NBER WORKING PAPER SERIES

THE COSTS OF FREE ENTRY: AN EMPIRICAL STUDY OF REAL ESTATE AGENTS IN GREATER BOSTON

Panle Jia Barwick
Parag A. Pathak

Working Paper 17227
http://www.nber.org/papers/w17227

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2011

Jia gratefully acknowledges the research support of the National Science Foundation via grant SES-1061970. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

The Costs of Free Entry: An Empirical Study of Real Estate Agents in Greater Boston
Panle Jia Barwick and Parag A. Pathak
NBER Working Paper No. 17227
July 2011
JEL No. L0,L00,L1,L5,L8,L85,R0,R00

ABSTRACT

This paper studies the real estate brokerage industry in Greater Boston, an industry with low entry barriers and substantial turnover. Using a comprehensive dataset of agents and transactions from 1998-2007, we find that entry does not increase sales probabilities or reduce the time it takes for properties to sell, decreases the market share of experienced agents, and leads to a reduction in average service quality. These empirical patterns motivate an econometric model of the dynamic optimizing behavior of agents that serves as the foundation for simulating counterfactual market structures. A one-half reduction in the commission rate leads to a 73% increase in the number of houses each agent sells and benefits consumers by about $2 billion. House price appreciation in the first half of the 2000s accounts for 24% of overall entry and a 31% decline in the number of houses sold by each agent. Low cost programs that provide information about past agent performance have the potential to increase overall productivity and generate significant social savings.

Panle Jia Barwick
MIT Department of Economics
E52-243C
50 Memorial Drive
Cambridge, MA 02142-1347
and NBER
pjia@mit.edu

Parag A. Pathak
MIT Department of Economics
50 Memorial Drive
E52-391C
Cambridge, MA 02142
and NBER
ppathak@mit.edu

1 Introduction

For a large majority of U.S. households, purchasing or selling a home is one of their most important financial decisions. In 2007, two-thirds of households owned their homes, more than a quarter of national wealth was held in residential real estate, and there were 6.4 million sales of existing homes.[1] The National Association of Realtors (NAR) estimates that almost 80% of residential real estate transactions involve a realtor. Nationwide, brokers' sales commissions exceeded $100 billion annually during the mid-2000s.[2] In Greater Boston, the site of our study, more than $4.1 billion in commissions was paid to realtors in the ten-year period from 1998-2007.

Several distinct features of the brokerage industry have long attracted popular and academic attention. First, the commissions that realtors charge are sizable, typically representing 5-6% of a property's transaction price. A property in our sample sells for $472,000 on average (in 2007 dollars).[3] At 5-6% of the sale price, the commission fee constitutes a significant transaction cost for most households, and is more than 40% of Massachusetts' average median household income in 2006. Second, the commission rate displays little variation over time and across regions, and does not seem to reflect changes in the cost of selling houses. According to a recent federal government report (GAO 2005), "commission rates within a market do not appear to vary significantly on the basis of the price of a home ... and do not appear to have changed much in response to rapidly rising home prices in recent years." These facts have been described in several studies and contrast with the declining fees of intermediaries in other industries (see, e.g., Hsieh and Moretti (2003) and Levitt and Syverson (2008a)). With rising housing prices, a fixed commission rate translates into a higher commission fee per transaction and makes working as a broker more lucrative.

The absence of price competition with regards to commission fees is counter-intuitive given that there are low barriers to enter or exit the brokerage industry. Although all US states require licensing of brokers and salespersons, these requirements do not appear to present significant constraints. For instance, when the national average housing price increased by 83% from 1997 to 2006, NAR membership surged from 716,000 to 1,358,000 during the same period. In subsequent years, house prices fell slightly, but the total number of housing transactions plunged. Many agents stopped working as realtors and NAR reported a 20% reduction in its membership from 2006 to 2009. Similarly, entry peaked in 2004 in our Greater Boston sample when sales prices were highest, and dropped by about one-half by 2007 after prices declined.

In this paper, we quantify the social costs of free entry and examine implications of the current fixed commission structure. Our rich micro-level data set contains all properties in the Multiple Listing System (MLS) network from January 1998 to December 2007 in all cities and towns within a 15-mile radius of downtown Boston. There are 10,088 agents and 257,923 housing observations, each with detailed property attributes and transaction information. We are able to exploit cross-market

[1] Sources for all numbers cited in this section can be found in the data appendix.

[2] While a real-estate broker usually supervises an agent, often as the owner of the firm, and is subject to stricter licensing requirements, we use the terms agent, broker and salesperson interchangeably.

[3] All dollar values in this paper are in terms of 2007 dollars, deflated using the urban CPI (series CUUR0100SA0).

variation across low and high-income suburbs of Boston and time-series variation from growing and declining local housing markets. The first part of the paper documents that an increase in the number of agents neither improves the likelihood of sale nor reduces the amount of time required for properties to sell, but rather results in decreased market share for experienced agents. In addition, listings by inexperienced agents are 9% less likely to sell than listings by an agent with six or more years of experience.

Having documented a strong "business stealing" effect, we next formulate an econometric model to quantify the social costs of free entry, building on existing dynamic discrete choice models (see, e.g., Aguirregabiria and Nevo (2010)). Entry and exit decisions of agents, together with their observed commission revenue, allow us to identify the amount of income entrants could have alternatively earned had they not worked as agents. This foregone income is a measure of inefficiency since agents' entry mostly dilutes the business of existing agents without increasing the total output of the brokerage industry. Our estimates imply that agents' foregone income is about 80% of their observed revenue.

By micro-founding the model on agent's choices, we then compute three counterfactuals taking into account agent re-optimization. First, motivated by FTC's investigations on rigid commissions, we present results from regulated across-the-board reductions in the commission rate.[4] Reducing the commission rate by one-half decreases entry by a third, increases the number of houses sold per agent by 73%, and raises the average sales likelihood by 2%. For our time period, this translates into $2 billion of savings for consumers and $900 million of savings in opportunity costs and entry costs from fewer real estate agents.

The second counterfactual measures the alternative market structure if agents are compensated by their costs of selling properties, a potential outcome with free entry and flexible commissions. Since these costs are unobserved, we treat the commissions that agents charged in 1998 as a conservative upper bound. If agents had been compensated by this upper bound, there would be 24% fewer agents and the number of houses each agent sold would increase by 31%. Total commissions would be reduced by $1.1 billion, and the social savings from fewer agents would be $525 million. Alternatively, this counterfactual can be interpreted as measuring how incentives to work as a realtor change with housing prices. Our estimates imply that only a small fraction of the benefits from housing price appreciation is passed onto agents because of the business stealing from entrants. If there were no housing price appreciation during our sample period, the average commission would be $59,700; by comparison, when housing prices rose 1.5 times during the same period, the average commission only increased to $63,300.

Finally, a fixed commission rate makes it difficult for consumers to distinguish good agents from mediocre ones based on the prices that they charge. Our last counterfactual simulates the

[4] The brokerage industry has been investigated for a number of reasons since at least the 1950s, including the case *U.S. vs National Association of Real Estate Boards* in 1950, which first prohibited coordination of realtor fees. Recent FTC and Department of Justice investigations have also examined internet and virtual real estate offices. See http://www.ftc.gov/bc/realestate/index.htm (last accessed February 2011) for additional information on the FTC's investigations of the real estate industry.

consequence of providing consumers with more information on the past performance of incumbent agents, a policy with low implementation costs given the records contained in the current MLS. By empowering consumers with more information, this policy reduces incentives for inexperienced agents to enter, resulting in an increased number of houses sold per agent and the potential for significant social savings.

The remainder of the paper is structured as follows: Section 2 provides industry background, describes our data source, and reviews the related literature. Section 3 presents an initial empirical analysis of the real estate brokerage industry in Greater Boston. Section 4 develops our econometric model and Section 5 outlines the estimation approach. Section 6 describes our empirical results, while Section 7 presents the counterfactual analyses. The last section states our conclusions. Jia and Pathak (2011) (hereafter JP2) provides supplementary material on the sample construction, computational details, and additional results not reported here.

2 Industry Background and Data

2.1 Industry Background

Real estate agents are licensed experts specializing in real estate transactions. They sell knowledge about local real estate markets and provide services associated with the purchase and sale of properties on a commission basis. For home sellers, agents are typically involved in advertising the house, suggesting listing prices, conducting open houses, and negotiating with buyers. For home buyers, agents search for houses that match their clients' preferences, arrange visits to the listings, and negotiate with sellers. In addition, they sometimes provide suggestions on issues related to changes in property ownership, such as home inspections, obtaining mortgage loans, and finding real estate lawyers.

All U.S. states require real estate brokers and agents to be licensed, but these requirements are minimal. In Massachusetts, applicants for a salesperson license need to take twenty-four hours of classroom instruction and pass a written exam. The qualifications for a broker's license involve a few additional requirements: one year of residence in Massachusetts, one year of active association with a real estate broker, completion of thirty classroom hours of instruction, passing a written exam, and paying a surety bond of five thousand dollars. Agents, or salespersons, can perform most of the services provided by a broker, except that they cannot sell or buy properties without the consent of a broker. All licenses need to be renewed biennially, provided the license holder has received six to twelve hours of continuing education and has paid appropriate fees for renewal (currently, \$93 for a salesperson and \$127 for a broker). The general perception that these requirements are not significant deterrents to working as realtors is confirmed in our dataset where entrants account for about 13% of active agents each year.

2.2 Data

The data for this study come from the MLS network for Greater Boston. We collected information on all listed non-rental residential properties for all towns within a fifteen-mile radius of downtown Boston, with a total of 18,857 agents and 290,738 observations.[5] The list of 31 markets are shown in Figure 1, where we group together some towns and cities with few agents. The record for each listed property includes: listing details (the listing date and price, the listing firm and agent, commissions offered to the buyer's agent, and so on), property characteristics, as well as transaction details (the sale price, date, the purchasing agent and firm) when a sale occurs. The number of days on the market is measured by the difference between the listing date and the date the property is removed from the MLS database. We merge this data set with a database from the Massachusetts Board of Registration on agents' license history which we use to measure their years of experience. Agents' gender is provided by List Services Corporation, which links names to gender based on historical census tabulations. We exclude observations with missing cities or missing listing agents.

Information on commissions charged by real-estate agents is difficult to obtain. Even though our data does not contain commissions paid to listing agents, it does contain commissions paid to buyer's agents. Jia and Pathak (2010) report that the buyer's agent commission is 2.0% or 2.5% for 85% of listings in the sample. Since we expect this to be a lower bound on commissions paid to the listing agent, in the analysis to follow, we assume that the total commission rate is 5% in all markets and years, and is split evenly between the seller's and buyer's agent. According to the 2007 NAR survey, most agents are compensated under a revenue sharing arrangement, with the median agent keeping 60% of his commissions and submitting 40% to his firm. We subsequently discuss how this assumption of a 60%-40% split impacts our analysis.

The MLS dataset does not indicate whether working as a broker is an agent's primary occupation. To eliminate agents who may have briefly obtained access to the MLS system to buy and sell for themselves, we only keep agents who either work as a buyer's agent or a listing agent for more than 1.5 properties per year. This sample restriction leaves us with 10,088 agents listing 257,923 properties, about 90% of the original records. JP2 provides more details on the sample construction.

Our analysis benefits from three sources of variation present in the data: time-series variation in the housing market (from an up market to a down market), cross-sectional variation among agents (from "green" realtors to established agents with decades of experience), and geographical variation (the median household income of the most affluent town is more than three times higher than that of the poorest one). Table 1 shows that the number of listings varied from 20,000 to 23,000 in the late 1990s and early 2000s, but increased to 32,500 in 2005. There was a sharp decline in the number of listed houses in the following years when the housing market suffered from the

[5] To verify MLS's coverage of transactions in the cities that we study, we compared it to the Warren Group's changes-of-ownership file based on town deeds records, which we have access to from 1999-2004. This dataset is a comprehensive recording of all changes in property ownership in Massachusetts. The coverage was above 70% for all cities except Boston, which was around 50%. This fact, together with concerns about data quality in Boston, lead us to exclude the city of Boston from the empirical analysis.

decline in the aggregate economy. The weakness of the housing market in the latter part of the sample is apparent in the fraction of properties sold: before 2005, 75-80% of listed properties were sold; in 2007, only 50% were sold. The third column of Table 1 shows that average sales price was about \$350,900 (in 2007 real dollars) in 1998 and peaked at \$529,200 in 2005. It dropped slightly to \$489,800 - \$502,400 in 2006 and 2007. The amount of time it takes for properties to sell appears to lead the trend in sales prices: it sharply increases in 2005, and by the end of the sample a listed property requires about two months longer to sell than in 1998.

2.3 Related Literature

This paper is related to research on real estate brokers and their impact on the housing market. An important precursor to our work is Hsieh and Moretti (2003) which presents evidence consistent with socially inefficient numbers of real estate agents. Using variation across 282 metropolitan areas from the 1980 and 1990 Census of Population and Housing, they document that the average earnings of real estate agents are similar despite large differences in housing prices, suggesting that agent entry dilutes rents associated with house price differences under fixed commissions. Moreover, agents in cities with high housing prices have lower productivity (measured by houses sold per agent) compared to agents in cities with low housing prices. Using the more recent five percent sample of the 2000 Census of Population and Housing, Han and Hong (2011) examine agents' variable costs of selling houses in a static entry model. Their estimates suggest that free entry leads to a loss of economies of scale: a 10% increase in the number of realtors increases the average variable cost of selling houses by 4.8%. Due to data limitations, they assume all agents are identical and ignore the opportunity cost of entry.

Further afield, there are a number of related papers on real estate agents. Kadiyali, Prince, and Simon (2009) study dual agency issues in real-estate transactions. Levitt and Syverson (2008b) compare home sales by agents who own the property to home sales by agents hired to sell the property. Hendel, Nevo, and Ortalo-Magné (2009) contrast traditional multiple-listing services with for-sale-by-owner platforms. Jia and Pathak (2010) study the impact of buyer's commissions on home sales.

