NBER WORKING PAPER SERIES

HETEROGENEOUS REAL ESTATE AGENTS AND THE HOUSING CYCLE

Sophia Gilbukh
Paul Goldsmith-Pinkham

Working Paper 31683
http://www.nber.org/papers/w31683

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2023

We thank Luis Cabral, Stijn Van Nieuwerburgh, Petra Moser, David Backus, and John Lazarev for their invaluable support and guidance. Gara Afonso, Edward Coulson, Diego Dariuch, Hanna Halaburda, Andrew Haughwout, Julian Kozlowski, Virgiliu Midrigan, Brian Peterson, Pau Roldan, Thomas Sargent, and Lawrence White provided helpful discussions and suggestions. We also greatly bene ted from comments by Vadim Elenev, Jihye Jeon, Nic Kozeniauskas, David Lucca, Song Ma, Joe Tracy, Jacob Wallace, and anonymous referees. In addition, we thank seminar participants at NYU Stern in Macro and IO, the Federal Reserve Bank of New York, the Federal Reserve Board, the Bank of Canada, Baruch College, London School of Economics, Harvard Business School, Colorado Finance Summit, Pre-WFA Real Estate Meeting, Society of Economic Dynamics, AREUEA International conference, GBRUES, Haverford College, Hong Kong University, SITE Financial Regulation Conference, University of Toronto IO and CREFR groups for valuable comments and suggestions. Sonia Gilbukh gratefully acknowledges the hospitality of the Federal Reserve Bank of New York where she spent the summer of 2017. Part of the work on this paper was completed while Goldsmith-Pinkham was employed by the Federal Reserve Bank of New York. The views expressed are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of New York or the Federal Reserve Board. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Heterogeneous Real Estate Agents and the Housing Cycle
Sophia Gilbukh and Paul Goldsmith-Pinkham
NBER Working Paper No. 31683
September 2023
JEL No. G5,R3

ABSTRACT

The real estate market is highly intermediated, with 90 percent of buyers and sellers hiring an agent to help them transact a house. However, low barriers to entry and fixed commission rates result in a market where inexperienced intermediaries have a large market share, especially following house price booms. Using rich micro-level data on 8.5 million listings and a novel instrumental variables research design, we first show that houses listed for sale by inexperienced real estate agents have a lower probability of selling, and this effect is strongest during the housing bust. We then study the aggregate implications of the distribution of agents' experience on housing market liquidity by building a dynamic entry and exit model of real estate agents with aggregate shocks. We find that 3.7 more listings would have been sold in a flexible commission equilibrium. It would require a six-fold increase in entry costs for real estate agents to achieve this level of liquidity within the fixed commission framework.

Sophia Gilbukh
Baruch College
sophia.gilbukh@gmail.com

Paul Goldsmith-Pinkham
Yale School of Management
165 Whitney Avenue
New Haven, CT 06511
and NBER
paulgp@gmail.com

1 Introduction

The U.S. housing market is subject to strong boom-bust cycles. The collapse prior to the Great Recession provides a particularly severe illustration: from 2006 to 2008, house prices dropped by 18 percent. What is less known is that the drop in price was accompanied by an even larger fall in liquidity: the probability of a house selling within a year of listing fell by 29 percentage points from 66% in 2005 to 47% in 2008.[1] This paper focuses on the role of intermediaries, a prominent feature in this market, in this liquidity collapse. Real estate agents are central to the matching process between buyers and sellers – 88% of home buyers and 89% of home sellers use an agent (National Association of Realtors, 2017b) – but low barriers to entry and fixed commission rates result in a market where inexperienced intermediaries have a large market share, especially following house price booms.

Using micro-level empirical evidence and a dynamic model of entry and exit, we show that listings of inexperienced agents are less likely to sell than those of experienced agents, and that the ubiquity of these inexperienced agents has aggregate implications for the average sale probability of listed properties, which we call housing market liquidity. Moreover, these effects are amplified in the downturns that follow housing booms. Downturns are particularly affected for two reasons: first, the disadvantage of inexperienced agents in selling listings is highest during housing busts. Second, the distribution of experience is skewed towards low experience agents during downturns. This happens because the preceding housing boom attracts new agents into the profession, intensifying competition for clients and hindering the accumulation of experience. Many of these new agents remain during the onset of the downturn, capturing a sizeable market share of listings.

We begin by documenting two empirical facts using a rich micro-level dataset of 8.5 million transactions over the 2001–2014 period on 60 different Multiple Listing Service (MLS) platforms. First, an agent's work experience is highly predictive of how successfully and quickly they can sell homes. All else equal, listings with agents in the 10th percentile of experience sell with a 10 percentage point (pp) lower probability than those listed by agents in the 90th percentile. Second, this difference varies significantly over the housing cycle, ranging from 8.2 pps in the boom to 12 pps in the bust. When compared to the respective average sale probability of 69.1 and 50.1 percent in those periods, the effects correspond to a 11.9 percent and 24.0 percent advantage in liquidity.

[1] Source: authors' calculations using the S&P/Case-Shiller U.S. National Home Price Index and CoreLogic Multiple Listing Service database.

A key challenge in this empirical exercise is the lack of random assignment between listings and agents. As a result, two types of selection bias could confound our results: selection on property (or listing) characteristics and selection on listing client characteristics. For example, a more experienced agent might select to work with easier-to-sell properties or more motivated clients. To partially address these concerns, we control for a rich set of housing characteristics as well as zip-code-by-list-year-month fixed effects, and we present subsample analyses where the selection effects are less likely to be a concern.

Our main solution to selection uses a novel instrumental variables (IV) strategy. This approach takes advantage of two features of the market: 1) homeowners tend to list their home with the same agent that helped them buy the home; and 2) homeowners whose buyer agents have exited tend to draw a new listing agent whose experience, on average, is representative of the overall pool of agents. This first feature – stickiness – means that initial buyer agent experience is highly predictive of subsequent listing agent experience. The second feature – regression-to-the-mean – means that the average experience of listing agents for sellers whose buyer agent has exited the market tends to be the same, irrespective of the (exited) buyer agent's experience. Both of these features on their own may face the selection issues, so we control for the direct effect of each and exploit the combination of the two channels.[2] Our instrument is very predictive of listing agent experience and the resulting estimates from this IV strategy are highly statistically significant and qualitatively similar to our OLS estimates, but are 30% smaller in the bust and 48% smaller in the boom.

We also show the consequences of experience beyond the sale probability of the initial listing. During the housing bust, the ability to quickly sell a home was crucial for homeowners who had difficulty making their mortgage payments. Those who fell delinquent on their mortgages and failed to sell were forced into foreclosure. Listed homes that failed to sell in 2008 had a 5.5 percent chance of going into foreclosure in the next two years as compared to close to zero percent for sold properties. Highlighting the importance of experience in real estate agents, we find that houses that listed in the bust years with inexperienced agents are 0.9 pps more likely to subsequently foreclose (30 percent of the average probability of subsequent foreclosure during that period) compared to those listed with experienced agents. Thus, not only did the inexperienced agents affect individual sale outcomes, but they also contributed to negative externalities on the neighboring properties through the foreclosure channel.[3]

The experience effect on the probability of sale could be a combination of several mechanisms. One salient mechanism is strategic pricing. Since, ceterus paribus, properties with lower list prices are more likely to sell,

[2] This empirical approach is similar to Abaluck et al. (2020), and we provide a theoretical microfoundation for this approach in the Appendix.

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

if experienced agents list properties with lower list prices, they will then have higher listing liquidity. Using repeat sales data, we show that on average, more experienced agents do list properties for lower list prices, leading to slightly lower sale prices. However, the difference in markup on a similar property is very small relative to the overall effect of experience on the probability of sale. Using a back-of-the-envelope calculation, we estimate that the price channel makes up roughly 20 percent of the overall impact of experience on listing liquidity. Hence, in our theoretical model, we focus on the overall effect and do not to distinguish between the mechanisms affecting the experience advantage.

The prevalence of inexperienced real estate agents arises from fixed commission rates and is exacerbated by low entry costs. Commission fees are perceived as “standard" and do not vary over time nor by experience.[4] This causes inefficiencies in several ways: first, the overall fixed level of commissions leads to an inefficiently high level of entry without the benefit of competitive pricing. This effect has been documented in Hsieh and Moretti (2003). Second, the inefficiently high entry causes each agent to work with fewer clients, leading to a slower accumulation of experience. Third, the sluggish accumulation of experience leads to lower total earnings and a more likely exit (and hence permanently lost accrued experience). These last two channels are novel to our paper and highlight the inefficiency caused by crowd-out and exit, in addition to entry.

In the second half of the paper, we use a dynamic structural model to quantify the overall effect of fixed commission rates on housing liquidity and assess a policy intervention that could achieve an efficient liquidity level within the current fixed-commission framework. A structural model allows us to endogenize the experience distribution arising from the entry and exit decisions of real estate agents, as well as the equilibrium accumulation of experience. The model embeds housing search in a dynamic labor market framework of real estate agents with aggregate market fluctuations. Consistent with our empirical findings, entry and exit decisions are affected by house prices, volume of listings, and the market tightness. In addition, agents’ decisions respond to costs of entry, commission structure and their individual experience advantage in the market. We consider how policies change agents’ incentives to enter and exit, resulting in a shift of the equilibrium distribution of experience towards more experienced agents, and thus improving overall market liquidity.

The model features frictional search in the housing market, where agent earnings depend on their experience. Experience has three advantages. First, agents with higher experience work with more clients. We assume that some buyers and sellers 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. Second, experienced agents have access to a more efficient matching technology for their

[4] Barwick and Wong (2019) document remarkably stable real estate commissions over our sample period. Hatfield, Kominers, and Lowery (2020) discuss how this coordination can arise in a bilateral market. We take fixed commission rates as given in our paper.

seller clients and thus have a higher probability of finding them a buyer and earning a commission. Finally, the model assumes that agents with higher experience get to keep a higher portion of their commission when splitting it with the office where they work in. While we do not explicitly model offices, the commission splits are important for agents' pay, so we assume that only a fraction of the earnings is retained by each agent.

We then embed the matching market of housing into an entry and exit model of real estate agents with aggregate market fluctuations. Our setup includes three aggregate states: bust, boom, and medium. Each state corresponds to the number of sellers willing to sell their house and the valuation for houses by the buyers. Agents' decisions to participate as intermediaries depend on aggregate market conditions, competition they face for clients, their success in earning commissions, and the value of accumulating experience and remaining in the industry in the future years. These features generate empirically realistic fluctuations in the overall entry and exit patterns of agents.

The distribution of agent experience depends on the entire history of aggregate state realizations and is a payoff-relevant variable on which real estate agents base exit and entry decisions. Keeping track of the full distribution of experience effectively makes the state space infinite. To address this, 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 by conditioning the experience distribution on the aggregate state in the current and previous period.

Using this dynamic model calibrated to our empirical moments, we consider an efficiency benchmark equilibrium that estimates the model without the main market failure of fixed commissions. In this benchmark equilibrium, agents are allowed to compete on commissions. We assume that agents with no experience work on their first listing for free, whereas higher experienced agents charge a commission rate that reflects higher expected sale probability. We find that in boom states, working with the highest experience agent is worth a commission of 1.5% to the seller (in addition to 3% offered to the buyer side), while in the bust and medium states the premium is as high as 2.6% and 2.8% respectively. In this equilibrium, entry is less attractive due to low starting commissions, but agents who enter are quick to accumulate experience. This leads to an overall higher distribution of experience and an 3.7% aggregate improvement of housing liquidity.

Finally, we consider how policy makers could achieve this efficient benchmark housing liquidity within the current fixed commission framework by directly restricting entry of real estate agents. We find that entry costs would have to increase to $124,000 (more than six times the calibrated baseline value) in order to have the same average impact on liquidity as competitive flexible commissions.

This paper contributes to a literature applying search-and-matching framework to models of housing to understand aggregate housing market fluctuations (Diaz and Jerez, 2013; Head, Lloyd-Ellis, and Sun, 2014; Ngai and Tenreyro, 2014; Anenberg, 2016; Guren, 2018; Hedlund, 2016a,b; Garriga and Hedlund, 2020). Our key contribution relative to this literature is to incorporate the heterogeneity in match technology due to real estate agents' differential experience. This builds upon a large literature, summarized in Han and Strange (2015), which studies the role of real estate agents in search models.[5]

Our paper most closely relates to Barwick and Pathak (2015), who study data from the Greater Boston area for years 1998–2007 and examine inefficiencies associated with cheap entry of real estate agents. Our paper expands on their work in three important ways: first, because our sample covers the full boom and bust period (2001 to 2014), we are able to compare the effect of agent experience during the bust period to the boom. There are large differences, with agent experience being much more important during busts. This also lets us study the impact of agents on subsequent foreclosure. Moreover, our sample covers a nationwide sample, rather than focusing on just one market, ensuring we are not capturing Boston-specific features. Second, we use a new instrumental variables research design to solve issues of selection, which is unique among papers studying the role of real estate agents. Third, our model incorporates several additional endogenous mechanisms arising from the distribution of agent experience. Besides competition for clients, our paper models the accumulation of experience such that the excessive number of agents in the market not only leads to inefficient business-stealing, but also precludes accumulation of experience as agents work with fewer clients. Our model additionally features endogenous entry of home buyers, where a more experienced pool of agents draws in more buyers as the market becomes more efficient at matching.

More broadly, this paper contributes to a literature on the value of real estate agents. 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. Hendel, Nevo, and Ortalo-Magné (2009) compare listing outcomes from an FSBO (for sale by owner) platform to those who were facilitated by an agent. They find that agents provide little value added. Levitt and Syverson (2008) find that agents can obtain a better price when they are selling their own homes rather than those of their clients. These papers abstract from agent heterogeneity, which has a significant impact on home sales. We also present a new identification strategy using the stickiness of agents from purchase to sale to identify the impact of real estate agents on listing outcomes.

This paper also connects to a macrofinance literature studying the significance of expectations, financial conditions, and other frictions in generating and amplifying the housing cycle (see Davis and Van Nieuwer-

[5] Buchak et al. (2020) also study the importance of intermediaries in housing liquidity, but focus on the role of the new emerging "iBuyers" who provide a source of liquidity for sellers.

burgh (2015) and Guerrieri and Uhlig (2016) for literature review on financial frictions and the housing cycle). While the aggregate movements in the housing cycle play an important role in our empirical and theoretical analysis, we take the overall aggregate movements in liquidity across the boom and bust as given, and instead focus on the relative contribution of intermediaries to market liquidity within each period.

The rest of the paper is structured as follows. Section 2 describes industry background. Section 3 describes data and our choice of measure of experience. In Section 4, we present the empirical analysis. Section 5 outlines the model and the calibration exercise. Section 6 presents counterfactual analysis for both the efficiency benchmark equilibrium and the policy exercise. We conclude in Section 7.

2 Real estate agents in the United States

Despite the existence of numerous for-sale-by-owner platforms, the housing market in the US remains highly intermediated, with 87 percent of buyers and 89 percent of sellers hiring an agent to facilitate buying or selling a home (National Association of Realtors, 2017b). There are many reasons why consumers may find agents valuable. First, an agent has access to the local MLS database, which provides detailed information on all the listings currently available in the area and allows sellers to advertise to potential buyers.[6] Second, an agent plays an invaluable role as an adviser. For example, a listing agent suggests improvements, or "staging," to make the property more attractive to buyers, provides input on an appropriate listing price, and advises on whether to accept the incoming offers. Last, an agent gives a client representation in a negotiation process in the final stages of the transaction, making an agreement with the counterparty more likely. Through these three channels, hiring an agent gives access to a more efficient matching technology between home buyers and sellers. Thus, a listing agent not only attracts more buyers to the listing but also makes buyers more likely to bid on the property and facilitates the transaction once a buyer is found.

Despite the important role of real estate agents, the costs to enter the profession are as low as 30 hours of classes and nominal exam and licensing fees.[7] While these classes familiarize agents with essential terminology and state laws, they provide little insight into local real estate markets or into the most effective ways to create transactions. Hence, agents have a substantial room for improvement after entry. In addition to learning about the local housing market and the tacit knowledge of selling, agents accrue an accumulated network of former clients, other agents, and a long list of useful professionals, such as construction workers, plumbers, electricians, mortgage brokers, appraisers, photographers, and interior designers. Tapping into these networks

[6] The creation of web platforms such as Zillow and RedFin has reduced agents' monopoly over the information on available listings, but agents maintain the exclusive ability to list on the MLS to advertise for-sale properties to other agents.

[7] The requirements vary somewhat across states, with class time ranging from 30–90 hours, the exam fees from $25-$150, and the cost of a license from $50-$300.

makes a sale more likely due to an increase of potential counterparties for their clients and by ensuring that the property is "fixed up" and is more desirable for a buyer. Hence, the inexperience of brand-new agents will likely make them worse at selling properties when compared to incumbent experienced agents.[8] This is a key empirical issue that we assess in Section 4.

While there are potentially large differences in the experience of agents, the compensation paid by buyers and sellers to real estate agents does not appear to vary across agents. As highlighted in other work studying agents, commissions in the market appear to be relatively fixed across agents, regardless of agent quality (Hsieh and Moretti, 2003; Barwick and Pathak, 2015; Barwick, Pathak, and Wong, 2017; Barwick and Wong, 2019). The ease of entry and fixed pricing results in many agents entering the industry for short periods of time.

Despite being paid the same commissions as experienced agents, inexperienced agents appear able to attract clients. In 2017, the National Association of Realtors (NAR) found that 74 percent of sellers and 70 percent of buyers signed a contract with the first agent they interviewed (National Association of Realtors, 2017b). While the first agent contacted is not always chosen at random, the survey indicates that clients do not approach the choice decision with much care. One reason may be that clients do not realize the importance of choosing the right agent or find it difficult to gauge experience. Alternatively, with so many people in the profession, clients may personally know someone who is a licensed agent and hire them to avoid social consequences. As a result, as we show below in Section 3, these inexperienced agents have a non-negligible share of the market.

It is not just clients who are affected by the prevalence of new and inexperienced agents. The industry has raised alarms about this phenomenon. In 2015, real estate agents identified the number one challenge to their industry to be "Masses of Marginal Agents Destroy Reputation" in a report commissioned by the NAR: "[t]he real estate industry is saddled with a large number of part-time, untrained, unethical, and/or incompetent agents. This knowledge gap threatens the credibility of the industry." In another report commissioned by *Inman*, an industry periodical, 77 percent of agents responded "low-quality agents" to the question "what are the challenges that the real estate industry is currently facing?"[9]

An inefficiently high number of inexperienced real estate agents is a result of a fixed commission structure combined with low barriers to entry. New agents are quick to enter the market in response to more favorable

[8] Indeed, conversations with brokers reveal a series of anecdotes that suggest a myriad of mistakes made by inexperienced brokers. The absence of these mistakes may be one of the key advantages benefiting experienced agents.

[9] A relevant respondent quote in the *Inman* report: "A great many agents are part-time. Other than the few transactions they finagle out of their family/ friends yearly they have very little to do with the industry and don't care to educate themselves or increase their skills. This is a disservice to their clients and gives real estate professionals a bad name." For more information about the Danger Report commissioned by the NAR, see their website: https://www.dangerreport.com/usa/. The *Inman* report is available here: https://www.inman.com/2015/08/13/special-report-why-and-how-real-estate-needs-to-clean-house/.

market conditions and easily recoup their low cost of entry from limited transactions they intermediate. More competition among agents leads to fewer clients per agent. The excess entry leads to sunk entry costs and decrease in average productivity with no benefit to consumers as there is no competitive pressure on commissions. This effect is described in Hsieh and Moretti (2003). In our paper, we emphasize two additional sources of inefficiency that stem from the importance of agent experience. First, lower agent productivity arising from excess entry means slower accumulation of experience, meaning lower quality service from real estate agents. Second, lower productivity leads to lower profits for agents and a more likely exit from the profession - a permanent loss of accrued experience in the market.

3 Data and measurement

In this section, we describe our data sources and the various sample restrictions that we use. We then discuss how we measure real estate agent experience and summarize our measure.

3.1 Data sources

For our main empirical analysis, we use a comprehensive listing-level dataset on residential properties for sale collected by CoreLogic. The data come from MLS platforms operated by regional real estate boards. Each MLS varies in size but, on average, covers a geographical area that is approximately equal to a commuting zone. Each observation in the data represents a listing on an MLS platform, with a large number of variables describing the property and the status of the listing. These include the date the property is listed, the associated listing agent (as well as secondary agent in some cases), the original list price, the last observed list price, and detailed property characteristics such as the living area, number of bedrooms and bathrooms, number of parking spaces, and age of the structure. If the listing sells, we observe the date of sale, the sale price, and the associated buyer agent. If the property fails to sell, we also observe when the property is de-listed. Crucial for our analysis is that each real estate agent in an MLS is given a unique identifier such that we can track them throughout the sample.

The full CoreLogic MLS dataset has information on over 150 MLS platforms. However, the history for each MLS in this dataset begins at different times due to variation in CoreLogic's contracts with each MLS, with some data beginning as late as 2009. Since we are interested in studying the boom period starting in 2001, we restrict our analysis to the subsample of MLS whose data goes back to 2001. Additionally, due to data quality issues, we drop several MLS whose data have large jumps in the number of listings during the sample period from 2001–2014 (more than 100 percent growth in the number of listings in a given year). This final restriction drops an additional 10 MLS and leaves 60 MLS platforms in our sample. Within these MLS, we exclude listings with asking prices below $1,000. Finally, we focus on real estate agents who have never had more than 200 transactions in one year to avoid potential measurement error in agent identifiers. This leaves us with 8.5 million observations. Appendix Figure J1 shows the coverage map of the final sample. A key feature of our dataset is that while we do not have full coverage of the United States, we have near-exhaustive coverage *within* a geographic location, ensuring that we observe all recorded activity by real estate agents in an area. Over the sample period from 2001 to 2014, we observe 567,230 different agents, with an average of 183,668 active agents in each year.

In Panel A of Appendix Table J1, we compare our selected sample of MLS platforms to the excluded MLS platforms in the raw Corelogic data. We focus on the period 2009-2014 where the two samples have comparable coverage. This exercise provides guidance on how to extrapolate our findings to other locations, but as our identification strategy relies on within-MLS variation, the internal validity of our estimates is not affected by differences documented across locations.

The samples are similar, though listings in our sample tend to sell with lower probability and for lower average prices. The share of listings that sold within one year was 53% in our sample and 58% in the excluded MLS, and the average list price in our sample was 258k dollars while the average listing price in the excluded MLS was 322k dollars. To check whether this was a composition effect within 2009-2014, we also examine the listing differences separately for years 2009 and 2014, and find similar differences.

To further examine external validity of our sample, we compare the counties covered in our data to those in the rest of the United States in Panel B of Appendix Table J1. We consider a county to be in our sample if our data have at least 1000 listings in that county over the period of 2009-2014. We report several demographic and housing related characteristics of the counties (note that these means are equal-weighted). On average, our sample is higher income, with a median household average of 40.2 thousand dollars, compared to 34.1 thousand dollars for the rest of the United States. Our sample's counties experienced a slightly larger housing wealth decline and have population that is slightly more educated and urban.

In addition to the MLS data, our robustness analysis makes use of two additional datasets. First, we use proprietary deed-level data purchased from CoreLogic, which contain information on housing transactions and their associated transaction prices recorded at county deeds offices. Using this data allows us to supplement our analysis in two ways: first, we identify properties that subsequently fall into foreclosure. Second, we identify the previous sale price for a listing, which gives us a way to control for unobserved heterogeneity of properties.

Our second dataset is Zillow's publicly available zip-code-level house price index. We use this data to construct a measure of "inferred price" for listings of previously transacted properties. To do so, we take the listed properties' previous sale price and use the realized house price appreciation in the listing's zip code to identify the approximate market price for the listing.

In Table 1, we present basic summary statistics about our analysis sample. The table contains summary statistics for both our main ordinary least squares (OLS) sample which includes our full 8.5 million observations, and the instrumental variables (IV) sample of 1.2 million observations (the approach discussed in further detail in Section 4.1). This is true for both the main sample and the IV sample. Roughly 30 percent of listings sell within 90 days, 50 percent sell within 180 days, and 61 percent sell within a year. On average, a home is on the market for 147 days. In terms of prices, we find that on average, homes sell roughly 8% above list price, on average (conditional on sale), have list prices about 10% higher than the inferred price based on Zillow indices, and sell for 2% below the inferred price. Roughly one percent of the sample is subsequently in foreclosure in the following 2 years after listing.

Table 1: Summary Statistics

	Main Sample			IV Sample		
	Mean	Median	SD	Mean	Median	SD
Listing sold w/in 30 days	0.03	0.00	0.16	0.02	0.00	0.15
Listing sold w/in 90 days	0.29	0.00	0.46	0.26	0.00	0.44
Listing sold w/in 180 days	0.50	0.00	0.50	0.45	0.00	0.50
Listing sold w/in 365 days	0.61	1.00	0.49	0.56	1.00	0.50
Days to Sale	124.09	95.00	94.65	127.18	98.00	96.18
Days on Market	147.52	112.00	122.23	156.53	122.00	126.69
Log(List Price)	12.12	12.13	0.76	12.11	12.13	0.78
Log(Sale Price)	12.00	12.04	0.80	11.96	12.02	0.84
Log(Sale Price/List Price)	0.08	0.04	0.12	0.09	0.05	0.13
Log(List Price/Inferred Price)	0.10	0.09	0.23	0.09	0.08	0.22
Log(Sale Price/Inferred Price)	-0.02	0.00	0.29	-0.03	0.00	0.27
Foreclosure in 2 years	0.01	0.00	0.10	0.02	0.00	0.12

Note: This table reports summary statistics for our main OLS sample and our subsample used for IV.

3.2 Measurement of experience

We next describe how we measure real estate agent experience. Ideally, our measure captures three features of real estate agent activity. First, our measure should be consistent over the sample period. Thus, a fully backward-looking measure, such as time spent as a real estate agent, will be inaccurate because our information about agents' history is censored in 2001 at the beginning of sample. Second, our measure should be consistent over locations. Hence, using an income-based measure will inaccurately assign higher experience to agents who work in high price areas. Third, the measure should capture as many sources of potential experience as possible.

Our preferred measure is the number of clients an agent had in the previous calendar year, as it closely matches those requirements. This measure captures three types of transactions: the number of listings sold by the agent in the previous year, the number of listings unsold by the agent in the previous year, and the number of buyers represented by this agent in a transaction that closed in the previous year.[10] Our measure of experience is in terms of recent *output*, rather than calendar time since entry, and has a high discount rate so that any clients who were served two or more years prior do not count toward the current experience. This provides a consistent measure that can be calculated across all time periods, except 2001, in our sample. Moreover, our 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 toward the listing agent experience. This helps ensure that markets with higher and lower levels of sales and prices will be counted equally and also uses all transactions that we observe in the data.

In Figure 1(A), we plot the distribution of experience of active agents, pooling across all years in our sample. Notably, almost 30 percent of all agents are completely inexperienced, with no previous clients.[11] In Figure 1(B), we again plot the distribution of experience, this time weighted by the agents' active listings in that year. While inexperienced agents represent fewer listings when compared to their unweighted presence in the market, they still hold considerable market share. Twenty-five percent of listings are handled by agents who had 4 or fewer clients in the past year, and 50 percent are listed with agents with an experience of 12 clients or fewer. In other words, the majority of sellers used a listing agent who worked with one client a month (or less) in the past year. Hence, if experience matters for liquidity, the prevalence of inexperienced agents could have large aggregate effects in the housing market.

In Appendix A, we discuss alternative measures and approaches to measuring experience, such as weighting listings differently depending on sale outcome, discounting older listing differently, or using years since entry for agents where we observe entrance. We find that these alternative measures do not materially affect our results but either limit our sample (due to the longer required time period) or complicate the mapping to a theoretical measure of experience in our model.

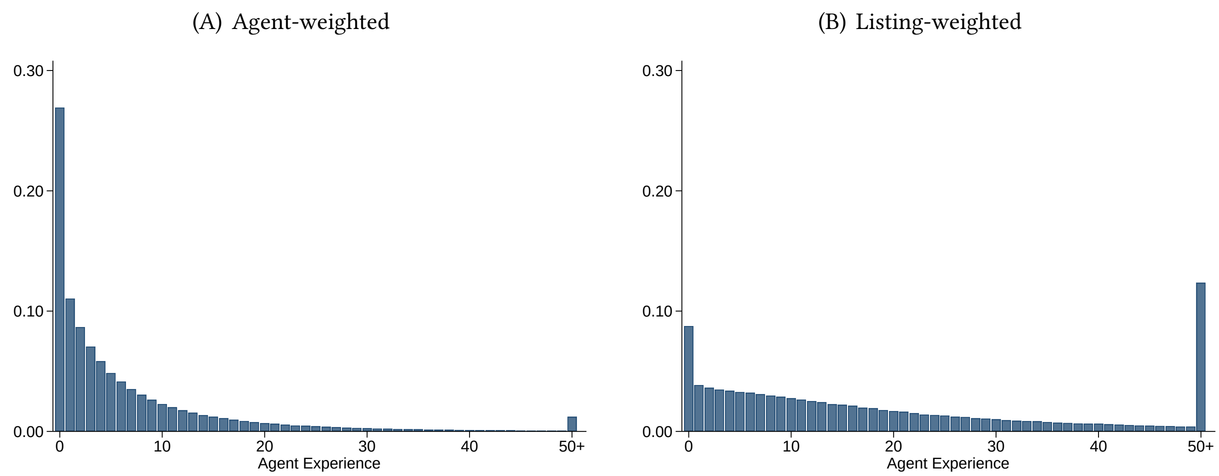
In Appendix Section A, we also explore the properties of our experience measure to make sure that it reflects accumulation in knowledge rather than random variation associated with housing conditions. We find that the experience level for an agent who enters the market gradually grows, conditional on not exiting, but the

[10] We are unable to measure clients with buyers agents who do not buy.

[11] Of these agents with zero experience, 82% are measured as new entrants by our definition of entry (unobserved in our data in the last two years). The remainder were not new entrants, but had no listings last year nor successful purchases. We discuss the definition of entry in more detail in Appendix Section C.

dispersion in experience also grows. Experience increases quicker in the beginning, with slower accumulation later on, reaching a relatively stable level on average three years post-entry.

