. "Toward an Understanding of Learning by Doing: Evidence from an Automobile Assembly Plant." *Journal of Political Economy* 121 (4):643–681.

Levitt, Steven D and Chad Syverson. 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.

Lin, Zhenguo, Eric Rosenblatt, and Vincent W Yao. 2009. "Spillover effects of foreclosures on neighborhood property values." *The Journal of Real Estate Finance and Economics* 38 (4):387–407.

Mian, Atif, Amir Sufi, and Francesco Trebbi. 2015. "Foreclosures, house prices, and the real economy." *The Journal of Finance* 70 (6):2587–2634.

Moen, Espen R. 1997. "Competitive Search Equilibrium." *Journal of Political Economy* 105 (2):385–411.

Rogerson, Richard, Robert Shimer, and Randall Wright. 2005. "Search-Theoretic Models of the Labor Market: A Survey." *Journal of Economic Literature* 43:959–988.

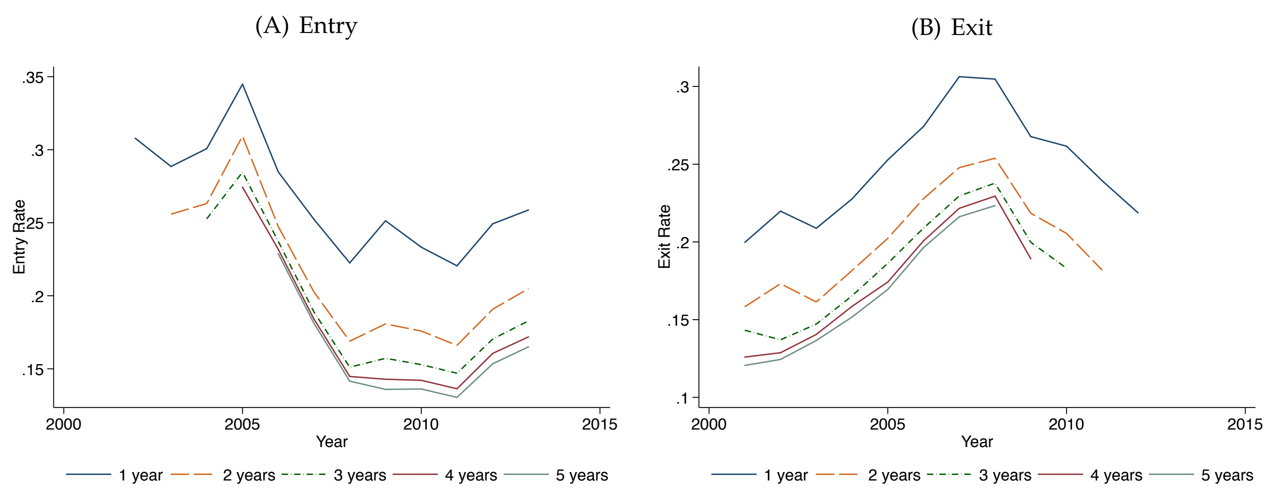
Shimer, Robert. 1996. "Contracts in a frictional labor market." Tech. rep., MIT mimeo.

Weintraub, Gabriel Y, C Lanier Benkard, and Benjamin Van Roy. 2010. "Computational methods for oblivious equilibrium." *Operations research* 58 (4(2)):1247–1265.

# A Entry and Exit Rates

Our data lets us observe selected activity of agents (listings on the seller side and successfully purchased homes on the buyer side) and we do not directly know whether an "inactive" agent has exited the market or was unsuccessful at getting any clients. We also acknowledge that some real estate agents might leave the market temporarily and then come back when housing conditions are more favorable for intermediaries. To examine these channels Figures 1(A) and 1(B) plot entry/exit rates defined as a fraction of currently active agents who are not active in the previous/next $n$ years. A wider window lets us more accurately define exit and avoid marking re-entering agents as new. It also limits the amount of data that we can use. Moreover, as discussed in the paper, if there is significant discounting in accumulation of knowledge (such as being familiar with contemporary market conditions, having a client base and being connected to a network of professions), a re-entering agent might not necessarily have an advantage over a newly licensed one. Taking into account the costs and the benefits (both rates change significantly from $n = 1$ to $n = 2$, but change less for larger $n$'s), we settle on choosing a 2 year window for our definition of entry and exit for both our descriptive analysis and model calibration.