Aside from real-estate brokers, another paper which investigates the impact of free-entry is Berry and Waldfogel (1999)'s study of the radio industry. Their paper is based on cross-sectional data on the number of stations, radio listening, and advertising prices. More recent contributions build dynamic econometric models of entry in imperfectly competitive industries (e.g., Collard-Wexler (2008), Dunne, Klimek, Roberts, and Xu (2009), Ryan (2010), and Xu (2008).) While sharing a methodological approach with these papers, our paper differs from these studies on several dimensions. To capture the main features of the housing market while still allowing estimation to be feasible, we work with a model with six state variables. Rather than following the common approach of discretizing the state space, we treat it as continuous, approximate the value function using basis functions, and cast the Bellman equation as model constraints. We adopt a similar procedure in the counterfactual simulations. Our estimator falls into the class of estimators studied by Ai and Chen

(2003) and Chen and Pouzo (2009). Bajari, Chernozhukov, Hong, and Nekipelov (2009) provide identification results for dynamic games that apply to our application. In an independent study, Kristensen and Schjerning (2011) show that maximum likelihood estimation of dynamic models with value functions being approximated by basis functions has desirable properties.

Finally, our paper is related to an extensive literature on dynamic models of occupational choice and job matching (see e.g., Keane and Wolpin (2009) for a recent survey). Related to our econometric approach, Keane and Wolpin (1994) develop a finite-period dynamic model and use backward induction to estimate the value function. In each step, they solve the value function for a subset of state points, and extrapolate to all other state points.[6] Our method may be applicable in similar problems with the advantage that it avoids discretization and is computationally much less demanding.

3 Initial empirical analysis

3.1 Descriptive statistics

Greater Boston's housing market exhibits significant time-series variation in the number of properties listed, the likelihood of sale, and sales prices over our sample period. Total volume of home sales increased from 18,094 in 1999 to 21,432 in 2004, an increase of 18%. It subsequently decreased to 76% of the 1999 level within three years. The average real sales price of homes went from $385,900 in 1999 to $529,200 in 2004, down to $489,800 in 2007. Since the expected revenue of an agent depends on each of these factors, it is unsurprising that agent entry and exit follow these market-wide trends. Table 2A shows that the number of active agents increased from around 3,800 in 1998 to a peak of more than 5,700 in 2005, just as house price appreciated during the same period. The number of agents who left the industry was around 400-500 during the early period, but rose to 700-800 in the latter part of our sample, when housing market conditions deteriorated and the fraction of listed properties that were sold dropped.

Agent performance is also related to overall trends in the housing market. During the run-up in prices in the early part of the 2000s, the number of properties each agent intermediated was about eight per year. By 2007, the average agent conducted a little more than five transactions. The distribution of agents' transactions is highly skewed: both the number of listings sold per agent and the number of houses bought per agent at the 75$^{\text{th}}$ percentile is four to six times that of the 25$^{\text{th}}$ percentile. During the down markets of 2006 and 2007, a significant fraction of real estate agents were hit hard: more than 25% of the listing agents did not sell any properties at all.

Home sales, agent entry and exit, and agent performance also vary across markets within Greater Boston, as shown in Table 2B. The most expensive town in our sample is Wellesley, where the average house sold for more than $1 million. On the other end, in Randolph, the average sold price is $290,000. Quincy, a town with over 10,000 housing transactions, has significant turnover:

[6] For a few other examples (including discussions) of value function approximation, see Ericson and Pakes (1995), Judd (1998), Farias, Saure, and Weintraub (2010), and Fowlie, Reguant, and Ryan (2011).

it is home to the most entries and has the second largest number of exits. Cambridge has about the same number of properties, but there are considerably fewer agents and much less turnover. This translates into a higher number of properties sold and bought per agent in Cambridge than in Quincy: 10.46 versus 7.27. In general, agents in higher-priced towns are involved in fewer transactions, and the correlation coefficient between the average housing price and the number of transaction per agent is around -0.43 across all markets.

An important component of performance differences between agents is their experience. Panel A of Table 3 reports the average annual commissions of agents based on the number of years they worked as a broker. The category of nine or more years of experience has 19,210 observations with a total of 3,146 agents. These agents were active at the beginning of our sample. All other categories (one to eight years of experience) are mostly comprised of brokers who entered during our sample period. Agents who have worked for one year earn $20,000 on average. They sell about 61% of their listed houses and generate a larger share of their income from working as a buyer's agent. In contrast, agents with the most experience are 13% more likely to sell their listed properties, earn about $73,000 in commissions, and earn more of their commission income working as a seller's agent than as a buyer's agent. There is a clear monotonic pattern between measures of agent experience, sales probabilities, and commissions. Finally, more experienced agents appear to sell faster, although the difference in days on market is modest.

Performance differences are also closely related to agent skill, which we measure by the number of transactions they brokered in the previous year. Panel B of Table 3 reports sales probabilities, days on the market, and commissions by deciles of agent skill. Since this measure is highly correlated with years of experience, Panel B displays similar patterns as Panel A. Agents in top deciles have higher sales probabilities, earn higher commissions, and a larger portion of their income comes from the selling side.

Next we examine agents' performance over time. We assign agents who were present in 1998 (the 1998 cohort) into four groups based on their 1998 commissions, and plot their annual commissions from 1999 to 2007 in Figure 2. Results using other cohorts are similar. The top quartile agents consistently earned $100,000 or more for most of the years, while the bottom quartile barely earned $30,000 in commissions even when housing prices were at their peak. Moreover, agents in the top quartile earn significantly more than those in the 2^{nd} or 3^{rd} quartile. The earning gap between the 2^{nd} and the 3^{rd} quartile agents is much smaller and also compresses in down markets.

Earning differences also influence an agent's decision on whether to work as a broker. Figure 3 follows the same 1998 cohort and reports the fraction of agents who continue working as a broker for each quartile in each year. There are stark differences in the exit rates among the four groups. Only 25% of the top quartile agents left by 2007. In contrast, about three quarters of the agents in the bottom quartile exited at some point during the ten-year period. Figure 3 presents our identification argument in a nutshell: differences in the exit rate of agents earning different commissions allow us to identify their opportunity cost of being a broker.

3.2 Descriptive regressions

We next investigate how competition affects agent performance. The correlations we document here inform the modeling choices we make in the next section. In the regressions below, we measure the extent of competition by the number of competing real estate agents who work in the same market and year.

We report estimates of agent performance, y_{imt}, for agent i working in market m in year t from the following equation:

$$y_{imt} = \delta_m + \lambda_t + \alpha s_{it} + \beta \log(N_{mt}) + \epsilon_{icmt}. \tag{1}$$

where y_{imt} is one of two measures of performance: commissions and the number of transactions. δ_m and λ_t are market and year fixed effects. s_{it} is agent skill, proxied by the number of properties agent i intermediated the previous year following Table 3B.[7] The parameter of central interest is β, the coefficient of the competition measure, which reveals the impact of an increase in the number of competing agents on agent i's performance, adjusting for his skill.

We estimate equation (1) for a fixed cohort, defined as the set of agents who are active or have entered as of a given initial year. This formulation avoids changes in the composition of agents, which would confound our results. For instance, the 1998 cohort analysis includes all agents active in 1998. Agents who entered in subsequent years are excluded from the regression, but they contribute to the competition variable $\log(N_{mt})$.[8] Table 4A reports estimates of β for the first seven cohorts (the 1998- to the 2004- cohort). Estimates for later cohorts 2005-2007 are much less precise as sample sizes drop. To allow for differential competition effects for established agents and others, we assign each agent in a cohort to four groups according to his commission in the initial cohort year, and estimate equation (1) separately for each group. Results of the commission regression and the number-of-transaction regression are reported in columns (1)-(5) and columns (6)-(10) of Table 4A, respectively.

Since agent entry is cyclical and our competition measure increases during a booming market, we anticipate that our estimates are biased downward. Despite this, the β estimates are significantly negative and sizable for almost all regressions we estimated, implying that incumbent agents receive lower commissions and conduct fewer transactions when competition intensifies. For example, a 10% increase in the level of competition is associated with a 1.1%-6.7% decrease in average commissions and a 2.5-7.3% reduction in the number of transactions, with the largest impact born by the 2004 cohort. Across agent quartiles, the estimates tend to be larger for the 2^{nd} and 3^{rd} quartiles, implying that competitors steal more business from middle-tier agents, and have a smaller impact on superstar agents.[9]

Having documented a strong business stealing effect, we now examine whether home sellers

[7] We also estimated models using the number of years worked as a broker and found similar estimates.

[8] Changes in the composition of cohorts do arise when agents exit. We also estimated equation (1) on the subset of agents who are active in all years (a balanced panel). These estimates produce larger estimates of the negative impacts of competition, although the differences are not significant.

[9] The coefficients for the bottom quartile agents are less precise since many have zero or one transaction in any given year.

benefit from more competition among agents. Let h_{imt} be a measure of the sales experience (the likelihood of sale, days on the market, or sales price) of a home intermediated by agent i who works in market m in year t. We estimate a property level regression of the following form:

$$h_{imt} = \delta_m + \lambda_t + \alpha s_{it} + \gamma' X_{ht} + \rho \log(N_{mt}) + \upsilon_{imt}. \tag{2}$$

where δ_m, λ_t, s_{it}, and N_{mt} are defined as in equation (1). X_{ht} represents a vector of property attributes including zip code fixed effects, the number of bedrooms, bathrooms, and other rooms, the number of garages, age, square footage, lot size, architectural style, whether it has a garden, type of heating, whether it is a condominium, a single family or a multi-family dwelling, and sometimes the list price. We report estimates of ρ in Table 4B.

When the number of competing agents in a market increases, the likelihood that a property sells decreases, even though more agents are typically associated with a booming market. The point estimate in column (1) implies that a 10% increase in the number of agents is associated with nearly a 1% reduction in the sales probability, whose sample average is 69%. It is possible that when there are more agents, sellers are more likely to list their property at a higher price to 'fish' for a buyer. We add list price in column (2) to control for seller's preference. With this adjustment, the negative impact reduces to 0.6%, but still significant. Another possible explanation is that the composition of properties changes with the market condition: houses that are harder to sell (due to unobserved attributes) are more likely to be listed in a booming market. To examine this possibility, we interact the competition measure with indicators for before or after 2005 in column (3). The estimates for both periods remain negative, suggesting that the negative impact of competition is not solely driven by unobserved changes in the composition of properties listed in an up market.

The impact of more agents on days on the market is negative, but insignificant. It is possible that variation in days on the market are mostly driven by market-wide conditions, captured by market and year fixed effects in equation (2). Competition does seem to be associated with an increase in the sales price of a property, but the impact is modest when we control for the list price, as documented in columns (8) and (9). A 10% increase in competition generates a 0.14% increase in the sales price, which translates to about $600 for a typical home. The estimate is similar when we allow for differential competition effects before and after 2005. Since a higher sales price is a transfer from buyers to sellers and has a negligible impact on aggregate consumer surplus, in subsequent sections we do not focus on the impact agents have on sales prices. In summary, results from estimates of equation (2) suggest that the benefit home sellers receive from more agent competition is limited at best.

4 Econometric Model

The patterns in the previous section show that there is a strong business stealing effect and that competition from more agents does not improve agents' quality of service as measured by sales likelihood and time to sale. In this section, we incorporate competition among agents in modeling their entry and exit decisions. These decisions, together with observed commission revenue, allow us to estimate their opportunity cost of being a broker. We first describe various elements of the model: the state variables, the revenue (or payoff) function, and the transition process of state variables. Then we present the Bellman equation and the value function and discuss some limitations of the model.

4.1 State variables

To model the evolution of the housing market and how it affects the entry and exit decisions of agents, we need to represent the housing market in terms of state variables. Since our data includes information on the attributes of each property that an agent intermediates, in principle, we could model how agents are matched to particular properties, and how this would impact their commission revenue. We do not pursue this rich representation and instead work with a more stylized version of the housing market.

There are two main reasons for this simplification. First, including property-specific features in the state space substantially increases its dimension. A large number of payoff relevant state variables in dynamic models is challenging for estimation and in counterfactual analyses that involve solving for a new equilibrium. In particular, it is difficult to estimate the joint transition process of many state variables and compute a high-dimensional integral of an unknown value function. Second, even if it were possible to surmount these computational and estimation hurdles, our data do not include information on the characteristics of home sellers and buyers, making it formidable to model the matching process between households and agents without ad hoc assumptions. As a result, we choose a parsimonious representation of the housing market that still allows for a reasonable fit of the main moments of the data.

We assume that agents' commissions are determined by two sets of (payoff-relevant) variables: aggregate variables and their individual characteristics. The aggregate variables are the total number of houses listed on the market, the average housing price, the ratio of inventory over the number of properties sold in the previous year, and the number of competing agents. The total number of listed houses H_{mt} counts all houses for sale as of the first day of year t together with all new listings throughout the year in market m. The average house price P_{mt} is the equal-weighted price of all houses that are sold in market m in year t. To construct the inventory-sales ratio, at the beginning of each month, we take the ratio of the number of listed properties in inventory (which includes new listings and unsold properties) and the number of properties sold in the previous year. Next, we average over 12 months to compute the inventory-sales ratio in year t, denoted by inv_{mt}. This ratio is included as a proxy for market tightness, considered an important factor by the NAR who publishes a similar Market Tightness Index. In our application, this ratio is an important predictor of whether a listed property is sold and the amount of time it takes to sell.

Individual characteristics include an agent's gender, firm affiliation, the number of years they have worked as a broker, and a count of their past transactions. We assume that the first three aggregate state variables describing the housing market – the number of houses for sale, the average price of houses, and the inventory-sales ratio – transition exogenously. That is, we abstract away from potential feedback between agent entry and the aggregate housing market state variables since it seems unlikely that this accounts for a significant fraction of market-level variation.

4.2 Revenue function

Realtors earn commissions either from sales (as listing agents) or purchases (as buyer's agents) of homes. We model these two components of agent revenue separately.

Agent i's commissions from sales depends on his share of houses listed for sale and the probability that these listings are sold within the contract period. Since the aggregate variables are the same for all agents in market m and year t, the listing share only depends on individual and rival characteristics (we omit the market subscript m throughout this section). The following listing share equation can be derived from a static home seller's discrete choice model (presented in the appendix):

$$ShL_{it} = \frac{\exp(X_{it}^L \theta^L + \xi_{it}^L)}{\sum_k \exp(X_{kt}^L \theta^L + \xi_{kt}^L)}. \tag{3}$$

The variables X_{it}^L include agent i's demographics, work experience, firm affiliation, and proxies for agent skill. The variable ξ_{it}^L represents his unobserved quality (observed by all agents, but unobserved by the econometrician), much like unobserved quality variables in Berry, Levinsohn, and Pakes (1995) and other discrete choice models. We report estimates assuming that ξ_{it}^L is independent across periods. In Section 6.1, we present evidence that correlated unobserved state variables may not be important once we include our proxy for agent skill. This assumption is also needed because of computational difficulties of incorporating correlated state variables in dynamic discrete choice models.