Figure 1: Distribution of agent experience



Note: This figure plots the distribution of agent experience. Panel A plots the distribution of experience at the agent-year-level. Panel B plots the distribution of experience at the agent-year-level, weighted by the number of listings that an agent participated in that year. In both panels, agents with experience greater than 50 are pooled with agents who have experience of 50. Agent experience is defined as the number of clients an agent worked with in the previous calendar year. See Section 3 for more details on the data sample and definition of experience.

4 Empirical results

In this section, we use our measure of experience to show a strong link between real estate agent experience and listing liquidity that varies over the housing cycle. We then highlight how the effect of experience on liquidity affected foreclosures during the housing bust of 2008–2010. Finally, we discuss the challenge of counterfactually changing agent experience. We show how agent experience itself varies over the cycle and responds endogenously to market conditions, demonstrating the need for a structural model that accounts for agents' endogenous acquisition of experience.

4.1 Estimation approach

To examine the effect of agent experience on listing outcomes, we estimate versions of the following regression:

$$y_{i,t} = \alpha_{i,t} + \sum_{p \in \text{periods}} \beta_p \log(1 + \text{experience}_{i,t}) + \delta W_{i,t} + \epsilon_{i,t}, \tag{1}$$

where $y_{i,t}$ is the outcome for listing i in time t, $\text{experience}_{i,t}$ is the experience of the listing agent for listing i in time t, $W_{i,t}$ is a vector of property-specific controls such as square footage and number of bedrooms, and $\alpha_{i,t}$ denotes time and location fixed effects based on the listing's location (e.g. zipcode-by-listing-year-month fixed effects). For most outcomes, time t indicates the year-month of the listing, except for sale outcomes, where time t denotes the year-month of the sale. To account for the highly skewed distribution of experience, we use log of one plus experience as our main explanatory variable.[12] In all regressions, unless noted otherwise, errors are clustered at the MLS level to account for within-MLS correlation between our experience measure and unobservable shocks (Bertrand, Duflo, and Mullainathan, 2004; Abadie et al., 2017).

In our estimation, we allow the effect of experience to vary by time period. We do this in two ways. First, for graphical illustration, we allow the effect of experience to vary year-by-year and then plot the effect for each year. Second, in anticipation of the calibration of our model in Section 5, we define three time periods—boom, medium, and bust—that reflect the aggregate state of the housing market in each year. The assignment of each year to period is based on 12-month real house price growth, as measured from 1960 to 2017 by the Case-Shiller index, deflated by the Consumer Price Index less costs of shelter. Years with growth rates above the 75th percentile are identified as booms, those below the 25th percentile are busts, and those in between are assigned to a medium period. Appendix Figure J2 illustrates this assignment procedure.[13] In our main tabular results, we report estimates pooled into each of the three time periods.

The challenge for this exercise is lack of random assignment between listings and agents. Selection may confound our OLS results through unobservable characteristics of the listing property or the client. For example, a more experienced agent might work with easier-to-sell properties or with more motivated-to-sell clients. As a result, a regression of probability of sale on agent experience would be biased upward, capturing other features of the homes or clients instead. One approach to alleviate selection is examining the set of alternative explanations and analyze subsamples of data where those selection effects are less likely to be a concern. We include the description of this analysis in Section 4.5 with more details in Appendix Section B.

Our preferred approach to addressing the selection problem is an instrumental variable strategy. This approach uses the two features of the market: first, homeowners are much more likely to list their homes with the agent that they bought the house with; and second, when the real estate agent with whom they bought has exited the market, homeowners tend to get new agents who look like the average population of agents. We now describe this approach in more detail.

Our research design uses a subsample of listings where we observe the previous purchase of the home. Among these listings, the sellers have substantial inertia in agent choice, working with the same agent who represented them in the purchase of the home 33% of the time if that agent is still active. However, if their

[12] To account for data errors and outliers, we focus on agents who had less than 200 clients in the previous year.

[13] Years 2007, 2008, 2009, 2010, and 2011 are assigned to the bust period; years 2006 and 2012 are in the medium period; and years 2002, 2003, 2004, 2005, and 2013 correspond to the boom period.

buying agent is no longer active, the seller is forced to pick a new agent from the market. Importantly, irrespective of the previous buyer's agent experience, the new agent selected by the client tends to have experience representative of the average experience distribution. As a result, if the initial buyer agent was relatively inexperienced compared to the average agent in the market, the seller of the home will experience a positive shock in listing agent experience if their buyer agent has exited between the homeowner's purchase and listing decision. In contrast, if their buying agent was relatively experienced, the seller of the home will experience a downward shock in experience if their buying agent has exited.

In Figure 2(A), we illustrate the impact of inertia and mean reversion in our quasi-experimental design. On the x-axis, we plot the experience of the agent with whom the homeowner initially bought the property, measured in the most recent period ($X_{i,t}$).[14] On the y-axis, we plot the experience of the listing agent used to list the property. The two lines represent the estimated relationship between previous agent's experience and the listing agent's experience, split based on whether the previous agent is still active in the market ($T_{i,t}$). If every homeowner used their buying agent as their listing agent, the "still active" plot would be a 45 degree line. While not quite a 45 degree line, the previous agent's experience is remarkably predictive of the listing agent's experience.

In contrast, for homeowners whose buying agent has exited the market, the line is almost flat, reflecting the mean reversion in the market. This means that buyer agent's experience weakly predicts the subsequent listing agent experience if the buyer agent has exited. This highlights two important features for us: first, that there is the mean-reversion in agent experience that we use for our identification strategy; second, it shows evidence against a theory in which there is an unobserved client characteristic which sorts into high or low experience agents – if there were a characteristic which caused clients to pick high experience agents, then we would see that the subsequent listing agent experience would be highly predicted by the buyer agents' experience if they exited. This offers us some confidence that our assumption of conditional random assignment of agent experience in our OLS estimation is reasonable. However, given that the slope is not zero, there may be some amount of unobserved heterogeneity that we are not able to control for, which drives some of the difference in our IV and OLS estimates.

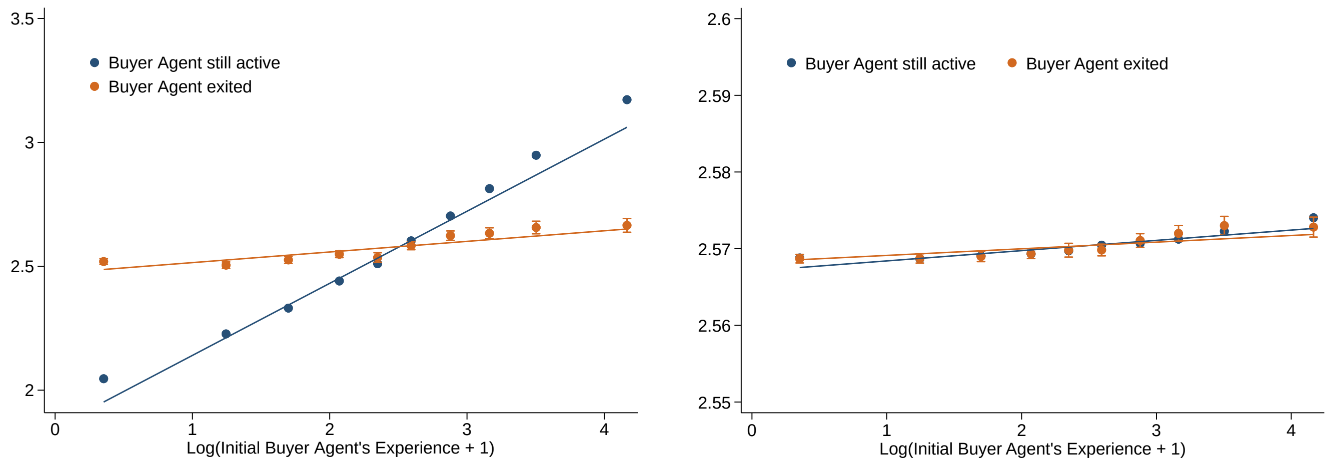
Our identification approach exploits the difference in the slopes of the two lines in Panel A of Figure 2(A). Specifically, we define our instrument as $Z_{i,t} = X_{i,t} \times T_{i,t}$, and control for the direct effect of X_{it} and T_{it}. Intuitively, we control for the direct differences across listings driven by differences in initial buyer agent's experience, and difference across listings where the buyer's agent exits or stays in the market prior to sale.

[14]If the agent has exited, we use their last observed experience.

Figure 2: Instrumental variable strategy with buyer's agent experience & exit

(A) First stage: effect on listing agent experience

(B) Exclusion restriction: effect on predictable agent experience



Note: This figure illustrates the validity of our quasi-experimental design. In Panel A, we plot listing agent experience against the most recent experience of a buyer agent who worked with the seller in the original purchase of the property. In Panel B we do a similar analysis where instead of the actual listing agent experience we plot the predicted experience based on observable housing characteristics. Data in both panels are split by whether the original buying agent is still active at the time of the listing.

The instrument exploits the interaction of the two effects to isolate variation in the listing agent experience. This is analogous to a difference-in-differences approach, where the identification is driven by the interaction, after controlling for the baseline marginal effects.

In Appendix Section I, we lay out a simple model that gives sufficient conditions in which these two features — stickiness and mean-reversion — can identify the effect of experience using IV when OLS is biased due to unobserved seller heterogeneity. In our model, clients are characterized by a preference for experience which is also correlated with housing outcomes. In fact, the interpretation of the small positive slope in Figure 2(A) for clients whose buyer agent has exited is that there exists a positive preference for experience among clients. If this preference for experience varies across clients in a way that is correlated with sale outcomes, simple comparisons of high vs. low experience agents will be biased. In the IV specification, we assume a random exit of buyers agents that is independent of clients' preference for experience, conditional on the initial buyers' agent experience. We then compare the outcomes of clients who had similarly experienced buyer agents when they bought the property, but may have a new listing agent due to a quasi-random exit by their original buyer's agent. Thus, insofar as the endogeneity problem is client or property-specific, it will be controlled for by the experience of the buyer agent who initially worked with the client. We illustrate our IV strategy using a simple example with two experience levels and two types of clients with built-in selection, but this proof can easily extend to more levels of experience or client types.

An important assumption for this approach is that selection into agents is not time-varying in a way that is correlated with the outcomes. Suppose homes A and B were transacted with buyer agents of the same experience. If house B became somehow harder to sell AND the buyer agent for house B has exited the market, then the owner of house B might only be able to contract with a low experience agent for the subsequent sale. Our strategy would then be comparing the outcome of that for-sale listing with the listing of house A that was plausibly listed with the same experienced buyer agent. Then the decreased sale probability would be falsely attributed to differences in the selling agent experience rather than the change in attractiveness between house A and house B that occurred over time. For our IV strategy to be valid, we thus must assume that any selection that occurs at the client or property level is time-invariant.

As in Equation 1, we are interested in the period-by-period effect of experience on different listing outcomes. Since we have three time periods, we have three endogeneous variables, which requires three instruments. We will mimic the setup of Equation 1, and interact our instrument $Z_{i,t}$ and direct controls $X_{i,t}$ and $T_{i,t}$ with time period fixed effects. Formally, this will give us three first stage equations, and one second-stage. For simplicity, as the first stages are symmetric, we present a representative first stage equation and the second stage below:

$$\begin{aligned}\log(1+\text{experience}_{i,t}) \times 1_{t\in p} &= \tilde{\alpha}_{i,t} + \sum_{s\in\text{periods}} \pi_{0,s} Z_{i,t} \times 1_{t\in s} \qquad (2)\\ &\quad + \pi_{1,s} X_{i,t} \times 1_{t\in s} + \delta_{2,s} T_{i,t} \times 1_{t\in s} + \delta_3 W_{i,t} + u_{i,t}\\ y_{i,t} &= \alpha_{i,t} + \sum_{p\in\text{periods}} \beta_p \log(1+\text{experience}_{i,t}) \times 1_{t\in p} \qquad (3)\\ &\quad + \delta_{1,p} X_{i,t} \times 1_{t\in p} + \delta_{2,p} T_{i,t} \times 1_{t\in p} + \delta_3 W_{i,t} + \epsilon_{i,t},\end{aligned}$$

Note that we directly control for $X_{i,t}$ (experience of the agent that the homeowner purchased the home with) and T_{it} (whether that agent exited) in both the first and second stage equations. The excluded variable, $Z_{i,t}$, provides our identifying variation. We include purchase-year-by-listing-year-by-zipcode fixed effects, such that we are comparing two homeowners who have purchased and subsequently listed in the same years, with similar housing market dynamics. The remaining controls, $W_{i,t}$, are similar to Equation 1.

As with all IV regressions, the necessary assumptions are relevance — the instrument predicts experience — and exclusion — this process only affects housing market outcomes through agent experience. Our approach, as shown in Figure 2(A) clearly satisfies the relevance assumption. We formally report the first stage coefficients for Equation 2 in Appendix Table J2, which has a first-stage F-statistic of 130, 112 and 131 for each of the endogeneous variables, satisfying formal cut-offs for a strong first stage set of instruments (Lee et al., 2020; Staiger and Stock, 1994). While it is fundamentally impossible to prove the exclusion restriction, we provide a test of the assumption in Figure 2(B) by examining whether listing characteristics correlate with our instrument. We regress the experience of listing agents on housing characteristics, and take the fitted value. We then replicate Figure 2(A), but replace the listing agent experience with the predicted values of the experience (excluding the housing controls from the right-hand side for this regression). If we found systematic differences between the fitted values and our instrument, we would be concerned that there may be other, additional unobservable characteristics that are unbalanced across our instrument. Instead, the plot shows that the predictable component of liquidity for these listings does not differ systematically with our instrument, lending support to the exclusion restriction assumption.

4.2 Effect of experience on listing liquidity

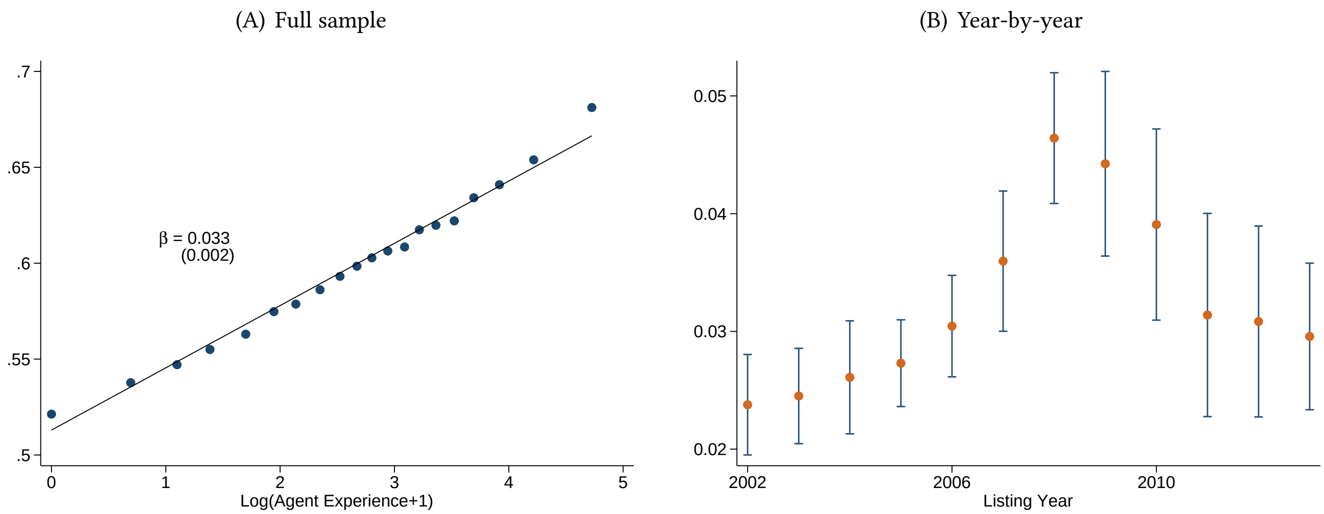
We begin by examining the effect of experience on the probability of sale within 365 days of listing. In Figure 3(A), we present a binned scatterplot of the relationship between listing liquidity and agent experience over our full sample period. The y-axis is the probability that a listing sells within 365 days, and the x-axis is our measure of agent experience, $\log(1 + \text{experience})$. This figure represents the pooled effect of experience on sale probability over the full sample. The relationship is strikingly linear and positive. A one log point increase in agent experience corresponds to approximately a 3.3 pp difference in the probability of sale within a year, or 5.4 percent of the average probability of sale. This corresponds to approximately 10 pp difference in the probability of sale between listings whose agents were in the 10th percentile of the experience (0 clients in the past year) and those of agents in the 90th percentile (21 clients in the past year).

In Figure 3(B), we let the effect of experience vary by listing year, using the same set of zip-code-by-list-year-month fixed effects as in Figure 3(A), and plot the corresponding coefficients with 95 percent confidence intervals. There are large changes in the effect of experience on listing liquidity, with an initial smallest effect of 2.4 pps (standard error (se) = 0.2 pps) in 2002, the largest coefficient of 4.6 pps (se = 0.3 pps) in 2008, falling again to 3 pps (se = 0.3 pps) in 2013.

We formally present estimates results from Equation 1 in Table 2. In each column, we report the effect of experience on the probability of a listing's sale within 365 days. We have two sets of analyses: our main sample in Columns 1–3 in Panel A, where we use all observations, and our IV sample in Columns 4–5 in Panel B, where we implement and evaluate the IV strategy outlined in Section 4.1.

We first focus on the full sample in Panel A. In Column 1, we report the overall pooled effect of experience with zip-code-by-list-year-month fixed effects, corresponding to the estimated effect from Figure 3(A). In Column 2, we repeat the same exercise but allow the effect to vary by our three aggregate time periods, with

Figure 3: Agent experience and listing's probability of sale in 365 days



Note: Panel A plots a binned scatterplot (with 20 bins) of the probability that a listing sells within 365 days against the listing agent's experience (measured as $\log(1 + \text{experience})$). The binned values and fitted line are residualized for zipcode-by-list-year-month fixed effects (the same controls as Column 1 in Table 2). The slope of the fitted line (the reported coefficient) corresponds to the coefficient on β of Equation 1, holding β fixed across time periods. Panel B plots the year-by-year effect of agent experience (measured as $\log(1 + \text{experience})$) on whether a listing sells within 365 days. The reported coefficients correspond to β of Equation 1, allowing β to vary by listing year. The bands correspond to the 95% confidence interval for each coefficient. The regression controls for zipcode-by-list-year-month fixed effects (the same controls as Column 1 in Table 2). Standard errors are clustered at the MLS-level. See Section 3 for more details on the data sample and definition of experience.

the base period of the housing boom. In Column 3, our preferred specification, we add the following housing controls to capture property-level characteristics: number of bedrooms, bathrooms, garages, living area, and type of cooling system and indicators for waterfront property, view, and fireplaces.[15]

Our baseline results show a strong positive effect of experience on listing liquidity. Split out by time period in Column 2, the effect is 2.65 pps (se of 0.19 pp) during the boom periods, 3.06 pps in the medium house price growth periods, and 3.96 pps during the housing bust periods. After adding housing controls in Column 3 the effect sizes remains similar.

In terms of the overall distribution of experience, listings of an agent in the 90th percentile (corresponding to an experience measure of 21) sold with a 8.2 pp higher probability than listings of agents in the 10th percentile (corresponding to an experience of 0) during the boom period. In the bust period, this gap increased to 12.0 pps. Compared to the average probability of sale of 69.1 percent during the boom period and 50.1 percent during the bust, this implies an increase of 11.9 percent of the mean during the boom and 24.0 percent of the mean during the bust. Thus, not only is agent experience an important factor in whether a listing sells, but the importance grows as the housing market contracts.

[15]For each discrete characteristic, we dummy out the values to nonparametrically control for their effect. We censor the top 1 percent of values in our controls to account for outliers.

Table 2: Effect of experience on probability of sale in 365 days

	Panel A: Main Sample			Panel B: IV Sample	
	(1)	(2)	(3)	(4)	(5)
Log(Exp + 1)	0.0326***	0.0265***	0.0273***	0.0158**	0.0299***
	(0.0021)	(0.0019)	(0.0018)	(0.0066)	(0.0021)
Bust × Log(Exp + 1)		0.0131***	0.0128***	0.0184**	0.0202***
		(0.0016)	(0.0016)	(0.0085)	(0.0021)
Medium × Log(Exp + 1)		0.0042***	0.0040***	0.0089	0.0048**
		(0.0015)	(0.0015)	(0.0094)	(0.0021)
Time-by-Zipcode FE	Yes	Yes	Yes	Yes	Yes
House Char.	No	No	Yes	Yes	Yes
Bust Effect		0.0396	0.0401	0.0342	0.0500
Bust p-value		0.0000	0.0000	0.0000	0.0000
Medium Effect		0.0306	0.0313	0.0247	0.0346
Medium p-value		0.0000	0.0000	0.0014	0.0000
Observations	8457612	8457612	8457612	1217983	1217983
Estimation Method	OLS	OLS	OLS	IV	OLS

Note: This table reports estimates of the effect of listing agent's experience (using the $\log(1 + \text{experience})$) on a listings' probability of sale in 365 days. All five columns use different versions of the specifications outlined in Equation 1 and 3. All columns include zipcode-by-listing-year-month fixed effects, and Columns 3-5 add controls for house characteristics. Columns 4 and 5 include purchase-year-by-listing-year-by-zipcode fixed effects. Panel A reports results using the main sample of listings. Panel B uses the IV sample of listings, restricted to observations where we observe the initial purchase of the listing. Column 4 shows results from the IV estimation while Column 5 repeats the specification in Column 3 using the IV sample. Details of the IV estimation are discussed in Section 4. Standard errors are clustered at the MLS level. *,**, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

In Panel B, we report the estimates from our IV approach. In Column 4, we find that the largest effects come during the bust (3.31 pps) and medium (2.63 pps) periods, and the smallest effect during the boom periods of 1.51 pps. Since the IV sample is substantially smaller (due to the restriction of observing the initial purchase), we rerun our OLS specification from Column 3 in Column 5, restricted to the IV sample. We find slightly larger estimates compared to Column 3.

Qualitatively, our OLS and IV estimates imply the same relationship between agent experience and listing sale probability – there is a significant effect of experience in all three periods, with a substantially higher effect during the bust periods. However, the IV point estimates are smaller, with the test statistic comparing the two sets of estimates rejecting the null of no difference at the 10% level (p-val = 0.069).[16] We consider two potential explanations for the difference between the IV and OLS estimates in Columns 4 and 5 of Table 2.

The first is omitted variable bias in our OLS regression; that selection between high and low experience agents is correlated with features that make it easier to sell a property. There is some evidence of this in Figure

[16] This test is done using a Sargan-Hansen test in Stata's `ivreg2`.

2(A), where the experience of the listing agent among sellers whose buyer agent exited is positively correlated with the experience of initial buyer agent. This suggests some amount sorting between agents and listings, and may explain part of the difference between the OLS and IV results since the IV approach will account for this sorting by exploiting the exit of the agents. In Section 4.5, we explore additional tests that examine different potential selection mechanisms, but find limited evidence on what those mechanisms might be.

The second possible explanation is heterogeneity in the effect of experience on listing outcomes. With heterogeneity in treatment effects, the IV approach will capture the Local Average Treatment Effect (LATE) of the sample of *compliers* induced by the empirical strategy (Imbens and Angrist, 1994), and this estimate may different from the average effect of experience in the whole population.[17] We explore this possibility by following Bhuller et al. (2020) and characterizing compliers based on time period and predicted probability of sale conditional on observable characteristics. We split our sample into twelve mutually exclusive subgroups based on time period (boom, bust and medium) and the predicted probability of sale (four quartiles). We then re-estimate our first-stage regression in each subgroup, using an indicator variable of listing agent experience above the listing-weighted median value (roughly experience equal to 10) as the dependent variable.[18] Using each first stage estimate, we construct the share of compliers within each subgroup, and re-weight the subgroups so that the proportion of compliers in a given subgroup matches the share of the estimation sample, and re-estimate our OLS regressions. We report these estimates in Appendix Table J3, and note that for the sale probability within 365 days, this re-weighting has a negligible effect. As a result, we do not find conclusive evidence that heterogeneity in treatment effects is driving our effects.

In sum, the larger magnitude in OLS estimates compared to IV estimates suggests that there is some selection between agents and listings, but the relative magnitude between the boom and bust period is nearly identical across estimation approaches.

We have chosen sale within 365 days as our main definition of sale, but we could have chosen other alternative cutoffs. In Appendix Section D.1, we re-estimate Table 2 using sale cutoffs at 30, 90 and 180 days, and find similar significant effects for experience at each horizon. However, the effects grow in magnitude, with the largest impact at 365 day horizon. This suggests that the benefits of experience are largest for those properties that do not sell immediately, consistent with listing agent experience benefiting "marginal" properties that are more difficult to transact. The gap between boom and bust is also largest at the longer horizon, consistent with experience mattering the most during busts.

[17] An additional assumption, like monotonicity, is necessary when there are heterogeneous effects of experience, in order to ensure that the IV estimate is a properly weighted average of the effects (Imbens and Angrist, 1994).

[18] We use this binary variable, rather than the full continuous value, to ensure that we can construct complier shares.

4.3 Agent experience and listing prices

So far, we have focused on the overall effect of experience on sale probability but not on the mechanisms by which experience increases the match probability. There are many mechanisms by which an experienced agent could improve the chances of a listing selling. For example, agents with more experience are likely more connected to other agents and also former clients. Thus, they can attract more matches for a listing by reaching out to potential buyers, or by tapping into their network of buyer agents. Additionally, an experienced agent may more effectively market a property to attract viewings and increase desirability for buyers who view the house.

One channel that is of more ambiguous value to clients is that experienced agents could set lower list prices for their properties, both attracting more buyers and making the purchase more likely. While the seller will benefit from their agent's network and expertise in the selling process, they face an important trade-off when it comes to the property price. Since properties with lower list prices are more likely to sell, ceterus paribus, if experienced agents list properties at lower prices, then that will lead to higher listing liquidity. We note here that the relevant trade-off for the seller is in the sales price and not the list price *per se*.

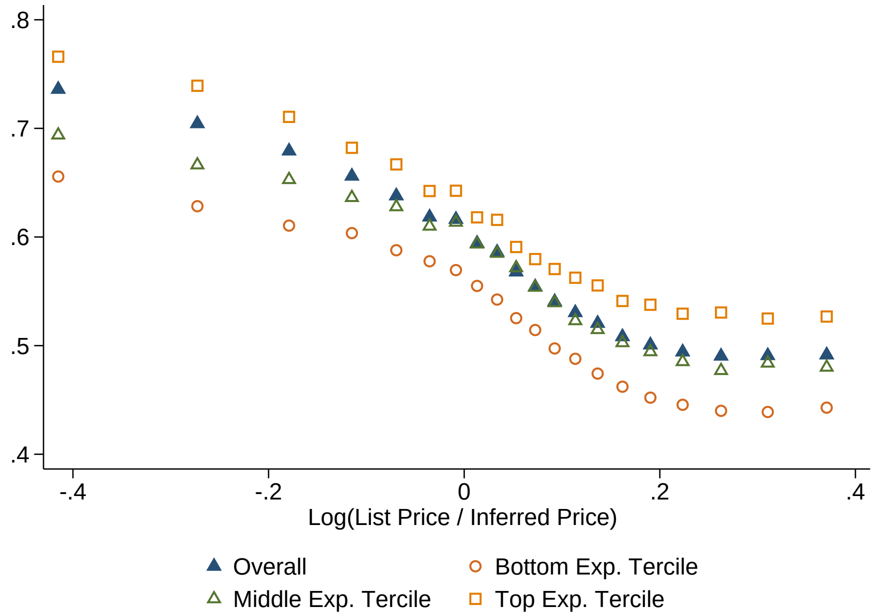
In this section, we explore whether agents' choice of list price drives the liquidity advantage of experience. We first consider how experience affects the liquidity of a property *conditional* on the initial list price decision. Then, we examine how the listing agent's experience correlates with the initial list price decision. Finally, we explore the sale price difference across experience levels. To compare list prices across properties, we construct a measure of the inferred home value. We do this by taking the last sale price of the property, and appreciating the price of the property forward to the current list date using the Zillow zipcode and tier-level house price index.[19]

In Figure 4, we plot the estimated average probability of sale in 365 days against binned values of the list price, scaled by the inferred value of the home, controlling for zip-code-by-list-year-month fixed effects and our housing controls. We plot two relationships on this plot. First, in solid triangles, we plot the overall relationship for all agents, which comes from pooling all agents together in a single regression. As expected, this relationship is negative. Listing for a higher price decreases the chance of sale and vice versa. We find an approximate elasticity of -0.55 for sale probability from changes in the normalized list price, with a change from -0.114 to 0.114 in the log normalized list price leading to a decline of 12.6 percent in the probability of

[19] A note on our sample: each observation requires not only a previous sale observation for the current listing, but also the Zillow price index for the corresponding zip code over that time period in order to estimate the inferred price. As a result, our sample is slightly smaller due to limited coverage of zipcodes at the beginning of our sample.

sale.[20]

Figure 4: Pricing and sale probability



Note: This graph plots coefficients from a regression of the expected sale probability against twenty equally sized bins of the log of normalized list price – list price scaled by our measure of inferred price. The regression is run both pooled and split by tercile of agent experience. We compute the inferred price as the last historical price that the property has sold for, appreciated to current list date using the Zillow zipcode and tier-level house price index. The regression controls for zipcode-by-year-month fixed effects and housing controls (the same controls as Column 3 in Table 2). See Section 3 for more details on the data sample and definition of experience.

We then split our estimates by agent experience terciles (weighted by listings) and show that there is a stark level difference in the probability of sale across experience levels, holding fixed the value of the list price markup. For all experience levels, a lower list price corresponds to a higher probability of sale, with a similar estimated elasticity for each experience tercile of around -0.5. Additionally, there is also an upward shift in the probability of sale for different experience levels across all levels of normalized list price. For a markup of zero, the difference in sale probability between the top and bottom tercile is 7.3 percentage points. These results imply a large experience effect holding fixed the pricing decisions, potentially due to the channels mentioned previously.