**Figure A1:** Entry and Exit




**Note: Panels A and B plot entry and exit rates respectively for various definitions of thereof. For, $n \in \{1, 2, .., 5\}$ we define entry/exit rates as a fraction of currently active agents who are not active in the previous/next $n$ years.**

# B On Experience Measure

Explored here are different measures of experience available in the data. For each agent, we observe their activity in every year - the number of listings they originated in that year, a fraction of those listings that sold, and the number of buyers that they have represented in a sale closed in that year[18].

[18] All of these statistics can be computed by location and property characteristics as well. This suggests that to assess an outcome for a particular property, one might weight the relevant experience (in same neighborhood or same type of property) more than other. We address this by exploring a neighborhood where all houses are near identical (priced within 10% of each other) in Appendix F. Agents operating in this neighborhood have experience almost exclusively with these homogeneous properties, thus our baseline experience measure is equivalent to the location- and type- specific measure.

We are interested in constructing a measure that is most predictive of our variables of interest: the number of clients that each agents gets each year, and the outcomes of the listings. In addition, we are interested in a measure that makes most use of the data available.

Table A1 illustrates an exercise where we regress the number of clients that an agent has in a particular year on several measures of experience. First column represents out preferred specification, which measure experience as the number of clients that an agent had in the previous year. In Column 2 we explore whether it matters that some of these clients were buyer and some sellers. While seller activity seems to weigh more in predicting the number of clients in the subsequent year, the coefficients are similar, and the fit does not improve much from our preferred specification. We next consider whether it is important to differentiate sellers into those who successfully sold their home and those who didn't. Regression in Column 3 suggests that unsold properties seem to influence current activity less than successful sales. However, again, the predictive power of this regression does not improve enough to justify considering unsold listings separately. In Columns 4 and 5 we test whether activity prior to last year has predictive power for current activity. The results suggest that both clients in the past year and in the past two and three years have predictive power, however the coefficients on second and third lag variables are small and the explanatory power of this regressions is almost identical to the preferred specification. Another measure of experience we could explore for a subsample of the data is the number of years since entry. Excluded in this subsample would be agents that we do not observe entering in the data. We add this measure to our comparison analysis in Column 6 and for a fair comparison re-do out preferred specification on the same subsample in Column 7[19]. Years since entry does not capture nearly as much variation as the baseline specification.

To see how the choice of experience measure affects our prediction for probability of sale, we construct different measures of experience and repeat the baseline regression on the ovariable of interest - probability of sale. Table **??** presents the results. We regress sale probability on the log of experience measure plus one, controlling for housing characteristics, and adding zip code by list month fixed effects. Eight experience measures are as follows: 1) baseline measure, sum of all clients in the previous year, 2) sum of all clients in the previous two years, 3) sum of all clients in the previous three years, 4) discounted sum of clients in the previous two years (discount factor 0.5), 5) discounted sum of clients in the previous three years (discount factor 0.5), 6) number of listings in the previous year, 7) number of sales in the previous year, 8) number of active years since entry in our data. Using the subsample of data used in Column 8, we re-run our preferred specification in Column 9.

All of the measures have almost identical explanatory power ($R^2$ in Column 8 is best comparable to one in Column 9). Since the baseline specification allows us to use the most of our data and is easy to implement in the model, we confirm that is it the best choice of experience measure for our analysis.