The denominator in equation (3)

$$L_t \equiv \sum_k \exp(X_{kt}^L \theta^L + \xi_{kt}), \tag{4}$$

is sometimes called the "inclusive value" (see, e.g., Aguirregabiria and Nevo (2010)). It measures the level of competition agents face in obtaining listings. Rather than tracking all rivals' decisions, they behave optimally against the aggregate competition intensity L_t. This approximation of competition can be motivated by the large number of agents per market.

Agents only receive commissions when listings are sold. The probability that agent i's listings are sold is assumed to have the following form:

$$\Pr(\text{sell}_{it}) = \frac{\exp(X_{it}^S \theta^S)}{1 + \exp(X_{it}^S \theta^S)},$$

where X_{it}^S includes measures of aggregate housing market conditions (total number of houses listed, the inventory-sales ratio, etc.), as well as his own characteristics. Since we treat the sales price as exogenous, this formulation does not allow for a trade-off between the probability of sale and the sales price. An agent's total commission from selling listed houses is:

$$R_{it}^{Sell} = r * H_t * P_t * ShL_{it} * \Pr(\text{sell}_{it}),$$

where r is the commission rate, H_t is the aggregate number of houses listed, and P_t is the average price index.

We develop the model for an agent's commissions from representing buyers in a similar way:

$$R_{it}^{Buy} = r * H_t^B * P_t * ShB_{it},$$

where H_t^B is the total number of houses bought by all home buyers, P_t is the same as before, and ShB_{it} is agent i's share of the buying market:

$$ShB_{it} = \frac{\exp(X_{it}^B \theta^B + \xi_{it}^B)}{\sum_k \exp(X_{kt}^B \theta^B + \xi_{kt}^B)}. \tag{5}$$

Here, X_{it}^B and ξ_{it}^B are his observed and unobserved characteristics, respectively. Similar to the listing share, the inclusive value on the buyer side is:

$$B_t \equiv \sum_k \exp(X_{kt}^B \theta^B + \xi_{kt}^B).$$

To reduce the number of state variables, we make the simplifying assumption that $H_t^B = 0.69 H_t$, where 0.69 is the average probability that houses are sold. In our sample the correlation between H_t^B and H_t is 0.94, so this simplification allows us to reduce the state space. Since an agent's revenue depends on $H_t * P_t$, we group these two variables together as HP_t, a single state variable that measures the aggregate size (in dollars) of a housing market.

Finally, agents earn commissions as both buyer's and seller's agents. As a result, agent i's earnings at any given set of payoff-relevant state variables $S_{it} = \{X_{it}^L, X_{it}^S, X_{it}^B, HP_t, inv_t, L_t, B_t\}$ is:

$$\begin{aligned} R(S_{it}) &= R^{Sell}(S_{it}) + R^{Buy}(S_{it}) \\ &= 0.015 * HP_t * (ShL_{it} * \Pr_{it}^{Sell} + ShB_{it} * 0.69). \end{aligned}$$

Despite this stylized representation of the housing market, the correlation between the model's predicted revenue and the observed revenue is 0.70. The model also captures well the upward and downward trend of observed revenues. We provide details on the model's fitness in Section 6.4.

4.3 Transition process of state variables

When agents consider entry and exit, they factor in both their current revenue and their future prospects as realtors, which are determined by the exogenous state variables as well as rival agents' entry and exit decisions. Table 2A shows that entry nearly doubled in 2005 and then dropped substantially afterward. We do not explicitly model agent's beliefs on how the aggregate state variables evolve. Instead, we adopt a standard AR(1) model with trend breaks before and after 2005, when house prices peaked in our sample.

The aggregate state variables are assumed to evolve according to the following equation:

$$S_{mt+1} = T_0 * 1\left[t < 2005\right] + T_1 * 1\left[t \geq 2005\right] + T_2 * S_{mt} + \alpha_m + \eta_{mt}, \tag{6}$$

where T_0 and T_1 are coefficients of the trend break dummies, $1\{\cdot\}$ are indicator functions, T_2 is a matrix of autoregressive coefficients, α_m is the market fixed effect, and η_{mt} is a mean-zero multivariate normal random variable. Market fixed effects in equation (6) are included to control for size differences across markets: the largest 5 markets have twice as many listings as the smallest 5 markets. In JP2, we show that omitting market fixed effects leads to biased estimates for the autoregressive coefficients.

We also investigated splitting the sample at year 2005 and estimating a separate transition process for each sub-sample without much success. The R^2 for the second part of the sample is very low, since we have only a few periods per market after 2005. An alternative to the structural break is to add multiple lags and high-order polynomials. We prefer equation (6) given that its R^2 is high (ranging from 0.77 to 0.96) and that our panel is relatively short. An agent's skill is also modeled as an AR(1) process, including the trend break as above.

4.4 Entry and exit decisions

In the econometric model, agents can make career adjustments each period: some incumbent agents continue to work as realtors, others leave it (exit), and new agents become brokers (entry). At the beginning of a period, agents observe the exogenous state variables, their own characteristics, as well as two endogenous variables L_{t-1} and B_{t-1} at the end of the previous period. L_t and B_t are measures of the competition intensity and are determined by all agents' entry and exit decisions jointly: they increase when more people become realtors and decrease when realtors quit and seek alternative careers. Agents also observe their private idiosyncratic income shocks and simultaneously make entry and exit decisions.

Since agents can start earning income as soon as they find clients, we assume that there is no time lag between entry (becoming an agent) and earning commissions. This assumption contrasts with the literature on the dynamic entry and exit decisions of firms, which assume that firms pay an entry cost at period t and start generating revenues in period $t+1$ after a one-period delay due to installing capital and building plants (see, e.g., Ericson and Pakes (1995)).

Let Z denote exogenous state variables and individual characteristics and Y denote endogenous state variables L and B. The Bellman equation for an active agent i is:

$$\tilde{V}(Z_{it}, Y_{t-1}) = E_{\tilde{\varepsilon}} \ \max \left\{ \begin{array}{c} E\left[R(Z_{it}, Y_t)|Z_{it}, Y_{t-1}\right] - c + \tilde{\varepsilon}_{1it} + \delta E\tilde{V}(Z_{i,t+1}, Y_t|Z_{it}, Y_{t-1}) \\ \tilde{\varepsilon}_{0it} \end{array} \right. \tag{7}$$

where $E\left[R(Z_{it},Y_t)|Z_{it},Y_{t-1}\right]$ is his expected commission revenue conditional on observed state variables. Conditioning on state variables, the revenue function also depends on ξ^L_{it} and ξ^B_{it} which we integrate out using their empirical distributions.[10] Since income shocks are private, agent i does not observe Y_t that is determined by all rivals' entry and exit at period t. Instead, he forms an expectation of his commission revenue for the coming period if he continues working as a broker.

The variable c captures agent i's costs of brokering house transactions. It includes his foregone labor income from working in an alternative profession, as well as the per-period fixed cost of being an agent due to the expense of renting office space, the cost of maintaining an active license, and resources devoted to building and sustaining a customer network. We assume that the cost of being a broker does not depend on the number of houses he handles, because the marginal *monetary* cost of listing more properties is likely swamped by the fixed costs. We report counterfactual results under different assumptions on marginal cost as a robustness check. In all specifications we consider, c differs across markets, but is the same for agents within a market. In the main specification, c is fixed throughout the sample period, but we also present results allowing it to vary over time.

The econometric model treats "exit" as a terminating action. Re-entering agents account for about 9% of our sample. Relaxing this assumption would require estimating two value functions and substantially increase the complexity of the model.[11]

Private shocks $\tilde{\varepsilon}_0$ and $\tilde{\varepsilon}_1$ are assumed to be i.i.d. extreme value random variables with standard deviation $\frac{1}{\beta_1}$, where $\beta_1 > 0$. Denoting the expected commission revenue $E\left[R(Z_{it},Y_t)|Z_{it},Y_{t-1}\right]$ as $\bar{R}(Z_{it},Y_{t-1})$, and multiplying both sides of equation (7) by β_1, the original Bellman equation can be rewritten as:

$$V(Z_{it},Y_{t-1}) = E_\varepsilon \;\max \begin{cases} \beta_1\bar{R}(Z_{it},Y_{t-1}) - \beta_1 c + \varepsilon_{i1t} + \delta EV(Z_{it+1},Y_t|Z_{it},Y_{t-1}) \\ \varepsilon_{i0t}, \end{cases}$$

where $V(Z_{it},Y_{t-1}) = \beta_1\tilde{V}(Z_{it},Y_{t-1})$ and $\varepsilon_{ikt} = \beta_1\tilde{\varepsilon}_{ikt}$, for $k = 0,1$. Given the distributional assumptions on ε, the Bellman equation is simplified to the usual log-sum form:

$$V(Z_{it},Y_{t-1}) = \log\left[1+\exp\left(\bar{R}(Z_{it},Y_{t-1},\beta) + \delta EV(Z_{it+1},Y_t|Z_{it},Y_{t-1})\right)\right],$$

where we have replaced $\beta_1\bar{R}(Z_{it},Y_{t-1}) - \beta_1 c$ with $\bar{R}(Z_{it},Y_{t-1},\beta)$ to keep the notation simple. The main focus of the empirical exercise is estimating $\beta = \{\beta_1,\beta_2\}$, with $\beta_2 = -\beta_1 c$.

The probability that incumbent agent i is active at the end of period t is:

$$\Pr(\text{stay}_{it}|Z_{it},Y_{t-1},\beta) = \frac{\exp\left(\bar{R}(Z_{it},Y_{t-1},\beta) + \delta EV(Z_{it+1},Y_t|Z_{it},Y_{t-1})\right)}{1+\exp\left(\bar{R}(Z_{it},Y_{t-1},\beta) + \delta EV(Z_{it+1},Y_t|Z_{it},Y_{t-1})\right)}. \tag{8}$$

[10] We ignore the dependence of L_t and B_t on ξ_{it}, which we suspect is negligible given the large number of agents included in L and B.

[11] The estimation strategy would be similar, except that we need to use the exit choice probability to recast one of the choice-specific value functions as a fixed point of a Bellman equation following Bajari, Chernozhukov, Hong, and Nekipelov (2009).

Let $W_{it} = 1$ be an indicator that agent i remains active at t. The log likelihood (for incumbent agents) is:

$$LL(\beta) = \sum_{i,t} 1\left[W_{it} = 0\right] * \log[1 - \Pr(\text{stay}_{it}|\beta)] + \sum_{i,t} 1\left[W_{it} = 1\right] * \log[\Pr(\text{stay}_{it}|\beta)]. \tag{9}$$

Provided we are able to solve for EV and calculate the choice probability $\Pr(\text{stay}_{it}|\beta)$, we can estimate β by maximizing equation (9).

Potential entrants must pay a fee (entry cost) to become a broker. They enter if the net present value of being an agent is greater than the entry cost κ (up to some random shock). The Bellman equation for potential entrant j is:

$$\begin{aligned} V^E(Z_{jt}, Y_{t-1}) &= E_\varepsilon \ \max \begin{cases} -\kappa + \bar{R}(Z_{jt}, Y_{t-1}, \beta) + \varepsilon_{j1t} + \delta EV(Z_{jt+1}, Y_t | Z_{jt}, Y_{t-1}) \\ \varepsilon_{j0t} \end{cases} \\ &= \log\left[1 + \exp\left(-\kappa + \bar{R}(Z_{jt}, Y_{t-1}, \beta) + \delta EV(Z_{jt+1}, Y_t | Z_{jt}, Y_{t-1})\right)\right]. \end{aligned}$$

Just as in equation (8), the probability of entry is:

$$\Pr(\text{entry}_{jt}|Z_{jt}, Y_{t-1}, \beta, \kappa) = \frac{\exp\left(-\kappa + \bar{R}(Z_{jt}, Y_{t-1}, \beta) + \delta EV(Z_{jt+1}, Y_t|Z_{jt}, Y_{t-1})\right)}{1 + \exp\left(-\kappa + \bar{R}(Z_{jt}, Y_{t-1}, \beta) + \delta EV(Z_{jt+1}, Y_t|Z_{jt}, Y_{t-1})\right)}.$$

Let $E_{jt} = 1$ be an indicator that potential entrant j enters at t. The log likelihood of observing $N_t^E = \Sigma_j E_{jt}$ new entrants out of a maximum of $\bar{N}^E$ potential entrants is:

$$LL^E(\beta) = \sum_{j \leq \bar{N}^E, t} 1[E_{jt} = 1] * \log[\Pr(\text{entry}_{jt}|\beta, \kappa)] + \sum_{j \leq \bar{N}^E, t} 1[E_{jt} = 0] * \log[1 - \Pr(\text{entry}_{jt}|\beta, \kappa)]. \tag{10}$$

Since the entry cost estimate $\hat{\kappa}$ is sensitive to the assumption of the maximum number of potential entrants $\bar{N}^E$, we estimate equation (10) separately from the main model (9). We report estimates of entry costs under three different assumptions on $\bar{N}^E$.

4.5 Discussion on modeling assumptions

The main structural parameter of interest is c, the average agent's total costs of brokering transactions. The current formulation of the model does not allow c to depend on state variables. This is constrained by the fact that we only observe one action for each active agent (stay or exit) and cannot separately identify the impact of a state variable working through c versus its impact working through revenue R on agent actions. Likewise, we cannot allow for a variable cost component in c (which depends on the number of transactions) because agent revenue is proportional to his total number of transactions. However, the model does allow c to vary across markets, as might be expected if outside opportunities are related to market conditions.

Some real estate brokers work part time. According to NAR (2007), 79% of realtors report that real estate brokerage is their only source of income. For the other 21% of agents holding more than one job, we do not observe their income from other sources. However, our estimate $\hat{c}$ is the relevant measure of agents' time devoted to working as brokers. Suppose an agent has two jobs, earning \$35,000 as a broker and \$10,000 from a second job. If we observe him exit after earning \$30,000, then his opportunity cost of working as a broker is between \$30,000 and \$35,000 (ignoring the optional value of future earnings), even though the value of his total working time is higher. Having part-time agents reduces $\hat{c}$, but this correctly measures the average value of time that agents devoted to being a broker. We attempted to formally address part-time agents using a discrete mixture model that allows two types of agents with different opportunity costs, but the likelihood was flat in a large region of parameter values.