Next, in Table 3 we consider the impact of real estate agent experience on several price measures. In Panel A we use the preferred empirical specification from Column 3 of Table 2 while in Panel B we report the estimates using our quasi-experimental IV approach. In all cases, we consider log outcomes.

In Column 1 of Table 3, we find that that a one log point increase in real estate agent experience is asso-

[20] Our version of this relationship is much more monotonic compared to the ordinary least squares (OLS) figures in Guren (2018). We discuss the difference in Appendix D.2.

ciated with approximately a 0.9 percent decline in list prices during boom periods and a 1.2 percent decline during busts. In Column 2, we see that these declines in list prices correspond to decline (although much smaller) in sale prices. During boom periods, a one log point increase in experience corresponds to a 0.8 percent decline in sale prices and in busts, a 0.7 percent decline. In Column 3, we show that experience does not have a significant effect on the "discount" taken off of list prices, by estimating the effect of experience on the ratio of list price to sale price (e.g. the gap between the initial list price and the subsequent sale price). This indicates that inexperienced agents are not eventually selling for a discount on the list prices relative to experienced agents. Note that for both Column 2 and 3, this sale price is *conditional* on a successful sale.

In Panel B, we report the analogous estimates of Panel A using the quasi-experimental IV design outlined in section 4.1. While these estimates are smaller and noisier, the coefficients on the regressions are comparable to those in the OLS specification in Panel A. We cannot reject the null hypothesis that the IV and OLS estimates are the same.

Table 3: Experience and prices

	Panel A: OLS Approach			Panel B: IV Approach		
	List / Infer.	Sale / Infer	List / Sale	List / Infer.	Sale / Infer	List / Sale
	(1)	(2)	(3)	(4)	(5)	(6)
Log(Exp + 1)	-0.0091***	-0.0084***	0.0002	-0.0056	-0.0052	-0.0013
	(0.0021)	(0.0032)	(0.0009)	(0.0060)	(0.0064)	(0.0026)
Bust × Log(Exp + 1)	-0.0031*	0.0012	-0.0014	0.0068	0.0109	-0.0014
	(0.0017)	(0.0030)	(0.0009)	(0.0045)	(0.0077)	(0.0045)
Medium × Log(Exp + 1)	0.0001	0.0018	-0.0004	-0.0102	-0.0087	0.0005
	(0.0009)	(0.0012)	(0.0006)	(0.0077)	(0.0089)	(0.0048)
Time-by-Zipcode FE	Yes	Yes	Yes	Yes	Yes	Yes
House Char.	Yes	Yes	Yes	Yes	Yes	Yes
Bust Effect	-0.0123	-0.0072	-0.0012	0.0013	0.0058	-0.0027
Bust p-value	0.0000	0.0477	0.3459	0.7457	0.5704	0.4832
Medium Effect	-0.0090	-0.0066	-0.0002	-0.0158	-0.0138	-0.0008
Medium p-value	0.0002	0.0479	0.8457	0.0004	0.1426	0.8540
Observations	2203966	1318153	1291368	740460	454767	447604
Estimation Method	OLS	OLS	OLS	IV	IV	IV

Note: This table reports estimates of the effect of listing agent's experience (using the $\log(1 + \text{experience})$) on listings' price outcomes. The first three columns use the specification outlined in Equation 1, and include zipcode-by-listing-year-month fixed effects and controls for house characteristics. Column 1 reports the effect of agent experience on list price normalized to inferred price (measured using the previous sale price, appreciated using zipcode- and price-tier-specific Zillow house price appreciation) for all listings. Column 2 reports the effect on sale prices normalized to inferred price. Column 3 reports the discount that a property sells at relative to its list price. Columns 4,5 and 6 report the analogues of Columns 1,2, and 3 using the IV strategy outlined in section 4.1. All measures are in logs (after taking ratios), and censored (ratios at the 1st and 99th percentile, levels at the 99th percentile). Standard errors are clustered at the MLS level. *,**, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

These results let us consider a simple back-of-the-envelope calculation. The effect of one log point increase in experience is a reduction on markups of roughly 1 percent, suggesting that the effect of list price differences would lead to an increase in the probability of sale by about 0.5 percent, given the elasticities outlined above in Figure 4. Since the effect of experience on sale probability is roughly 2.7 pps during the boom and 4 pps during the bust in Column 3 of Table 2, the listing price effect is between 13 to 19 percent of the overall impact of experience on listing liquidity, depending on the period.[21] This suggests that listing prices, while important, play a limited role in the effect of agent experience on listing liquidity. Thus, for the rest of the paper and in the model, we abstract from differing pricing strategies and focus on the overall effect of experience on liquidity.

4.4 Foreclosure consequences of illiquidity

We have shown that real estate agent experience significantly affects the probability of sale. Why does the ability to sell a home matter? First, many people change homes to accommodate the size of their household and to be closer to a job, friends, or family. Inability to sell the current house thus impedes the purchase of a home that better serves their needs. This channel is valuable across all time periods. Second, listing liquidity can be important in the ability to reallocate financial resources from housing to more pressing needs, which can be particularly valuable during a recession. During the recent housing crisis, many households found themselves with expensive mortgages that they could not refinance due to tightening credit. Many attempted to sell their properties but could not do so, and some ended up in foreclosure.

Foreclosures result in a significant financial burden for people who lose their homes. A likely outcome is a substantially lower credit score that limits borrowing ability for years to come. Foreclosures are also socially inefficient because vacant properties tend to depreciate faster, either due to lack of upkeep or through a higher chance of looting and crime, which reduces the value of the property and puts downward pressure on prices for all houses in the neighboring areas.[22]

In our listings data, we observe properties that enter foreclosure after being listed for sale as non-foreclosure or non-REO properties. We focus on the outcome of whether a non-foreclosure and non-REO listing is associated with a future foreclosure sometime in the next two years. As one might expect, listings that successfully sold did not experience subsequent foreclosure; however, as we show in Figure 5(A), listings that *failed* to sell in 2008 had a 4.5 pp chance of subsequent foreclosure. Hence, an increased probability of sale for a given listing could play an important role in avoiding foreclosures.

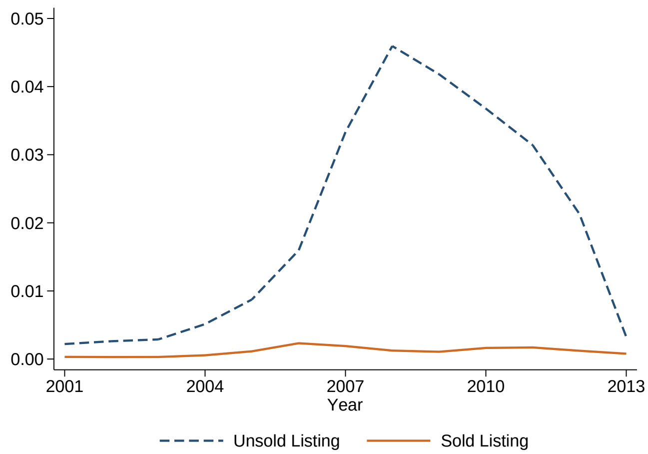
We examine the effect of agent experience on foreclosure probability using the same specifications in

[21] Compared to our IV estimates in Column 4 of Table 2, it would be between 14 and 33 percent.

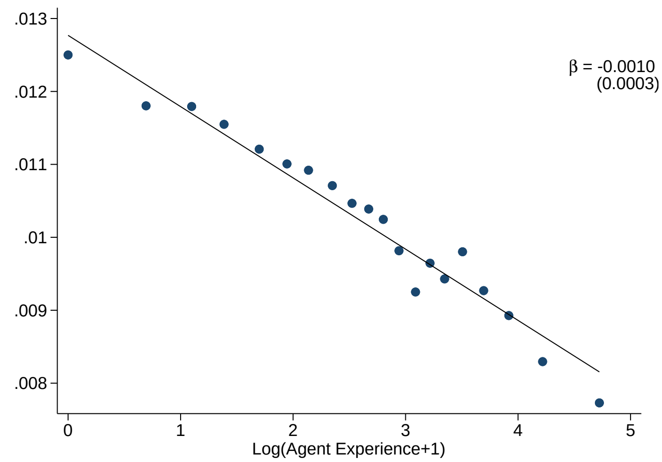
[22] Some examples of papers examining foreclosure externalities include Lin, Rosenblatt, and Yao (2009), Campbell, Giglio, and Pathak (2011), Mian, Sufi, and Trebbi (2015), Gupta (2016), and Guren and McQuade (2019).

Figure 5: Listing sale, subsequent foreclosure, and agent experience

(A) Probability of subsequent foreclosure by listing sale



(B) Effect of experience on subsequent foreclosure



Note: Panel A plots the fraction of listed properties that we observe going into foreclosure in the next two years. The sample is split into listings that did not sell within a year versus those that did. The sample is restricted to non-REOs and non-foreclosure listings. Panel B plots a binned scatterplot (with 20 bins) of the probability that a listing goes into foreclosure in the subsequent two years against the listing agent's experience (measured as $\log(1 + \text{experience})$). The binned values and fitted line include controls for zipcode-by-list-year-month fixed effects (the same controls as Column 1 in Table 2). The slope of the fitted line corresponds to the coefficient on β of Equation 1, holding β fixed across time periods. Standard errors are clustered at the MLS-level. See Section 3 for more details on the data sample and definition of experience.

Figure 5(B) and Table 4. In Figure 5(B), we plot the binscatter of subsequent foreclosure in the next two years against the log of listing agent's experience. We see a negative and significant effect of agent experience; a one log point increase in an agent's experience leads to a 0.10 pp reduction in the subsequent foreclosure probability (this probability was roughly 2.5 pps at the peak in 2008 in our sample). In Table 4, we estimate the effect of experience on foreclosure across periods. In Column 3 of Panel A, we see that the effect of experience is economically significant during the housing bust, with a one log point increase in experience leading to a reduction in the probability of subsequent foreclosure by 0.2 pps, or almost 10 percent of the average rate of subsequent foreclosure during the bust. This result is even larger in the IV approach in Panel B, but we cannot reject the null that the OLS and IV estimates are the same. This suggests that the effect of experience on foreclosure is not a selection effect by agents into certain homes or sellers, but instead an important channel for real estate agent experience's effect in alleviating foreclosures.

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 has not been put on the market due to a backlog (Mian, Sufi, and Trebbi, 2015). Second, if the lender takes ownership of the property, they might not necessarily put it up for sale right away,

Table 4: Effect of experience on foreclosure in next two years

	Panel A: Main Sample			Panel B: IV Sample	
	(1)	(2)	(3)	(4)	(5)
Log(Exp + 1)	-0.0010***	-0.0002**	-0.0001**	0.0006	-0.0000
	(0.0003)	(0.0001)	(0.0001)	(0.0006)	(0.0001)
Bust × Log(Exp + 1)		-0.0019***	-0.0019***	-0.0054**	-0.0034***
		(0.0006)	(0.0006)	(0.0027)	(0.0009)
Medium × Log(Exp + 1)		-0.0004**	-0.0004**	-0.0043	-0.0010**
		(0.0002)	(0.0002)	(0.0027)	(0.0004)
Time-by-Zipcode FE	Yes	Yes	Yes	Yes	Yes
House Char.	No	No	Yes	Yes	Yes
Bust Effect		-0.0021	-0.0021	-0.0048	-0.0034
Bust p-value		0.0008	0.0008	0.0556	0.0003
Medium Effect		-0.0006	-0.0006	-0.0037	-0.0010
Medium p-value		0.0267	0.0297	0.1398	0.0293
Observations	8014291	8014291	8014291	1140519	1140519
Estimation Method	OLS	OLS	OLS	IV	OLS

Note: This table reports estimates of the effect of listing agent's experience (using the $\log(1 + \text{experience})$) on a listings' probability of foreclosure in the next two years. All columns include zipcode-by-listing-year-month fixed effects, and Columns 3-5 add controls for house characteristics. Panel A reports results using the main sample of listings. Panel B uses the IV sample of listings, restricted to observations where we observe the initial purchase of the listing. Column 4 shows results from the IV estimation while Column 5 repeats the specification in Column 3 using the IV sample. Details of the IV estimation are discussed in Section 4. Standard errors are clustered at the MLS level. *,**, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

again excluding a foreclosure observation from our data.

4.5 Additional robustness checks

While the quasi-experimental design outlined in Section 4.1 is our main approach to addressing selection issues in the OLS specification, another approach is to rule out alternative theories through robustness tests. We do so by examining subsamples of the data where the specific selection concerns are not likely to play a role. We briefly outline the main theories we tested here, and defer a broader discussion to Appendix Section B.

We first consider the alternative mechanism that agents with higher experience choose to work with properties that look observably similar but have unobserved qualities that make them of higher value and, as a result, easier to sell. We test this in two ways: controlling for a proxy of the inferred value of the home in our main specification, and restricting our analysis to a housing market where houses are nearly identical. We next consider whether agents with higher experience choose to work with clients whose properties are easier to sell. We test this in two ways: first, we control for the clients' home equity at the time of the listing, as proxied by the amount of house price appreciation experienced by the seller since the house was last transacted (Guren, 2018). Second, we examine a subsample of listings that followed a deed transfer that we assume proxies for a life-changing event (Kurlat and Stroebel, 2015). Finally, we consider how our results change if we include listing agent fixed effects. However, we note that these results are likely biased: those agents who were unsuccessful in selling properties when they had low experience are less likely to continue as agents and build experience, which will bias our within-agent effect of experience downwards.

In sum, we find that our estimated effect of experience on listing outcomes is remarkably robust across different analyses. Moreover, we are unable to identify a particular channel that explains the gap between our OLS and IV estimates in Table 2.

4.6 Naive counterfactual and entry and exit patterns

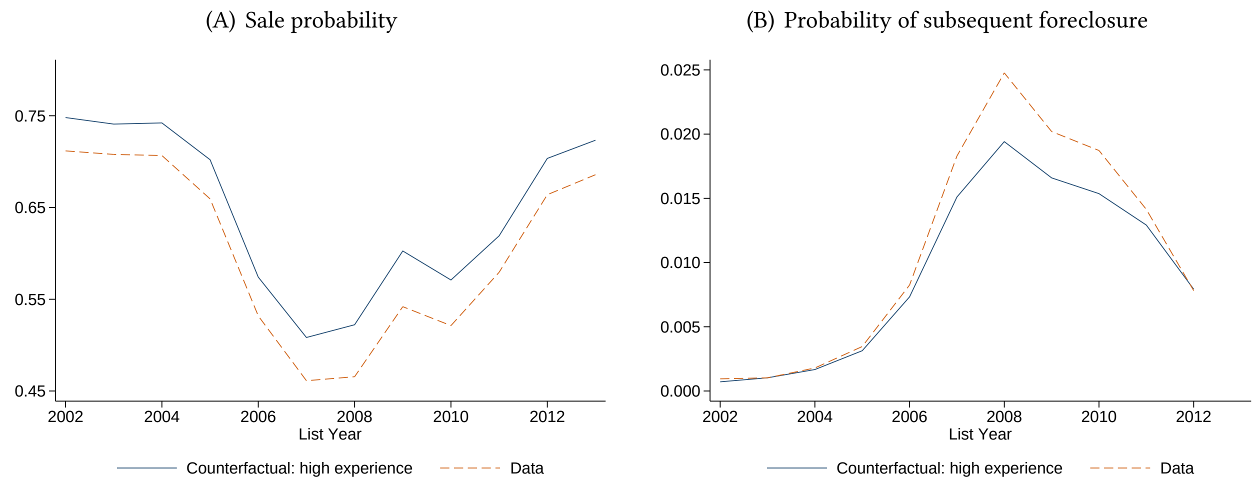
Given our estimates, can we say how much real estate agent experience contributed to the drop in listing liquidity in the recent housing bust? One naive approach to this question is to use our regression model from Section 4.2 and compute the predicted sale probability for the counterfactual, where all variables are fixed except for the experience of the listing agent. For the counterfactual, we split all agents in terciles according to their experience (listings weighted) and compute the average experience within each tercile. For all agents whose experience is below the average of the top tercile, we replace experience with that average. We then calculate the predicted probability of sale and subsequent foreclosure using our preferred specification (e.g., including house controls and zip-by-year-month) and allowing the effect of experience to vary by year.

Figure 6(A) plots the observed average yearly probability of sale and the predicted counterfactual. We see an increase in the probability of sale for all years. In Appendix Table J7, we report the year-by-year numbers, which show that the effect is highest in the bust. In 2008, the naive counterfactual leads to a 12.2 percent increase in the probability of sale, and in 2004 it improves liquidity by only 5.0 percent. A similar exercise for our measure of subsequent foreclosure probability (illustrated in Figure 6(B)) suggests that roughly 20 percent of listings that subsequently foreclosed could have avoided foreclosure between years 2004 and 2010.

However, this counterfactual is not achievable in practice. Agent experience is endogenous and depends on agents' entry and exit decisions as well as on their opportunities to accrue experience. The churn for low experience agents in this market is substantial, making it difficult for newly entered agents to become experienced. In Figure 7(A), we plot the aggregate entry and exit rates for real estate agents in the US, where the entry rate is the share of currently active agents who had zero activity in the previous two years and exit rate is the share of currently active agents who we do not observe as active in the following two years.[23] In the boom years of 2003 to 2006, more than a quarter of all active agents were brand-new and between 15 percent

[23]See Appendix C for a discussion on alternative definitions of entry and exit.

Figure 6: Naive counterfactuals



Note: This figure plots the results of the naive counterfactual discussed in Section 4.6. In Panel A, we consider the probability of sale in 365 days as the outcome. In Panel B, we consider the probability of a listing subsequently going into foreclosure in the next two years. In each panel, the empirical time series is plotted in the orange dashed line. We then regress the outcome on housing controls, zipcode-list-year-month fixed effects, as well as listing agent experience agent interacted with each calendar year. Using the coefficients of this regression, we then predict sale probability for a counterfactual where all agents are in the top experience tercile. The blue solid line plots the average counterfactual outcome using the predicted values.

and 22 percent of all agents subsequently exited each year. Starting in 2008, the share of new entrants had plunged from its previous peak of 30 percent but remained as high as 17 percent. As the entry of agents fell, the exit rate of agents grew steadily, peaking in 2008.[24]

The high exit rates are concentrated among inexperienced agents. In Figure 7(B), we plot the exit rates at each experience level, broken out by time periods. In all settings, inexperienced agents have far higher exit rates, near 30 percent, while the exit rates for agents with experience above 30 dip below 5 percent. During the bust periods, inexperienced agents have the highest exit rates, but all agents' exit rates shift upwards.

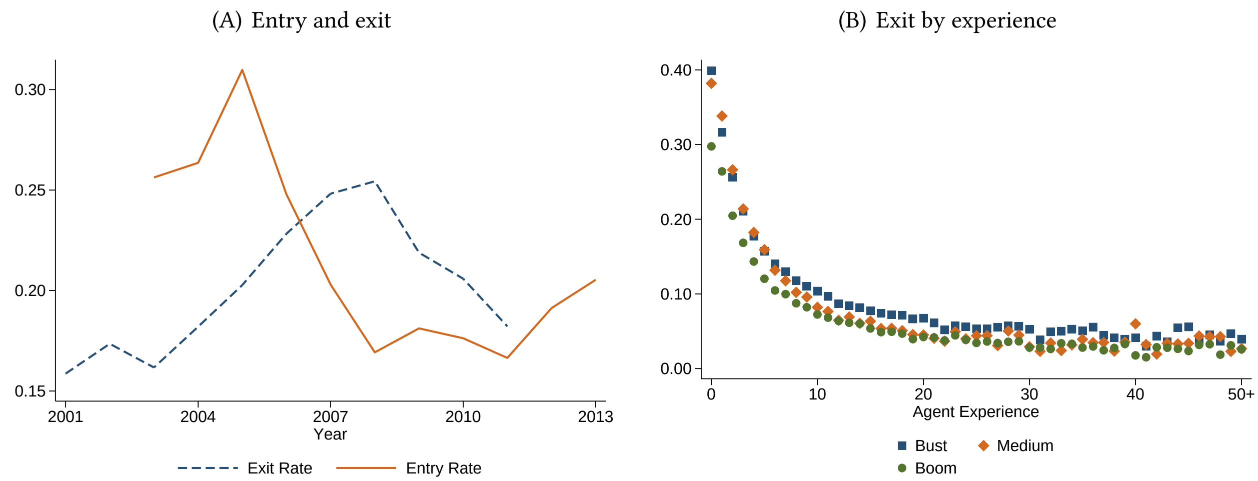
This churn is heavily driven by market conditions. Since commissions paid to listing agents tend to be a fixed percentage of the sale price, this creates tremendous incentives to enter (and exit) the market as the house prices change.[25] In addition, agent earnings are directly related to listing volume (the opportunity to make a sale) and the ease with which transactions are made (whether the sale occurs). We now show that housing market conditions also influence the *distribution* of agent experience.

To examine how the real estate agent's entry, exit, and experience shifts in response to market conditions, we assign each agent to a home market (as measured by the county in which they have the largest share of

[24] For comparison, according to the US Census Bureau's Business Dynamics Statistics, the entry and exit rates of the establishments in the US range between 8 percent to 12 percent in the same time period (2000–2015), where exit is defined as the fraction of establishments with positive employment who had/will have zero employment in the previous/following year. A similar definition for agents (one-year window) delivers an even larger churn than is described in this section (see Appendix C).

[25] The influence of housing market conditions on real estate agent entry has been documented previously in Hsieh and Moretti (2003).

Figure 7: Entry and exit rates



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, with experience greater than 50 pooled with agents who have experience of 50.

activity). We define entry rate in a particular county as the fraction of corresponding agents currently active who we have not observed in our data (including in other counties) in the previous two years. Similarly, exit rate is the share of agents who are currently active in the county who we do not observe in the following two years. Appendix Table J6 summarizes the number of counties in the data as well as the mean and standard deviation of the number of active agents, exit rates, and entry rates in each county. We observe from 663 to 869 distinct counties per year.

We estimate county-level regressions of the following form:

$$Y_{it} = \alpha_i + \text{Sales / Listings}_{it}\gamma_1 + \Delta\text{Sales Price}_{it}\gamma_2 + \Delta\text{Listing Volume}_{it}\gamma_3 + \epsilon_{it}, \tag{4}$$

where Sales / Listings$_{it}$ measures the market tightness in county i and year t, ΔSales Price$_{it}$ measures the percentage change in average sale price, and ΔListing Volume$_{it}$ measures the percentage change in the number listings. Y_{it} corresponds to several measures of agent entry and exit within the market as well as measures of the experience distribution. α_i controls for county fixed effects to allow for county-specific time-invariant heterogeneity. We weight these regressions by the number of listings in a county in a given year.

In Table 5, we report the estimates of the effect of market conditions on agents' entry, exit, and experience. In Column 1, we see that easier markets (high sales relative to listings), increase in prices, and increase in listings volume all lead to higher real estate agent entry. In fact, the change in listing volume is a larger predictor of agent entry than changes in sale price or market tightness. On the other hand, in Column 2, we see that market tightness is the only statistically significant predictor of exit. Conditional on Sales / Listings, neither the change in prices nor the change in listings leads to an increase in exit rates. In Column 3–7, we examine how market conditions affect the distribution of experience. Interestingly, with easier markets, the average experience in the market increases, but the average log experience declines. This occurs because the experience distribution skew increases, with the 25th and 50th percentile decreasing and the 75th percentile increasing. In contrast, with an increase in listing volume, the experience distribution shifts leftward and both the average experience and log experience fall. The distribution is not affected in a statistically significant way due to shifts in the average price, suggesting that the change in listing volume and, to a lesser extent, sale/listings capture the main effect on experience.

Table 5: Turnover rates and market conditions

	Probability of		Experience Summary Statistic				
	Entry (1)	Exit (2)	Mean (3)	Mean (Log) (4)	25th perc. (5)	50th perc. (6)	75th perc. (7)
Sales / Listings	0.12***	-0.19***	0.65*	-0.11***	-0.77***	-0.70**	0.87*
	(0.02)	(0.02)	(0.33)	(0.04)	(0.17)	(0.30)	(0.45)
Δ Sales Price	0.07***	-0.04	0.43	0.03	-0.23	0.32	0.72
	(0.02)	(0.04)	(0.41)	(0.05)	(0.23)	(0.31)	(0.59)
Δ Listing Volume	0.24***	-0.02	-3.33***	-0.50***	-1.73***	-2.75***	-3.75***
	(0.02)	(0.01)	(0.35)	(0.04)	(0.13)	(0.24)	(0.42)
R^2	0.5819	0.6881	0.8953	0.8469	0.6361	0.8153	0.8679
County FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	5364	4694	5751	5751	5751	5751	5751

Note: In this table, we report how agent entry and exit, along with the distribution of experience, varies with county-level housing market conditions. We assign each active agent in the data to a fips code in which they have the most activity. We report the estimated coefficients from Equation 4 in each column for different outcomes, where Sales / Listings$_{it}$ measures the market tightness in county i and year t, ΔSales Price$_{it}$ measures the percentage change in average sale price and ΔListing Volume$_{it}$ measure the percentage change in the number listings. In Column 1, we report the effects for agent entry rates. For Column 2, we report the effects for agent exit rates. In Columns 3-6, we report the effects on different components of the agent experience distribution at the county level. In all regressions, we control for county-level fixed effects, and weight by the number of listings in a county in a given year. Standard errors are clustered at the county-level. *,**, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

A policymaker interested in influencing listing liquidity cannot directly manipulate the experience of agents. However, our results suggest that economic incentives play an important role in the accrual of experience. Thus, by changing the incentives of the agents through realistic policies, such as increasing the certification cost to become an agent, a policymaker might hope to affect the experience distribution. To accurately assess the impact of these policies on the overall market, we develop a structural model of real estate intermediaries that will capture the effect of policies on the distribution of experience as well as on the aggregate listing liquidity in the housing market.

5 Theoretical model of real estate agents

This section first describes the setup for our structural dynamic model of real estate agents. We then characterize the dynamic equilibrium. Finally, we numerically calibrate the model and evaluate the fit to the data.

5.1 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 market experience, e. Consistent with our empirical analysis, an agent's experience is defined as the number of their listings in the previous year plus the number of successful transactions they facilitated when representing a buyer. We revisit the formal definition when we describe how experience is updated.

Time is discrete $t \in \mathbf{N}$ ($\mathbf{N} = \{0, 1, 2, ...\}$), and all 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 $n^a_{-i,t}(e) = n_t^a(e) - 1$ if $e = e_{it}$ and $n^a_{-i,t}(e) = n_t^a(e)$ 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 home. We assume that the industry state evolves according to a Markov process with transition probabilities P and takes on three values $z_t \in \{z_1, z_2, z_3\}$ representing bust, medium, and boom activity in the housing market. Finally, we denote n_t^b as the total number of buyers (determined endogenously) that search for a house in period t.

In the beginning of each period t, the industry state $z_t = (n_t^s, v_t)$ is realized and competition level n_t^a is observed. There is an infinite pool of potential real estate agents who 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 and enter the market.

Next, all buyers and sellers are paired with an agent. We assume that a fraction ϕ of clients contact an agent at random and the remaining fraction gets a referral and is matched with an agent with a probability proportional to the agent's experience share. The number of seller and buyer clients are Poisson random variables with means and variances both equal to $s(e, n_t^s; n_t^a)$ and $b(e; n^a, n_t^b)$, respectively, where the average number of sellers an agent with experience e is expected to work with is

$$s(e, n_t^s; n_t^a) = \phi n_t^s \frac{1}{\sum_{\tilde{e}} n_t^a(\tilde{e})} + (1-\phi) n_t^s \frac{e}{\sum_{\tilde{e}} n_t^a(\tilde{e})\tilde{e}}. \tag{5}$$

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

$$b(e; n^a, n_t^b) = \phi n_t^b \frac{1}{\sum_{\tilde{e}} n_t^a(\tilde{e})} + (1-\phi) n_t^b \frac{e}{\sum_{\tilde{e}} n_t^a(\tilde{e})\tilde{e}}. \tag{6}$$

An experienced agent can then expect to have more clients on both the seller and the buyer sides. While a linear relationship between experience and number of listings might seem ad hoc, it is a surprisingly accurate representation of what we observe in the data. Appendix Figure J3 plots the median and the 25th and 75th percentiles of the number of clients we observe in the data (this includes all listings and successful buyers) at each value of agent experience (recall that this measure uses *historical* information, so the linear relationship is not mechanical). Appendix Table J8 explores this relationship more formally in a regression. The coefficient on agent experience is one of the moments matched in the calibration exercise.

Clients fully delegate the housing search process to their agents and thus have no further role in the model. 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 cannot easily 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 the labor, finance, and industrial organization literature. In this setting, buyer agents can direct their search toward houses whose listing agents have a particular experience. This effectively creates different submarkets that are indexed by the experience of selling agents operating in that submarket.[26]

In each submarket j, with s seller agents and b buyer agents, $s(1 - e^{-b\nu(e_j)/s})$ matches are realized, where e_j is experience level of listing agents in that market.[27] The function $\nu(e)$ captures the overall experience

[26] While our model's setup and solution method echoes the standard directed search model (see Moen (1997) and Shimer (1996)), it differs in a significant way. The standard directed search model 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 their 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.