[19] We also tried exploring non linear relationship between current clients and years since entry. For that we treated years since entry as a categorical variable. It did not change the results or the conclusion

**Table A1:** Experience Measures and Number of Clients

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| Clients (t-1) | 0.77*** | | | | | | 0.75*** |
| | (0.00) | | | | | | (0.00) |
| Buyers (t-1) | | 0.70*** | 0.72*** | 0.64*** | 0.64*** | | |
| | | (0.00) | (0.00) | (0.00) | (0.00) | | |
| Sellers (t-1) | | 0.80*** | 0.88*** | 0.76*** | 0.76*** | | |
| | | (0.00) | (0.00) | (0.00) | (0.00) | | |
| Failed Sellers (t-1) | | | -0.12*** | | | | |
| | | | (0.00) | | | | |
| Buyers (t-2) | | | | 0.10*** | 0.09*** | | |
| | | | | (0.00) | (0.00) | | |
| Sellers (t-2) | | | | 0.04*** | 0.02*** | | |
| | | | | (0.00) | (0.00) | | |
| Buyers (t-3) | | | | | 0.01*** | | |
| | | | | | (0.00) | | |
| Sellers (t-3) | | | | | 0.03*** | | |
| | | | | | (0.00) | | |
| Years Active | | | | | | 0.78*** | |
| | | | | | | (0.00) | |
| $R^2$ | 0.5155 | 0.5161 | 0.5213 | 0.5172 | 0.5173 | 0.1336 | 0.4438 |
| Fips Effect | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| $N$ | 1843865 | 1843865 | 1521838 | 1843865 | 1843865 | 1177450 | 1177450 |

**This table shows regressions of number of client and agent has in the current period on several possibly informative variables on prior activity. In Column 2 the dependent variable is the sum of all clients (both buyers and sellers) in the previous year, Column 2 regresses current activity on lagged buyer and seller client count separately. Column 3 adds unsuccessful sales. In Columns 4 and 5 we test whether more than one lag matters for additional explanatory power. In Column 6 we instead look at how many years the agent has been active since entry in our data. Column 7 repeats specification of Column 1 with a subsample of data used in Column 6.**

**Table A2:** Experience Measures and Sale Probability

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| Log (Exp1 + 1) | 0.029*** | | | | | | | | 0.025*** |
| | (0.002) | | | | | | | | (0.002) |
| Log (Exp2 + 1) | | 0.026*** | | | | | | | |
| | | (0.001) | | | | | | | |
| Log (Exp3 + 1) | | | 0.025*** | | | | | | |
| | | | (0.001) | | | | | | |
| Log (Exp4 + 1) | | | | 0.028*** | | | | | |
| | | | | (0.001) | | | | | |
| Log (Exp5 + 1) | | | | | 0.028*** | | | | |
| | | | | | (0.001) | | | | |
| Log (Exp6 + 1) | | | | | | 0.062*** | | | |
| | | | | | | (0.003) | | | |
| Log (Exp7 + 1) | | | | | | | 0.029*** | | |
| | | | | | | | (0.002) | | |
| Log(Years Active +1) | | | | | | | | 0.030*** | |
| | | | | | | | | (0.004) | |
| $R^2$ | 0.3433 | 0.3434 | 0.3434 | 0.3434 | 0.3434 | 0.3503 | 0.3432 | 0.4436 | 0.4448 |
| Time X Zip Effect | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| House Characteristics | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| $N$ | 1547580 | 1547580 | 1547580 | 1547580 | 1547580 | 1547580 | 1547580 | 770048 | 770048 |

**In Column 1 we regress sale probability on the log of experience measure plus one, controlling for housing characteristics, and adding zip code by list month fixed effects. Next collumns correspond to the same analysis for different experience measures: 2) sum of all clients in the previous two years, 3) sum of all clients in the previous three years, 4) discounted sum of clients in the previous two years (discount factor 0.5), 5) discounted sum of clients in the previous three years (discount factor 0.5), 6) number of listings in the previous year, 7) number of sales in the previous year, 8) number of active years since entry in our data. Using the subsample of data used in Column 8, we re-run our preferred specification in Column 9.**

## C Outcomes

In this section we present regression results with the fulls set of specifications for all outcome variables of interest.