Two other features of the model are also due to data constraints that we do not observe all aspects of the interaction between home-sellers, agents, and the housing market. First, the model is silent on possible unobserved benefits that consumers derive from competition among realtors. For instance, we do not measure gifts that agents give away in marketing their services. In addition, when there is enhanced competition from entry, agents might work harder to satisfy requests from their clients and provide better services. The results in Section 3 suggest that these possibilities did not translate into gains for sellers on the likelihood of sale and days on the market. It is possible, however, that buyers benefit from the variety afforded by a large pool of agents. If this is an important source of consumer surplus, then our counterfactual results miss the losses to consumers with fewer agents.

Second, since we do not observe the actual contract terms between agents and their firms, we assume that agents keep 60% of total commissions, based on the 2007 national survey conducted by NAR. This assumption affects our estimate proportionately: if the average commission is under-estimated by $\alpha\%$, then β_1 will be over-estimated by the same amount. As a result, the opportunity cost $c = -\frac{\beta_2}{\beta_1}$ will be under-estimated by $\alpha\%$.

5 Solution Method

As explained above, the unknown value function $V(\cdot)$ is implicitly defined by a functional Bellman equation. The ability to quickly compute the value function is a crucial factor in most empirical dynamic models and in many cases is a determining factor in model specification. In this section, we describe our solution algorithm. JP2 contains additional computational details and Monte-Carlo results. To simplify notation, we omit subscripts throughout this section, and use S to denote the vector of state variables.

5.1 Difficulties with existing approaches

We began our analysis with the traditional approach of discretizing the state space, but met with substantial memory and computational difficulties when we tested our model with four state variables. One of the challenges involves calculating the future value, $EV(S'|S)$, a high-dimensional integral of an unknown function. The quadrature rules require evaluating the value function $V(S)$

at quadrature points that do not overlap with grid points. Since $V(S)$ is unknown at any point outside grids, we need to interpolate $V(S)$ from grid points to quadrature points. With four state variables and ten grids each, more than 95% of our computing time was spent on interpolation. As a result, solving the value function using the Bellman iteration $V^k(S) = \Gamma(S, E(V^{k-1}(S'|S))$ for a given parameter value was slow and often took a couple of hours. Moreover, the memory requirements of discretization increases exponentially.[12]

Another factor that discouraged us from discretization is that there are far fewer data points than the size of the state space when the number of state variables is large. Discretizing the state space and solving the value function for the entire state space implies that most of the time in estimation is spent solving value function $V(S)$ for states that are never observed in the data (and hence not directly used in the estimation). In addition, both discretization and interpolation introduce approximation errors that grow with the number of state variables.

The alternative method we pursue approximates the value function $V(S)$ using sieves where unknown functions are approximated by parametric basis functions (see, e.g., Chen (2007)). For our application, this approach has several benefits. First, the sieve approximation eliminates the need to iterate on the Bellman equation to solve the value function, and therefore avoids the most computationally intensive part of estimation. The Bellman equation is instead cast as a constraint of the model that has to be satisfied at the parameter estimates. This formulation reduces the computational burden and makes it feasible to solve for the equilibrium of models with medium to high dimensions. In addition, the algorithm does not spend time calculating the value function in regions of the state space not observed in the sample. The method has the potential to improve upon methods that require calculating the value function for the entire state space, whose number of elements is often an order of magnitude larger than a typical sample size. For example, with six state variables (which is the number of state variables in our base specification) and ten grids for each, there are 10^6 elements in the state space. There are two main downsides of our approach: a) finite-sample biases from the approximation and b) the non-parametric approximation converges to the true value function at a rate slower than the square root of the sample size. JP2 documents Monte-Carlo evidence that the method works well in our application: with a reasonable number of basis functions, the value function approximation error is small, the bias in parameter estimates is negligible, and the computation is very fast. We now present the estimation procedure in detail.

5.2 Sieve estimation of the value function

Recall that our Bellman equation is:

$$V(S) = \log\left(1 + \exp\left[\bar{R}(S, \beta) + \delta EV(S'|S)\right]\right). \tag{11}$$

[12] We ran out of memory on a server with 32GB of RAM when we experimented with 20 grid points for each of the four state variables.

Kumar and Sloan (1987) show that if the Bellman operator is continuous and $EV(S'|S)$ is finite, then sieve approximation approaches the true value function arbitrarily close as the number of sieve terms increases.[13] This fact provides the theoretical foundation for using basis terms to approximate for the value function $V(S)$.

Specifically, let $V(S)$ be approximated by a series of J basis functions $u_j(S)$:

$$V(S) \simeq \sum_{j=1}^{J} b_j u_j(S), \tag{12}$$

with unknown coefficients $\{b_j\}$. Substitution of equation (12) into equation (11) leads to a nonlinear equation:

$$\sum_{j=1}^{J} b_j u_j(S) = \log\left(1 + \exp\left[\bar{R}(S,\beta) + \delta \sum_{j=1}^{J} b_j * Eu_j(S'|S)\right]\right).$$

This equation should hold at all states observed in the data. Our approach is to choose $\{b_j\}$ to best-fit this non-linear equation in "least-squared-residuals":

$$\left\{\hat{b}_j\right\} = \arg\min_{\{b_j\}} \left\| \sum_{j=1}^{J} b_j u_j(S_{(k)}) - \log\left(1 + \exp\left[\bar{R}(S_{(k)},\beta) + \delta \sum_{j=1}^{J} b_j Eu_j(S'|S_{(k)})\right]\right) \right\|_2 \tag{13}$$

where $\left\{S_{(k)}\right\}_{k=1}^{K}$ denotes state values observed in the data, and $\|\cdot\|_2$ is the L^2 norm. Essentially, $\{b_j\}$ are solutions to a system of first order conditions that characterize how changes in $\{b_j\}$ affect violations of the Bellman equation. There are many possible candidates for suitable basis functions $u_j(S)$ including power series, Fourier series, splines, and neural networks. Jia and Genz (2011) compare a group of popular basis functions using Monte-Carlo simulations. In general, the best basis function is application specific, but well-chosen basis functions should approximate the shape of the value function. A large number of poor basis functions can create various computational problems and estimation issues such as large bias and variance.

Since we observe agents' revenue directly, we exploit information embodied in the revenue function to guide our approximation of the value function. In general, if the revenue function $\bar{R}(S)$ increases in S and the transition process TS also increases in S, then the value function $V(S)$ increases in S.[14] This property suggests the following approach: use basis functions that fit the revenue function $\bar{R}(S)$ to approximate the value function in the Bellman equation. Since these basis functions are chosen to preserve the shape of $\bar{R}(S)$, they should also capture the shape of the value function.

Choosing basis terms in high-dimensional models is not a simple matter; hence, we want an adaptable procedure to economize on the number of terms to reduce numerical errors and parameter

[13] We thank Alan Genz for suggesting this reference.

[14] The formal proof of this fact follows from the Contraction Mapping Theorem and is contained in the appendix. In our application, this observation applies since $R(\cdot)$ increases in $-L$ and $-B$, and $-L'$ and $-B'$ increase in $-L$ and $-B$ (i.e., the transition process increases in $-L$ and $-B$).

variance. We adopt the 'Multivariate Adaptive Regression Spline' (MARS) method popularized by Friedman (1991) and Friedman (1993) to find spline terms that approximate the revenue function to a desired degree. MARS repeatedly splits the state space along each dimension, adds spline terms that improve the fitness according to some criterion function, and stops when the marginal improvement of the fit is below a threshold ($1.0 * 10^{-3}$, for example).[15] Once we obtain a set of spline basis terms $\{\hat{u}_j(S)\}$ that best fit our revenue function $\bar{R}(S)$, we substitute them for $\{\hat{u}_j(S)\}$ in equation (13). The estimated coefficients $\{\hat{b}_j\}$ are those that minimize the squared difference between the left-hand-side and right-hand-side of the Bellman equation, where the value function is approximated by $\sum_{j=1}^{J} \hat{b}_j u_j(S_{(k)})$ for each point in the state space $S_{(k)}$. As in other applications of Mathematical Programming with Equilibrium Constraints (see, e.g., Judd and Su (2008) and Dube, Fox, and Su (2009)), we impose equation (13) as a constraint and do not explicitly solve for $\{\hat{b}_j\}$ in each iteration of the estimation procedure.

The number of spline terms J is an important component of estimation. We propose a data driven method to determine J. Let $\hat{\beta}^J$ denote the parameter estimates when the value function is approximated by J spline terms. We increase J until parameter estimates converge, when the element by element difference between $\hat{\beta}^J$ and $\hat{\beta}^{J-1}$ is smaller than half of its standard deviation (which we estimate using non-parametric bootstrap simulations).

5.3 Identification

Identification of β_1 and β_2 follows from the identification argument of a standard entry model. Substantial exit following a moderate reduction in revenue implies a relatively large value of β_1, the coefficient which measures sensitivity to revenue. On the other hand, if exit does not vary much with reductions in revenue, then β_1 is small. The coefficient β_2 is identified from the level of revenue at which exits start to occur. Identification of the spline coefficients b follows from Hotz and Miller (1993), which proved that differences in choice-specific value functions can be identified from observed choice probabilities. In our application, the value function associated with the outside option is set to 0. With this normalization, choice probabilities directly lead to identification of the value function and the spline coefficients b.

6 Estimates

We first examine estimates of the revenue function and state variables' transition process, then present opportunity cost estimates and discuss the model's fit. Throughout this section, we bring back the market subscript m. Following Hajivassiliou (2000), we standardize all state variables to avoid computer overflow errors. The aggregate state variables, HP_{mt}, inv, L_{mt}, and B_{mt} are standardized with zero mean and 1 standard deviation; the skill variable s_{it} is standardized with

[15] We use the R package 'earth' (which implements MARS, written by Stephen Milborrow), together with the L^2 norm as our criterion function. We search for spline knots and spline coefficients that minimize the sum of the square of the difference between the observed revenue and the fitted revenue at each data point.

zero mean and 0.5 standard deviation. JP2 includes additional details and alternative specifications not presented below.

6.1 Revenue function

The revenue function contains three elements: the listing share equation, the buying share equation, and the probability that an agent's listings are sold. De-meaning the log of the listing share (3), we obtain:

$$\ln ShL_{imt} - \overline{\ln ShL_{\cdot mt}} = (X^L_{imt} - \overline{X^L_{\cdot mt}})\theta^L + (\xi^L_{imt} - \overline{\xi^L_{\cdot mt}}) = (X^L_{imt} - \overline{X^L_{\cdot mt}})\theta^L + \tilde{\xi}^L_{imt}, \qquad (14)$$

where $\overline{\ln ShL_{\cdot mt}} = \frac{1}{K}\sum_{k=1}^{K} \ln ShL_{kmt}$. The other two averages are defined similarly.

We estimated equation (14) using different control variables X_{imt}: gender, firm affiliation, the number of years as a realtor, and an agent's total number of transactions in the previous period, which is used as a proxy for his skill s_{it}. We excluded observations with 0 shares, or entrants and second-year agents since their s_{it} is either undefined or biased downward.[16] There are 32,237 agent-year observations.

The number of transactions an agent intermediates in the previous year is an important predictor of listing shares, partly because agents with many past transactions are more likely to receive referrals and attract new customers. When s_{it} is the sole regressor, the R^2 of the listing-share regression is 0.44, a high value given the extent of potential agent heterogeneity. The coefficient on s_{it} is also economically large: increasing s_{it} by one standard deviation increases agent i's listing share by more than sixty percent. In contrast, conditioning on past transactions, gender or affiliation with the top three firms (Century 21, Coldwell Banker, and ReMax) does not improve the R^2. Experience is also an important predictor of listing shares, but it has a limited explanatory power once s_{it} is included. Our preferred specification is column (1) of Table 5A, which only uses s_{it} as a regressor; alternative specifications are reported in JP2.

Given that our proxy s_{it} cannot fully capture all aspects of an agent's skill, residuals $\tilde{\xi}^L_{imt}$ could be positively serially correlated: a good agent consistently out-performs his peers with the same observed value of s_{it}. To investigate this issue, we regressed the residual estimate $\widehat{\tilde{\xi}}^L_{imt}$ on its lags. Interestingly, these residuals exhibit little persistence over time. The R^2 of the OLS regression is 0.002, and the coefficient of lagged $\widehat{\tilde{\xi}}^L_{imt}$ is small and negative (about -0.04), which seems to suggest a "mean reversion" phenomenon. We repeated the analysis with the Arellano-Bond estimator that accommodates agent fixed effects. The coefficient of lagged $\widehat{\tilde{\xi}}^L_{imt}$ is slightly larger in the absolute value but again with a negative sign: -0.15, which indicates the possibility of a "luck" component in agents' performance: a good year is often followed with a bad year. These results suggest that unobserved persistent attributes, which induce a positive serial correlation, are unlikely to be important given our controls.

[16] Including first- or second-year agents only slightly reduces s_{it} coefficient.

Once we have estimated the listing share equation, we compute the state variable

$$\hat{L}_{mt} = \sum_k \exp(X^L_{kmt}\hat{\theta}^L + \hat{\tilde{\xi}}^L_{kmt}),$$

for all markets and periods. Since we cannot estimate $\tilde{\xi}^L_{imt}$ for agents with $ShL_{imt} = 0$, we replace these missing $\tilde{\xi}^L_{imt}$ with the average $\tilde{\xi}^L_{kmt}$ among agents with the same experience.[17] Results of the purchasing share (5) are similar, with a slightly lower R^2 of 0.3. We construct state variable $\hat{B}_{mt}$ analogously as $\hat{L}_{mt}$.

The third element in the revenue function is the probability that agent i's listings are sold:

$$\Pr(\text{sell}_{imt}) = \frac{\exp(X^S_{imt}\theta^S)}{1 + \exp(X^S_{imt}\theta^S)},$$

where X_{imt} includes both aggregate state variables and agent attributes. Assuming whether listed properties get sold are independent events conditioning on X_{imt}, the probability that agent i with a total of L_{it} listings sells S_{it} properties is:

$$\Pr(S_{it}|L_{it}) = \binom{L_{it}}{S_{it}} \Pr(\text{sell}_{imt})^{S_{it}} \left(1 - \Pr(\text{sell}_{imt})\right)^{L_{it}-S_{it}}.$$

We report MLE estimates of θ^S in column (3) of Table 5A. A linear probability model delivers similar results. A standard deviation change in the inventory-sales ratio reduces the probability of sales by 11-16%, while a standard deviation change in s_{it} increases the probability of sales by 3-6.5%. Market fixed effects are included to control for aggregate conditions in different housing markets that affect whether a property gets sold.