[27] This matching function is an approximation of an urn-and-ball matching function for a large number of agents. The formulation is convenient because it restricts the probability of match to be between zero and one. In addition, match probabilities for each side exhibit constant return to scale, which allows us 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).

advantage of attracting clients to a property and making the match more likely. We impose ν to have the following functional form: $\nu(e) = \nu_1 e^{\nu_2}$. Power functions are useful in this setting, as they allow for a decreasing returns to scale, meaning faster "learning" by inexperienced agents observed in the data.[28]

Then, the match probability for a buyer and a seller is a function of listing agents experience e and the market tightness, $\theta = b/s$:

$$\eta(e,\theta) = \frac{1}{\theta}\left(1 - e^{-\nu(e)\theta}\right) \qquad \text{Buyer Match Probability}$$
$$\mu(e,\theta) = 1 - e^{-\nu(e)\theta} = \theta\eta(e,\theta) \qquad \text{Seller Match Probability}$$

Once a meeting occurs, prices are determined via Nash bargaining with bargaining parameter γ for the buyer. We assume that a seller of an unsold house, and a buyer of a house, identically value the future changes in resale price. As a result, the total surplus of a transaction will not be affected by the continuation value of holding on to the property and is simply ν_t. The prices will then be the same in each submarket and is equal to

$$p(\nu_t) = \gamma\nu_t. \tag{7}$$

Buyer agents choose the submarket to enter to maximize buyer valuation:

$$V^B = -c_b + \max_j \eta(e_j, \theta_{j,t})(\nu_t - p_t). \tag{8}$$

Since prices do not differ by submarket, it must be that the probability of purchase, $\eta(e_j, \theta_{j,t})$, is also constant in equilibrium. Otherwise, only markets with highest $\eta(e_j, \theta_{j,t})$ would 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 until $V^B = 0$. The free entry condition, combined with the equilibrium result of equal match rates, determines the technology queue trade-off for the buyers:

$$\eta(e_j, \theta_{j,t}) \equiv \frac{1}{\theta_{j,t}}(1 - e^{-\nu(e)\theta_{j,t}}) = \frac{c_b}{(1-\gamma)\nu_t} = \eta(\nu_t). \tag{9}$$

The left-hand side is decreasing in θ, while the right-hand side is constant in θ. Thus, there is a unique $\theta_{j,t}$ for each market that satisfies the equilibrium conditions for free entry and submarket indifference. Solving for

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

$\theta_{j,t} = \theta(e_j, \nu_t)$ allows us to compute the equilibrium match probabilities for the seller side

$$\mu(e_j, \theta_{j,t}) = 1 - e^{-\nu(e_j)\theta(e_j,\nu_t)} = \mu(e_j, \nu_t). \tag{10}$$

While in equilibrium $\eta(\nu_t)$ is constant across markets, $\mu(e_j, \nu_t)$ is increasing in the experience of a listing agent operating in submarket j through the $\nu(e_j)$ 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 to abstract from heterogeneity on both sides of the search market, but we think it is quite realistic. While the marketing effort and expertise is often crucial in whether a house finds a buyer, the buyer agent mainly engages in scheduling viewings for existing homes for sale, which arguably requires less know-how. For simplicity, we subsequently drop the j subscripts from equilibrium equations since every submarket j is uniquely identified by the experience e of listing agents of that submarket.

After the matches are realized, buyers pay p_t, of which 3 percent goes toward the buyer agent earnings, 3 percent goes toward the seller agent earnings, and the remaining 94 percent is taken by the seller. In reality, agents only get to keep a percentage of the commission, while the remaining share is taken by the office where they work. Moreover, more experienced agents, who bring in more business to the office, get to keep a higher fraction of their earnings, while new agents have a less favorable split. While we do not explicitly model real estate offices, we assume that agents in the model get to keep a fraction of their commission as a function of their earnings. We parameterize this function to be consistent with survey evidence on commission splits: $f(x) = 0.1498x^{0.1455}$ so that an agent who receives x dollars in commissions takes $f(x)x$ in profits.[29]

Next, for a particular distribution n_t^a of experience across agents, we compute the total number of buyers n_t^b in equilibrium:

$$n_t^b = \sum_e n_t^a(e) s(e, n_t^s; n_t^a)\theta(e, \nu_t). \tag{11}$$

This equation aggregates the buyers who are present in each market, using the equilibrium market tightness multiplied by the number of listings (sellers) allocated to the corresponding experience group.

We can now construct the per-period expected profit function for each agent of experience e:

$$E[\pi(e)|z_t, n_t^a, n_t^b] = E\left[0.1498\left(s(e, n_t^s; n_t^a)\mu(e, \nu_t)\psi p(\nu_t) + b(e; n_t^b, n_t^a)\eta(\nu_t)\psi p(\nu_t)\right)^{1.1455}\right], \tag{12}$$

where agents expect to get $s(e, n_t^s; n_t^a)$ listings that will sell with probability $\mu(e, \nu_t)$ as well as $b(e; n_t^b, n_t^a)$

[29] Appendix F describes the survey evidence.

buyers who buy with probability $\eta(\nu_t)$. All transacted properties will earn the agent a fraction of the total commission $\psi = 3$ percent on the sale price $p(\nu_t)$.

The experience of all agents is updated at the end of the period. Consistent with the empirical analysis, we assume that all listings contribute to experience equally, no matter if they are sold, while only successful buyers count toward experience. Then the expected experience level of an agent entering time t with experience e_t is

$$E[e_{t+1}|e_t, z_t; n_t^b, n_t^a] = s(e, n_t^s; n_t^a) + b(e; n_t^b, n_t^a)\eta(\nu_t). \tag{13}$$

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, with $\log(c_{i,t}) \sim N(\mu_{fc}, \sigma_{fc})$. If the drawn cost exceeds the agents' 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

$$V_t(e_{i,t}, z_t; n_t^b, n_t^a) = E[\pi(e_{i,t})|z_t, n_t^a, n_t^b] + \beta E_t[\max\{0, -c_{i,t} + V_{t+1}(e_{i,t+1}, z_{t+1}; n_{t+1}^b, n_{t+1}^a)\}]. \tag{14}$$

A value of an entrant entering time t is similarly

$$V_t(0, z_t; n_t^b, n_t^a) = -c_e + E[\pi(0)|z_t, n_t^a, n_t^b] + \beta E_t[\max\{0, -c_{i,t} + V_{t+1}(e_{i,t+1}, z_{t+1}; n_{t+1}^b, n_{t+1}^a)\}]. \tag{15}$$

Since both the 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:

$$\rho_t(e_{i,t+1}, c_{i,t})) = \begin{cases} 1 & \text{if } c_{i,t} > E_t[V_t(e_{i,t+1}, z_{t+1}; n_{t+1}^b, n_{t+1}^a)] \\ 0 & \text{otherwise.} \end{cases} \tag{16}$$

The free entry condition for real estate agents implies that if any agents find it profitable to enter, agents will keep entering until the value of entry is driven down to zero. If, however, no entry happens, then the value of entry must be negative. Formally, if λ_t is the entry rate at time t, then $\lambda_t V_t(0, z_t; n_t^b, n_t^a) = 0$.[30]

5.2 Model equilibrium

We allow the exogenous aggregate state $z_t = (n_t^s, \nu_t)$ to take on three different pairs of values corresponding to boom, bust, and medium periods of the housing market, as in our empirical analysis. The endogenous

[30] 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 λ_t.

measure of buyers n_t^b is a function of v_t, n_t^s, and n_t^a, as described in Equation 11, so it is not a distinct state variable. The main challenge is n_t^a, the distribution of agents across all experience groups. Allowing agents to keep track of n_t^a makes the state space essentially infinite since each value of the function $n_t^a(e)$ is a state variable itself. 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 its long-run average value corresponding to a recent history of aggregate states z_t. Adopting the notation of the original paper, let $\{w_t = (z_t, z_{t-1})\}$ be a Markov chain adopted to the filtration generated by $\{z_t : t \geqslant 0\}$. 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 under entry rates λ and policy function ρ. We now define agent's value function $\tilde{V}(e, w|\rho', \rho, \lambda)$ as the expected present value for an agent of experience e in aggregate state w given that they follow an exit strategy ρ', while the competitors follow a common strategy ρ and enter at rate λ [31]:

$$\tilde{V}(e, w|\rho', \rho, \lambda) = E[\pi(e, w)] + \beta E[\max\{0, -c + \tilde{V}(e', w')|e, w, \rho', \rho, \lambda]. \tag{17}$$

Similarly, an entrant's value is

$$\tilde{V}(0, w|\rho', \rho, \lambda) = -c_e + E[\pi(0, w)] + \beta E[\max\{0, -c + \tilde{V}(e', w')|0, w, \rho', \rho, \lambda]. \tag{18}$$

In both,

$$E[\pi(e)|w, \tilde{n}^a_{\lambda,\rho}, n^b] = E\left[0.1498\left(s(e, n^s; \tilde{n}^a_{\lambda,\rho})\mu(e, v)\psi p(v) + b(e; n^b, \tilde{n}^a_{\lambda,\rho})\eta(v)\psi p(v)\right)^{1.1455}\right], \tag{19}$$

Where n^s and v are defined by the state z (i.e., are a function of w); total buyers for each state are defined in Equation 11; functions s and b defining the distribution of clients are defined by Equations 5 and 6; match probabilities η and μ are defined in Equations 9 and 10; and price $p(v)$ is defined in Equation 7. Finally, the w is updated via adopting the Markov process for aggregate state z and agent experience updates according to

[31] Equations 17 and 19 are slightly abusing notation since ρ' is built in the value function, as we already showed that all firms will follow a cut-off strategy. This is, however, an equilibrium result, so we choose to stay consistent with the original formulation of the problem.

Equation 13.

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 their exit strategy using 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):

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

2. $n^b(w)$, 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 a slightly modified version of the solution method described in Weintraub, Benkard, and Van Roy (2010). The full algorithm is described in detail in Appendix G.[32]

5.3 Calibration

Calibrating 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 aggregate states for each $z_t = (v_t, n_t^s)$ to match prices (that directly correspond to the housing valuation) and the overall number of sellers looking to sell their property that we see in the

[32]The intractability of a distribution as a state variable could also be tackled by a commonly used algorithm introduced in Krusell and Smith (1998). There, agents' decisions are allowed to depend on a finite set of moments that describe the underlying distribution. These moments evolve according to a parameterized law of motion that is approximated to best fit the model generating process. While this approach solves a similar problem, the oblivious equilibrium concept differs in an important way. It allows agents to internalize an entire approximate distribution (rather than estimated moments of the distribution). Thus, instead of keeping track of several moments to base their decisions on, the agent keeps track of past few realizations of some aggregate state and bases their decisions on the approximate distribution implied by the corresponding history. If the distribution in question has a nonregular shape (and thus is difficult to summarize by a few moments), the oblivious equilibrium approach might be a better way to address the issue of high dimensionality.

data. Next, for a given state z_t, we calibrate the directed search model to match the sale probabilities for each agent experience group. Finally, given the parameters from the previous two steps, we fit the entry and exit parameters to match the observed entry and exit rates for every state $w_t = \{z_t, z_{t-1}\}$ and agent experience level.

For the first step, we define three states for z_t using the historical series of the Case-Shiller house price index for years 1940–2017 in the same way as we did in the empirical section. We first deflate the index by the Consumer Price Index (less shelter) and then compute the annual average of the 12-month growth rate. We define years with growth rates in the bottom and top quartile of the data to be bust and boom years, respectively. The remaining years correspond to the medium state. Figure (J2) plots the adjusted growth rates together with our approximation for the state process. The evolution of states in this dataset allows us to compute a Markov transition probability matrix P for the aggregate state z_t (in step three, we use P to infer the transition probability matrix for recent state history, w_t).

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 in the data. 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 the Nash bargaining parameter, $\gamma = 0.5$, 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 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 $\mu(e, z_t) = 1 - e^{-\nu_1(z_t)e^{\nu_2}\theta(e,z_t)}$. In equilibrium, $\theta(e, z_t)$ is a function of $c_b, v(z_t)$ and γ due to free entry of the buyers (Equation (9)). Since the cost of entry for the buyer, c_b, identifies the overall level of sale probabilities across all states, we normalize $\nu_1(bust) = 1$ such that $\nu_1(medium)$ and $\nu_1(boom)$ measure the differences in sale probabilities across aggregate states. Lastly, ν_2 governs the differences in sale probability across experience levels within states. Formally, let $\Theta_1 = (c_b, \nu_1(medium), \nu_1(bust), \nu_2)$ be the parameters of interest, while the set of moments are $g(e, z, \Theta) = (\mu_{model}(e, z_t, \Theta) - \tilde{\mu}^{obs}(e, z_t))/\tilde{\mu}^{obs}(e, z_t)$, the vector of normalidifferences between observed and model predicted sale probabilities by each state and experience level. The chosen parameters $\hat{\Theta}_1$ are then

$$\hat{\Theta}_1 = argmin_{\Theta_1} \sum_{e,z} g(e, z, \Theta_1)^2. \tag{20}$$

Finally, we estimate the remaining parameters, c_e, μ_{fc}, σ_{fc}, and ϕ governing the entry and exit rates of real estate agents, as well as the client distribution across agents. Computing the entry and exit rates implied by these parameters involves a computation of the equilibrium that also uses the calibrated values for aggregate states $z_t = (n^s(z_t), \nu(z_t))$, P, and the parameters from the previous step, $\hat{\Theta}_1$. We choose parameters c_e, μ_{fc}, and σ_{fc} to minimize the difference between the observed entry and exit rates corresponding to each experience and state history $w_t = (z_t, z_{t-1})$, $\Lambda(w_t)$, and $\rho^{obs}(e, w_t)$ and their counterparts in the model. As a final set of moments we use the empirical distribution of agents across experience levels.

While there are a total of nine values for w_t in the model (corresponding to pairwise combinations of the three values for z_t), we can match them with only six in the data. In addition, for two of the six states, we cannot identify exit rates because they appear late in the sample, and so we can not identify if the agent leaves the market for the following two years or not. In total we have 6 moments for entry rates, 4 $\times$ 50=200 moments for exit rates, and 6 $\times$ 50=300 moments for experience distributions. In order to give the three categories of moments equal weight, we weight each set of moments according to model-implied probability distribution of being in each of the aggregate states and a particular experience level. $W_1(e, w)$ is a product of the long term probability that the aggregate state is w and the fraction of agents of experience e in that state w. While $W_1(w)$ is simply the long term probability that the aggregate state is w.

Formally, let $\Theta_2 = (c_e, \mu_{fc}, \sigma_{fc}, \phi)$ be the parameter space, $g_1(e, w, \Theta) = (\rho_{model}(e, w_t, \Theta) - \tilde{\rho}^{obs}(e, w_t))/\tilde{\rho}^{obs}(e, w$ be the exit rate and the distribution moments for each state and experience level and $g_2(w, \Theta) = (\Lambda_{model}(w_t, \Theta) - \tilde{\Lambda}^{obs}(w_t))/\tilde{\Lambda}^{obs}(w_t)$ be the entry rate moments for each available state. Then,

$$\hat{\Theta}_2 = \text{argmin}_{\Theta_2} \sum_{e,w} \left(W_1(e, w) g_1(e, w, \Theta)^2\right) + \sum_{w} \left(W_2(w) g_2(w, \Theta)^2\right). \tag{21}$$

We summarize the parameter values and the calibration strategy in Table 6.

5.4 Model fit

We next evaluate how well the model fits the data. To do so, we compare several moments in the model, both explicitly targeted in the calibration exercise and those not targeted, to their counterparts in the data.

The first set of moments identify four parameters to target the probability of sale in each state z_t for each experience group e. Figure 8 plots the values predicted by the model and the equivalent counterpart in the data. The model captures these rates quite well.

The next set of moments identify three parameters that govern the entry and exit rates of real estate agents for every state and experience level. Entry and exit, together with experience accumulation, are the three key

Table 6: Model calibration

Parameter	Value	Identifying Moment
P	$\begin{array}{cccc} & \text{Bust} & \text{Medium} & \text{Boom} \\ \text{Bust} & 0.65 & 0.16 & 0.19 \\ \text{Medium} & 0.23 & 0.58 & 0.19 \\ \text{Boom} & 0.12 & 0.25 & 0.63 \end{array}$	historical price data
$n^s(z)$	$[810,245 \quad 862,357 \quad 781,919]$	number of listings
v	$[\$202,915 \quad \$223,306 \quad \$224,863]$	price level
γ	0.5	-
β	0.9	-
$v_1(z)$	$[1 \quad 0.96 \quad 1.07]$	norm / average sale probability by state
v_2	0.03	sale probability by experience
c_b	\$162,788	overall sale probability
c_e	\$20,000	entry rates
μ_c	8.6	exit rates across experience groups
σ_c	1.7	
ϕ	0.3	experience accumulation

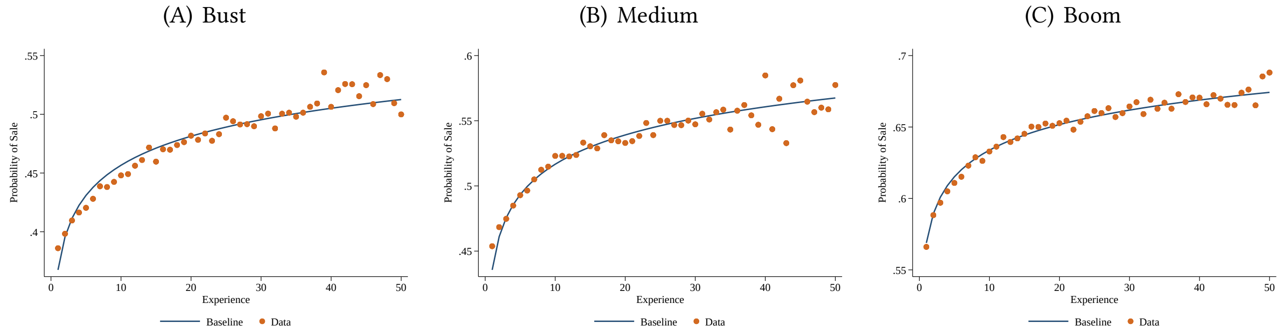
Note: This table reports the calibrated parameter values for the model, together with the description of the identifying moment in the data. See Section 5.3 for more details on the calibration procedure.

dynamic features that shape the experience distribution of real estate agents. To see how well our model fares against data, we first compare model fit by averaging all values across aggregate states observed in our sample weighted by probability of those states occurring in the model. Panel A of Figure 9 plots the average empirical and model exit rates at each level of experience. Next, Panel B compares the average changes in experience at each level of experience (i.e., the experience accumulation if the agent were to stay in the market). Last, Panel C plots the empirical and model generated distributions of experience. We see that the distribution of experience fits reasonably well but underfits the rate of entry (agents with experience of zero) and the fraction of inexperienced agents. The model captures the shape of the experience accumulation but predicts larger decay in experience than in the data.[33] Finally, the exit rates by experience match closely.

Recall that under our equilibrium concept, agents make their entry and exit decisions based on the recent history, namely the last two values, of aggregate states. In Table J9, we report the model fit for entry and exit rates as well as the experience accumulation and distribution in each realization of the aggregate state history that we observe in the data and for various experience levels. Interestingly, the model predicts no

[33] The experience accumulation in the model is quite low on average, but it varies by aggregate state. During very competitive periods struggle to find clients, while in less competitive periods they gain experience at every point in the experience distribution.

Figure 8: Sale probability: calibrated model vs. data

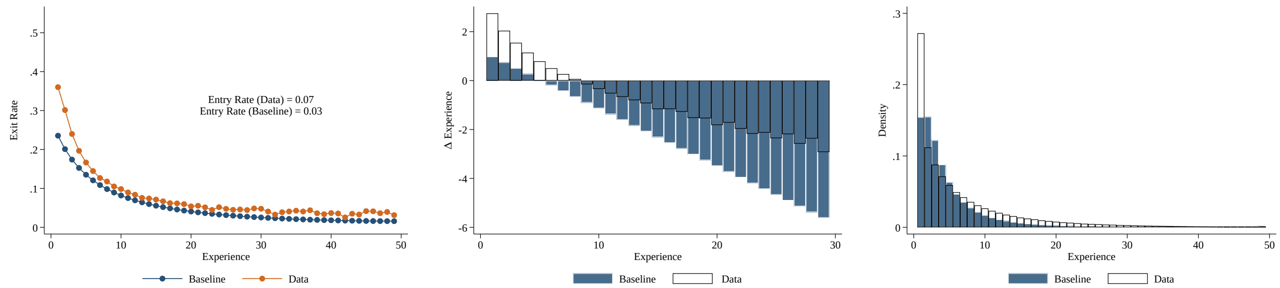


Note: This figure plots the sale probability for each agent experience level from the model and the data counterpart. In the model, these values vary only by aggregate state z, corresponding to housing boom, a medium state and the housing bust. The empirical counterpart plots the coefficients semiparametric estimates of the effect of experience on sale probability, from a regression of sale outcome variable on housing controls, zip-by-year-month fixed effects, and a separate dummy for each experience level of the listing agent (relative to experience of zero). The reported estimates are the estimated coefficient, plus the overall average sale probability for experience level at zero.

entry in periods that follow big spikes in entry in the previous period. The model matches exit rates fairly well in all states. To capture how fast agents accumulate experience, we compute the change in experience of agents conditional on staying in the market and present the expected change in experience for different experience points. To capture the distribution of agents, we compute the 25th, 50th, and 75th percentile of agent experience. With our calibrated model in hand, we can now consider various counterfactual changes to the model policy parameters and evaluate the change in market equilibrium.

Figure 9: Entry, exit, experience accumulation and distribution: calibrated model vs. data

(A) Exit Rates by Experience (B) Experience Accumulation (C) Distribution of Experience



Note: This figure plots the baseline model fit against the observed data. Panel A plots the aggregate exit rates across different experience bins in the equilibrium of the model and as observed in the data. It also reports the average entry rates for the model and the data. Panel B plots average experience accumulation. Panel C plots the average distribution of agents across experience levels, comparing the predicted model distribution against the observed experience distribution. As discussed in the calibration section, we do not observe states bust-boom, medium-medium, and boom-bust. In addition, we only observe bust-medium and medium-boom in the last two years, so it is not possible to identify exit probability for agents in those states, since we can not rule out them coming back to the market in the following two years.

6 Counterfactual Analysis

In this section, we propose a counterfactual without the main market failure of fixed commissions. We call this equilibrium our efficiency benchmark. We then estimate how large the entry cost adjustment would have be under the status quo of fixed commissions in order to reach the level of liquidity implied by the efficiency benchmark. This gives us a way to quantify the size of the distortion caused by the fixed commissions and agent heterogeneity in experience.

6.1 Efficiency Benchmark Equilibrium

The central failure in this market is that the real estate commission rate is fixed, both overall and across agent experience levels. This causes inefficiencies in several ways: first, as in Hsieh and Moretti (2003), the overall fixed level of the commission leads to inefficiently high level of entry without the benefit of competitive commission rates. Second, the inefficiently high entry causes each agent to work with fewer clients, leading to a slower accumulation of experience (i.e. slower improvement in matching technology). Third, the sluggish accumulation of experience leads to lower total earnings and a more likely exit (and hence permanently lost accrued experience). To quantify the effect of the fixed commission on market efficiency, we propose a counterfactual equilibrium in our model with flexible, competitive commissions.

In this competitive pricing equilibrium, agents compete for clients by setting competitive commission rates. A home seller trades-off between selling their home with a high experience agent (and higher probability of sale) for a higher commission, and a low experience agent (and lower sale probability) for a lower commission rate. We solve for the most competitive equilibrium where the new entrants charge no commission. This is akin to an apprenticeship set-up where professionals provide their first service for free in exchange for the experience they receive. While the competitive seller commissions are endogenous, we assume that each listing will still offer the standard 3% commission to a buyer agent who brings a successful buyer. This is a standard practice even for discount brokerages, and deviating from this component of the industry structure is beyond the scope of this paper.[34]

Formally, we define a seller commission function $\psi^s(e, w_t)$ to solve the seller's indifference condition of working with agents of different experience levels:

$$\mu(e,z_t)p(z_t)(1-3\%-\psi^s(e,w_t))+(1-\mu(e,z_t))\beta E_t V^S(w_{t+1}) = \mu(0,z_t)p(z_t)(1-3\%)+(1-\mu(0,z_t))\beta E_t V^S(w_{t+1}) \tag{22}$$

The left hand side is the value of working with an agent of experience e in aggregate state w_t[35]. With probability $\mu(e, z_t)$ a seller is able to successfully sell their home for the prevailing price $p(z_t)$. They will owe 3% commission to the buyer agent and pay their listing agent $\psi^s(e, w_t)$. With the complementary probability, they will receive the expected continuation value $E_t V^S(w_{t+1})$ of entering the market in the following period discounted at rate $\beta = 0.95$. The right hand side is the value of working with a new agent whose commission is set to 0: $\psi^s(0, w_t) = 0$. We solve for the continuation value recursively as if the seller always chooses to work with a new agent (not paying any commission):[36]

$$\tilde{V}^s(w_t) = \mu(0,z_t)p(z_t)(1-3\%) + (1-\mu(0,z_t))\beta E_t \tilde{V}^s(w_{t+1}) \tag{23}$$

The equilibrium of this model is defined in the same way as in the baseline model, with the exception of the agent profit equation, which now features a flexible commission $\psi^s(e, w_t)$ for the listing clients, as opposed to a constant commission ψ. Thus, equation 12 becomes:

$$E[\pi(e)|w_t, n_t^a] = E\left[0.1498\left(s(e,n_t^s;n_t^a)\mu(e,v_t)\psi^s(e,w_t)p(v_t) + b(e,n_t^b;n_t^a)\eta(v_t)\psi p(v_t)\right)^{1.1455}\right] \tag{24}$$

In computing this benchmark efficiency equilibrium, we are able to uncover what the commission rates

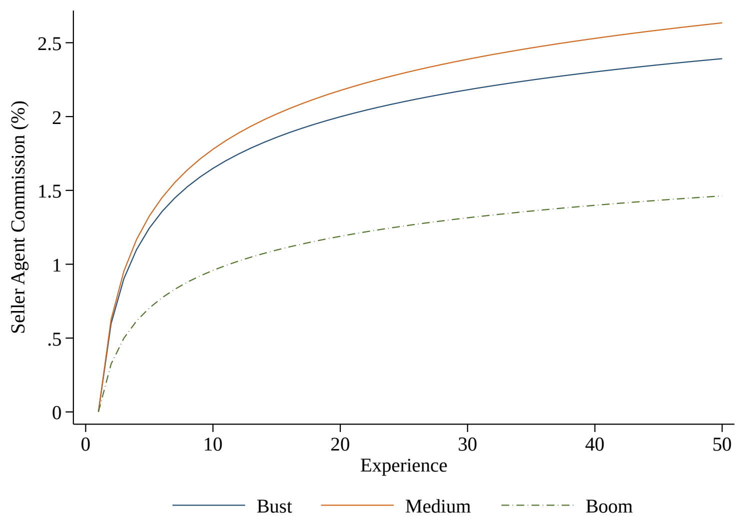
[34]See Barwick, Pathak, and Wong (2017) and Gilbukh, Goldsmith-Pinkham, and Sinkinson (2021) for a discussion of the institutional features, and Hatfield, Kominers, and Lowery (2020) for a discussion of the current market equilibrium between buyer and seller agents.

[35]Recall that w_t is the recent history of the aggregate state z_t, z_{t-1}. While the history does not affect sale probabilities and prices, it matters for state transitions when computing the seller's continuation value.

[36]We do not explicitly model the state transitions for sellers and buyers. Formally they enter and exit according to an exogenous process. However in this set up the continuation value for the seller is crucial to place a realistic value on increasing sale probability. Assuming that the seller has a chance to enter the search market every period provides a natural way to compute this continuation value using model fundamentals.

would be in a competitive market for intermediation. Figure 10 plots model-implied commission rates against agent experience level in the three aggregate states. Experienced agents can charge a premium because they offer their clients a more experienced matching technology. In the boom, the most experienced listing agents can charge about 1.5% commission in addition to the 3% offered to the buyer agent. In the medium and bust states they can charge about 2.8% and 2.6% respectively, for a total commission of close to 6% in both states - the rate that prevails in the status quo. The experience premium is especially large in the bust and medium states for two reasons. First, in these bad states of the market experience matters most for probability of sale. Second, sellers particularly value selling their homes in the bust and medium markets because their continuation value is lower then, due to the persistence of these states.

Figure 10: Benchmark Equilibrium Flexible Seller Commission

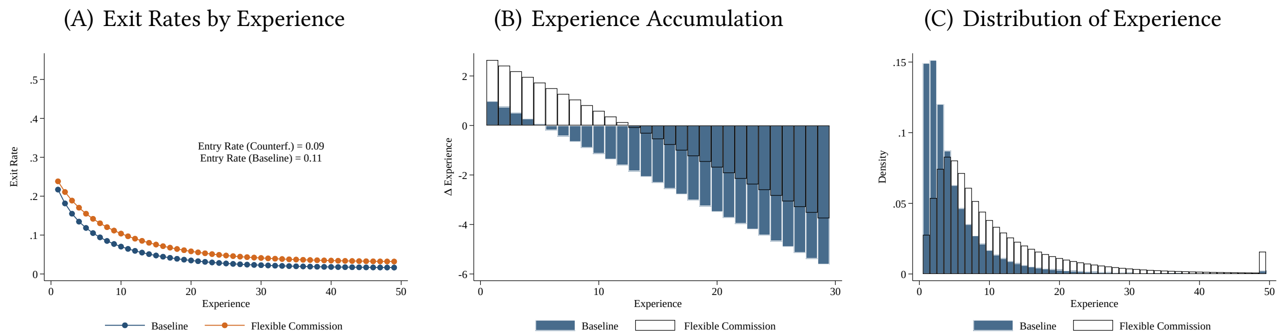


Note: This figure plots equilibrium commission levels of our benchmark efficiency model specification. Here new agents do not charge commission, while more experienced agents charge a competitive commission rate. A competitive rate insures that sellers get the same utility no matter who they work with. They trade off higher commission for a higher probability of sale provided by more experienced agents.

Figure 11 plots the equilibrium exit rates, experience accumulation, as well as the distribution of agents across experience for the flexible commission equilibrium and the baseline specification. Competitive commissions make it less profitable for new agents to enter the market as inexperienced agents make very low commissions on listing clients. However, with fewer entrants, the existing agents are able to work with more clients and accumulating experience faster. Although exit rates are higher in this equilibrium and some experience is lost with the exiting agents, the faster learning compensates for this channel. Overall, the experience distribution shifts significantly towards higher experience.