**Table A3:** Days on Market

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Log(Exp+1) | -2.035*** | -2.438*** | -3.063*** | -3.088*** | -3.072*** | -3.061*** |
| | (0.432) | (0.514) | (0.565) | (0.750) | (0.760) | (0.751) |
| Bust X Log(Exp+1) | -3.375*** | -2.729*** | -2.700*** | -2.991*** | -2.952*** | -3.013*** |
| | (0.630) | (0.393) | (0.464) | (0.736) | (0.733) | (0.744) |
| Medium X Log(Exp+1) | -1.931*** | -1.109*** | -1.252*** | -0.963 | -0.812 | -0.971 |
| | (0.467) | (0.314) | (0.371) | (0.877) | (0.851) | (0.876) |
| Inferred Price | | | | 4.115** | | |
| | | | | (1.959) | | |
| Client Equity | | | | | -35.168*** | |
| | | | | | (8.392) | |
| $R^2$ | 0.0887 | 0.2968 | 0.3283 | 0.3840 | 0.3864 | 0.3837 |
| Time Effect | Yes | - | - | - | - | - |
| Zip Effect | Yes | - | - | - | - | - |
| Time X Zip Effect | No | Yes | Yes | Yes | Yes | Yes |
| House Characteristics | No | No | Yes | Yes | Yes | Yes |
| $N$ | 2524494 | 2524494 | 1957519 | 680991 | 680991 | 680991 |

**Note: This table displays several specification of regression outlined in equation 1. Column one includes fixed effects for zip code and list month, Column 2 instead includes zip code by list month fixed effects, Column 3 adds controls for house characteristics. In Column 4 we consider unobserved heterogeneity by computing inferred price for repeat sales (previous price appreciated using zip code and price tier specific Zillow appreciation rates). Column 5 includes a proxy for client equity (the percent appreciation since last purchase). Column 6 runs specification of Column 3 with the repeat sale sample so that it is comparable to columns 4 and 5.**

**Table A4:** Days to Sale

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Log(Exp+1) | -1.093*** | -1.469*** | -2.108*** | -2.178*** | -2.209*** | -2.155*** |
| | (0.291) | (0.357) | (0.425) | (0.582) | (0.605) | (0.585) |
| Bust X Log(Exp+1) | -3.114*** | -2.230*** | -1.828*** | -1.637* | -1.643* | -1.656* |
| | (0.602) | (0.381) | (0.462) | (0.851) | (0.828) | (0.857) |
| Medium X Log(Exp+1) | -1.702*** | -0.872** | -0.720* | -0.635 | -0.519 | -0.632 |
| | (0.507) | (0.360) | (0.400) | (0.863) | (0.866) | (0.869) |
| Inferred Price | | | | 4.109** | | |
| | | | | (1.731) | | |
| Client Equity | | | | | -36.888*** | |
| | | | | | (7.588) | |
| $R^2$ | 0.0743 | 0.3611 | 0.3893 | 0.4726 | 0.4759 | 0.4722 |
| Time Effect | Yes | - | - | - | - | - |
| Zip Effect | Yes | - | - | - | - | - |
| Time X Zip Effect | No | Yes | Yes | Yes | Yes | Yes |
| House Characteristics | No | No | Yes | Yes | Yes | Yes |
| $N$ | 1429458 | 1429458 | 1206408 | 417255 | 417255 | 417255 |

**Note: This table displays several specification of regression outlined in equation 1. Column one includes fixed effects for zip code and list month, Column 2 instead includes zip code by list month fixed effects, Column 3 adds controls for house characteristics. In Column 4 we consider unobserved heterogeneity by computing inferred price for repeat sales (previous price appreciated using zip code and price tier specific Zillow appreciation rates). Column 5 includes a proxy for client equity (the percent appreciation since last purchase). Column 6 runs specification of Column 3 with the repeat sale sample so that it is comparable to columns 4 and 5.**