Once we have estimated payoff parameters $\theta = \left\{\theta^L, \theta^B, \theta^S\right\}$, we construct our revenue function as follows:

$$R(S_{imt};\theta) = 0.015 * HP_{mt} * \left(\Pr(\text{sell}_{imt}) * ShL_{imt} + 0.69 * ShB_{imt}\right),$$

where S_{imt} denotes state variables $\{HP_{mt}, inv_{mt}, L_{mt}, B_{mt}, s_{it}, \text{ whether } t < 2005\}$. Note that agents do not observe their revenue in the coming period t, because L_{mt} and B_{mt} are determined by all agents' decisions simultaneously and are unknown ex ante. We calculate expected revenue by integrating out L_{mt} and B_{mt} using their empirical distributions estimated in Section 6.2.

6.2 State variable's transition

There are four stochastic aggregate state variables HP, inv, L, and B. We estimate equation (6) with market fixed effects using the Arellano-Bond GMM-IV estimator to accommodate size differences across markets. Market fixed effects in these autoregressions are incidental parameters and cannot be consistently estimated; yet they are necessary for our second stage estimation when we

[17] We also experimented with replacing missing $\tilde{\xi}^L_{imt}$ with zero. Both procedures lead to very similar estimate of L_{mt}.

forecast future state variables. We compute the average residual within each market during the ten-year sample period as our estimate of market fixed effects.

We add the lag of HP in inv's autoregressions because a large number of listings in the previous year is likely to generate an upward pressure on the inventory-sales ratio. Similarly, the lag of HP and inv are added to L and B's autoregressions, as both L and B are endogenous and respond to market conditions: a growing housing market with a larger HP attracts more agents, while a deteriorating market with a higher inventory-sales ratio leads to fewer agents. The lag of HP in inv's regressions and the lag of HP and inv in L and B's autoregressions are treated as predetermined.

As shown in Table 5B, there is a sizeable level shift in the housing market before and after 2005, and the trend-break dummies are significantly different from each other in the regressions for HP and inv. On the contrary, such a level shift is not pronounced in L and B's regressions, suggesting that conditioning on aggregate housing market conditions, there are no structural breaks in the amount of competition agents face in each market. Finally, the adjusted R^2 is high, ranging from 0.77 to 0.96.

We estimated a variety of AR(1) models for s_{it}. As in the listing share regression, agent gender and firm affiliation have no impact on R^2, but a different constant term before and after 2005 produces noticeable differences. Our preferred specification (column (5) of Table 5B) includes the lag of skill as well as trend-break dummies.

6.3 Structural Parameters

As explained in Section 4.4, we allow the cost parameter c to differ across markets. Specifically, we choose $\beta = \{\beta_1, \beta_{2m}\}_{m=1}^{M}$ and $b = \{b_j\}_{j=1}^{J}$ to maximize the following constrained log-likelihood:

$$\max_{\beta,b} LL(S;\beta,b) \text{ such that}$$
$$\{b_j\}_{j=1}^{J} = \arg\min \left\| \sum_{j=1}^{J} b_j u_j(S_{imt}) - \log\left[1+\exp\left(\bar{R}(S_{imt},\beta)+\delta\sum_{j=1}^{J} b_j E u_j(S_{imt+1}|S_{imt})\right)\right]\right\|_2$$

where S denotes state variables, S_{imt} is the vector of state variables for agent i in market m and period t, and the log-likelihood function $LL(\cdot)$ is defined in equation (9).[18]

Since we use a data dependant approach to determine the number of spline basis functions that approximate the value function, we estimate parameters eight times, with an increasing number of spline terms ranging from 24 to 45. For each set of parameters, the standard errors are computed using 100 non-parametric bootstraps.[19] Then we choose the set of parameter estimates $\hat{\beta} = \left\{\hat{\beta}_1^k, \hat{\beta}_{2m}^k\right\}_{m=1}^{M}$ whose element by element difference from the previous set $\left\{\hat{\beta}_1^{k-1}, \hat{\beta}_{2m}^{k-1}\right\}$ is

[18] To minimize potential issues with numerical computing, we use the KNITRO optimization procedure for all estimation (including bootstrap simulations), provide analytic gradients for both the objective function and the nonlinear constraints, experimented with different starting values, and use 10^{-6} for all tolerance levels.

[19] In these bootstrap estimations, we hold estimates of the revenue function and state variables' transition process fixed, because re-estimating them in bootstrap samples requires recomputing all elements of the model and would take too long to compute for all of the specifications we present in Section 6.

smaller than half of their standard deviation:

$$k = \min \left\{ \tilde{k} \in (1, ..., 8) \ : \ |\hat{\beta}_j^{\tilde{k}} - \hat{\beta}_j^{\tilde{k}-1}| \leq 0.5 * \text{std} \left(\hat{\beta}_j^{\tilde{k}} \right), \ \forall j \right\}.$$

Our parameters stabilize when the number of spline terms increases to 39. Once we have $\hat{\beta}$, we take the ratio of $\hat{\beta}_{2m}$ to $\hat{\beta}_1$ to compute opportunity cost estimates: $\hat{c}_m = -\frac{\hat{\beta}_{2m}}{\hat{\beta}_1}$. The standard errors of $\hat{c}_m$ are calculated from the empirical sample of the bootstrap estimates. We report $\hat{c}_m$ and their standard errors, the number of observations, and the number of spline terms in the first two columns in Table 6A. $\left\{ \hat{\beta}_1^k, \hat{\beta}_{2m}^k \right\}$ for $k = 1, ..., 8$ are reported in JP2.

There is a total of 41,856 observations. All estimates have the right sign and are significant at the 0.01 level.[20] On average, the opportunity cost is \$49,000 and accounts for 80% of observed commissions. There is a substantial variation across markets, from \$30,000 for poor towns like Revere to above \$60,000 for wealthier towns such as Newton and Wellesley, which might be expected if residents in richer towns have better outside options.

An important premise of our econometric model is that entry and exit decisions are based in part on the future path of state variables. To examine whether or not agents consider their future earnings, we estimate the model with discount factor δ equal to zero and report $\hat{c}$ in columns (3)-(4) of Table 6A. For a third of the markets, the estimates are negative or insignificantly different from zero; they average \$9,850 for the remaining markets. Compared to Massachusetts' per capita income of \$46,000 in 2006, these numbers appear to be too small and suggest that agents are not entirely myopic. As in most empirical dynamic studies, we do not estimate δ and instead set it to $\delta = 0.90$. In columns (5)-(8), we examine how results vary with different discount factors. Everything else equal, a smaller $\delta = 0.85$ leads to a lower discounted stream of future income and reduces the incentive to work as a broker. To offset the change in δ, the model relies on smaller opportunity cost estimates, which could be interpretated as diminished payoff from an alternative career. The opportunity costs vary from 85% to 90% of the original estimates in column (1) for most markets. With a higher value of $\delta = 0.95$, the average opportunity cost is about \$56,300, or 15% larger than our original estimate.

In our preferred specification, we measure an agent's skill or ability by his past number of transactions s_{it}. One might be concerned about using the lagged outcome variable as a regressor. To address this issue, we re-estimate our model replacing past transactions with an agent's experience in columns (9)-(10). These estimates are similar to those in column (1): the correlation between these two sets of estimates is 0.89, and the average is \$47,300 and \$49,000, respectively. Despite the similarity in $\hat{c}$, this alternative measure of skill delivers a much worse fit of the data. The sample log-likelihood is -14,645, compared with -12,883 using past transactions. Using this measure also reduces the model's fit of observed commissions considerably.

There is a shift in the aggregate economy at the end of our sample, and it is possible that agents outside options are impacted by this change. The last variation on the model allows two opportunity

[20] We restrict $\{\beta, b_j\}$ to between -150 and 150 to prevent overflow or underflow of the exponentiation.

costs per market, $c_{t\geq 2005}$ and $c_{t<2005}$.[21] Results are reported in columns (11)-(14). All but one parameter is significant at the 0.01 level. Interestingly, the opportunity costs are generally higher prior to 2005, although the differences are significant for only a few markets. Mechanically, lower opportunity cost estimates are driven by the fact that conditioning on observed commissions which decreased substantially post 2005, observed exit rates were actually *smaller* than those prior to 2005, ceteris paribus. Results from our preferred specification (column (1)) are in general between $\hat{c}_{m,t<2005}$ and $\hat{c}_{m,t\geq 2005}$, and are closer to $\hat{c}_{m,t<2005}$ on average.

Aside from specifications reported in Table 6A, we estimated a large number of alternative models. For instance, we experimented with imposing a common c across markets, and found that market specific values of c significantly improved the model's fit. Restricting c to be the same for all markets leads to an estimate of $\$41,300$. However, the fit as measured by log-likelihood is worse (-14,088 compared with -12,883 when c varies across market). In addition, the difference between the observed and fitted Pr(stay) is greater than 0.02 for more than half of the markets. In light of large differences across cities, we also estimated our model separately for each market, with or without variations in c over time. Finally, we experimented with different revenue functions that control for both agent experience and skill. Our opportunity cost estimates display consistent patterns across all specifications we analyzed. These results lead us to conclude that the estimates presented here are driven by entry and exit patterns observed in our data and are quite robust. Our preferred specification is column (1), and is the basis of the discussions below.

We report entry cost estimates in Table 6B. As mentioned in Section 4.4, the entry cost is sensitive to the assumption on the maximum number of entrants $\bar{N}_m^E$. We examine three different scenarios. The first one assumes that $\bar{N}_m^E$ is equal to the largest number of entrants ever observed, $\max(N_{mt}^E)$, which has been used in a number of studies (see., e.g., Seim (2006).) The second one assumes that $\bar{N}_m^E$ is twice the number of $\max(N_{mt}^E)$. The third scenario recognizes that markets with many listed houses are more likely to attract potential realtors and assumes that $\bar{N}_m^E$ is proportional to the average number of listings $\bar{N}_m^E = \frac{H_m}{25}$, where $H_m = \frac{1}{T}\Sigma_{t=1}^T H_{mt}$. We also experimented with several other measures ($2 * \text{mean}(N_{mt}^E)$, $\frac{H_m}{10}$, $\frac{H_m}{20}$, etc.) with similar results.

The entry cost κ, its standard deviation, and the probability of entry (defined as $\frac{N_{m,t}^E}{\bar{N}_m^E}$) are reported for each market in all three scenarios.[22] Entry costs increase mechanically with the assumed number of potential entrants. Under the assumption $\bar{N}_m^E = \max(N_{mt}^E)$, the average entry rate across all markets is 61%, leading to the lowest average entry cost estimates of \$18,000 among the three cases reported. Three markets have negative entry costs, which are necessary to justify the high entry rates observed in these markets. The estimated entry costs in the other two scenarios exhibit similar patterns, except that higher costs are associated with a larger value of $\bar{N}_m^E$. The average entry cost is \$79,000 assuming $\bar{N}_m^E = 2 * \max(N_{mt}^E)$ and \$26,800 assuming $\bar{N}_m^E = \frac{H_m}{25}$. These

[21] We also estimated the model using $c_{t\geq 2006}$ and $c_{t<2006}$. This introduces an additional state variable (a trend break dummy at year 2006). Results are similar, but estimates of $c_{t\geq 2006}$ are less stable since our sample ends in 2007.

[22] We do not report the fitted probability of entry, because in this simple MLE estimation (without any constraint), the model is able to match the average entry rate exactly market by market.

numbers might seem high given the general perception of low entry barriers of the realtor brokerage industry. We use the most conservative estimates of entry cost in the counter-factual analysis, and report welfare estimates both with or without incorporating entry costs.

6.4 Model's fit

We compare our model's predictions to the observed data in two ways: information used directly in estimation vs. information not exploited in estimation. We start with the difference between observed revenues and our fitted revenues. Given that agents' commissions are the main driving force of their entry/exit decisions, it is important that our model can approximate the observed commissions.

Observed and predicted revenues may differ because our model abstracts away from differences in properties' physical attributes and only exploits measures of the aggregate housing market. Moreover, observed commissions are agents' realized *ex post* revenue, while fitted commissions are *ex ante* revenue that agents expect to earn: $E[R(Z_{imt}, Y_{mt}; \theta)|Z_{imt}, Y_{mt-1}]$. Any deviation in realized competitive measure from the expected one would cause $E(R)$ to differ from R.

Given these considerations, it is encouraging to observe that the correlation coefficient between R and $E(R)$ is as high as 0.70. The first two columns of Table 7 tabulate observed vs. predicted commissions by year. The model is able to replicate both the upward and the downward trend without year-specific effects. The average observed commission is \$63,300, and the average fitted expected commission is \$63,900. Columns (1) and (2) in Table 8 report observed vs. predicted commissions by market, with small differences for all markets except Arlington and Revere.[23]

Since agents' exit patterns identify their opportunity costs, it is important that the model fits this pattern. Columns (3) and (4) in Table 7 and 8 document observed vs. fitted probability of staying by year and by market, respectively. We are able to match exit probabilities for all years except 2005 when the difference is around 0.02. In particular, the model captures the U-shaped exit probability accurately, even though the only time-series control is the trend-break dummy in state variables' transition matrix. On average, 12% of incumbents exit in any given year, varying from 15% in down markets to 10% in up markets.

Exit rates exhibit bigger differences across markets and range from 16% to 8%. As reported in columns (3) and (4) in Table 8, the model closely approximates the average exit rate for almost all markets, and the biggest difference between the model's prediction and the observed exit rate is 0.02. One reason for this tight relation is the market-specific cost parameter, $\hat{c}_m$, though the fit is notable given the model's nonlinear structure and complex equilibrium constraints.

To benchmark our estimate of agent's opportunity cost $\hat{c}_m$, we searched for other suitable measures that are not in our dataset. By construction, opportunity costs depend on what agents would have earned in an alternative profession and are never observed in reality. We looked for data on earnings by professions, but are not able to obtain such information at the city level. As

[23] The gap is \$10,000 for Arlington and \$7000 for Revere. There is a big discrepancy between the observed and predicted L and B for these two markets, which contributes to the large gaps.

a result, we compare our estimates to each city's 2007 median household income available from http://www.city-data.com/. Figure 4 plots the estimated opportunity cost $\hat{c}_m$, from the smallest to the largest, together with the median household income for each market in our sample. As the opportunity cost increases from the left to the right, the median household income also rises: in general, the foregone income is lower in poor cities and higher in rich towns. The correlation coefficient between $\hat{c}_m$ and the median household income is reasonably high: 0.74. There is some variation in the gap between a realtor's opportunity cost and a typical household income across towns, but on average, his foregone income is slightly higher than half the median household income. These results – high correlation coefficients and comparable magnitudes – reassure about the sensibility of the opportunity cost estimates.