We next evaluate the improvement in the probability of sale and seller welfare in the efficiency bench-

Figure 11: Entry, exit, experience accumulation and distribution: baseline vs. flexible commission benchmark



Note: This figure plots model results for the flexible commission equilibrium and compares them to the baseline model. Panel A plots the aggregate exit rates across different experience bins. It also reports the average entry rates for the two models. Panel B plots average experience accumulation. Panel C plots the average distribution of agents across experience levels.

mark equilibrium as compared to the baseline model. Note that because of the buyer free entry condition, the probability of purchase remains the same in all specifications and consequently the buyer welfare is unaffected by the change in commission structure. We believe that the free entry assumption for buyers is a reasonable one. While a seller's decision to sell is often accompanied by extraneous circumstances, potential buyers often have more flexibility (for example rent for an additional year). Thus marketing the property to be particularly attractive can actually bring more buyers into seriously considering to buy. We explore relaxing the buyer free entry assumption in Appendix H. To evaluate welfare consequences for the sellers we make assumptions about continuation values in the case of no-purchase of no-sale.[37] We assume that sellers who do not transact in a given period return to the market the next period and repeat the effort to sell. A seller's value function is the solution to the following value function:

$$\tilde{V}^s(w) = \sum_{\tilde{e}} \underbrace{\left(\phi \frac{n^a(w,\tilde{e})}{\sum_e n^a(w,e)} + (1-\phi)\frac{\tilde{e}n^a(w,\tilde{e})}{\sum_e n^a(w,e)e}\right)}_{\text{Match prob. with exp. } e} \times \\ \left(\underbrace{\mu(\tilde{e},v(w))(1-\psi)p(v(w))}_{\text{Sell this period}} + \underbrace{(1-\mu(\tilde{e},v(w)))\beta E[\tilde{V}^s(w')|w]}_{\text{Try to sell next period}}\right). \tag{25}$$

The first part measures the expected value of selling a home in the current period (including the cost of the commission), conditional on matching with an agent of a given experience; the second part is the value of moving into the next period with an unsold house, which is scaled by the probability of *not* selling the home this period with an agent of a given experience. These values are then integrated over the relative probability

[37]Recall that entry and exit of buyers and sellers is exogenous in the model.

of matching with an agent of experience e.

Panel B of Table 7 presents the results for the efficiency benchmark equilibrium. The probability of sale improves more relative to the baseline model in the bust and medium states. This is unsurprising because experience has a larger effect in those periods. Overall, the efficiency benchmark delivers a 54.3 pp of sold listings relative to 52.4 pp, an improvement of 3.7%. Given our empirical estimate that 4.5% of unsold properties went into foreclosure in the bust period, we estimate the improvement in foreclosure rate from the efficient equilibrium to be 3.4%. [38] Seller valuation increases due to two factors. First, the increased probability of sale delivers a more likely income from sale. Second, lower commissions allow for savings on all sales. On average, seller value increases to $199,344 from $195,638 in the baseline calibration.

As noted earlier, the probability of purchase as well as buyer valuations are unchanged, because new buyers entering the market crowd out the benefits of improved matching technology. In Appendix H we explore an alternative specification of the model where we fix the number of buyers across states and thus allow both buyers and sellers to benefit from improvements in technology. We find that the improvement in probability of sale for the sellers is significantly smaller if no new buyers are allowed to enter the market. Instead of a 3.7% improvement in probability of sale, the fixed buyer specification of the efficiency benchmark delivers only about 1.1% increase in the sale probability (Appendix Table H1). However, buyers then also benefit with about 1.1% increase in probability of purchase.

6.2 Policy Analysis

The flexible commission benchmark equilibrium allows us to quantify the inefficiencies associated with the main market failure – the fixed commission rate. We now examine how large a policy intervention needs to be to achieve this benchmark level of liquidity under the status quo of fixed commissions. We consider a counterfactual where policy makers directly raise the cost of entry either through increasing licensing fees or mandating longer training.

We solve the model for a grid of entry cost parameters, each corresponding to a $1000 increase in c_e from the baseline value of $20,000 calibrated as described in section 5.3. We find that to achieve the agent distribution that delivers the efficiency benchmark level of liquidity, the entry cost needs to be as high as $124,000, more than six times higher than the baseline value of $20,000. Figure 12 shows the equilibrium exit and entry rates, experience accumulation, as well as the overall average experience distribution of this policy. Free entry condition implies that to compensate for this increased entry cost, new agents have to work with

[38] A 1.9 pp improvement in sale probability would lead to a 0.019 * 0.045 =0.09 pp improvement in foreclosure rate. This would bring down the overall foreclosure rate of 2.5pp in 2008 by about 3.4 percent.

Table 7: Counterfactual Results

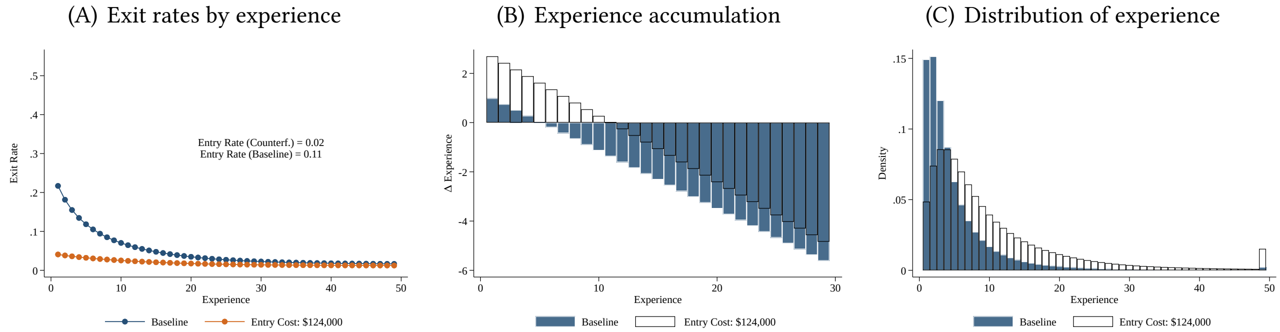
Panel A: Baseline				
	Sale Prob	Buy Prob	Seller Value	Buyer Value
Bust	0.441	0.802	185,360	201,790
Medium	0.502	0.729	198,170	218,000
Boom	0.622	0.724	203,160	219,660
Total	0.524	0.752	195,638	213,166
Panel B: Flexible Commission Benchmark				
	Sale Prob	Buy Prob	Seller Value	Buyer Value
Bust	0.462	0.802	188,543	201,790
Medium	0.522	0.729	201,553	218,000
Boom	0.638	0.724	207,783	219,660
Total	0.543	0.752	199,344	213,166
Panel C: Increased Entry Cost				
	Sale Prob	Buy Prob	Seller Value	Buyer Value
Bust	0.463	0.802	185,870	201,790
Medium	0.523	0.729	198,990	218,000
Boom	0.640	0.724	203,780	219,660
Total	0.543	0.752	196,268	213,166

Note: This table reports sale and buy probabilities as well as the seller and buyer value, in each of the three periods and overall (weighted by each state's ergodic probability). The values are reported for three models. Panel A reports results from the baseline specification. Panel B reports results from the flexible commission counterfactual discussed in section 6.1. Finally, Panel C reports results from the counterfactual policy of increased entry costs that targets sale probability improvements delivered by the flexible commission efficiency benchmark (as described in Section 6.2.

more clients to earn more profits. As a result, entrants learn faster and are less likely to exit the market. With little exit and somewhat faster learning, the overall distribution of agents shifts significantly towards higher experience.

In this counterfactual, probability of sale increases about the same as in the efficiency benchmark equilibrium (by construction). However, seller valuation does not improve as much as it does in the benchmark efficiency scenario because here sellers still face a total of 6% commissions on the sale, as opposed to paying competitive rates. Panel C of Table 7 presents the results. On average, seller value improves to \$196,268 from \$195,638. In Appendix H, we repeat the exercise for an alternative model specification where the number of buyers is fixed for the counterfactual analysis. With fixed number of buyers, the counterfactual entry cost that matches the efficiency benchmark improvement in sale probability is \$108,000. It provides a similar increase in sale and buy probability as the efficiency benchmark equilibrium with buyer number fixed, with a modest improvement in buyer and seller welfare.

Figure 12: Counterfactual experiment: increase of entry cost to $124,000 dollars



Note: This figure plots the baseline model fit against the counterfactual of setting entry costs to $124,000 dollars. Panel A plots weighted average exit rates across different experience bins in the baseline and counterfactual equilibrium of the model. Panel A also reports the average entry rates for the baseline and counterfactual model. Panel B plots average experience accumulation. Panel C plots the average distribution of agents across experience levels, comparing the baseline model distribution against the counterfactual distribution.

7 Conclusion

The experience of real estate agents affects the sale probability of homes listed for sale, and this effect aggregates to influence housing liquidity over the housing cycle through the distribution of experience. Downturns are particularly affected for two reasons: first, not only are inexperienced agents worse at selling listings, but they are especially bad during housing busts. Second, due to low barriers to entry and fixed commission rates, the housing boom attracts many new agents into the profession, intensifying competition for clients and thus hindering experience accumulation. These new agents remain in the market for the onset of the downturn, resulting in a distribution skewed toward lower experience.

The main market failure that results in sub-optimal agent entry is the fixed commission structure of the real estate intermediation market. To quantify how much this market failure contributes to low sale probability due to the prevalence of inexperienced agents, we build an entry and exit model of real estate agents with aggregate shocks. We find that allowing for competitive commissions improves sale probability by 3.7%. To achieve a comparable improvement in sale probability within the status quo of fixed commissions, the policy markers would have to increase the entry cost by more than six times the calibrated value.

References

Abadie, Alberto, Susan Athey, Guido W Imbens, and Jeffrey Wooldridge. 2017. "When should you adjust standard errors for clustering?" Tech. rep., National Bureau of Economic Research.

Abaluck, Jason, Mauricio M Caceres Bravo, Peter Hull, and Amanda Starc. 2020. "Mortality Effects and Choice Across Private Health Insurance Plans." Tech. rep., National Bureau of Economic Research.

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.

Barwick, Panle Jia, Parag A Pathak, and Maisy Wong. 2017. "Conflicts of interest and steering in residential brokerage." *American Economic Journal: Applied Economics* 9 (3):191–222.

Barwick, Panle Jia and Maisy Wong. 2019. "Competition in the Real Estate Brokerage Industry: A Critical Review." .

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

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How much should we trust differences-in-differences estimates?" *The Quarterly Journal of Economics* 119 (1):249–275.

Bhuller, Manudeep, Gordon B Dahl, Katrine V Løken, and Magne Mogstad. 2020. "Incarceration, recidivism, and employment." *Journal of Political Economy* 128 (4):1269–1324.

Buchak, Greg, Gregor Matvos, Tomasz Piskorski, and Amit Seru. 2020. "iBuyers: Liquidity in Real Estate Markets?" *Available at SSRN 3616555* .

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

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

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

Garriga, Carlos and Aaron Hedlund. 2020. "Mortgage debt, consumption, and illiquid housing markets in the great recession." *American Economic Review* 110 (6):1603–34.

Gilbukh, Sonia, Paul Goldsmith-Pinkham, and Michael Sinkinson. 2021. "How to Make the Market for Real Estate Agents More Competitive." *promarket.org* .

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, Adam M. 2018. "House price momentum and strategic complementarity." *Journal of Political Economy* 126 (3):1172–1218.

Guren, Adam M and Timothy J McQuade. 2019. "How do foreclosures exacerbate housing downturns?" Tech. rep., National Bureau of Economic Research.

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.

Han, Lu and William C Strange. 2015. "The microstructure of housing markets: Search, bargaining, and brokerage." In *Handbook of regional and urban economics*, vol. 5. Elsevier, 813–886.

Hatfield, John William, Scott Duke Kominers, and Richard Lowery. 2020. "Collusion in Brokered Markets." Working Paper 20-023, Harvard Business School Entrepreneurial Management. URL https://ssrn.com/abstract=3450767. Available at SSRN: https://ssrn.com/abstract=3450767 or http://dx.doi.org/10.2139/ssrn.3450767.

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.

Hedlund, Aaron. 2016a. "The cyclical dynamics of illiquid housing, debt, and foreclosures." *Quantitative Economics* 7 (1):289–328.

———. 2016b. "Illiquidity and its discontents: Trading delays and foreclosures in the housing market." *Journal of Monetary Economics* 83:1–13.

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.

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.

Imbens, Guido W and Joshua D Angrist. 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica: Journal of the Econometric Society* :467–475.

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.

Krusell, Per and Anthony A Smith, Jr. 1998. "Income and wealth heterogeneity in the macroeconomy." *Journal of political Economy* 106 (5):867–896.

Kurlat, Pablo and Johannes Stroebel. 2015. "Testing for information asymmetries in real estate markets." *The Review of Financial Studies* 28 (8):2429–2461.

Lee, David L, Justin McCrary, Marcelo J Moreira, and Jack Porter. 2020. "Valid t-ratio Inference for IV." *arXiv preprint arXiv:2010.05058* .

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 and Amir Sufi. 2014. "What explains the 2007–2009 drop in employment?" *Econometrica* 82 (6):2197–2223.

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.

National Association of Realtors. 2017a. “2017 Member Profile.” Tech. rep., National Association of Realtors.

———. 2017b. “2017 Profile of Home Buyers and Sellers.” Tech. rep., National Association of Realtors.

Ngai, L Rachel and Silvana Tenreyro. 2014. “Hot and cold seasons in the housing market.” *American Economic Review* 104 (12):3991–4026.

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

Saiz, Albert. 2010. “The geographic determinants of housing supply.” *The Quarterly Journal of Economics* 125 (3):1253–1296.

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

Staiger, Douglas and James H Stock. 1994. “Instrumental variables regression with weak instruments.” Tech. rep., National Bureau of Economic Research.

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

Online Appendix for

Heterogeneous Real Estate Agents and the Housing Cycle

Sonia Gilbukh   Paul Goldsmith-Pinkham

Table of Contents

A Measuring experience

Explored here are different measures of experience available in the data as well as detailed analysis of our baseline experience measure. 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.[39] We are interested in constructing a measure that is most predictive of our variables of interest: the number of new clients that each agent 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. We conclude that our preferred measure is best suited for our analysis.

A.1 Predicting New Clients

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[40]. Years since entry does not capture nearly as much variation as the baseline specification. In Table A2, we report the pairwise correlations for each of these measures. From a pure correlation standpoint, the most highly correlated measure with current period clients is last period, followed closely by the number of sellers (a subset of total clients).

A.2 Affecting Listing Outcomes

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 probability of sale. Appendix Table A3 presents the results. We regress sale probability on the log of experience measure plus one, controlling for housing characteristics, and adding zip-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

[39] 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 B. 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.

[40] 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

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 consider it the best choice of experience measure for our analysis.

A.3 What does experience capture?

We now document more details on the nature of our experience measure and how it varies over agents' tenure. Note that we can only observe the full lifecycle of experience for agents that enter and exit during our sample period. For others, we miss their experience accumulation pattern after entry or leading up to exit, or both.

We highlight four empirical facts that are broadly consistent with our model of agents' experience. First, the experience level for an agent who enters the market is slow moving but gradually grows conditional on not exiting. As time progresses, the dispersion in experience becomes wider. Second, for low experience agents the transition rate to higher experience is particularly low; agents with higher experience fluctuate but have a low probability of transitioning to early-stage experience levels. Third, for agents that do eventually exit, on average our experience measure declines in the years prior to exit. Fourth, because of these life-cycle features, the within-agent variation is substantially smaller than the cross-agent variation, with a within-agent standard deviation of 5.8 compared to an overall standard deviation of 10.3. Agent fixed effects explain 68 percent of the overall variation. If we focus on a sample of agents who are either seasoned or several years after their entry, the within-agent variation in experience is even smaller relative to the overall variation (5.4 within vs. 11.9 overall), with agent fixed effects explaining 79 percent of the overall variation.

Transition Probabilities In Appendix Figure A1, we document how the experience measure changes over time. This heat map plots the share of agents who for a given experience level this year (x axis) transition to a given experience level next year (y axis), with each column summing to one. For low levels of experience, the transition outcomes are tightly clustered close to 1, with few agents gaining more experience. As experience level grows, we see the tendency to cluster around the same level again, with some variation around that mean. The white line denotes the y = x line, which shows that experience accumulation is a slow process with some mean reversion downwards, possibly driven by decline in experience prior to subsequent exit, as we explore below. It's important to note that many agents have low experience: conditional on being past the first year since entry, the median agents' experience is 5.

Life Cycle To further explore the life cycle of experience, we next focus on the change in experience over time for new entrants, conditional on survival. First, we plot their level of experience over time (Appendix Figure A2). We see that after a few years, the center of the distribution stabilizes at a median of around 5 with a large right skewed tail. In Appendix Figure A3, we plot the change in experience over time. While the change in experience can be big, the 5th and 95th percentiles are -9 and 10, respectively. After the first three years, the median change is zero. This suggest that there is significant experience accumulation early on, but over time, and conditional on survival, the within-agent fluctuation in experience is small with the modal agent having a large degree of persistence around their current experience measure.

Experience can also decline if an agent is either less active or unlucky (gets fewer listing clients, or is unable to buy new homes with buying clients). We view this decline as equally important, since it implies less connections and potential leads for the future. This decline is apparent among those agents who eventually exit. In Appendix Figure A4 and A5, we plot the average experience level for agents who eventually exit the market, split by entrants (those who we see enter) and seasoned agents (incumbent agents who were in the market at the beginning of our sample). Over time, there is a slowdown in experience accumulation that turns slightly negative prior to exit.

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

Note: This table shows regressions of number of listings (successful or not) and successful purchases an agent has in the current period on several measures of prior activity. In Column 1, the right hand side variable is the sum of all clients (both buyers and sellers) in the previous year. In Column 2, the regression splits on lagged buyer and seller client count separately. Column 3 adds unsuccessful sales. In Columns 4 and 5 we add additional lags of buyers and sellers. In Column 6, we instead look at how many years the agent has been active since entry in our data. Column 7 repeats Column 1 with a subsample of data used in Column 6.

Table A2: Correlation between Experience measures

	Clients (t)	Clients (t-1)	Buyers (t-1)	Sellers (t-1)	Failed Sellers (t-1)	Buyers (t-2)	Sellers (t-2)	Buyers (t-3)	Sellers (t-3)	Years Active
Clients (t)	1.00									
Clients (t-1)	0.73	1.00								
Buyers (t-1)	0.23	0.30	1.00							
Sellers (t-1)	0.72	0.98	0.12	1.00						
Failed Sellers (t-1)	0.68	0.93	0.04	0.96	1.00					
Buyers (t-2)	0.18	0.24	0.76	0.11	0.04	1.00				
Sellers (t-2)	0.42	0.82	0.18	0.82	0.75	0.21	1.00			
Buyers (t-3)	0.16	0.19	0.60	0.09	0.04	0.76	0.18	1.00		
Sellers (t-3)	0.20	0.26	0.24	0.23	0.10	0.31	0.47	0.37	1.00	
Years Active	0.10	0.15	0.31	0.10	0.05	0.37	0.20	0.40	0.35	1.00

Note: This table reports the bivariate correlations between the different experience measures.

Table A3: 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

Note: 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. The next columns 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.

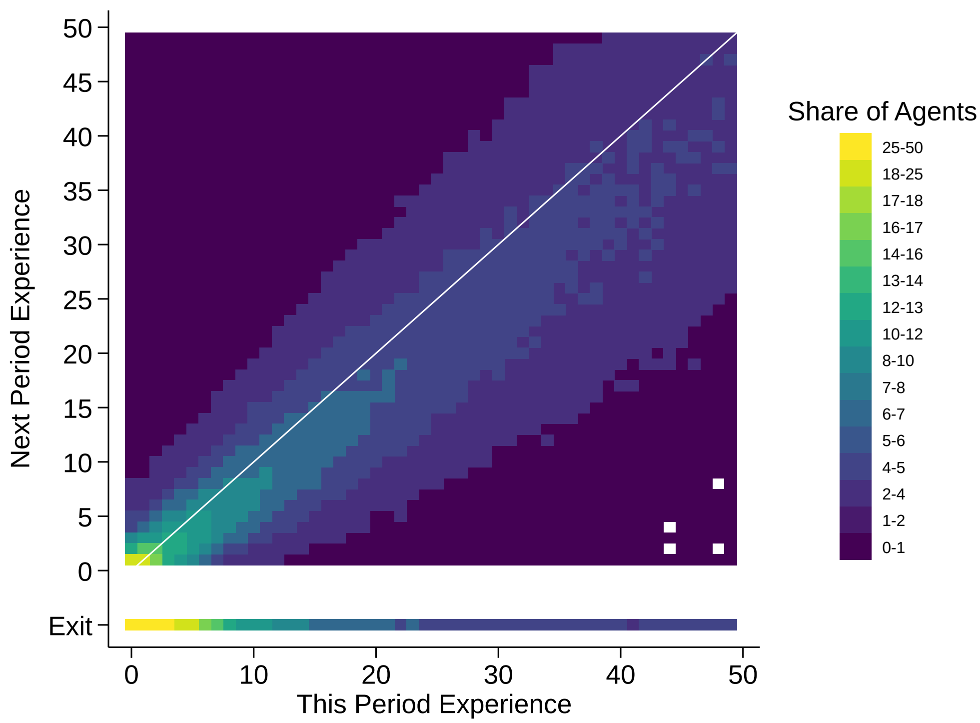


Figure A1: Transition probabilities for experience levels. Cell probability corresponds to the Markov probability of transitioning to next period's experience, conditional on this year's experience.

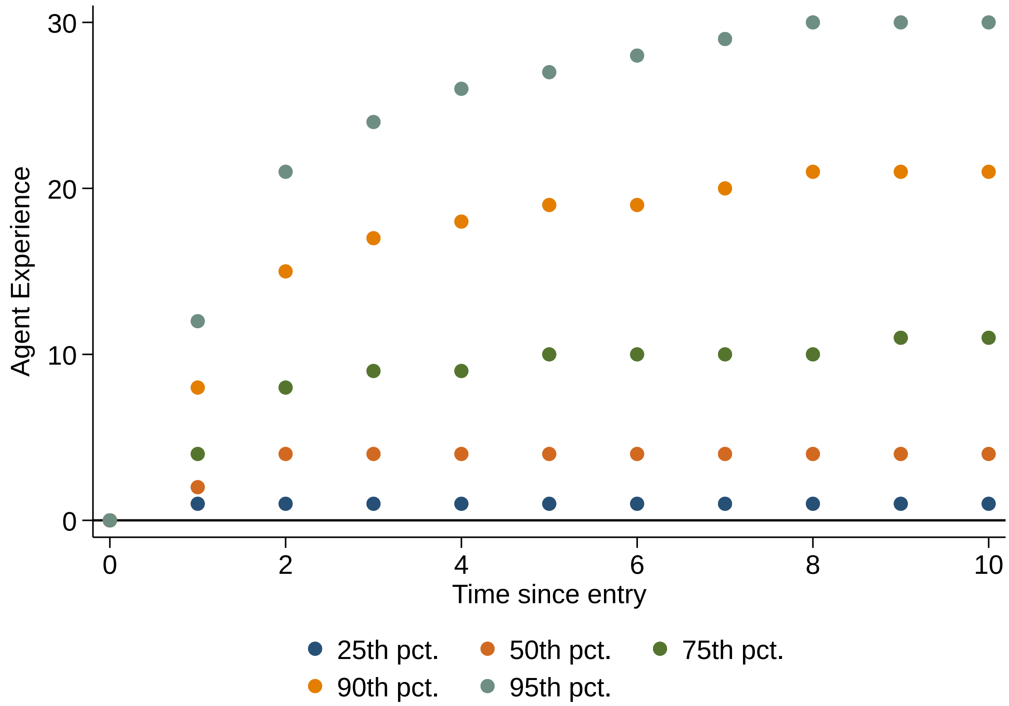


Figure A2: Life Cycle of Agent Experience Levels

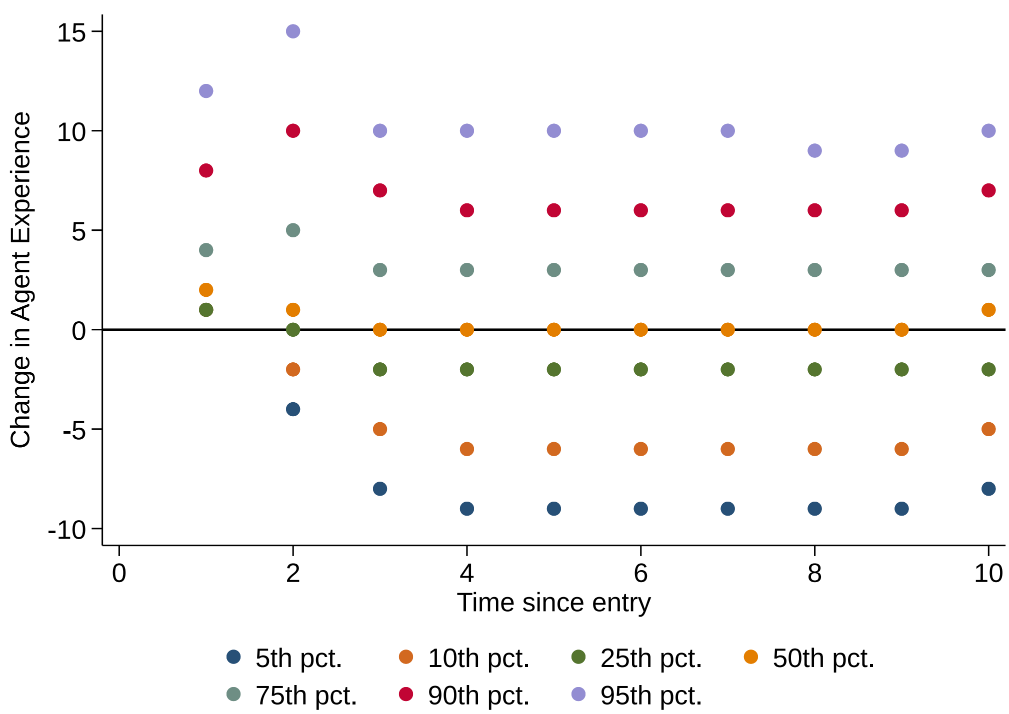


Figure A3: Change in Agent Experience over Lifecycle

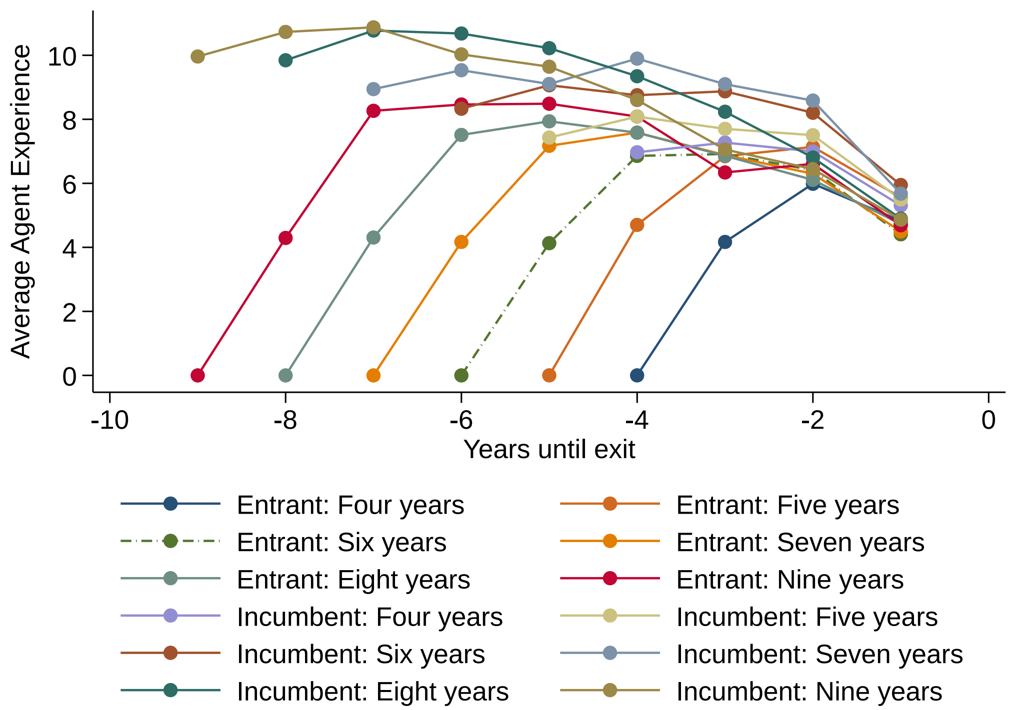


Figure A4: Life Cycle of Agent Experience Levels Prior to Exit

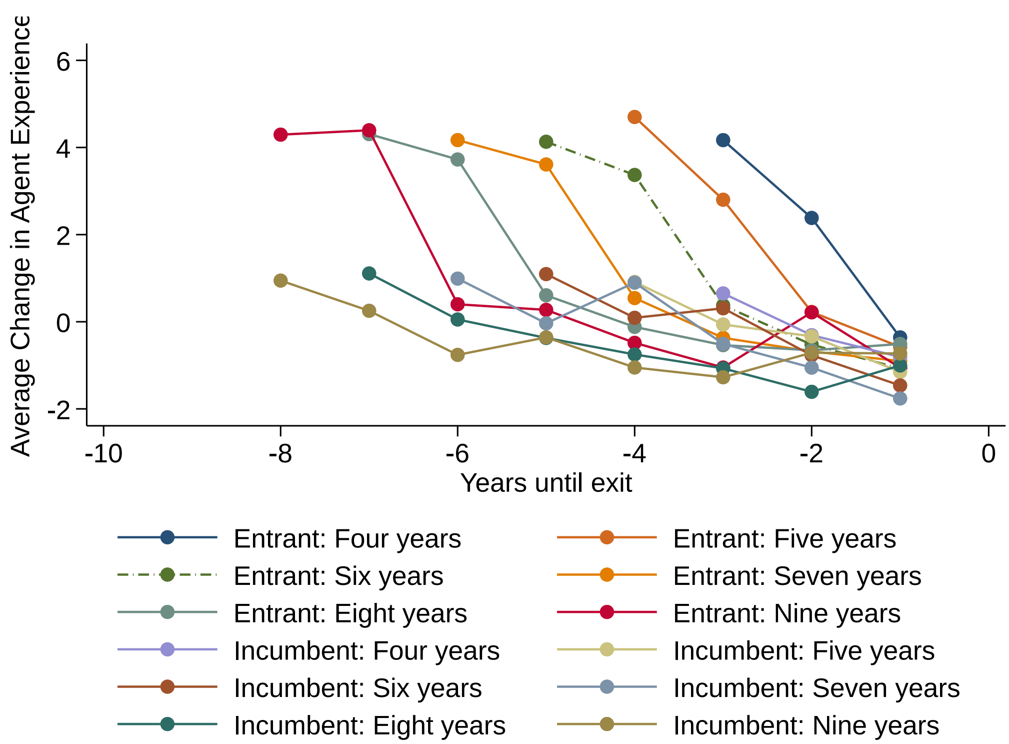


Figure A5: Change in Agent Experience over Lifecycle Prior to Exit

B Additional robustness analysis

In this section, we elaborate on the additional robustness tests that we use to rule out alternative theories of selection between listings and agents.