**Table A5:** List Price

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Log(Exp+1) | -0.000 | 0.005 | -0.008*** | -0.015*** | -0.014*** | -0.014*** |
| | (0.003) | (0.003) | (0.002) | (0.004) | (0.004) | (0.004) |
| Bust X Log(Exp+1) | -0.014* | -0.016*** | -0.012*** | -0.012*** | -0.012*** | -0.012*** |
| | (0.007) | (0.002) | (0.002) | (0.003) | (0.003) | (0.003) |
| Medium X Log(Exp+1) | -0.008** | -0.005*** | -0.004* | -0.002 | -0.002 | -0.002 |
| | (0.004) | (0.002) | (0.002) | (0.004) | (0.004) | (0.004) |
| Inferred Price | | | | 0.133*** | | |
| | | | | (0.023) | | |
| Client Equity | | | | | -0.013 | |
| | | | | | (0.026) | |
| $R^2$ | 0.5429 | 0.6625 | 0.8700 | 0.8928 | 0.8865 | 0.8865 |
| Time Effect | Yes | - | - | - | - | - |
| Zip Effect | Yes | - | - | - | - | - |
| Time X Zip Effect | No | Yes | Yes | Yes | Yes | Yes |
| House Characteristics | No | No | Yes | Yes | Yes | Yes |
| $N$ | 2520159 | 2520159 | 1956672 | 672828 | 672828 | 672828 |

**Note: This table displays several specification of regression outlined in equation 1. Column one includes fixed effects for zip code and list month, Column 2 instead includes zip code by list month fixed effects, Column 3 adds controls for house characteristics. In Column 4 we consider unobserved heterogeneity by computing inferred price for repeat sales (previous price appreciated using zip code and price tier specific Zillow appreciation rates). Column 5 includes a proxy for client equity (the percent appreciation since last purchase). Column 6 runs specification of Column 3 with the repeat sale sample so that it is comparable to columns 4 and 5.**

**Table A6:** Close Price

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Log(Exp+1) | 0.005 | 0.010** | -0.003 | -0.010* | -0.009* | -0.009* |
| | (0.003) | (0.004) | (0.002) | (0.005) | (0.005) | (0.005) |
| Bust X Log(Exp+1) | -0.021** | -0.017*** | -0.010*** | -0.009** | -0.010*** | -0.010*** |
| | (0.008) | (0.004) | (0.002) | (0.004) | (0.003) | (0.003) |
| Medium X Log(Exp+1) | -0.007* | -0.004* | -0.001 | 0.000 | -0.000 | 0.000 |
| | (0.004) | (0.003) | (0.001) | (0.002) | (0.002) | (0.002) |
| Inferred Price | | | | 0.135*** | | |
| | | | | (0.024) | | |
| Client Equity | | | | | 0.045** | |
| | | | | | (0.020) | |
| $R^2$ | 0.5665 | 0.7216 | 0.8906 | 0.9177 | 0.9129 | 0.9128 |
| Time Effect | Yes | - | - | - | - | - |
| Zip Effect | Yes | - | - | - | - | - |
| Time X Zip Effect | No | Yes | Yes | Yes | Yes | Yes |
| House Characteristics | No | No | Yes | Yes | Yes | Yes |
| $N$ | 1431852 | 1431852 | 1211627 | 414291 | 414291 | 414291 |

**Note: This table displays several specification of regression outlined in equation 1. Column one includes fixed effects for zip code and list month, Column 2 instead includes zip code by list month fixed effects, Column 3 adds controls for house characteristics. In Column 4 we consider unobserved heterogeneity by computing inferred price for repeat sales (previous price appreciated using zip code and price tier specific Zillow appreciation rates). Column 5 includes a proxy for client equity (the percent appreciation since last purchase). Column 6 runs specification of Column 3 with the repeat sale sample so that it is comparable to columns 4 and 5.**

# D Experience Advantage and Probability of Sale

Suppose there are $s$ houses for sale and $b$ buyers who each decide to view one house at random. The probability that any particular house is visited by at least one buyer is $1 - \left(1 - \frac{1}{s}\right)^b$ - the complimentary probability to that of an outcome where every buyer chooses to view another house. An approximation of this match probability for large numbers of $s$ and $b$ is $1 - e^{-\theta}$, where $\theta = b/s$. The number of total matches that will be made, or match function, is $m(b,s) = s(1 - e^{-\theta})$. As $\theta \to \infty$ or $\theta \to 0$, this function approaches a Leontief formulation. Intuitively, if there are very few houses relative to the number of buyers, most houses will be visited and $s$ matches will be made. Similarly, if there are very few buyers relative to the number of houses, each buyer is likely to visit a distinct house, so the number of matches will be $b$. For $\theta$'s outside the extreme range however, there are inefficiencies associated with the lack of coordination among the buyers. Since they can not ex-ante agree to each visit a separate house, there will be houses that have multiple buyers and some that will end up with none.