7 Counterfactual analyses

7.1 Methods

In this section, we use the parameter estimates to simulate alternative market structures when agents' payoff functions change. We first describe how we solve for the counterfactual equilibrium, and then present our results. All standard errors of the counterfactual analyses are calculated using 100 nonparametric bootstrap simulations.

The main issue in simulating counterfactuals involves finding the new transition process of future L and B. These two endogenous state variables are determined by all agents' joint entry and exit decisions. In estimating structural parameters, we obtain their transition process directly from data; in a counterfactual, we need to find the equilibrium transition process for L' and B' that is consistent with changes in the payoff function.

To explain our approach, consider the thought experiment of realtors facing reduced payoffs for their services. After forming a belief of the distribution of L' and B' based on current state variables, agents individually solve the new Bellman equation and choose an optimal decision. These decisions jointly determine the distribution of L' and B'. Given the large number of agents ($\geq$100 per market) and the assumption of i.i.d private random shocks $\{\varepsilon_{i0},\varepsilon_{i1}\}$, the distribution of L' and B' (conditional on current state S) can be approximated by that of a normal random variable with two parameters: the mean and the variance.

Agents beliefs are consistent when they are the same as the distribution of L' and B' generated by all realtors' optimal behavior (which in turn depends on their beliefs). In other words, the distribution of L' is a fixed point of the new Bellman equation, with its mean determined by the following equation:

$$\begin{aligned} E\left(L'(S)\right) &= \sum_i E\left[1\left\{\tilde{R}(S_i,\beta)+\delta E_{L',B'}\left[V(S_i')|S_i\right]+\varepsilon_{i1}>\varepsilon_{i0}\right\}\right]\exp(\overline{X_i^L\theta^L}) \\ &+\sum_{j\leq \bar{N}^E} E\left[1\left\{-\kappa+\tilde{R}(S_j,\beta)+\delta E_{L',B'}\left[V(S_j')|S_j\right]+\varepsilon_{j1}>\varepsilon_{j0}\right\}\right]\exp(\overline{X_j^L\theta^L}), \end{aligned}$$

where the first and second term sums over incumbents and potential agents, respectively.[24] $\tilde{R}(S_i, \beta)$ is agent i's expected revenue in the counterfactual, $E_{L',B'}\left[V(S_i')|S_i\right]$ is the expectation of his value function $V(S_i')$ over the distribution of future state variables (L', B', and other exogenous state variables), and $\exp(\overline{X_i^L\theta^L})$ is $E_\xi\left[\exp(X_i^L\theta^L + \xi_i)\right]$. Replacing $E\left[1\left\{\cdot\right\}\right]$ with choice probabilities, simplifying the summation over entrants who are identical ex ante (with attributes X), and omitting the dependence of L' on S, we have:

$$\begin{aligned} E\left(L'\right) &= \sum_i \Pr(\text{stay}_i; L', B')\exp(\overline{X_i^L\theta^L}) + \bar{N}^E \Pr(\text{enter}; L', B')\exp(\overline{X^L\theta^L}) \\ &= \sum_i \Pr(\text{active}_i; L', B')\exp(\overline{X_i^L\theta^L}) \end{aligned} \tag{15}$$

where we write $\Pr(\text{active}; L', B')$ to emphasize that agent i's optimal decision depends on his belief about future competition intensity L' and B'. Equation (15) is similar to equation (4), except that the former takes expectation over agents' entry and exit decisions. The variance of L' is:

$$\text{Var}(L') = \sum_i \Pr(\text{active}_i; L', B')(1 - \Pr(\text{active}_i; L', B'))\exp(2 * \overline{X_i^L\theta^L}). \tag{16}$$

The equilibrium conditions for B' are defined analogously. To summarize, computing the equilibrium is equivalent to searching for the mean and variance of L' and B' in each period for each market.

We show in the appendix that equation (15) has a unique fixed point when L' and B' can be approximated as a normal random variable. Hence, an iterative approach can be employed to compute the equilibrium in the counterfactual.[25] Starting from an initial guess of

$$T^0 = \left\{E^0(L_t'), E^0(B_t'), \text{Var}^0(L_t'), \text{Var}^0(B_t'), \forall t\right\},$$

we compute expected revenue $\tilde{R}(S, \hat{\beta})$ (which depend on T^0 and our estimate $\hat{\beta}$) and search for $\{b_j\}_{j=1}^J$ that minimizes the Bellman constraint:

$$\{b_j\}_{j=1}^J = \arg\min \left\| \sum_{j=1}^J b_j u_j(S) - \log\left(1 + \exp\left[\tilde{R}(S, \hat{\beta}) + \delta \sum_{j=1}^J b_j * E_{L',B'}\left[u_j(S')|S; T^0\right]\right]\right)\right\|_2.$$

With $\{b_j\}_{j=1}^J$ at hand, we calculate the choice probability $\Pr(\text{active}_i)$ for both incumbents and potential entrants, and update our initial guess T^0 using equations (15) and (16).

In practice, solving this problem using an iterative procedure is cumbersome. We use the MPEC

[24] The transition process of exogenous state variables remains unchanged throughout the counterfactual exercises. The dependence of EV on the transition process of exogenous state variables is understood and not explicitly stated for notational simplicity.

[25] The right-hand-side of equation (15) monotonically decreases in $E(L')$, so the iterative approach is guaranteed to converge.

framework to cast the counterfactual analysis as another problem of constrained optimization:

$$\min_{\substack{E(L'),\text{Var}(L'),\\ E(B'),\text{Var}(B')}} \left\| \begin{array}{c} E(L_1') - \sum_i \Pr(\text{active}_{i1}; L', B') \exp(\overline{X_{it}^L \theta^L}) \\ \text{Var}(L_1') - \{\sum_i \Pr(\text{active}_{i1}; L', B')(1 - \Pr(\text{active}_{i1}; L', B')) \\ \exp(2 * \overline{X_{i1}^L \theta^L})\} \\ \ldots \\ E(B_T') - \sum_i \Pr(\text{active}_{iT}; L', B') \exp(\overline{X_{iT}^B \theta^B}) \\ \text{Var}(B_T') - \{\sum_i \Pr(\text{active}_{iT}; L', B')(1 - \Pr(\text{active}_{iT}; L', B')) \\ \exp(2 * \overline{X_{iT}^B \theta^B})\} \end{array} \right\|$$

$$s.t. \quad \{b_j\}_{j=1}^J = \arg\min \left\| \begin{array}{c} \sum_{j=1}^J b_j u_j(S) - \\ \log\left(1 + e^{\tilde{R}(S,\hat{\beta}) + \delta \sum_{j=1}^J b_j E_{L',B'}\{u_j(S'|S)\}}\right) \end{array} \right\|_2 . \quad (17)$$

7.2 Three counterfactuals and results

Several recent developments in the brokerage industry imply downward pressure on the current commission rate. There has been an increasing interest in using non-traditional methods to buy and sell properties (e.g., Levitt and Syverson (2008a)). Some home sellers list their houses on the MLS database for a flat fee (usually less than a thousand dollars) and sell properties on their own. Others use discount brokers who offer a la carte service or work on an hourly basis, often with reduced fees. Our first counterfactual asks the following question: if a regulator imposes a price cap on the commission rate, what is the market structure and social cost savings at various levels of the commission rate? We simulate the brokerage industry using ten different commission rates, and report results in Table 9.

The first row of Table 9 replicates the sample, where the average annual number of agents and entrants across all markets is 154 and 23, respectively. The average number of transactions per agent is 7.78. A typical agent earns \$63,300 per year, and sells 70% of his listings. The other rows of the table report simulations with commission rates ranging from 2.5% to 4.75%. The standard errors are in parentheses, most of which are reasonably tight. When the commission rate is as low as 2.5%, the number of entrants declines 31%, the exit rate initially doubles and gradually stabilizes, and the number of incumbents drops to 91, which is only 59% of what is observed in the sample. Agents' productivity increases by 73%, with a typical agent conducting 13.5 transactions annually. All of these numbers are significantly different from the first row. Note that a 50% reduction in commission rate only leads to a moderate change in the revenue per agent – \$54,400 vs. \$63,300 – due to fewer rivals that partially offsets the decline in the commission rate. This finding is consistent with Hsieh and Moretti (2003), who documented that free entry dissipates rents associated with housing price appreciation, leading the average agent to benefit much less from higher prices. As agents leave this industry, the average sales probability increases by 2%, because remaining agents are generally more experienced and better at striking a deal.

Perhaps the most noteworthy finding of this exercise is that the magnitude of social savings is substantial. Savings in opportunity cost amounts to $863 million, or 22% of total commissions paid by households during the same period. Using our most conservative estimate of entry cost, savings in entry cost is still sizeable, about $36 million. Hsieh and Moretti (2003) compares houses sold per agent across cities and uses a benchmark city to derive social losses from excess entry. Using the top 10th-quantile most productive city or the bottom 25th-quantile city as a benchmark, social loss is about 50% or 7% of brokers' earnings, respectively. While exploiting very different methods and datasets, our results are broadly consistent with their finding.

The last column of Table 9 presents the reduction in commissions paid by households. These figures do not constitute increases in social surplus, because commissions are transfers from households to realtors. To the extent that we care about the distributional effect of commissions, these benefits are huge and twice larger than the welfare cost of free entry. For example, when the commission rate is reduced by half, households would benefit by about $1.94 billion.

A rough back-of-the-envelope calculation suggests that reducing the commission rate by half at the national level would decrease opportunity costs and entry costs by as much as $23 billion, while lowering the commission payments of consumers by $50 billion. While there are many caveats with applying our results at the national level, it is clear that policies which encourage lower commissions could generate significant welfare gains.

Results for other commission rates reveal a similar pattern: lower commissions lead to fewer agents, higher productivity, and higher social savings. While our simulation rests on the assumption that agents are not capacity constrained, we believe that this assumption is defendable for situations considered here. For instance, agents in our study sold 80% more houses in earlier years than they did in later years. In addition, while the membership of NAR nearly doubled from 1998 to 2006, the number of national home sales only increased by 30%. These empirical patterns suggest that most agents are not capacity constrained and that reducing the number of agents by 40% may not have a major impact on the total number of properties brokered.

Under the assumption that free entry does not increase consumer surplus, supported by the descriptive regressions presented in Section 3.2, the socially optimal commission rate is one that minimizes agents' idle capacity. Since our data do not contain information on the number of hours it takes an agent to sell a property, we cannot directly measure the number of transactions an agent could manage at full capacity. If an agent could handle 5 listings or more at the same time and a property takes 12 weeks to sell, then a broker could conduct 20 or more transactions per year. At that level, the optimal fixed commission rate would be considerably lower than 2.5%.

The second counterfactual examines what would happen if agents were compensated by their cost of selling a house, a possible outcome under adjustable commission rates and free entry. Unlike housing prices that experienced significant appreciation in the last decade, the cost of selling houses is likely to have decreased with widespread internet use.[26] Since the cost of selling houses is not observed, we use the 1998 average commission ($18,118 per transaction) as a conservative upper

[26] According to Survey of California Home Buyers (C.A.R. 2008), 78% used the internet as an important part of their home-buying process.

bound and hold it fixed for 1999-2007. This exercise could also answer the question of what would have happened if there were no housing price appreciation throughout our sample period. Table 10 shows that if agents had been compensated at the 1998 level, there would be 24% fewer agents, each facilitating 31% more transactions. Total commissions paid by households would be reduced by \$1.1 billion, and the social savings in opportunity costs and entry costs would reach \$525 million.

So far our exercises are mainly concerned with the impact of lowering commissions. Our third counterfactual examines the benefit of providing consumers with more information. Under the current commission structure, households cannot use the price to distinguish good agents from mediocre ones. Many rely on referrals, which are often subjective and are sometimes difficult to obtain. In 2006, the FTC published *FTC Facts for Consumers* and explicitly advised consumers to "find out what types of properties, how many units, and where brokers have sold" to "determine how efficiently they're operating and how much experience they have" (FTC 2006). A potentially useful policy instrument, therefore, is an experience rating program, where a government makes public agent past performance or uses this information to certify good vs. subpar agents. Understanding the economic forces at play in such interventions is the focus of our third counterfactual analysis. Specifically, we simulate the model raising the skill coefficient by 20% to 100%. Larger coefficients represent higher premiums to skills and lead to bigger market shares for skilled brokers, a likely consequence when consumers can easily access information on agents' past performance.

Table 11 shows that when the skill coefficient is doubled, entry declines by 34% and the average number of exits decreases from 18 to 15. On net, the number of active agents drops from 154 to 127, a 17% reduction. Agent productivity is enhanced by 23%, the average sales probability increases to 72%, and the average commission income rises from \$63,300 to \$77,200. This is because large coefficients diminish entry, dampen the business stealing effect, and raise the skill premium for experienced agents. There is no direct benefit to consumers (except for a higher sales probability), but savings in opportunity costs and entry costs are still sizable at \$372 million. Increasing the skill coefficient by 20% to 80% generates similar benefits, though of a smaller magnitude.

Relative to a price cap, this proposal has the advantage of easy implementation with all information directly available in the Multiple Listing System. Moreover, our estimates suggest that it may have the support of incumbent agents whose average commissions could increase substantially.

We repeated Table 9-11 using other specifications discussed in Section 6 and obtained similar results. For example, the social saving in opportunity costs is roughly 10-12% less assuming the discount factor $\delta = 0.85$, and about 13-15% more with $\delta = 0.95$. Using two cost estimates per market, when the commission rate is reduced by half, savings in opportunity cost and entry cost would be \$880 million, compared to \$899 million in Table 9.

One might be concerned that our social cost savings are mismeasured because the marginal monetary cost of housing transactions is assumed to be zero. To address this issue, we simulated our model twice, assuming the marginal cost is either \$100 or \$500 per transaction.[27] Using agent i's observed number of transactions in our MLS data, we first backed out his fixed cost (FC_{imt}) by

[27] As explained in Section 4.5, the marginal cost cannot be seperately identified from the fixed cost.

subtracting the variable cost (VC_{imt}) from $\hat{c}_{mt}$:

$$FC_{imt} = \hat{c}_{mt} - VC_{imt} = \hat{c}_{mt} - MC * T_{imt},$$

where MC is the marginal cost of \$100 or \$500 and T_{imt} is the number of transactions by agent i. Then we re-computed the counter-factual analysis where agents incur both a fixed cost and a variable cost buying or selling properties.