We first consider the alternative mechanism that agents with higher experience work with properties that have unobserved (to the econometrician) qualities that make them of higher value and, as a result, easier to sell. To address this issue, we control for the inferred price of each home and rerun our main specification in Column 3 of Table 2. We measure the inferred price using the previous observed sale price (as measured using deeds data) for the property and appreciating the value of the home using Zillow zip-code- and tier-level house price appreciation indexes. We report these estimates for each of our main outcomes in Appendix Table B1, and find very similar results to our main specification.

Table B1: Effect of experience on outcomes controlling for inferred price

	Sale Pr.	Will Foreclose	List / Inferred	Sale/Infer	List / Sale
	(1)	(2)	(3)	(4)	(5)
Log(Exp + 1)	0.0342***	-0.0003**	-0.0096***	-0.0088***	0.0002
	(0.0030)	(0.0001)	(0.0022)	(0.0032)	(0.0009)
Bust $\times$ Log(Exp + 1)	0.0133***	-0.0030***	-0.0051***	-0.0008	-0.0014
	(0.0030)	(0.0010)	(0.0016)	(0.0028)	(0.0009)
Medium $\times$ Log(Exp + 1)	0.0028	-0.0008**	-0.0006	0.0010	-0.0004
	(0.0026)	(0.0003)	(0.0010)	(0.0014)	(0.0006)
Time-by-Zipcode FE	Yes	Yes	Yes	Yes	Yes
House Char.	Yes	Yes	Yes	Yes	Yes
Inferred House Price	Yes	Yes	Yes	Yes	Yes
Bust Effect	0.0475	-0.0034	-0.0147	-0.0096	-0.0013
Bust p-value	0.0000	0.0028	0.0000	0.0082	0.3146
Medium Effect	0.0371	-0.0011	-0.0102	-0.0078	-0.0002
Medium p-value	0.0000	0.0144	0.0000	0.0204	0.8201
Observations	2752831	2465516	2203966	1318153	1291368

Note: This table reports estimates for our outcomes using our main specification from Equation 1 with an additional control for inferred price. We measure the inferred price using the previous observed sale price (as measured using deeds data) for the property and appreciating the value of the home using Zillow zip-code- and tier-level house price appreciation indexes. Displayed are our preferred specification of regression outcomes in equation 1 for several variables: sale probability, future foreclosures, relative list price, relative sale price and the discount from the original list price. The regressions include zipcode-by-year-month fixed effects and housing controls (the same controls as Column 3 in Table 2). In Column 1, we report the effect of experience on the probability of sale in 365 days. In Column 2, we report the effect on subsequent foreclosures. Column 3 reports the effect of agent experience on list price normalized to inferred price (measured using the previous sale price, appreciated using zipcode- and price-tier-specific Zillow house price appreciation) for all listings. Column 4 reports the effect on sale prices normalized to inferred price. Column 5 reports the discount that a property sells at relative to its list price. Standard errors are clustered at the MLS-level. See Section 3 for more details on the data sample and definition of experience.

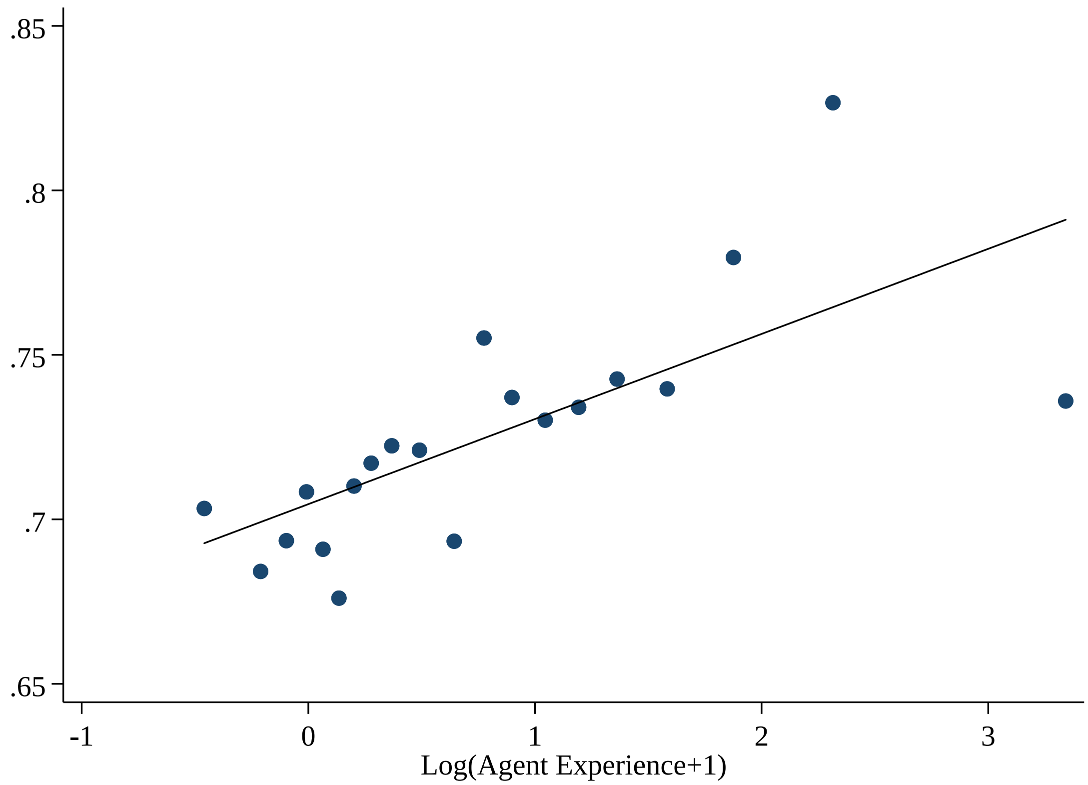
As an additional check for selection-on-properties by agents, we restrict our analysis to a homogeneous suburb of San Diego, Chula Vista, where houses are nearly identical. In this market, the standard deviation of prices for listings is less than 20 percent, and as a result, there is limited range for agents of differing experience to select into different types of home. Appendix Figure B1 shows the satellite view of this area, illustrating the homogeneity of properties.

Figure B1: Satellite view of Chula Vista, CA

Note: A satellite view of Chula Vista, CA from Google Maps.

Appendix Figure B2 repeats our main empirical results from Section 4.2 and find the same linear and monotonic relationship between agent experience and the probability of sale in Chula Vista. In Column 1 of Appendix Table B2, using our preferred regression specification from Column 3 of Table 2, we find that the effect of experience on the probability of sale is still positive but is smaller in magnitude during the boom period. However, the effect of experience in the medium and bust periods are large and significant, similar to what we find in Table 2. We report the estimates for our other outcomes in Columns 2-5 and find similar results to the main analysis.

Figure B2: Agent experience and probability of sale in Chula Vista, CA



Note: This figure focuses on the subsample of listings in Chula Vista, CA. This figure plots a binned scatterplot (with 20 bins) of the probability that a listing sells within 365 days against the listing agent's experience (using the $\log(1 + \text{agent experience})$). This plot and fitted line account for zipcode-by-year-month fixed effects and housing controls (the same controls as Column 3 in Table 2). The fitted line, average bin values, and the reported coefficient correspond to the coefficient on β of Equation 1, not allowing β to vary by time period. Standard errors are clustered at the zipcode-level. See Section 3 for more details on the data sample and definition of experience.

Table B2: Effect of experience on outcomes in Chula Vista, CA

			Log Prices		
	Sale Pr.	Will Foreclose	List / Inferred	List / Sale	Sale/Infer
	(1)	(2)	(3)	(4)	(5)
Log(Exp + 1)	0.0047	-0.0001	0.0074*	0.0057*	0.0015
	(0.0054)	(0.0002)	(0.0030)	(0.0024)	(0.0016)
Bust $\times$ Log(Exp + 1)	0.0376**	-0.0167**	0.0003	0.0117*	-0.0107**
	(0.0138)	(0.0051)	(0.0050)	(0.0051)	(0.0029)
Medium $\times$ Log(Exp + 1)	0.0470***	-0.0018	-0.0036	0.0022	0.0003
	(0.0093)	(0.0014)	(0.0042)	(0.0040)	(0.0016)
Time-by-Zipcode FE	Yes	Yes	Yes	Yes	Yes
House Char.	Yes	Yes	Yes	Yes	Yes
Bust Effect	0.0424	-0.0167	0.0077	0.0174	-0.0092
Bust p-value	0.0376	0.0150	0.0956	0.0212	0.0052
Medium Effect	0.0517	-0.0019	0.0038	0.0079	0.0017
Medium p-value	0.0010	0.1615	0.2845	0.0314	0.3801
Observations	11128	10258	5740	4114	4104

Note: This table reports estimates for our outcomes using our main specification from Equation 1, focusing on the homogeneous subsample of listings in Chula Vista, CA. The regressions include zipcode-by-year-month fixed effects and housing controls (the same controls as Column 3 in Table 2) In Column 1, we report the effect of experience on the probability of sale in 365 days. In Column 2, we report the effect on subsequent foreclosures. Column 3 reports the effect of agent experience on list price normalized to inferred price (measured using the previous sale price, appreciated using zipcode- and price-tier-specific Zillow house price appreciation) for all listings. Column 4 reports the effect on sale prices normalized to inferred price. Column 5 reports the discount that a property sells at relative to its list price. Standard errors are clustered at the zipcode-level. See Section 3 for more details on the data sample and definition of experience.

We next consider whether agents with higher experience choose to work with clients whose properties are easier to sell. To test this, we control for the client equity at the time of the listing, as proxied by the amount of house price appreciation experienced by the seller since the house was last transacted. As argued in Guren (2018), there are two reasons why clients with lower equity are likely to be less flexible in the selling process. First, low equity sellers are likely to be cash constrained, especially if they are looking into purchasing another property and need money for down payment. Second, sellers who have a higher equity in the property are less likely to experience loss aversion from selling at a lower price than what they initially paid. We control for this house price appreciation and report the estimates in Appendix Table B3. Again, we find similar results to our main estimates.

As an additional check for selection-on-clients by agents, we examine a subsample of listings that followed a deed transfer that we assume proxies for a life-changing event (Kurlat and Stroebel, 2015). Specifically, we look at listings that occur within two years of a previous transaction where both parties have the same last name but have a different first name. These transactions likely capture a transfer of property from a married couple to one partner, which likely happens in a case of divorce or death of one of the spouses. Sellers in this sample are likely more motivated in getting rid of the property than an average seller because they either cannot afford maintaining it or do not have use for it altogether. Due to a smaller sample size across locations, we are unable to control for zip-code-by-list-year-month fixed effects and instead include county-by-list-year-month fixed effects. We first replicate our main figure and find the same linear and monotonic relationship

Table B3: Effect of experience on outcomes controlling for equity stake

	Sale Pr.	Will Foreclose	List / Inferred	Sale/Infer	List / Sale
	(1)	(2)	(3)	(4)	(5)
Log(Exp + 1)	0.0341***	-0.0003**	-0.0092***	-0.0085***	0.0001
	(0.0030)	(0.0001)	(0.0021)	(0.0031)	(0.0009)
Bust × Log(Exp + 1)	0.0132***	-0.0031***	-0.0030*	0.0012	-0.0014
	(0.0030)	(0.0010)	(0.0016)	(0.0030)	(0.0009)
Medium × Log(Exp + 1)	0.0025	-0.0008**	0.0003	0.0020	-0.0003
	(0.0026)	(0.0003)	(0.0009)	(0.0012)	(0.0006)
Time-by-Zipcode FE	Yes	Yes	Yes	Yes	Yes
House Char.	Yes	Yes	Yes	Yes	Yes
Equity Stake	Yes	Yes	Yes	Yes	Yes
Bust Effect	0.0473	-0.0033	-0.0122	-0.0073	-0.0012
Bust p-value	0.0000	0.0026	0.0000	0.0478	0.3255
Medium Effect	0.0366	-0.0010	-0.0089	-0.0065	-0.0001
Medium p-value	0.0000	0.0126	0.0003	0.0503	0.8721
Observations	2752831	2465516	2203966	1318153	1291368

Note: This table reports estimates for our outcomes using our main specification from Equation 1 with an additional control for seller equity stake. We proxy equity stake by house price appreciation since the previous sale.
Displayed are our preferred specification of regression outcomes in equation 1 for several variables: sale probability, future foreclosures, relative list price, relative sale price and the discount from the original list price. The regressions include zipcode-by-year-month fixed effects and housing controls (the same controls as Column 3 in Table 2). In Column 1, we report the effect of experience on the probability of sale in 365 days. In Column 2, we report the effect on subsequent foreclosures. Column 3 reports the effect of agent experience on list price normalized to inferred price (measured using the previous sale price, appreciated using zipcode- and price-tier-specific Zillow house price appreciation) for all listings. Column 4 reports the effect on sale prices normalized to inferred price. Column 5 reports the discount that a property sells at relative to its list price. Standard errors are clustered at the MLS-level. See Section 3 for more details on the data sample and definition of experience.

between agent experience and the probability of sale in Figure B3.

We then reestimate our main specification in Appendix Table B4 and find a significant and positive effect of experience on sale probability in Column 1, with a similar magnitude to our main estimates. However, we do not find significant differences in the effect of experience across boom and bust periods. We replicate our main outcomes in the remaining tables.

Next, we address the concern that agents may differ by more than their measured experience level. We test this in two ways. We first consider the most natural approach to this in Appendix Table B5, where we rerun our preferred regression specification from Column 3 of Table 2 and include listing agent fixed effects. In Column 1, one log point increase in listing agent's experience increases the probability of sale by 0.8 pps during the boom period, and 1.1 and 1.8 pp during the medium and bust periods, respectively. These effects are smaller in magnitude than in Table 2 but the relative value of experience in the bust is much higher.

However, including agent fixed effects creates bias in our estimates of experience. Examining the effect of experience *within-agent* is complicated by the fact that agents who continue to work (and build experience) were more likely to be successful early on. Those agents who were unsuccessful in selling properties when they had low experience are less likely to continue as agents and build experience. As a result, the *within-agent* effect of experience on listing liquidity will be flattened, as those agents who continue on will have been most

Table B4: Effect of experience on outcomes in motivated seller sample

			Log Prices		
	Sale Pr.	Will Foreclose	List / Inferred	Sale/Infer	List / Sale
	(1)	(2)	(3)	(4)	(5)
Log(Exp + 1)	0.0327*	0.0000	-0.0040	-0.0012	0.0051
	(0.0163)	(0.0001)	(0.0125)	(0.0173)	(0.0062)
Bust × Log(Exp + 1)	0.0066	-0.0044**	-0.0201*	-0.0458*	0.0015
	(0.0220)	(0.0017)	(0.0114)	(0.0236)	(0.0170)
Medium × Log(Exp + 1)	0.0087	-0.0007	-0.0100	-0.0345	0.0022
	(0.0228)	(0.0027)	(0.0154)	(0.0288)	(0.0192)
Time-by-Zipcode FE	Yes	Yes	Yes	Yes	Yes
House Char.	Yes	Yes	Yes	Yes	Yes
Bust Effect	0.0393	-0.0044	-0.0241	-0.0471	0.0066
Bust p-value	0.0000	0.0134	0.0010	0.0507	0.6963
Medium Effect	0.0415	-0.0007	-0.0140	-0.0358	0.0074
Medium p-value	0.0007	0.7831	0.2123	0.3718	0.7267
Observations	12196	11957	2794	1305	1305

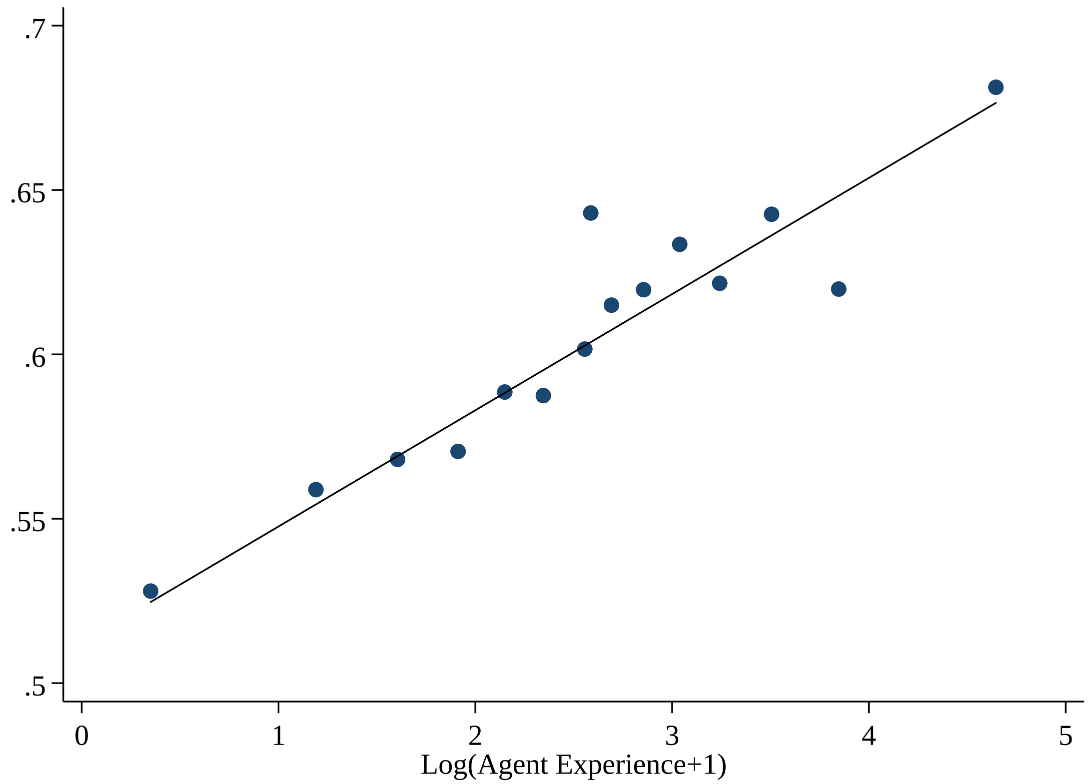
Note: This table reports estimates for our outcomes using our main specification from Equation 1, focusing on a subsample of motivated sellers who have likely inherited the property or gone through a divorce. Specifically, these listings occur within two years after a deeds record of a transaction between two people who 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. The regressions include county-by-year-month fixed effects and housing controls (the same controls as Column 3 in Table 2). In Column 1, we report the effect of experience on the probability of sale in 365 days. In Column 2, we report the effect on subsequent foreclosures. Column 3 reports the effect of agent experience on list price normalized to inferred price (measured using the previous sale price, appreciated using zipcode- and price-tier-specific Zillow house price appreciation) for all listings. Column 4 reports the effect on sale prices normalized to inferred price. Column 5 reports the discount that a property sells at relative to its list price. Standard errors are clustered at the MLS-level. See Section 3 for more details on the data sample and definition of experience.

Table B5: Effect of experience on outcomes with agent fixed effects

			Log Prices		
	Sale Pr.	Will Foreclose	List / Inferred	Sale/Infer	List / Sale
	(1)	(2)	(3)	(4)	(5)
Log(Exp + 1)	0.0082***	0.0008***	-0.0030***	-0.0020***	-0.0013***
	(0.0004)	(0.0001)	(0.0004)	(0.0005)	(0.0001)
Bust $\times$ Log(Exp + 1)	0.0100***	-0.0017***	-0.0025***	0.0034***	-0.0027***
	(0.0005)	(0.0001)	(0.0005)	(0.0007)	(0.0002)
Medium $\times$ Log(Exp + 1)	0.0032***	-0.0003***	-0.0007	0.0015**	-0.0008***
	(0.0005)	(0.0001)	(0.0005)	(0.0007)	(0.0001)
Time-by-Zipcode FE	Yes	Yes	Yes	Yes	Yes
House Char.	Yes	Yes	Yes	Yes	Yes
Agent FE	Yes	Yes	Yes	Yes	Yes
Bust Effect	0.0182	-0.0009	-0.0055	0.0013	-0.0039
Bust p-value	0.0000	0.0000	0.0000	0.0184	0.0000
Medium Effect	0.0114	0.0005	-0.0037	-0.0005	-0.0021
Medium p-value	0.0000	0.0000	0.0000	0.4102	0.0000
Observations	8399120	7955319	2146647	1263059	4942005

Note: This table reports estimates for our outcomes using our main specification from Equation 1, with the addition of listing agent fixed effects. The regressions include zipcode-by-year-month fixed effects and housing controls (the same controls as Column 3 in Table 2). In Column 1, we report the effect of experience on the probability of sale in 365 days. In Column 2, we report the effect on subsequent foreclosures. Column 3 reports the effect of agent experience on list price normalized to inferred price (measured using the previous sale price, appreciated using zipcode- and price-tier-specific Zillow house price appreciation) for all listings. Column 4 reports the effect on sale prices normalized to inferred price. Column 5 reports the discount that a property sells at relative to its list price. Standard errors are clustered at the zipcode-level. See Section 3 for more details on the data sample and definition of experience.

Figure B3: Effect of experience on outcomes in motivated seller sample

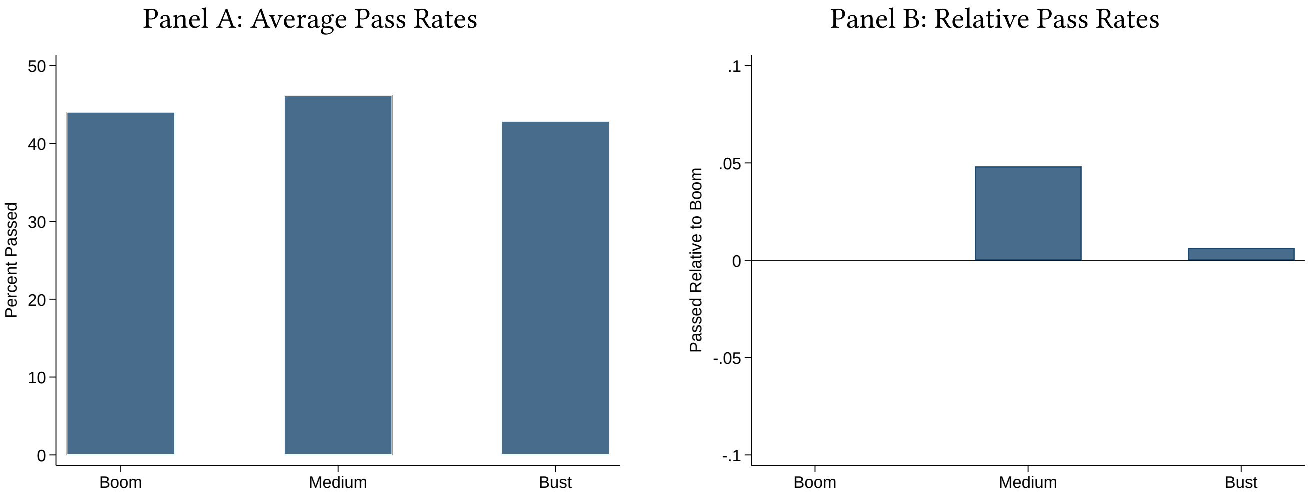


Note: This figure focuses on the subsample of listings where we can identify a recent death or divorce prior to the listing. This figure plots a binned scatterplot (with 20 bins) of the probability that a listing sells within 365 days against the listing agent's experience (using the $\log(1 + \text{agent experience})$). This plot and fitted line account for for county-by-year-month fixed effects and housing controls (the same housing controls as Column 3 in Table 2). The fitted line, average bin values, and the reported coefficient correspond to the coefficient on β of Equation 1, not allowing β_e to vary by time period. Standard errors are clustered at the MLS-level. See Section 3 for more details on the data sample and definition of experience.

successful in selling properties with less experience. Thus these estimates are best viewed as a lower bound of our effects, and reassuring evidence that even after controlling for time-invariant agent characteristics, there are strong positive effects on listing liquidity.

A related concern is that during the bust periods, the new (and inexperienced) agents that select into the real estate market are worse at selling properties in unobservable ways. If the quality of people entering the profession changes over the cycle, it would presumably be reflected in the pass rates for real estate license exams. In Appendix Figure B4, we show that exam pass rates are nearly identical across the cycle, with similar pass rates in boom and bust periods. This suggests that the incoming pool of interested agents is similar across time.

Figure B4: Exam pass rates across the cycle

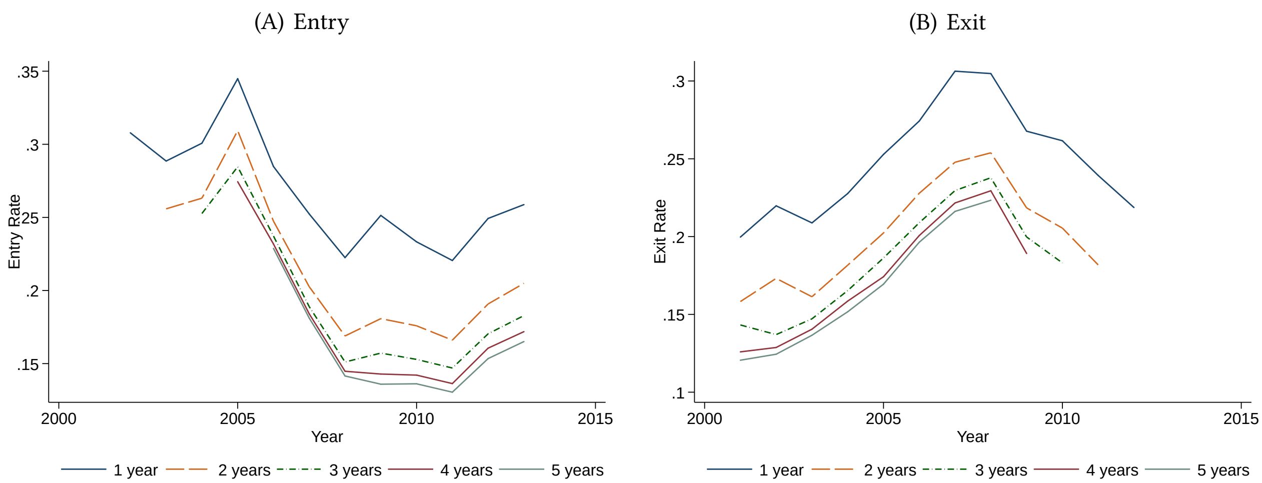


Note: This plot uses collected data on real estate salespersons exam statistics in each state from Arello, a data provider. In the first panel, we show the average of all pass rates in the United States corresponding to each period as defined in the model. The second plot we take the pass rate for each time period in each state and normalize it to the boom period in that state. We report the U.S. average of these relative pass rates weighted by the number of exams given in the corresponding state.

C 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 unable to get clients. 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 C1: Entry and exit at different horizons



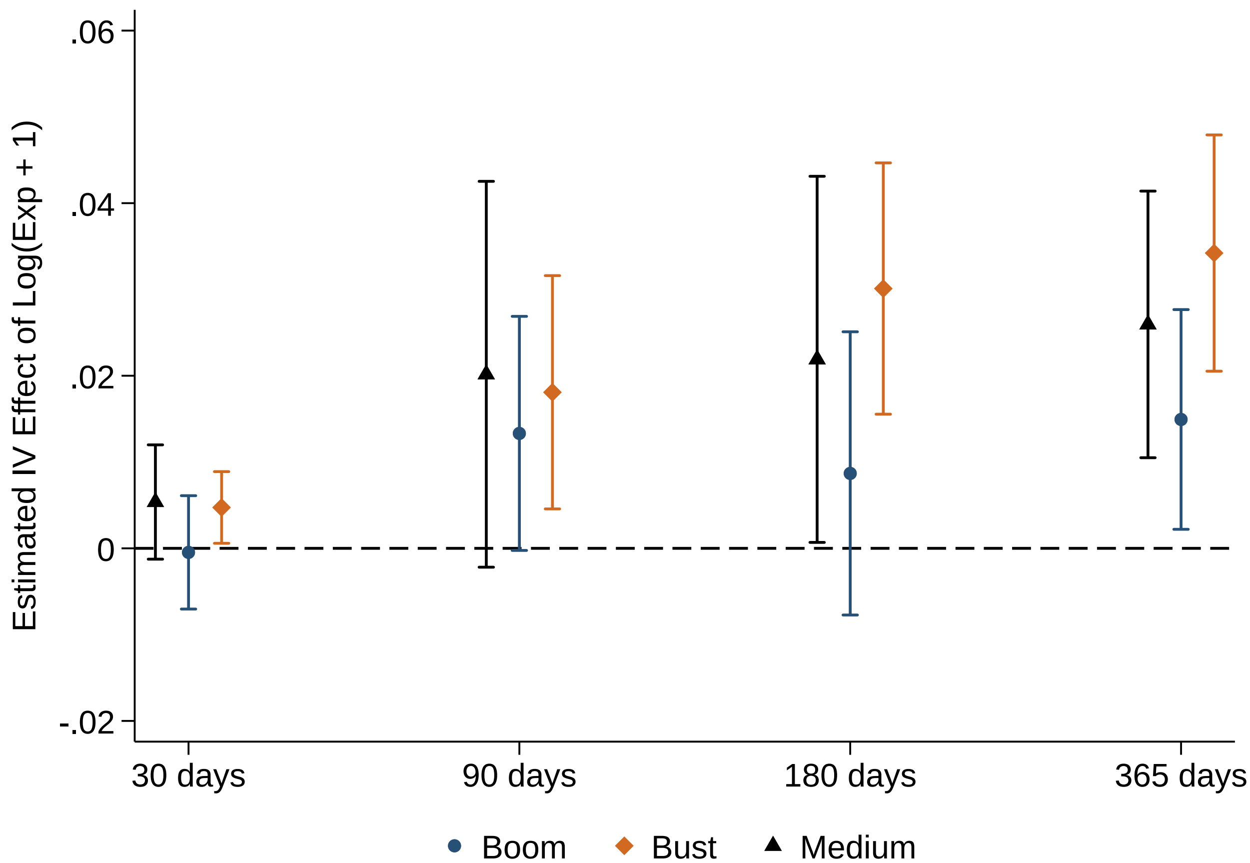
Note: Panels A and B plot entry and exit rates respectively for various definitions of thereof. For, $n \in \{1, 2, .., 5\}$ we define entry/exit rates as a fraction of currently active agents who are not active in the previous/next n years.