Imagine now that instead of visiting sellers, a buyer visits real estate agents. Then a real estate agent can schedule buyer visits to one house in their inventory. If the inventories consist of one seller per agent, the matching function resulting in this set up is exactly the same as in the buyer - seller matching problem. However if an agent has more then one house, the coordination inefficiency is reduced due to the ability of an agent to perfectly coordinate the buyers within their housing stock. At the extreme, if there is only one agent, the match function is Leontief for any ratio of buyers and sellers: an agent will assign one house per each buyer until either the buyers of houses run out. More generally, if there are $b$ houses and $a$ agents with $l$ listings each, and if $b$ and $a$ is a large number. We can approximate the probability of match for each seller as

$$\begin{aligned}
\mu^l(a,b) &= \sum_{i=1}^{l}\left(e^{-b/a}\frac{(b/a)^i}{i!}\frac{i}{l}\right) + \left(1 - \sum_{i=0}^{l}\left(e^{-b/a}\frac{(b/a)^i}{i!}\right)\right) \\
&= 1 - \sum_{i=0}^{l}\left(e^{-b/a}\frac{(b/a)^i}{i!}\frac{l-i}{l}\right)
\end{aligned}$$

**Proposition 1.** $m^1(a,b) < m^l(a/l,b), \forall l > 1$

*Proof.* We can restate the original problem by considering agents who have $l$ listings each, but buyers who are bypassing the agents and looking at houses directly. Then the probability of each particular house to be visited is as follows:

$$\mu(la,b) = \sum_{i=1}^{\infty} e^{-b/a}\frac{(b/a)^i}{i!}\left(1 - \left(1 - \frac{1}{l}\right)^i\right)$$

The arrival of buyers to agents is still a poisson distributed variable. For each realization of it, buyers are randomly landing on each house in the inventory, thus if $i$ buyers arrive for a particular agent, the conditional probability of at least on match is $1 - (1 - 1/l)^i$. If however the agents can direct the buyers, they can avoid the congestion of many buyers randomly deciding to visit the same house and instead either assign one buyer for each house or ration the houses among buyers. Thus the conditional probability of match is $min(i/l, 1)$

$$\mu^l(a,b) = \sum_{i=1}^{\infty} e^{-b/a} \frac{(b/a)^i}{i!} \min\left\{1, \frac{i}{l}\right\}$$

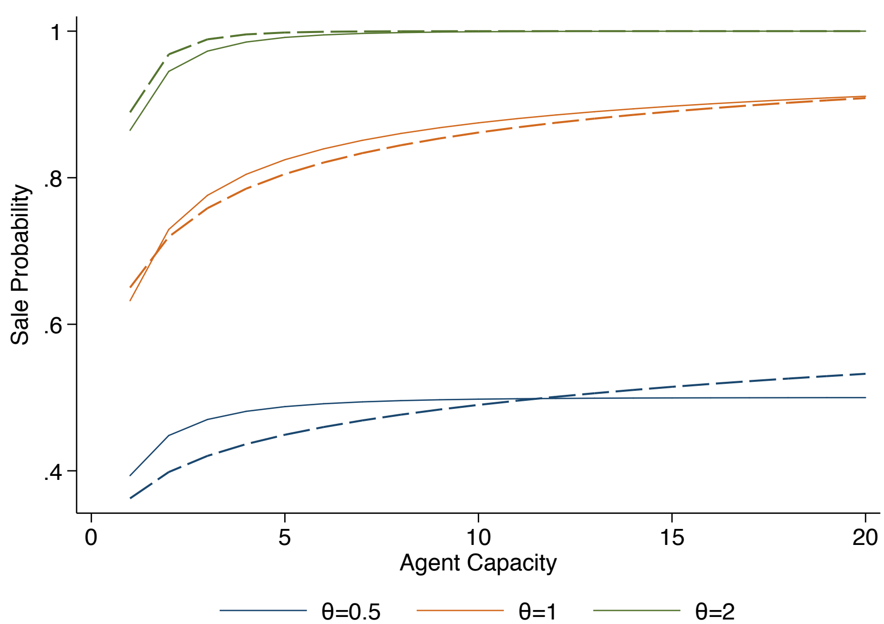
At $i = 0$, the expressions in the sum are the same and equal to 0. However as $i$ increases, $m^l(a,b)$ increases faster than $m(la,b)$. We can see that from computing the slope of the part that differs in the too expressions with respect to $i$.