A non-zero marginal cost introduces two countervailing forces. On the one hand, agents sell more properties in the counterfactual and hence incur a higher variable cost. These additional marginal costs are ignored in Table 9-11. Incorporating them leads to lower social cost savings. On the other hand, with these costs factored in, agents' net earning is lower. As a result, fewer agents would remain active, which translates into higher cost savings. In our application, these two forces largely cancel each other out. For example, with a \$500 marginal cost, when the commission rate is reduced by half, there are on average 86 active agents per market/year instead of 91 agents as reported in Table 9. However, the total cost savings are similar: \$902 million (with marginal costs) vs. \$899 million (without marginal costs).

8 Conclusion

In this paper we use a new dataset to document stylized facts of entry and exit among realtors in Greater Boston. Traditional arguments suggest that if the production process involves fixed costs, free entry could be socially inefficient, although such inefficiencies might be outweighed by benefits to consumers if free entry brings more variety, better products, or lower prices. Among real estate brokers, the absence of price competition between agents implies that agents compete on non-price dimensions. Yet, increased competition is not associated with higher sales probabilities or faster sales. While it is possible that competition among brokers improves the allocation between houses and buyers in ways that our data do not capture, on the dimensions that we do observe, the consumer benefits from increased competition appear limited. Our dynamic structural model of real estate agent entry and exit allows us to quantify the social costs of free entry. The method for estimating dynamic discrete choice models may have applications in other settings with rich state spaces.

It is important to emphasize that this paper is not against free entry. Creating additional barriers to entry might lead to adverse effects that are not captured by the model, for example, if real estate licenses are rationed inefficiently. Instead, our analysis measures social costs from competition among brokers under rigid commissions in the absence of observable benefits for households. Since most agents do not appear capacity constrained, alternative market structures with fewer agents seem unlikely to impact the total number of properties brokered and may lead to social savings.

Our counterfactuals are intended to investigate these alternatives. Although the econometric model we develop is stylized in certain dimensions, we are able to match key moments of the data. Hence, the counterfactuals may be useful for measuring quantitative changes in the market structure. Each of the three situations we investigate – regulated price caps, commissions based on costs, or improved information about agents' past performance – indicate large social costs with the current fixed percentage commission regime and the potential for interventions to generate social savings.

There are other benefits associated with lower or flexible commissions that are not captured by our model or the counterfactuals. For example, lower commissions reduce transaction costs, which might lead to a more liquid housing market, improved asset allocation, and better housing consumption. Flexible commissions also provide a channel for consumers to choose services tailored to their preferences. While we take the absence of price competition as given throughout this paper, an interesting issue for future work is understanding forces that sustain this market structure.

References

Aguirregabiria, V., and A. Nevo (2010): "Recent Developments in IO: Dynamic Demand and Dynamic Games," Econometric Society World Congress.

Ai, C., and X. Chen (2003): "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71, 1795–1843.

Bajari, P., V. Chernozhukov, H. Hong, and D. Nekipelov (2009): "Nonparametric and Semiparametric Analysis of a Dynamic Discrete Game," Working Paper, Stanford University.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, 63(4), 841–890.

Berry, S., and J. Waldfogel (1999): "Free Entry and Social Inefficiency in Radio Broadcasting," *Rand Journal of Economics*, 30(3), 397–420.

C.A.R. (2008): "2008 Survey of California Home Buyers," California Association of Realtors, available at: http://california.realestaterama.com/2008/09/19/car-2008-survey-of-california-home-buyers-ID0233.html.

Chen, X. (2007): "Large Sample Sieve Estimation of Semi-nonparametric Models, Chapter 76," in *Handbook of Econometrics (Volume 6B)*, ed. by J. J. Heckman, and E. Leamer, vol. 5. Elsevier.

Chen, X., and D. Pouzo (2009): "Estimation of Nonparametric Conditional Moment Models with Possibly Nonsmooth Generalized Residuals," Cowles Foundation Discussion Paper No. 1650R.

Collard-Wexler, A. (2008): "Demand Fluctuations in the Ready-Mix Concrete Industry," Working Paper, New York University.

Dube, J.-P., J. Fox, and C.-L. Su (2009): "Improving the Numerical Performance of BLP Static and Dynamic Discrete Choice Random Coefficients Demand Estimation," Working Paper, Chicago-Booth.

Dunne, T., S. D. Klimek, M. J. Roberts, and D. Y. Xu (2009): "Entry, Exit and the Determinants of Market Structure," Working paper, New York University.

Ericson, R., and A. Pakes (1995): "Markov Perfect Industry Dynamics: A Framework for Empirical Work," *Review of Economic Studies*, 62(1), 53–82.

Farias, V., D. Saure, and G. Weintraub (2010): "An Approximate Dynamic Programming Approach to Solving Dynamic Oligopoly Models," Working paper, MIT Sloan.

Fowlie, M., M. Reguant, and S. Ryan (2011): "Pollution Permits and the Evolution of Market Structure," Working Paper, MIT.

Friedman, J. H. (1991): "Multivariate Adaptive Regression Splines," *Annals of Statistics*, 19(1).

——— (1993): "Fast MARS," Technical Report, Department of Statistics, Stanford University.

FTC (2006): "Selling Your Home? Tips for Selecting a Real Estate Professional," FTC Facts for Consumers, Bureau of Competition, Available at: http://www.ftc.gov/bc/edu/pubs/consumer/homes/zrea01.pdf.

GAO (2005): "Real Estate Brokerage: Factors that May Affect Price Competition," Report to the Committee on Financial Services, Government Accountability Office, available at: http://www.ftc.gov/bc/realestate/resources/GAOreport.pdf.

Hajivassiliou, V. (2000): "Some Practical Issues in Maximum Simulated Likelihood," in *Simulation-Based Inference in Econometrics: Methods and Applications*, ed. by R. Mariano, T. Schuermann, , and M. J. Weeks, vol. 1. Cambridge University Press.

Han, L., and S.-H. Hong (2011): "Testing Cost Inefficiency under Free Entry in the Real Estate Brokerage Industry," Journal of Business and Economic Statistics, forthcoming.

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

Hotz, V. J., and R. Miller (1993): "Conditional Choice Probabilities and the Estimation of Dynamic Models," *Review of Economic Studies*, 60, 497–529.

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

Jia, P., and A. Genz (2011): "A Practical Solution to Medium-Sized Dynamic Models," Working paper, MIT.

Jia, P., and P. Pathak (2010): "The Impact of Commissions on Home Sales," *American Economic Review, Papers & Proceedings*, pp. 475–479.

——— (2011): "Supplemental Material to 'The Cost of Free Entry: An Empirical Study of Real Estate Agents in Greater Boston'," Available online.

Judd, K. (1998): *Numerical Methods in Economics*. MIT Press.

Judd, K. L., and C.-L. Su (2008): "Constrained Optimization Approaches to Estimation of Structural Models," Working paper, Chicago Booth.

Kadiyali, V., J. Prince, and D. Simon (2009): "Is Dual Agency in Real Estate Transactions a Cause for Concern?," Working paper, Indiana University.

Keane, M., and K. Wolpin (2009): "Empirical Applications of Discrete Choice Dynamic Programming Models," *Review of Economic Dynamics*, 12, 1–22.

Keane, M. P., and K. I. Wolpin (1994): "The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation and Interpolation: Monte Carlo Evidence," *Review of Economics and Statistics*, 76, 648–672.

Kristensen, D., and B. Schjerning (2011): "Maximum Likelihood Estimation of Discrete Markov Decision Models by Simulations and Sieves," Working paper, Columbia University.

Kumar, S., and I. Sloan (1987): "A New Collocation-Type Method for Hammerstein Integral Equations," *Math. Computation*, 48, 585–593.

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

——— (2008b): "Market Distortions when Agents are Better Informed: The Value of Information in Real Estate Transactions," *Review of Economics and Statistics*, 90, 599–611.

Ryan, S. (2010): "The Costs of Environmental Regulation in a Concentrated Industry," Working paper, MIT.

Seim, K. (2006): "An empirical model of firm entry with endogenous product-type choices," *Rand Journal of Economics*, 37(3), 619–640.

Sokey, N., R. Lucas, and E. Prescott (1989): *Recursive Methods in Economic Dynamics*. Harvard University Press.

Xu, D. Y. (2008): "A Structural Empirical Model of R&D, Firm Heterogeneity and Industry Evolution," Working paper, New York University.

A Seller's choice model

We derive the listing share (3) using a simple sellers' choice model. Suppose there are H home sellers, each with a unit of property to sell. All properties are identical. A total of N_t agents compete for the listing business. There are many more houses than agents: $H > N_t$. The utility of seller h listing with agent i at time t is assumed to have the following form:

$$U_{hit} = \begin{cases} X_{i,t}\theta^L + \xi_{i,t} + \tau_{hit}, \ i = 1, ..., N_t \\ \tau_{h0t}, \end{cases}$$

where $X_{i,t}$ is a vector of agent i's characteristics, including demographics, past experience, firm affiliation; $\xi_{i,t}$ represents agent i's unobserved quality; and $\tau 8_{hit}$ is the iid error term that captures idiosyncratic utility seller h derives from listing with agent i. If a seller is not matched with any listing agent, he consumes his outside option with utility τ_{h0t}. Assuming that $\{\tau_{hit}\}_{i=0}^{N_t}$ are mean zero iid extreme value random variables, agent i's listing sharing is:

$$S_{i,t}^L = \frac{\exp(X_{i,t}\theta^L + \xi_{i,t})}{\sum_k \exp(X_{k,t}\theta^L + \xi_{k,t})}$$

B Value function monotonicity

Claim. *If the revenue function $\bar{R}(S)$ and transition process TS increase in S, then the value function $V(S)$ increases in S.*

Let $S = \{HP, inv, -L, -B, skill\}$ denote our state variables, where $-L$ is the negative of L. We want to show that our value function $V(S)$ is monotonically increasing in S, where:

$$V(S) = \log(1 + e^{\pi(S) + \beta \int V(S') f(S'|S) dS'}).$$

It is straightforward to show that the operator

$$\Gamma(V) = \log(1 + e^{\pi(S) + \beta \int V(S') f(S'|S) dS'})$$

is a contraction mapping because $f \leq g$ implies $\Gamma(f) \leq \Gamma(g)$, and $\Gamma(V + a) \leq \Gamma(V) + \beta a$ for $a > 0, \beta \in (0, 1)$. According to Corollary 1 on page 52 in Sokey, Lucas, and Prescott (1989), if the contraction mapping operator satisfies $\Gamma[C'] \subseteq C''$, where C' is the set of bounded, continuous, and nondecreasing functions, while C'' is the set of strictly increasing functions, then its fixed point V is strictly increasing.

In our application, $\pi(S)$ strictly increases in S, and the transition matrix

$$S' = TS + \varepsilon,$$

also increases in S (HP' increases in HP, $-L'$ increases in $-L$, etc.). To prove that $\Gamma[C'] \subseteq C''$, we only need to show that if $V(S)$ is nondecreasing in S, then $\int V(S')f(S'|S)dS'$ is nondecreasing in S. Note that:

$$\begin{aligned} g(S) &= \int V(S')f(S'|S)dS' \\ &= \int V(S')f_\varepsilon(S'-TS)dS'. \end{aligned}$$

Let $Z = S' - TS$. Using change of variables, we have:

$$g(S) = \int V(Z+TS)f_\varepsilon(Z)dZ$$

which is nondecreasing in S because $V(S)$ is nondecreasing in S, TS increases in S, and $f_\varepsilon \geq 0$.

C Unique fixed point of equation (15) under normality

Conditional on a given vector of state variables S, if we can approximate L' by a normal random variable, then write

$$L' = \mu + \varepsilon,$$

where ε is a mean-zero normal random variable. We will show that there is a unique μ associated with any value of S.

Recall that μ is the fixed point of the following equation (we omit S since we are conditioning on S):

$$\begin{aligned} \mu &= E(L') = \sum_i \Pr(\text{active}_i)e^{\overline{X_i^L\theta^L}} \\ &= \sum_i \frac{e^{\pi_i(\mu)+\beta\int V(L')f(L';\mu)dL'}}{1+e^{\pi_i(\mu)+\beta\int V(L')f(L';\mu)dL'}}e^{\overline{X_i^L\theta^L}}. \end{aligned} \tag{18}$$

We first show that the right-hand-side of this equation is monotonic in μ. Since expected profits reduce with more competition, $\pi_i(\mu)$ decreases in μ. In addition,

$$\begin{aligned} \int V(L')f(L';\mu)dL' &= \int V(L')f_\varepsilon(L'-\mu)dL' \\ &= \int V(Z+\mu)f_\varepsilon(Z)dZ, \end{aligned}$$

where we replaced L' with $Z+\mu$ in the second equation. Since $V(Z+\mu)$ decreases in μ (because $V(\cdot)$ increases in $-L$, as shown above) and $f_\varepsilon > 0$, this completes the proof that the right-hand-side decreases in μ. Hence, equation (18) has at most one fixed point. At the boundary, when μ approaches 0 (so that few agents are active), $\Pr(\text{active}_i)$ is close to 1, so the right hand side of equation (18) exceeds μ; as $\mu \to \infty$, $\Pr(\text{active}_i) \to 0$, the right hand side of equation (18) is smaller than μ. Hence there is a unique fixed point.

D Data sources cited in Section 1

- In 2007, two-thirds of households owned their homes, more than a quarter of national wealth was held in residential real estate, and there were 6.4 million sales of existing homes.

  - U.S. Department of Housing and Urban Development, Office of Policy Development and Research, "U.S. Housing Market Conditions," 1st, 2nd, 3rd, 4th Quarters 2007, 1st Quarter 2008. Available: http://www.federalreserve.gov/releases/z1/current/accessible/b100.htm, B.100 Balance Sheet of Households and Nonprofit Organizations, from Flow of Funds Accounts of the United States published by the Federal Reserve.

- Brokers' commissions on the sale of real estate properties exceeded $100 billion annually during the mid 2000s.

  - Bureau of Economic Analysis, National Income and Product Accounts Table 5.4.5. Private Fixed Investment in Structures by Type. 1929-2008.

- The recent nationwide appreciation of housing prices by 83% from 1997 to 2006 corresponded to a substantial increase in agent entry.