D Alternative timing cutoff for sale probability

D.1 Change in experience effect depending on duration cutoff

We have chosen sale within 365 days as our main definition of sale. In this section, we consider how changing this 365th day cutoff affects our results on sale probability. We re-estimate Table 2 with sale cutoffs at 30, 90 and 180 days, and report the estimates in Panels A, B and C of Appendix Table D1. In Appendix Figure D1, we plot the effect of this cutoff for the different time periods. Two interesting facts emerge. First, the impact of experience is increasing with the length of the duration measure. This suggests that the benefits of experience are largest for those properties that do not sell immediately, consistent with listing agent experience benefiting "marginal" properties that may not sell instantaneously. Second, the differential effect of agent experience in boom vs. bust periods is largest at longer horizons, consistent with experience mattering even more during bust periods.

Figure D1: Effect of experience on probability of sale at different horizons



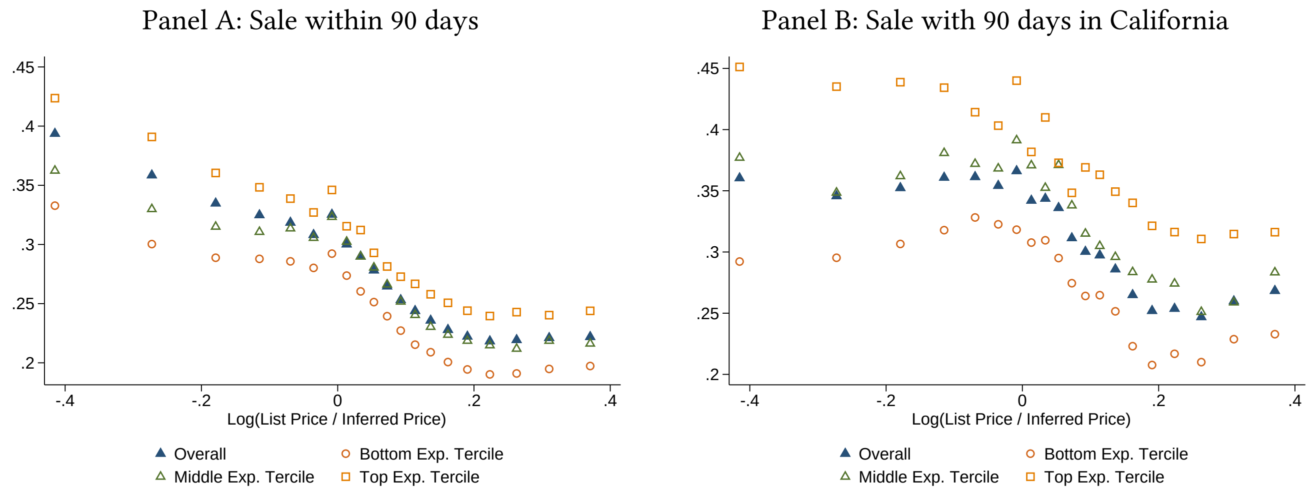
Note: Panels A and B plot entry and exit rates respectively for various definitions of thereof. For, $n \in \{1, 2, .., 5\}$ we define entry/exit rates as a fraction of currently active agents who are not active in the previous/next n years.

D.2 Reconciling cutoff vs. list price / inferred price for Guren (2018)

These results reconcile Figure 4 with Figure 1 from Guren (2018). Several things differ between our samples. First, in Guren (2018), the outcome focuses on the probability of sale in 13 weeks. In Panel A of Appendix

Figure D2, we replicate Figure 4 using probability of sale in 90 days to make it comparable. In Panel B, we additionally limit our sample to listings in the state of California, the same state that Guren (2018) focused on. In this subsample of Panel B, we see shape to the curve as in Figure 1 of Guren (2018).

Figure D2: Pricing and sale probability



Note: This graph plots a binned scatterplot (with 20 bins) of the expected sale probability against the log of normalized list price – list price scaled by our measure of inferred price. We compute the inferred price as the last historical price that the property has sold, appreciated to current list date using the Zillow zipcode and tier-level house price index. The regression controls for zipcode-by-year-month fixed effects and housing controls (the same controls as Column 3 in Table 2), and we plot this relationship split by tercile of agent experience. See Section 3 for more details on the data sample and definition of experience.

Table D1: Effect of experience on probability of sale at different horizons

	Panel A: Probability of sale in 90 days				
	Main Sample			IV Sample	
	(1)	(2)	(3)	(4)	(5)
Log(Exp + 1)	0.0027***	0.0029***	0.0030***	-0.0000	0.0023***
	(0.0002)	(0.0003)	(0.0003)	(0.0034)	(0.0003)
Bust × Log(Exp + 1)		-0.0002	-0.0003	0.0047	0.0005
		(0.0003)	(0.0003)	(0.0038)	(0.0004)
Medium × Log(Exp + 1)		-0.0002	-0.0003	0.0049	0.0004
		(0.0003)	(0.0003)	(0.0059)	(0.0004)
Time-by-Zipcode FE	Yes	Yes	Yes	Yes	Yes
House Char.	No	No	Yes	Yes	Yes
Bust Effect		0.0026	0.0027	0.0046	0.0028
Bust p-value		0.0000	0.0000	0.0342	0.0000
Medium Effect		0.0026	0.0027	0.0048	0.0026
Medium p-value		0.0000	0.0000	0.1583	0.0000
Observations	8457612	8457612	8457612	1217983	1217983
	Panel B: Probability of sale in 90 days				
	(1)	(2)	(3)	(4)	(5)
Log(Exp + 1)	0.0207***	0.0186***	0.0200***	0.0137**	0.0239***
	(0.0014)	(0.0013)	(0.0012)	(0.0066)	(0.0017)
Bust × Log(Exp + 1)		0.0049***	0.0044***	0.0047	0.0069***
		(0.0009)	(0.0009)	(0.0100)	(0.0017)
Medium × Log(Exp + 1)		0.0007	0.0003	0.0054	0.0002
		(0.0013)	(0.0013)	(0.0112)	(0.0020)
Time-by-Zipcode FE	Yes	Yes	Yes	Yes	Yes
House Char.	No	No	Yes	Yes	Yes
Bust Effect		0.0235	0.0244	0.0184	0.0308
Bust p-value		0.0000	0.0000	0.0092	0.0000
Medium Effect		0.0192	0.0204	0.0192	0.0241
Medium p-value		0.0000	0.0000	0.0821	0.0000
Observations	8457612	8457612	8457612	1217983	1217983
	Panel C: Probability of sale in 180 days				
	(1)	(2)	(3)	(4)	(5)
Log(Exp + 1)	0.0307***	0.0253***	0.0267***	0.0080	0.0302***
	(0.0019)	(0.0019)	(0.0018)	(0.0083)	(0.0022)
Bust × Log(Exp + 1)		0.0115***	0.0109***	0.0216**	0.0167***
		(0.0015)	(0.0015)	(0.0090)	(0.0023)
Medium × Log(Exp + 1)		0.0037**	0.0034**	0.0129	0.0032
		(0.0015)	(0.0015)	(0.0138)	(0.0021)
Time-by-Zipcode FE	Yes	Yes	Yes	Yes	Yes
House Char.	No	No	Yes	Yes	Yes
Bust Effect		0.0368	0.0376	0.0296	0.0469
Bust p-value		0.0000	0.0000	0.0001	0.0000
Medium Effect		0.0290	0.0301	0.0209	0.0333
Medium p-value		0.0000	0.0000	0.0476	0.0000
Observations	8457612	8457612	8457612	1217983	1217983
Estimation Method	OLS	OLS	OLS	IV	OLS

Note: This table reports estimates of the effect of listing agent's experience (using the $\log(1+\text{experience})$) on a listings' probability of sale in 30 days, 90 and 180 days. All five columns use different versions of the specifications outlined in Equation 1 and 3. All columns include zipcode-by-listing-year-month fixed effects, and Columns 3-5 add controls for house characteristics. Columns 4 and 5 include purchase-year-by-listing-year-by-zipcode fixed effects. Panel A reports results using the main sample of listings. Panel B uses the IV sample of listings, restricted to observations where we observe the initial purchase of the listing. Column 4 shows results from the IV estimation while Column 5 repeats the specification in Column 3 using the IV sample. Details of the IV estimation are discussed in Section 4. Standard errors are clustered at the MLS level. *,**, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

E Microfoundation of the theoretical matching function

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 θ'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 than 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 $min(i/l, 1)$

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

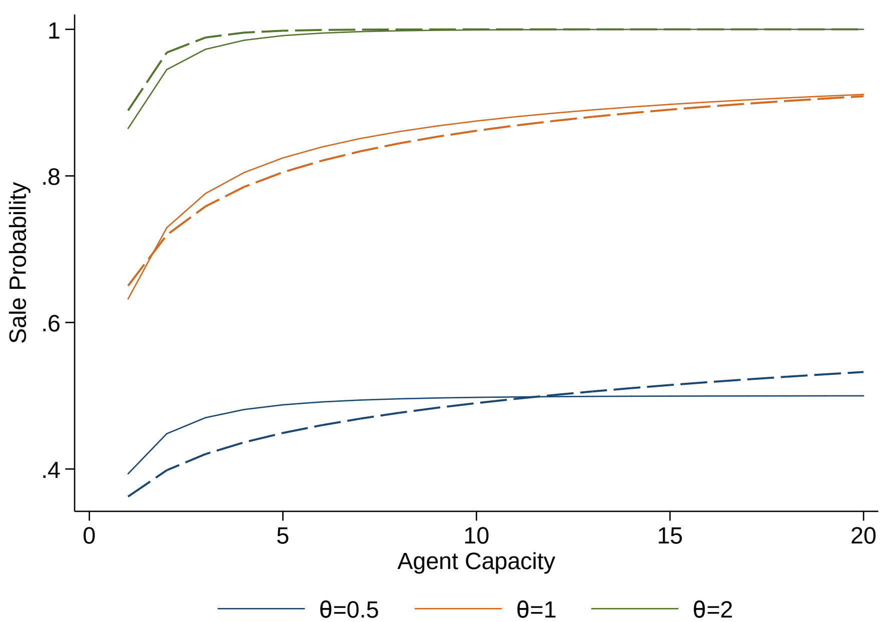
At $i = 0$, the expressions in the sum are the same and equal to 0. However as i increases, $m^l(a,b)$ increases faster than $m(la, b)$. We can see that from computing the slope of the part that differs in the too expressions with respect to i.

$$\frac{d}{di}\left(1 - \left(1 - \frac{1}{l}\right)^i\right) = -\left(1 - \frac{1}{l}\right)^i \log\left(1 - \frac{1}{l}\right) <^{41} \left(1 - \frac{1}{l}\right)^i \frac{1}{l} < \frac{d}{di}\frac{i}{l} = \frac{1}{l}$$

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 is bounded. Figure E1 plots the $\mu^l(s/l, b)/s$ for various values of $\theta = s/b$.

Figure E1: 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 θ. The dashed lines represent the matching function set up used in the model. We allow for θ to vary across l, and λ_2 vary across states.

For a fixed θ 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 λ_1 to change with the state. Here λ_2 represents the experience advantage. For the illustration above, we can calibrate $\lambda_1(z)$ and λ_2 to match the relationship that is delivered by the micro-founded model. While z represents varying θ 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 its own ratio of buyers to sellers which will be larger for more efficient agents. In the dashed lines, Figure E1 then plots the model specification where we allow for λ_1 to vary across the three levels, but within each level, θ increases with l. We can see that our model approximates well the micro founded model described above.

F Office commission splits

Real estate agents can not legally sign contracts with clients without being affiliated with a real estate broker. The agents are thus always affiliated with a real estate office (where there is a real estate broker). In return for an opportunity to work and other services, such as advertising and brand recognition, an agent typically gives an office a part of their commission. The commission split is a negotiable part of an agent-office contract and thus varies substantially. Unsurprisingly, agents who bring in more business to the office are able to negotiate a more favorable commission split, while new agents tend to give up about half of their commissions. National Association of Realtors survey conducts a study of real estate professionals (National Association of Realtors (2017a)) and documents the commission splits for each earning bin summarized in Table F1.

Figure F1: Survey evidence on commission splits

Exhibit 3-3 COMPENSATION STRUCTURES FOR REALTORS®, BY GROSS PERSONAL INCOME

(Percentage Distribution)

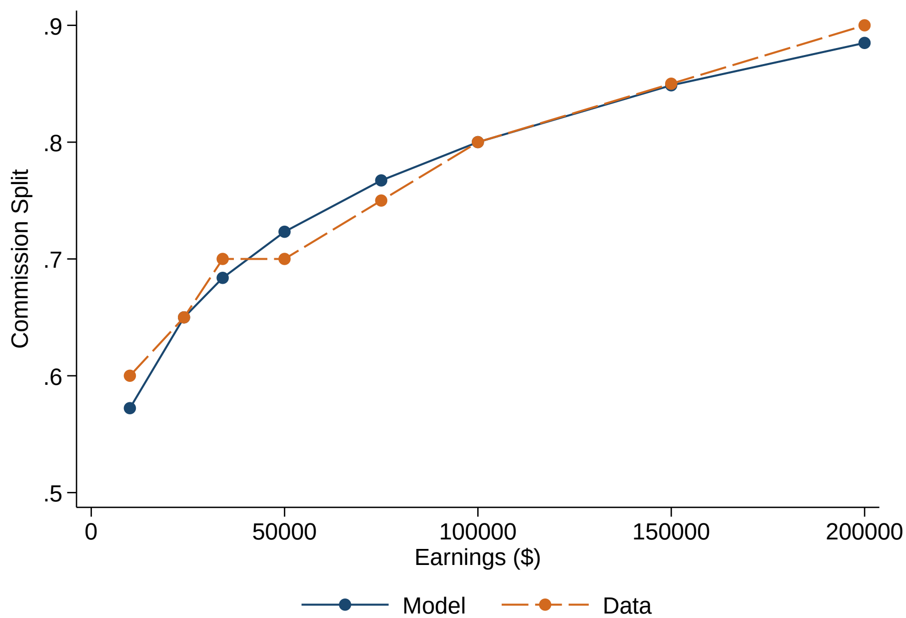
		GROSS PERSONAL INCOME							
	ALL REALTORS®	Less than $10,000	$10,000 to $24,999	$25,000 to $34,999	$35,000 to $49,999	$50,000 to $74,999	$75,000 to $99,999	$100,000 to $149,999	$150,000 or more
Percentage commission split	68%	77%	79%	73%	70%	61%	60%	51%	48%
100% Commission	18	15	13	15	15	21	21	26	29
Commission plus share of profits	3	1	2	2	3	3	2	4	6
Salary plus share of profits/ production bonus	3	*	1	2	2	4	5	9	6
Salary only	2	1	1	2	4	2	4	2	1
Share of profits only	1	1	1	*	1	2	1	2	2
Other	6	6	4	6	5	8	8	6	8
Median year-starting percentage commission split	70%	58%	60%	65%	70%	70%	70%	80%	80%
Median year-ending percentage commission split	70%	60%	65%	70%	70%	75%	80%	85%	90%

* *Less than 1 percent*

Note: This table summarizes the compensation structure of real estate professionals based on their income level. The commission split displayed in the first row is the average percent of the earned commission that an agent shares with their office.

In our model, we choose the commission split to be a function of earnings that matches this survey evidence. Using the functional form $f(x) = ax^b$, we find that $a = 0.1498$ and $b = 0.1455$ best approximates the data as shown in Figure F2.

Figure F2: Matching office commission split



Note: This figure plots the reported commission splits corresponding to different earning levels, as reported in the National Association of Realtor survey (National Association of Realtors (2017a)). On top of these survey values, we fit the best approximation of the function $f(x) = ax^b$.

G 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 ρ

$n = 0$

repeat

repeat

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

Given s, b, ρ, 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*

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

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

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

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

until $\Delta_2 \leqslant \epsilon_2$ and $\Delta_3 \leqslant \epsilon_3$

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*.

H Model Counterfactuals with Fixed Number of Buyers

The baseline version of the model has a free entry condition for buyers. As a result, any improvements in market technology due to agent experience only increase sell probability and do not affect buy probability. This is because new buyers crowd out the benefits for existing buyers. In this section we consider how our model results would change without the buyer free entry assumption. That is, we consider the counterfactual results of the efficiency benchmark equilibrium (discussed in section 6.1) and the increased entry cost policy equilibrium (discussed in section 6.2) if we keep the number of buyers fixed at the baseline level.

The solution algorithm to finding both counterfactual equilibria differs from the baseline equilibrium solution in a non-trivial way. The challenge is that for the counterfactual equilibrium analysis the $n^b(z_t)$ is exogenously given,[42] so changes in overall agent experience distribution triggers a re-allocation of buyers across experience levels in a way that affects probability of sale and purchase for all clients. We can no longer solve for the sale and purchase probabilities independently from the agent distribution. Recall that at every iteration of the convergence algorithm we guess entry and exit policies of real estate agents and compute the implied distribution of agents across experience levels. We then check that the implied value functions are consistent with the entry and exit decisions. In solving for a counterfactual equilibrium with fixed buyers we add two additional steps. First, we solve for changes in buyer valuation at every aggregate state so that the corresponding market tightness implies the correct number of buyers in each state, $n^b(z_t)$.[43] This has an effect on overall buyers' purchase as well as the sellers' sale probabilities. Second, using the updated sale and purchase probabilities, we solve for optimal commissions of listing agents using equations 23 and 22. Finally, we incorporate the changes into real estate agent's value functions through the profit equation 24.

This model specification allows existing buyers to benefit from increased market efficiency due to agent experience. We construct a measure of welfare for buyers similarly to that of sellers in Equation 25:

$$\tilde{V}^b(w) = \underbrace{\eta(v(w))(v(w) - p(v(w))}_{\text{Buy this period}} + \underbrace{(1 - \eta(v(w))\beta E[\tilde{V}^b(w')|w]}_{\text{Try to buy next period}}\Big).$$

Buyers enjoy value $v(w)$ of the house if they successfully purchase a house, which happens with probability $\eta(v(w))$. With complimentary probability they come back to the search market in the following period with uncertainty about which state w they might find themselves in.

We report results from the two counterfactuals in table H1. In the flexible commission benchmark equilibrium (Panel B of table H1) the improvement in probability of sale for the sellers is significantly smaller. Instead of a 3.7% improvement in probability of sale (table 7), the fixed buyer specification of the efficiency benchmark delivers only about 1.1% increase in the sale probability. However in that specification buyers also benefit with about 1.1% increase in probability of purchase. Overall this leads to an improvement in seller welfare of $3,084 which comes from improvement in sale probability and reduced commissions due to the competitive market. Buyers benefit more modestly by only about $166 dollars.

Within the status quo of fixed commissions, the policy markers would have to increase agent entry costs to $108,000 in order to achieve even this modest improvement in sale probability. Panel C of table H1 reports the

[42] Whereas before we solved for it ex-post to match market tightness and the distribution of sellers across markets, see Equation 11.

[43] This is computationally equivalent to raising the cost of entry for buyers in each state z_t to ensure that the same number of buyers enter under "free entry" in each sate as in the baseline equilibrium. The difference in the "entry costs" can then be interpreted as increased buyer valuation for existing buyers.

results for this counterfactual. The increase in sale probability is matched to that of the efficiency benchmark by construction. Because the number of buyers is fixed, the change in buyer purchase probability (and consequently buyer welfare) also match the flexible commission equilibrium. However in this equilibrium sellers do not benefit from competitive commissions, so the seller welfare only increases by $216.

Table H1: Counterfactual Results with Fixed Number of Buyers

Panel A: Baseline				
	Sale Prob	Buy Prob	Seller Value	Buyer Value
Bust	0.441	0.802	185,360	201,790
Medium	0.502	0.729	198,170	218,000
Boom	0.622	0.724	203,160	219,660
Total	0.524	0.752	195,638	213,166
Panel B: Flexible Commission Benchmark				
	Sale Prob	Buy Prob	Seller Value	Buyer Value
Bust	0.446	0.812	187,973	201,867
Medium	0.508	0.737	200,773	218,213
Boom	0.628	0.731	207,287	219,857
Total	0.530	0.761	198,722	213,332
Panel C: Increased Entry Cost				
	Sale Prob	Buy Prob	Seller Value	Buyer Value
Bust	0.447	0.812	185,510	201,873
Medium	0.508	0.738	198,423	218,227
Boom	0.629	0.732	203,390	219,873
Total	0.530	0.761	195,844	213,334

Note: This table reports sale and buy probabilities as well as the seller and buyer value, in each of the three periods and overall (weighted by each state's ergodic probability). The values are reported for three models. Panel A reports results from the baseline specification. Panel B reports results from the flexible commission counterfactual where we fix the number of buyers to be equal to the baseline equilibrium number. Finally, Panel C reports results from the counterfactual policy of increased entry costs that targets sale probability improvements delivered by the flexible commission efficiency benchmark, also fixing the number of buyers to that of the baseline equilibrium.

I Model motivating instrumental variables approach

In this section, we parameterize client sorting into agents based on experience that can confound OLS estimates of the effect of experience on house sale. Using a simplified set-up we then illustrate how to use the identification strategy from the main text to correctly estimate the effect.

Recall that the estimation uses the sample of individuals who purchase a property with a buyer agent and subsequently choose to sell it. For our identification strategy we rely on the fact that sellers are highly likely to sell with the same agent they bought with (if that agent is still active). If that agent is no longer in the market, they will draw a new agent from the pool of existing agents, causing mean reversion in the level of experience. This variation is enough to identify the effect of experience on sale outcomes.

I.1 Setup

Let the outcome of interest, sale, for client i working with agent j be $Y_{ij} = f(e_j, \alpha_i)$, a function of the agent's observable experience e_j and client's unobservable characteristic α_i. This α_i may denote a client's impatience or cost of time, or reflect unobservable qualities of the home.

A client's choice of an agent is parameterized in the following way:

$$U_{ij} = \underbrace{\beta_i}_{(\gamma_0+\gamma_1\alpha_i)} e_j + u_{ij} \tag{26}$$

where u_{ij} is an idiosyncratic taste shock for agent j by person i, and β_i is the individual's preference for experience. We parameterize $\beta_i = \gamma_0 + \gamma_1\alpha_i$, so that preference for experience is a function of the unobservable client characteristic α_i. For simplicity, assume that u_{ij} is logit and i.i.d. such that

$$\Pr(U_{ij} > U_{ij'}, \,\forall\, j' \neq j|e) = \frac{e^{\alpha_i e_j}}{\sum_{j'} e^{\alpha_i e_{j'}}}. \tag{27}$$

and hence the relative probability of client i choosing an agent j over an agent k is

$$\log(\Pr(U_{ij} > U_{ij'}, \,\forall\, j' \neq j|e) - \log(\Pr(U_{ik} > U_{ik'}, \,\forall\, k' \neq k|e) = \beta_i(e_j - e_k). \tag{28}$$

Note that if $\gamma_1 = 0$, then all clients value experience equally; if $\gamma_1 \neq 0$, then the preference for experience is correlated with the unobservable client characteristic. Finally, if $\gamma_0 = \gamma_1 = 0$, then clients exhibit no preference for experience.

Let $e^*(\alpha_i)$ be the experience of the agent chosen by client i. For each level of α, we can then define the frequency that an agent with experience e is used as the sum of the probabilities of choosing an agent with that experience level:

$$\Pr(e^*(\alpha_i) = e|\alpha_i) = \sum_{j, e_j = e} \frac{e^{\beta_i e_j}}{\sum_{j'} e^{\beta_i e_{j'}}}. \tag{29}$$

To illustrate how our IV strategy works, we use a simplified set-up where:

1. There are two levels of α_i, high and low, with equal probability
2. There are two levels of experience, e_k and $e_{k'}$, with an equal number of agents at each level.

It is straightforward to expand the proof to allow for more levels of α_i and e_j in a more general case.

Now, using Bayes' rule, it is straightforward to show that:

$$\pi(\alpha_i|e_k) \equiv \Pr(\alpha_i|e^*(\alpha_i) = e_k) = \frac{\left(\frac{e^{\beta_i e_k}}{e^{\beta_i e_k}+e^{\beta_i e_{k'}}}\right)}{\left(\frac{e^{\beta_L e_k}}{e^{\beta_L e_k}+e^{\beta_L e_{k'}}}\right)+\left(\frac{e^{\beta_H e_k}}{e^{\beta_H e_k}+e^{\beta_H e_{k'}}}\right)} \tag{30}$$

where $\beta_H = \gamma_0 + \gamma_1\alpha_H$ and $\beta_L = \gamma_0 + \gamma_1\alpha_L$.

Consider our estimand of interest for a given α_i: $\tau_{k,k'}(\alpha_i) = E(f(e_k, \alpha_i) - f(e_{k'}, \alpha_i))$. This is the effect on the sale outcome of shifting experience for a given α_i. We want to estimate this effect, weighted across α_i. We will first show that a standard OLS regression will fail to estimate a convex-weighted combination of these $\tau_{k,k'}$. Then, we will show that our indentification strategy does estimate a convex-weighted combination of these $\tau_{k,k'}$, under a set of assumptions.

I.2 OLS Estimation

Consider the OLS estimator of the effect of experience on the probability of sale (this would arise from taking the regression of Y_{ij} on a dummy for e_j and a constant):

$$\begin{aligned}
\tau_{OLS} &\equiv E(Y_i|e^*(\alpha_i) = e_k) - E(Y_i|e^*(\alpha_i) = e_{k'}) \\
&= E(f(e_k, \alpha_i)|e^*(\alpha_i) = e_k) - E(f(e_{k'}, \alpha_i)|e^*(\alpha_i) = e_{k'}) \\
&= \pi(\alpha_H|e_k)E(f(e_k, \alpha_H)) - \pi(\alpha_H|e_{k'})E(f(e_{k'}, \alpha_H)) \\
&+ \pi(\alpha_L|e_k)E(f(e_k, \alpha_L)) - \pi(\alpha_L|e_{k'})E(f(e_{k'}, \alpha_L)) \\
&= \underbrace{\sum_\alpha \pi(\alpha|e_k)\tau_{k,k'}(\alpha)}_{\text{weighted effect of experience}} \\
&+ \underbrace{\sum_\alpha (\pi(\alpha|e_k) - \pi(\alpha|e_{k'}))E(f(e_{k'}, \alpha))}_{\text{selection effect into experience}}.
\end{aligned}$$

The OLS estimator does not simply reflect the average effect on the outcome of changing e. This would only be the case in a special case where $f(e_k, \alpha_i) = f(e_k)$, that is if sale outcome is independent of the client or property characteristic. If we relax this assumption, but do not allow selection of clients into agent experience level, i.e. if $\pi(\alpha|e_k) = \pi(\alpha)$, then the OLS estimator simplifies to a weighted combination of $\tau_{k,k'}$. In a general case, however, if α_i affects the sale probability through f, and there is selection of clients into experience levels of the agents, the OLS estimator τ_{OLS} is biased.

I.3 Identification Strategy

Now consider the IV estimator that exploits the buyer agent choice. Let $h^*(\alpha_i)$ denote the initial experience level of the buyer's agent chosen by client i. We assume that clients choose the initial buyer agent with the same utility maximization process as they use to choose the seller's agent. With some probability p that buyer agent will exit prior to the subsequent sale, requiring the client to pick a new agent. Call the binary event that client i's agent exits N_i. For the sake of simplicity, we will assume that if the agent does not exit, the client always stays with the same agent.[44] The key additional assumption is that N_i is independent of α_i, conditional on $h^*(\alpha_i)$.[45]

We consider the following estimator for the effect of selling agent experience on the sale probability (effectively it is the IV estimator using the initial agent experience level interacted with the exit indicator as an instrument for experience, controlling for the initial experience level and the exit indicator):

$$\tau_{IV} \equiv = \frac{\Delta_{N=1} - \Delta_{N=0}}{\Delta^e_{N=1} - \Delta^e_{N=0}}, \tag{31}$$

where

$$\Delta_N = E(Y_i|h^*(\alpha_i) = e_k, N_i) - E(Y_i|h^*(\alpha_i) = e_{k'}, N_i) \tag{32}$$

$$\Delta^e_N = Pr(e^*(\alpha_i) = e_k|h^*(\alpha_i) = e_k, N_i) - Pr(e^*(\alpha_i) = e_k|h^*(\alpha_i) = e_{k'}, N_i). \tag{33}$$

Note that the numerator is the difference in the effect of initial experience on the sale probability for those whose agents exit and those who do not (e.g. the reduced form effect of the interaction) and the denominator is the difference in the effect of initial experience on eventual seller agent experience for those whose agents exit and those who do not. We will now show that this estimator gives a weighted combination of effects of experience across α.

We show below that

$$\tau_{IV} = \frac{\sum_\alpha \omega(\alpha)\tau_{k,k'}(\alpha)}{\sum_\alpha \omega(\alpha)}, \tag{34}$$

where

$$\omega(\alpha) = Pr(e_{k'}|\alpha, N=1)Pr(\alpha|h^*(\alpha_i) = e_k) + Pr(e_k|\alpha, N=1)Pr(\alpha|h^*(\alpha_i) = e_{k'}), \tag{35}$$

which is the share of clients whose experience will change from one value to the other when their buyer agent exits, across different α. This will capture mean reversion across unobserved types, and is guaranteed to be positive.