$$\frac{d}{di}\left(1 - \left(1 - \frac{1}{l}\right)^i\right) = -\left(1 - \frac{1}{l}\right)^i log\left(1 - \frac{1}{l}\right) <^{20} \left(1 - \frac{1}{l}\right)^i \frac{1}{l} < \frac{d}{di}\frac{i}{l} = \frac{1}{l}$$

Note than when $\min\left\{1, \frac{i}{l}\right\}$ reaches 1, it is always larger than $0 < \left(1 - \left(1 - \frac{1}{l}\right)^i\right) < 1$. Since $m^l(a,b) = la\mu^l(a,b)$ and $m(la,b) = la\mu(la,b)$, the inequality in the proposition holds. □

We have shown that markets where agents have larger networks are thus more efficient at producing matches. Let us now fix the number of sellers $s$ and buyers $b$ and explore how the probability of match $\mu^l(s/l,b)/s$ varies with capacity of agents $l$. Note first, that the coordination problem that agents solve is more of an issue then $s$ is similar to $b$, so improvement in efficiency will vary depending on the market tightness. Also, the maximum possible number of matches is the minimum of $s$ and $b$, so improvement in efficiency are bounded. Figure A2 plots the $\mu^l(s/l,b)/s$ for various values of $\theta = s/b$.

**Figure A2:** Agent Capacity and Efficiency Improvement




**Note: This plot graphs the probability of sale for houses in market with different agent capacity holding market tightness (the ratio of buyers to sellers) fixed. The three solid lines represent different values for buyer to seller ratios $\theta$. The dashed lines represent the matching function set up used in the model. We allow for $\theta$ to vary across $l$, and $\lambda_2$ vary across states.**

For a fixed $\theta$ the probability of sale for each value of agent capacity is a concave function approaching a constant. This relationship can be approximated by the functional form that we assume in the model: $\mu(exp) = 1 - e^{-\lambda_1 exp^{\lambda_2}\theta}$. Since different aggregate states imply different market tightness (ratio of buyers to sellers), we allow the curvature $\lambda_1$ to change with the state. Here $\lambda_2$ represents the experience advantage. For the illustration above, we can calibrate $\lambda_1(z)$ and $\lambda_2$ to match the relationship that is delivered by the micro-founded model. While $z$ represents varying $\theta$ in our toy model, in the baseline set up buyers have more incentives to go into markets that are more efficient, so for the overall market tightness $n_t^b/n_t^s$, each market will have it's own ratio of buyers to sellers which will be larger for more efficient agents. In the dashed lines, Figure A2 then plots the model specification where we allow for $\lambda_1$ to vary across the three levels, but within each level, $\theta$ increases with $l$. We can see that our model approximates well the micro founded model described above.

# E Solution Algorithm for the Baseline Model

$\lambda(w) = \tilde{\lambda}(w)$ for all $w$ : guess entry rate

$\rho(e,w) = \tilde{\rho}(e,w)$ for all $e,w$ : guess exit policy

$n^a(e,w) = \tilde{n}^a(e,w)$ for all $e,w$ : guess distribution of agents

$\tilde{V}_\rho(e,w)$, for all $w,e$: compute value functions consistent with $\rho$

$n = 0$

**repeat**

**repeat**

Given $n^a(e,w)$, compute $s(e,w), b(e,w)$ - distribution of clients

Given $s, b, \rho, T$ (transition probability matrix for $w$) compute transition probabilities over the entire state space $P$

Compute new distribution $n^{a*}(e,w) = \lambda[P^0 + P^1 + ... + P^{40}]$

$\Delta_1 = ||n^{a*} - n^a||$, update $n^a = n^{a*}$

**until** $\Delta_1 < \epsilon$

Solve for optimal prices and probabilities of sale

Compute expected profit and $V^*(e,w|\rho,\lambda) = E[\pi] + \beta E[\max\{0, -c + V(e',w'|\rho,\lambda)\}]$

$\lambda^*(w) = \lambda(w)\frac{V(0,w|\rho,\lambda)+c_e}{c_e}$ *for all* $w$

$\lambda = \lambda + (\lambda^* - \lambda)/(n^{\delta_1} + N_1)$

$$\rho^* = \begin{cases} 1 & \text{if } c > V^*(e,w|\rho,\lambda) \\ 0 & \text{if } c \leq V^*(e,w|\rho,\lambda) \end{cases}$$

$\rho = \rho + (\rho^* - \rho)/(n^{\delta_2} + N_2)$

$\Delta_2 = ||\rho - \rho^*||$, $\Delta_3 = ||\lambda - \lambda^*||$

**until** $\Delta_2 \leq \epsilon_2$ and $\Delta_3 \leq \epsilon_3$

We note here that uniqueness of extended oblivious equilibrium has not been proven. It well may be that there are multiple equilibria associated with the same set of parameters. However with multiple different starting points, we were unable to find more than one equilibrium. Furthermore, for our exercise we are only aiming at finding an equilibrium that is closest to the data and are not interested in multiplicity per se.

# F Homogeneous Market

This section repeats the empirical analysis for a homogeneous market of 3-bedroom houses in Chula Vista, California. We picked this market based on the following criteria: 1) each year the standard deviation of a list price is less than 20% of the mean price, indicating that the differences between properties are fairly small 2) we have a relatively large number of observations.

Figure A3 shows the satellite view of this area illustrating the homogeneity of properties.

**Figure A3:** Satellite View of Chula Vista, CA


**Note: This image shows a satellite view of Chula Vista, CA.**

Table A7 presents the results.

**Table A7:** Experience and outcomes: Chula Vista, CA

| | (1)<br>Sale Probability | (2)<br>Days on Market | (3)<br>Days to Sale |
|---|---|---|---|
| Log(Exp+1) | 0.008 | -3.548** | -3.061** |
| | (0.007) | (1.580) | (1.382) |
| Bust X Log(Exp+1) | 0.027*** | -13.386*** | -10.600*** |
| | (0.009) | (2.122) | (1.961) |
| Medium X Log(Exp+1) | 0.041*** | -1.697 | -1.587 |
| | (0.013) | (3.072) | (2.931) |
| $R^2$ | 0.156 | 0.115 | 0.134 |
| Time Effect | Yes | Yes | Yes |
| House Characteristics | Yes | Yes | Yes |
| $N$ | 10293 | 10202 | 7453 |

**Note: This table displays our preferred specification of regression outcomes in equation 1 for several variables: sale probability, days on market, and days to sale.**

**Table A8:** Experience and prices: Chula Vista, CA

| | Price (Log) | | |
|---|---|---|---|
| | List | Sale | Frac. Discount |
| Log(Exp+1) | 0.012*** | 0.009*** | 0.028*** |
| | (0.002) | (0.002) | (0.008) |
| Bust X Log(Exp+1) | -0.009*** | -0.003 | -0.042*** |
| | (0.003) | (0.003) | (0.012) |
| Medium X Log(Exp+1) | -0.001 | 0.004 | 0.011 |
| | (0.004) | (0.004) | (0.017) |
| $R^2$ | 0.828 | 0.839 | 0.152 |
| Time Effect | Yes | Yes | Yes |
| House Characteristics | Yes | Yes | Yes |
| $N$ | 10107 | 7339 | 7351 |

**Note: This table displays our preferred specification of regression outcomes in equation 1 for several variables: sale probability, days on market, and days to sale.**