[44] What we need is a strong degree of stickiness, but not necessarily 100 percent. Empirically, if the agent is still in the market, 33% of clients list with their buyer's agent. For the purposes of this note, we assume 100 percent for tractibility in this example.

[45] More generally, it is important that α_i is not time-varying in a way that is correlated with the outcomes and h^*.

I.4 Proof of Identification Strategy

First, note the following updated joint probability distributions given our assumptions about agent exit and the independence with α_i:

$$\Pr(\alpha_i, e^*(\alpha_i)|h^*(\alpha_i), N=1) = \Pr(e^*(\alpha_i)|\alpha_i, h^*(\alpha_i), N=1)\Pr(\alpha_i|h^*(\alpha_i), N=1) \tag{36}$$

$$= \Pr(e^*(\alpha_i)|\alpha_i, N=1)\Pr(\alpha_i|h^*(\alpha_i)) \tag{37}$$

$$\Pr(\alpha_i, e^*(\alpha_i)|h^*(\alpha_i), N=0) = \Pr(e^*(\alpha_i)|\alpha_i, h^*(\alpha_i), N=0)\Pr(\alpha_i|h^*(\alpha_i), N=0) \tag{38}$$

$$= \begin{cases} \Pr(\alpha_i|h^*(\alpha_i)) & e^*(\alpha_i) = h^*(\alpha_i) \\ 0 & e^*(\alpha_i) \neq h^*(\alpha_i) \end{cases} \tag{39}$$

Now consider the following restatement of the conditional expectations using these updates:

$$E(Y_i|h^*(\alpha_i), N=1) = \sum_{\alpha_i \in \{\alpha_L, \alpha_H\}} \Pr(\alpha_i|h^*(\alpha_i)) \sum_{e^*(\alpha_i) \in \{e_k, e_{k'}\}} E\left(f(e^*(\alpha_i), \alpha_i)\right) \Pr(e^*(\alpha_i)|\alpha_i, N=1) \tag{40}$$

$$E(Y_i|h^*(\alpha_i), N=0) = \sum_{\alpha_i \in \{\alpha_L, \alpha_H\}} \Pr(\alpha_i|h^*(\alpha_i)) \sum_{e^*(\alpha_i) \in \{e_k, e_{k'}\}} E\left(f(e^*(\alpha_i), \alpha_i)\right) 1(e^*(\alpha_i) = h^*(\alpha_i)) \tag{41}$$

$$\Pr(e^*(\alpha_i)|h^*(\alpha_i), N=1) = \sum_{\alpha_i} \Pr(e^*(\alpha_i)|\alpha_i, h^*(\alpha_i), N=1)\Pr(\alpha_i|h^*(\alpha_i), N=1) \tag{42}$$

$$= \sum_{\alpha_i} \Pr(e^*(\alpha_i)|\alpha_i, N=1)\Pr(\alpha_i|h^*(\alpha_i)) \tag{43}$$

$$\Pr(e^*(\alpha_i)|h^*(\alpha_i), N=0) = \sum_{\alpha_i} \Pr(e^*(\alpha_i)|\alpha_i, h^*(\alpha_i), N=0)\Pr(\alpha_i|h^*(\alpha_i), N=0) \tag{44}$$

$$= \sum_{\alpha_i} 1(e^*(\alpha_i) = h^*(\alpha_i))\Pr(\alpha_i|h^*(\alpha_i)) \tag{45}$$

We first consider the numerator of the IV estimator:

$$\Delta_{N=1} = \sum_{\alpha_i \in \{\alpha_L, \alpha_H\}} \left(\Pr(\alpha_i | h^*(\alpha_i) = e_k) - \Pr(\alpha_i | h^*(\alpha_i) = e_{k'})\right) \times \tag{46}$$

$$\sum_{e^*(\alpha_i) \in \{e_k, e_{k'}\}} E\left(f(e^*(\alpha_i), \alpha_i)\right) \Pr(e^*(\alpha_i) | \alpha_i, N = 1) \tag{47}$$

$$\Delta_{N=0} = \sum_{\alpha_i \in \{\alpha_L, \alpha_H\}} \Pr(\alpha_i | h^*(\alpha_i) = e_k) E\left(f(e_k, \alpha_i)\right) - \Pr(\alpha_i | h^*(\alpha_i) = e_{k'}) E\left(f(e_{k'}, \alpha_i)\right) \tag{48}$$

$$\Delta_{N=1} - \Delta_{N=0} = \sum_{\alpha} - \Pr(e_{k'} | \alpha, N = 1) \Pr(\alpha | h^*(\alpha_i) = e_k) E(f(e_k, \alpha)) \tag{49}$$

$$+ \Pr(e_k | \alpha, N = 1) \Pr(\alpha | h^*(\alpha_i) = e_{k'}) E(f(e_{k'}, \alpha)) \tag{50}$$

$$+ \Pr(e_{k'} | \alpha, N = 1) \Pr(\alpha | h^*(\alpha_i) = e_k) E(f(e_{k'}, \alpha)) \tag{51}$$

$$- \Pr(e_k | \alpha, N = 1) \Pr(\alpha | h^*(\alpha_i) = e_{k'}) E(f(e_k, \alpha)) \tag{52}$$

$$= \sum_{\alpha} - \left(\underbrace{\Pr(e_{k'} | \alpha, N = 1) \Pr(\alpha | h^*(\alpha_i) = e_k) + \Pr(e_k | \alpha, N = 1) \Pr(\alpha | h^*(\alpha_i) = e_{k'})}_{\omega(\alpha)} \right) \tau_{k,k'}(\alpha). \tag{53}$$

Now a similar exercise for the denominator:

$$\Delta^e_{N=1} = \sum_{\alpha_i} \Pr(e^*(\alpha_i) = e_k | \alpha_i, N = 1) (\Pr(\alpha_i | h^*(\alpha_i) = e_k) - \Pr(\alpha_i | h^*(\alpha_i) = e_{k'})) \tag{54}$$

$$\Delta^e_{N=0} = \sum_{\alpha_i} \Pr(\alpha_i | h^*(\alpha_i) = e_k) \tag{55}$$

$$\Delta^e_{N=1} - \Delta^e_{N=0} = \sum_{\alpha} - \Big(\Pr(e^*(\alpha_i) = e_{k'} | \alpha_i, N = 1)) \Pr(\alpha_i | h^*(\alpha_i) = e_k) \tag{56}$$

$$+ \Pr(e^*(\alpha_i) = e_k | \alpha_i, N = 1) \Pr(\alpha_i | h^*(\alpha_i) = e_{k'}) \Big) \tag{57}$$

$$= \sum_{\alpha} -\omega(\alpha). \tag{58}$$

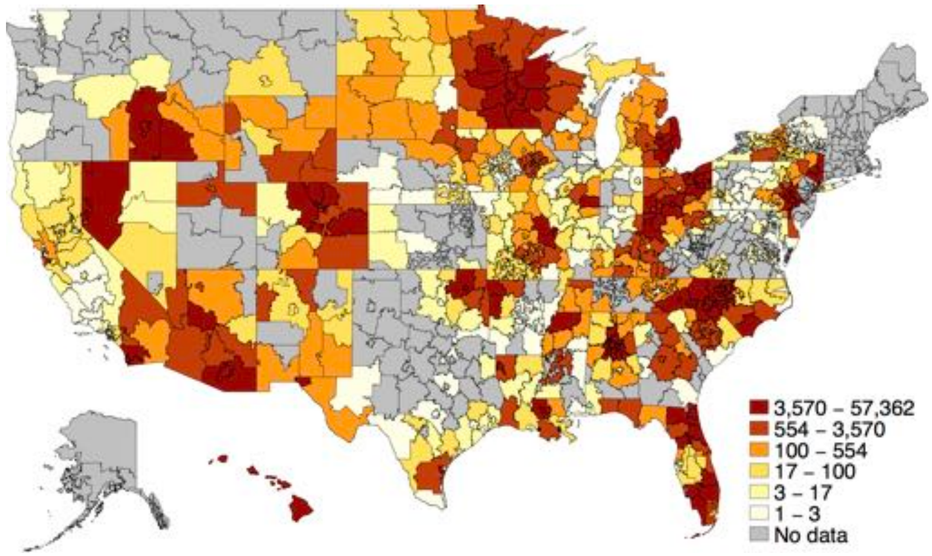
Hence, we have that:

$$\tau_{IV} = \frac{\sum_{\alpha} \omega(\alpha) \tau_{k,k'}(\alpha)}{\sum_{\alpha} \omega(\alpha)}, \tag{59}$$

where $\omega(\alpha)$ must be positive by the definition of probabilities. Note that the negative signs on the numerator and denominator will cancel – this is a feature of the fact that we are exploiting *mean reversion* – a high initial experience will induce lower subsequent experience, and vice versa. The first stage is actual a *negative* correlation between the instrument and the endogenous variable.

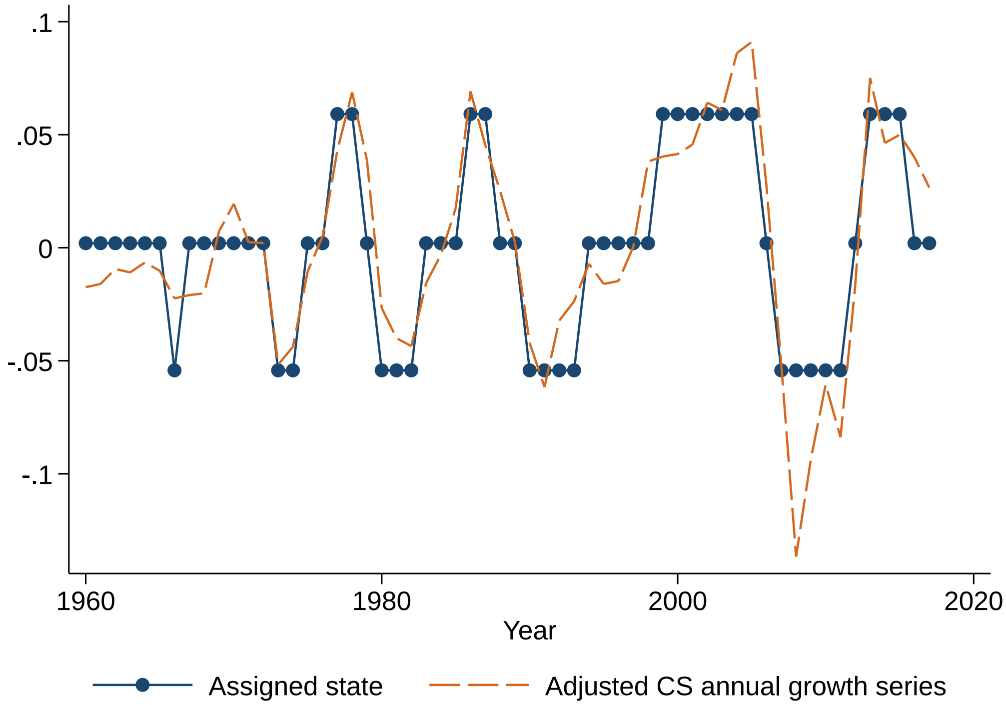
J Additional results

Figure J1: Coverage



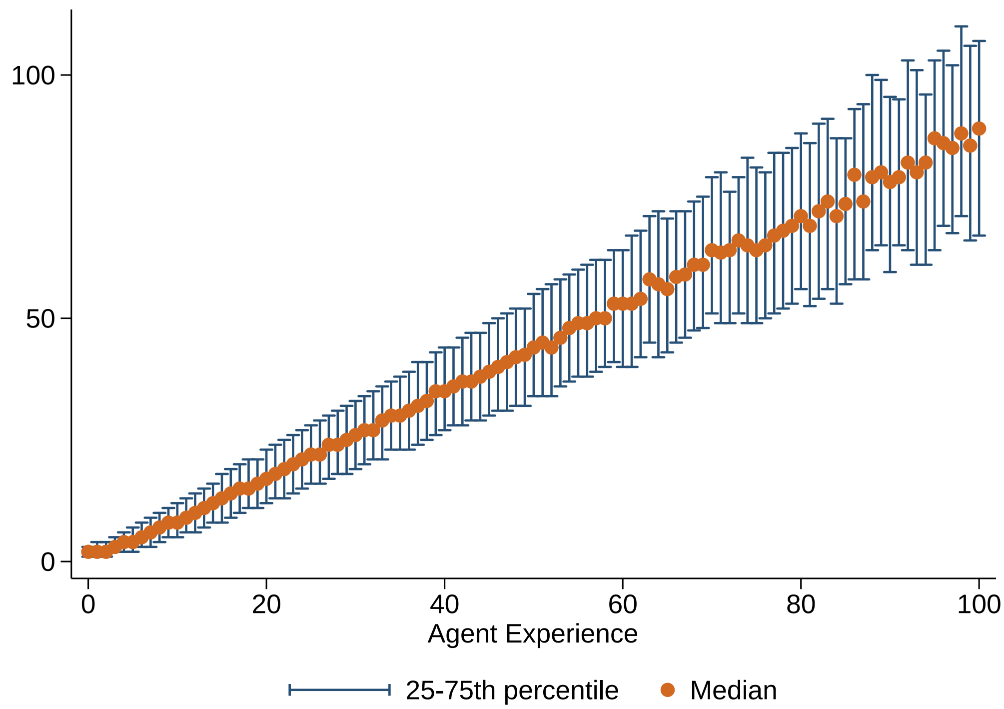
Note: This figure plots a choropleth map of the number of listings per three-digit zip in the main sample.

Figure J2: Construction of aggregate state variables using Case Shiller adjusted series



Note: This figure shows the construction of our three aggregate state variables. The dashed line plots the average annual 12-month growth rates of the Case-Shiller house price index deflated by the overall Consumer Price Index less shelter. The dots represent one of the three states assigned to each year.

Figure J3: Clients and experience



Note: This figure plots the number of clients (all listings and successful buyers) that an agent is observed working with in a given year, based 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 year. See Section 3 for more details on the data sample and definition of experience.

Table J1: Sampled MLS Comparison to Excluded MLS over 2009-2014

Panel A: Listing Characteristics

	Sampled MLS		Non-Sampled MLS	
	Mean	SD	Mean	SD
Listing sold w/in 30 days	0.03	0.16	0.03	0.17
Listing sold w/in 90 days	0.26	0.43	0.30	0.44
Listing sold w/in 180 days	0.44	0.48	0.49	0.49
Listing sold w/in 365 days	0.53	0.49	0.58	0.48
Listing sold w/in 365 days (2009)	0.47	0.48	0.54	0.44
Listing sold w/in 365 days (2014)	0.59	0.48	0.64	0.47
Days on Market	165.36	152.53	150.14	138.16
Days to Sale	135.93	122.02	129.27	116.20
Sale Price	212,601.95	196,691.58	273,549.50	268,477.88
List Price	258,015.59	320,621.31	322,658.44	429,557.44
Sale Price (2009)	198,851.22	176,628.39	252,558.97	238,266.91
List Price (2009)	261,325.66	315,319.06	317,254.94	410,854.12
Number of Listings (2009-2014)	5,952,304		17,206,380	

Panel B: County Characteristics

	Sampled MLS Counties			Rest of US		
	Mean	SD	N	Mean	SD	N
Change in housing net worth (county level), 2006-09	-0.07	0.07	175	-0.06	0.09	769
Saiz (2010) Housing supply elasticity	2.13	0.79	157	2.59	1.43	711
Total number of households in county (2000)	58,951.90	113,815.02	359	34,092.04	110,166.49	2,776
Median Household Income (2000)	40,238.70	8,575.77	359	34,997.21	9,047.59	2,776
Median Home Value (2000)	101,514.29	42,376.02	359	80,450.02	45,484.81	2,776
Owner-Occupied Share (2000)	0.74	0.08	359	0.74	0.07	2,776
College Educated Share (2000)	0.13	0.05	359	0.11	0.05	2,776
Urban Share (2000)	0.49	0.29	359	0.38	0.31	2,776

Note: This table reports characteristics of our sample compared to the non-sampled data. In Panel A, we compare listing characteristics for MLS in our sample vs. the non-sampled MLS for various basic listing characteristics. This analysis focuses on 2009-2014. In Panel B, we compare the counties covered in our data to those in the rest of the United States. We consider a county to be in out sample if our data have at least 1000 total listings in that country in 2009-2014. The demographic statistics come from the decennial Census, and the housing net worth and housing supply elasticity are calculated at the county level by Mian and Sufi (2014), with the original housing supply elasticity from Saiz (2010).

Table J2: First stage of instrumental variables

	Experience	Experience × Bust	Experience × Medium
	(1)	(2)	(3)
Log(Buyer Agent Exp + 1) × Inactive	-0.2658***	0.0002	0.0001
	(0.0165)	(0.0002)	(0.0001)
Bust × Log(Buyer Agent Exp + 1) × Inactive	0.0329***	-0.2335***	-0.0000
	(0.0094)	(0.0162)	(0.0001)
Medium × Log(Buyer Agent Exp + 1) × Inactive	0.0065	-0.0000	-0.2599***
	(0.0128)	(0.0003)	(0.0137)
First-stage F-statistic	130.5841	112.8275	131.9488
Observations	1217738	1217738	1217738

Note: This table reports estimates for our first stage regression from Equation 2. The regressions include zipcode-by-purchase-year-month-by-listing-year-month fixed effects and housing controls (the same controls as Column 3 in Table 2). Standard errors are clustered at the MLS-level. See Section 3 for more details on the data sample and definition of experience.

Table J3: Complier analysis

			Log Prices		
	Sale Pr.	Will Foreclose	List / Inferred	List / Sale	Sale/Infer
	(1)	(2)	(3)	(4)	(5)
Log(Exp + 1)	0.0298***	-0.0000	-0.0077***	-0.0050***	-0.0013***
	(0.0021)	(0.0001)	(0.0009)	(0.0014)	(0.0004)
Bust × Log(Exp + 1)	0.0206***	-0.0035***	-0.0021	0.0016	-0.0012
	(0.0022)	(0.0009)	(0.0019)	(0.0029)	(0.0009)
Medium × Log(Exp + 1)	0.0055***	-0.0011**	0.0004	-0.0002	0.0005
	(0.0021)	(0.0005)	(0.0009)	(0.0011)	(0.0005)
Time-by-Zipcode FE	Yes	Yes	Yes	Yes	Yes
House Char.	Yes	Yes	Yes	Yes	Yes
Inferred House Price	Yes	Yes	Yes	Yes	Yes
Bust Effect	0.0503	-0.0035	-0.0098	-0.0033	-0.0025
Bust p-value	0.0000	0.0004	0.0000	0.3178	0.0304
Medium Effect	0.0353	-0.0012	-0.0073	-0.0052	-0.0009
Medium p-value	0.0000	0.0396	0.0000	0.0113	0.2260
Observations	1217983	1140519	740460	454767	447604

Note: This table reports estimates for our outcomes using our main specification from Equation 1, reweighting 12 mutually exclusive subgroups so that the proportion of compliers from our IV analysis in a given subgroup matches the share of the estimation sample. See the text for details on the complier reweighting. The regressions include zipcode-by-year-month fixed effects and housing controls (the same controls as Column 5 in Table 2). In Column 1, we report the effect of experience on the probability of sale in 365 days. In Column 2, we report the effect on subsequent foreclosures. Column 3 reports the effect of agent experience on list price normalized to inferred price (measured using the previous sale price, appreciated using zipcode- and price-tier-specific Zillow house price appreciation) for all listings. Column 4 reports the effect on sale prices normalized to inferred price. Column 5 reports the discount that a property sells at relative to its list price. Standard errors are clustered at the MLS-level. See Section 3 for more details on the data sample and definition of experience.

Table J4: Effect of experience on probability of sale in 365 days, including experience > 200

	Panel A: Main Sample			Panel B: IV Sample	
	(1)	(2)	(3)	(4)	(5)
Log(Exp + 1)	0.0368***	0.0280***	0.0284***	0.0171**	0.0320***
	(0.0031)	(0.0026)	(0.0026)	(0.0067)	(0.0030)
Bust × Log(Exp + 1)		0.0178***	0.0172***	0.0183**	0.0274***
		(0.0019)	(0.0020)	(0.0090)	(0.0024)
Medium × Log(Exp + 1)		0.0064***	0.0060***	0.0102	0.0082***
		(0.0012)	(0.0012)	(0.0092)	(0.0019)
Time-by-Zipcode FE	Yes	Yes	Yes	Yes	Yes
House Char.	No	No	Yes	Yes	Yes
Bust Effect		0.0459	0.0456	0.0354	0.0594
Bust p-value		0.0000	0.0000	0.0000	0.0000
Medium Effect		0.0344	0.0344	0.0273	0.0402
Medium p-value		0.0000	0.0000	0.0002	0.0000
Observations	9003763	9003763	9003763	1279131	1279131
Estimation Method	OLS	OLS	OLS	IV	OLS

Note: This table reports estimates of the effect of listing agent's experience (using the $\log(1 + \text{experience})$) on a listings' probability of sale in 365 days, including agents whose experience is greater than 200. All five columns use different versions of the specifications outlined in Equation 1 and 3. All columns include zipcode-by-listing-year-month fixed effects, and Columns 3-5 add controls for house characteristics. Columns 4 and 5 include purchase-year-by-listing-year-by-zipcode fixed effects. Panel A reports results using the main sample of listings. Panel B uses the IV sample of listings, restricted to observations where we observe the initial purchase of the listing. Column 4 shows results from the IV estimation while Column 5 repeats the specification in Column 3 using the IV sample. Details of the IV estimation are discussed in Section 4. Standard errors are clustered at the MLS level. *,**, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Table J5: Experience and prices, including experience > 200

	Panel A: OLS Approach			Panel B: IV Approach		
	List / Infer.	Sale / Infer	List / Sale	List / Infer.	Sale / Infer	List / Sale
	(1)	(2)	(3)	(4)	(5)	(6)
Log(Exp + 1)	-0.0139***	-0.0141***	0.0014	-0.0056	-0.0051	-0.0013
	(0.0037)	(0.0049)	(0.0011)	(0.0064)	(0.0063)	(0.0029)
Bust × Log(Exp + 1)	-0.0038	-0.0015	-0.0007	0.0080	0.0104	-0.0006
	(0.0027)	(0.0050)	(0.0016)	(0.0050)	(0.0098)	(0.0056)
Medium × Log(Exp + 1)	-0.0007	-0.0008	0.0001	-0.0128	-0.0083	0.0006
	(0.0008)	(0.0012)	(0.0007)	(0.0084)	(0.0099)	(0.0055)
Time-by-Zipcode FE	Yes	Yes	Yes	Yes	Yes	Yes
House Char.	Yes	Yes	Yes	Yes	Yes	Yes
Bust Effect						
Bust p-value						
Medium Effect						
Medium p-value						
Observations	2399467	1462353	1428501	774444	481874	473494
Estimation Method	OLS	OLS	OLS	IV	IV	IV

Note: This table reports estimates of the effect of listing agent's experience (using the $\log(1 + \text{experience})$) on listings' price outcomes, excluding agents whose experience is greater than 50. The first three columns use the specification outlined in Equation 1, and include zipcode-by-listing-year-month fixed effects and controls for house characteristics. Column 1 reports the effect of agent experience on list price normalized to inferred price (measured using the previous sale price, appreciated using zipcode- and price-tier-specific Zillow house price appreciation) for all listings. Column 2 reports the effect on sale prices normalized to inferred price. Column 3 reports the discount that a property sells at relative to its list price. Columns 4,5 and 6 report the analogues of Columns 1,2, and 3 using the IV strategy outlined in section 4.1. All measures are in logs (after taking ratios), and censored (ratios at the 1st and 99th percentile, levels at the 99th percentile). Standard errors are clustered at the MLS level. *,**, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Table J6: County summary statistics

Year	Unique Counties	Agents Mean	Agents SD	Exit Rates Mean	Exit Rates SD	Entry Rates Mean	Entry Rates SD
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
2002	663	225	(656)	0.18	(0.22)	-	-
2003	713	228	(692)	0.17	(0.20)	0.31	(0.28)
2004	747	246	(762)	0.18	(0.22)	0.32	(0.28)
2005	808	266	(845)	0.20	(0.23)	0.35	(0.28)
2006	851	263	(832)	0.24	(0.26)	0.30	(0.27)
2007	853	254	(772)	0.26	(0.25)	0.27	(0.27)
2008	857	225	(683)	0.26	(0.25)	0.20	(0.24)
2009	858	209	(656)	0.23	(0.25)	0.19	(0.23)
2010	851	201	(637)	0.23	(0.25)	0.20	(0.25)
2011	869	186	(611)	0.21	(0.24)	0.20	(0.25)
2012	861	191	(632)	-	-	0.21	(0.26)

Note: This table presents summary statistics for our data at the county level. For each year, Column 1 counts the number of distinct counties observed in our data. Column 2 and 3 report the mean and standard deviation of number of agents active in the counties. Column 4 and 5 report the mean and standard deviation of exit rates. Columns 6 and 7 report the mean and standard deviation of entry rates.

Table J7: Naive counterfactuals

	Sale Probability			Foreclosure Probability		
	Data	Counterf.	% Δ	Data	Counterf.	% Δ
2002	0.71	0.75	5.1	0.001	0.001	-32.9
2003	0.71	0.74	4.7	0.001	0.001	-0.2
2004	0.71	0.74	5.0	0.002	0.002	-7.4
2005	0.66	0.70	6.5	0.003	0.003	-10.2
2006	0.53	0.57	8.0	0.008	0.007	-12.6
2007	0.46	0.51	10.2	0.018	0.015	-21.1
2008	0.47	0.52	12.2	0.025	0.019	-27.5
2009	0.54	0.60	11.2	0.020	0.017	-21.6
2010	0.52	0.57	9.5	0.019	0.015	-21.9
2011	0.58	0.62	6.9	0.014	0.013	-9.7
2012	0.66	0.70	6.0	0.008	0.008	1.7
2013	0.69	0.72	5.5	.	.	.

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: $y_{i,t} = \alpha_{i,t} + \sum_{p \in \text{periods}} \beta_p \log(1 + \text{experience}_{i,t}) + \delta W_{i,t} + \epsilon_{i,t}$, where $W_{i,t}$ are detailed property characteristics, $\alpha_{i,t}$ are zipcode-by-list-month fixed effects, and the β_p vary by year. For the counterfactual, we split all agents in terciles according to their experience (listings weighted) and compute the average experience within each tercile. For all agents whose experience is below the average of the top tercile, we replace experience with that average. 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.

Table J8: Number of clients

	(1)	(2)
Agent Experience	0.85***	0.91***
	(0.02)	(0.02)
Bust × Experience		-0.14***
		(0.03)
Medium × Experience		-0.03
		(0.03)
R^2	0.7152	0.7195
FIPS Code F.E.	Y	Y
N	1672032	1672032

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 and time fixed effects, where the county used for each observation is where an agent has the most number of clients in a particular year. The second specification includes three time periods for boom, bust and medium aggregate states interacted the experience measure. *,**, and *** denote statistical significance at the 10%, 5%, and 1% level, respectively.

Table J9: Model fit

Panel A:	Exit Rates						Entry Rates	
	Experience 0		Experience 10		Experience 40			
	Baseline	Data	Baseline	Data	Baseline	Data	Baseline	Data
Bust Bust	0.31	0.39	0.10	0.10	0.02	0.04	0.00	0.06
Bust Medium	0.22	.	0.02	.	0.02	.	0.50	0.06
Bust Boom	0.22	.	0.02	.	0.02	.	0.39	.
Medium Bust	0.22	0.41	0.04	0.11	0.01	0.04	0.09	0.08
Medium Medium	0.28	.	0.09	.	0.02	.	0.00	.
Medium Boom	0.22	.	0.03	.	0.01	.	0.35	0.06
Boom Bust	0.22	.	0.02	.	0.02	.	0.45	.
Boom Medium	0.22	0.38	0.04	0.08	0.01	0.06	0.11	0.10
Boom Boom	0.30	0.30	0.09	0.07	0.02	0.02	0.00	0.06

Panel B:	Distribution							
	25th Percentile		50th Percentile		75th Percentile		95th Percentile	
	Baseline	Data	Baseline	Data	Baseline	Data	Baseline	Data
Bust Bust	1	1	3	3	6	8	15	24
Bust Medium	0	0	0	3	3	8	10	24
Bust Boom	0	.	1	.	4	.	11	.
Medium Bust	1	0	3	3	5	8	15	23
Medium Medium	2	.	3	.	6	.	16	.
Medium Boom	0	0	1	3	4	8	12	24
Boom Bust	0	.	0	.	3	.	10	.
Boom Medium	1	0	2	3	5	8	14	23
Boom Boom	1	0	3	3	6	8	15	23

Panel C:	Learning							
	Experience 0		Experience 5		Experience 10		Experience 40	
	Baseline	Data	Baseline	Data	Baseline	Data	Baseline	Data
Bust Bust	-0.1	3.4	-3.1	0.7	-6.0	-0.4	-23.6	-4.7
Bust Medium	0.7	3.5	4.5	0.9	8.3	-0.0	31.2	-2.0
Bust Boom	0.7	4.0	2.6	1.4	4.6	0.3	16.3	-1.5
Medium Bust	0.9	3.4	0.8	0.7	0.7	-0.4	0.0	-4.7
Medium Medium	-0.0	3.5	-2.7	0.9	-5.4	-0.0	-21.6	-2.0
Medium Boom	0.7	4.0	2.0	1.4	3.3	0.3	11.0	-1.5
Boom Bust	0.9	3.4	4.9	0.7	8.8	-0.4	32.7	-4.7
Boom Medium	0.7	3.5	0.9	0.9	1.0	-0.0	1.6	-2.0
Boom Boom	-0.2	4.0	-3.0	1.4	-5.9	0.3	-23.0	-1.5

Note: This table reports the fit of the baseline calibrated model against the observed empirical data. Each panel reports the predicted baseline model values and the observed empirical values for pairs of aggregate states, corresponding to the previous year's aggregate state and the current aggregate state. Panel A reports the exit for different experience levels of agents, as well as the overall entry rates. Panel B reports the change in experience (denoted as the change in the experience level this period less the experience last period) for those individuals who did not exit the market. Panel C characterizes the experience distribution at different points in the distribution.