  - http://www.huduser.org/portal/periodicals/ushmc/spring10/hist_data.pdf, Table 10: Repeated Sales House Price Index: 1991-Present. U.S. Department of Housing and Urban Development.

- NAR membership surged from 716,000 to 1,358,000 during this period

  - http://www.realtor.org/library/library/fg003, "Field Guide to NAR Membership Statistics, 1908-Present," National Association of Realtors.

Table 1: Number of Properties, Prices, Days on the Market, and Total Commissions

Year	No. of Properties (1000)		Sales Price ($1000)		Days on Market		Amount Spent on Intermediation ($mill)
	Listed	Sold	mean	std. dev	mean	std. dev	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1998	23.7	18.3	350.9	295.7	70.4	38.5	281.3
1999	22.0	18.1	385.9	320.4	61.5	35.0	342.6
2000	20.9	17.2	436.5	367.3	54.4	35.0	367.6
2001	22.6	17.6	462.8	365.5	64.5	35.8	386.3
2002	23.2	17.9	508.0	375.2	67.7	40.5	437.0
2003	25.6	19.4	513.1	362.7	77.5	39.0	476.0
2004	28.6	21.4	529.2	363.0	73.7	41.1	547.9
2005	32.5	21.1	526.1	355.6	96.8	45.5	536.3
2006	31.5	17.2	502.4	361.0	131.9	51.0	417.0
2007	27.3	13.6	489.8	364.2	126.2	52.9	359.5
All	257.9	181.9	472.1	358.5	85.4	50.0	4151.6

Note: the numbers include all properties listed and sold by 10,088 agents in the Greater Boston Area. List of towns is in the supplemental material. All prices in 2007 dollars, deflated using urban CPI. Days-on-market is winsorized at 365.

Table 2A: Real Estate Agent Listings and Sales by Year

Year	Entrant	Incumbent Agent	Exiting Agents	Number of Properties Sold	Num. Sold per Listing Agent			Num. Bought per Buyer's Agent		
					mean	25th	75th	mean	25th	75th
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1998	0	3,840	0	18,256	4.75	1	6	3.76	1	5
1999	602	4,054	388	18,094	4.46	1	6	4.43	1	6
2000	462	4,013	503	17,235	4.29	1	6	4.15	1	6
2001	483	4,052	444	17,645	4.35	1	6	3.94	1	6
2002	696	4,344	404	17,872	4.11	1	5	3.91	1	6
2003	883	4,791	436	19,418	4.05	1	5	3.72	1	5
2004	1,005	5,328	468	21,432	4.02	1	5	3.70	1	5
2005	1,002	5,763	567	21,078	3.66	1	5	3.38	1	5
2006	691	5,671	783	17,198	3.03	0	4	2.75	1	4
2007	424	5,227	868	13,648	2.61	0	3	2.90	1	4
All	6,248	10,088	4,861	181,876	3.86	1	5	3.61	1	5

Note: data from the Multiple Listing Service for Greater Boston. An entrant is an agent who did not work in the previous year (either as a listing or a buyer's agent), an incumbent is one who worked as an agent in the year, and an exiting agent does not work in subsequent years.

Table 2B: Real Estate Agent Listings and Sales by Market

Town	Average Sold Price ($1000) (1)	Entrant (2)	Incumbent Agent (3)	Exiting Agents (4)	Number of Properties Sold (5)	Num. Sold per Listing Agent (6)	Num. Bought per Buyer's Agent (7)
WELLESLEY	1051.16	239	505	280	7,459	2.93	2.73
CONCORD	925.47	67	174	91	2,581	2.68	2.46
NEWTON	746.73	215	434	195	8,779	3.94	3.93
LEXINGTON	711.98	141	268	113	4,814	3.27	3.27
HINGHAM	701.88	142	261	132	3,715	2.78	2.75
WINCHESTER	694.65	76	161	90	2,980	3.48	3.36
NEEDHAM	692.15	82	175	71	3,347	3.48	2.97
BROOKLINE	616.98	129	244	104	6,346	4.94	4.60
CAMBRIDGE	582.23	262	417	159	10,763	5.18	5.28
MARBLEHEAD	550.26	107	238	109	5,769	4.33	4.38
WATERTOWN	528.37	157	259	106	5,229	4.10	3.98
DEDHAM	516.62	110	207	103	3,689	3.55	3.12
ARLINGTON	454.32	103	196	85	5,230	4.96	4.86
WALPOLE	446.14	218	369	193	5,496	3.36	2.73
SOMERVILLE	444.83	229	303	152	4,762	3.87	3.89
READING	430.76	128	244	124	4,918	3.95	3.45
WALTHAM	405.42	146	228	108	4,823	4.58	4.10
WILMINGTON	399.37	148	250	150	3,745	3.52	2.69
PEABODY	390.43	191	317	151	5,529	3.73	3.28
STOUGHTON	386.12	272	453	235	7,234	3.48	2.99
MEDFORD	385.21	113	191	86	4,826	5.20	4.10
WAKEFIELD	381.57	243	403	208	7,919	4.19	3.84
QUINCY	379.84	472	677	321	10,757	3.68	3.59
DANVERS	357.96	97	203	110	2,771	2.89	2.59
MALDEN	347.58	404	495	215	7,136	3.70	4.10
WOBURN	347.09	109	179	103	2,918	3.80	3.16
REVERE	325.41	408	520	216	8,454	4.04	4.10
WEYMOUTH	324.14	470	652	329	9,938	3.52	3.04
SALEM	303.79	173	268	134	5,103	4.18	3.75
LYNN	299.47	470	605	286	11,048	4.37	4.18
RANDOLPH	290.57	127	192	102	3,798	4.81	3.60
All	495.38	6,248	10,088	4,861	181,876	3.86	3.61

Note: data from the Multiple Listing Service for Greater Boston. An entrant is an agent who did not work in the previous year (either as a listing or a buyer's agent), an incumbent is one who worked as an agent in the year, and an exiting agent does not work in subsequent years. All sales prices in 2007 dollars, deflated using urban CPI.

Table 3: Days on market, Sales probability, and Commission by Agent Experience and Skill

		Sales Probability		Days on Market		Commissions ($1000)		Listing Commissions		Sales Commissions	
Experience	N	mean	median	mean	median	mean	median	mean	median	mean	median
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
					A. Agents, Sorted by Years of Experience						
1	6,729	0.61	0.67	73.1	54.4	19.9	13.4	7.7	4.2	12.2	8.3
2	6,635	0.63	0.67	71.9	56.0	35.8	24.3	14.0	7.6	21.8	15.0
3	5,237	0.64	0.67	74.6	58.0	40.8	27.7	17.4	9.7	23.4	15.5
4	4,184	0.64	0.68	72.2	55.7	45.1	30.5	21.0	12.0	24.1	16.1
5	3,366	0.67	0.75	70.9	55.7	50.2	33.3	24.4	14.8	25.8	17.4
6	2,657	0.70	0.76	68.9	53.7	55.1	37.4	28.0	17.2	27.1	18.5
7	2,138	0.70	0.78	69.1	54.8	59.6	39.9	31.4	19.2	28.1	18.8
8	1,788	0.70	0.75	70.9	56.5	63.7	42.2	34.2	19.2	29.5	19.5
9+	19,210	0.74	0.80	71.7	58.5	73.4	47.5	41.8	24.8	31.6	20.5
Skill					*B. Agents, Sorted by Deciles of Skill*						
Entrants	7,421	0.62	0.67	71.6	53.7	20.3	13.6	8.0	4.3	12.3	8.3
<10%	3,966	0.67	0.75	73.6	55.0	24.7	17.4	10.7	6.1	13.9	9.8
10-20%	3,966	0.67	0.75	72.8	55.0	26.8	18.4	11.8	6.8	15.0	10.0
20-30%	3,966	0.67	0.75	74.7	55.3	28.6	20.0	13.3	7.9	15.3	10.0
30-40%	3,966	0.68	0.75	75.1	57.0	34.4	24.5	16.1	10.1	18.3	12.7
40-50%	3,967	0.69	0.75	72.7	55.5	39.8	30.1	19.1	12.7	20.7	15.0
50-60%	3,966	0.70	0.75	71.2	56.0	47.0	35.6	23.4	16.5	23.6	17.3
60-70%	3,966	0.71	0.75	71.4	57.0	59.7	47.2	30.2	22.5	29.5	22.3
70-80%	3,966	0.71	0.75	71.2	57.8	73.5	60.0	38.0	28.6	35.5	27.5
80-90%	3,966	0.73	0.78	68.6	58.0	97.0	79.7	52.4	41.2	44.5	35.3
90%+	3,967	0.73	0.78	69.2	60.6	158.7	126.4	94.6	71.8	64.1	50.4

Note: data source is Multiple Listing Service for Greater Boston. All reported commissions are in $1000 2007 dollars, deflated using urban CPI. Skill is proxied by the number of transactions in the previous year. Days-on-market is winsorized at 365. Experience equals 1 means first year of entry.

Table 4A: Impact of Competition on Agent Performance Across Cohorts

Agent Cohorts	All Agents (1)	Top Quartile (2)	2nd Quartile (3)	3rd Quartile (4)	Bottom Quartile (5)	All Agents (6)	Top Quartile (7)	2nd Quartile (8)	3rd Quartile (9)	Bottom Quartile (10)
	I. Agent Commissions					II. Number of Transactions				
1998	-0.11**	-0.05	-0.17**	-0.25***	-0.07	-0.26***	-0.20***	-0.35***	-0.34***	-0.26***
	(0.04)	(0.07)	(0.07)	(0.08)	(0.11)	(0.04)	(0.06)	(0.06)	(0.08)	(0.09)
1999	-0.13***	-0.04	-0.07	-0.24**	-0.22*	-0.25***	-0.20***	-0.26***	-0.31***	-0.31***
	(0.05)	(0.06)	(0.07)	(0.09)	(0.12)	(0.04)	(0.05)	(0.07)	(0.08)	(0.10)
2000	-0.21***	-0.17***	-0.16**	-0.22**	-0.38***	-0.30***	-0.31***	-0.34***	-0.31***	-0.29***
	(0.05)	(0.06)	(0.08)	(0.09)	(0.11)	(0.04)	(0.06)	(0.07)	(0.09)	(0.11)
2001	-0.36***	-0.24***	-0.40***	-0.26**	-0.45***	-0.39***	-0.32***	-0.49***	-0.34***	-0.30**
	(0.05)	(0.07)	(0.08)	(0.11)	(0.14)	(0.05)	(0.07)	(0.08)	(0.10)	(0.13)
2002	-0.42***	-0.40***	-0.47***	-0.49***	-0.31*	-0.41***	-0.46***	-0.54***	-0.45***	-0.16
	(0.07)	(0.09)	(0.11)	(0.13)	(0.16)	(0.06)	(0.09)	(0.11)	(0.13)	(0.15)
2003	-0.50***	-0.47***	-0.61***	-0.73***	-0.52**	-0.49***	-0.52***	-0.66***	-0.70***	-0.47**
	(0.09)	(0.11)	(0.13)	(0.17)	(0.23)	(0.09)	(0.12)	(0.15)	(0.18)	(0.21)
2004	-0.67***	-0.50***	-0.75***	-1.01***	-0.30	-0.73***	-0.70***	-0.90***	-1.03***	-0.15
	(0.13)	(0.17)	(0.19)	(0.24)	(0.33)	(0.14)	(0.19)	(0.22)	(0.26)	(0.33)

Note: * significant at 10% level, ** significant at 5% level, and *** significant at 1% level. Dependent variable is the log of the total agent commissions in Panel I and log of number of transactions in Panel II. The regressors are: log of total number of agents in a given market/year, market/year fixed effects. Each cell reports coefficient on log of total number of agents in a given market/year, with robust standard errors clustered by agent. Each model is estimated by cohort, holding fixed the set of incumbent agents in the cohort year. The sample excludes 1,631 agent-years without transactions (out of 47,083, or 3%).

Table 4B: Impact of Competition on Property Sales

	Sales Probability			log(Days on Market)			log(Sales Price)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log(Nmt)	-0.096***	-0.061**		-0.048	-0.073		0.182***	0.014***	
	(0.03)	(0.02)		(0.07)	(0.07)		(0.03)	(0.00)	
Log(Nmt) Before 2005			-0.043*			-0.059			0.015***
			(0.03)			(0.08)			(0.00)
Log(Nmt) After 2005			-0.066**			-0.082			0.013***
			(0.03)			(0.07)			(0.00)
List price control	N	Y	Y	N	Y	Y	N	Y	Y
R2	0.0888	0.0979	0.0979	0.1169	0.1173	0.1173	0.8569	0.9933	0.9933
N	239462	239252	239252	171314	171217	171217	171228	171212	171212

Note: * significant at 10% level, ** significant at 5% level, and *** significant at 1% level. Each cell reports coefficients on log of total number of agents in a given market/year (log(Nmt)). All models include flexible controls for property characteristics, market/year fixed effects, zip code fixed effects, and number of transactions sold in the previous by listing agent. Robust standard errors clustered by market.

Table 5A: Revenue Function Regressions

	Listing Share (1)	Buying Share (2)	Sold Probability (3)
Skill	1.27***	0.90***	0.21***
	(0.01)	(0.01)	(0.01)
Inv			-0.35***
			(0.01)
L05			1.27***
			(0.03)
Ge05			0.83***
			(0.03)
Estimation Method	OLS	OLS	MLE
Market Fixed Effects	No	No	Yes
N	32237	30986	32237
R^2 adjusted	0.44	0.30	0.18

Table 5B: State Variable Autoregressions

	HP (1)	Inv (2)	L (3)	B (4)	Skill (5)
lag_HP	0.74***	0.21***	0.35***	0.35***	
	(0.05)	(0.06)	(0.02)	(0.02)	
lag_Inv		0.65***	-0.13***	-0.13***	
		(0.05)	(0.02)	(0.02)	
lag_L			0.79***		
			(0.02)		
lag_B				0.76**	
				(0.02)	
lag_Skill					0.75***
					(0.00)
L05	0.29***	-0.10**	0.03*	0.04**	0.04**
	(0.03)	(0.05)	(0.02)	(0.02)	(0.00)
Ge05	0.17***	0.62***	0.12***	0.09***	0.00
	(0.05)	(0.07)	(0.03)	(0.03)	(0.00)
Estimation Method	GMM-IV	GMM-IV	GMM-IV	GMM-IV	OLS
Market Fixed Effects	Yes	Yes	Yes	Yes	No
N	279	279	279	279	30648
R^2 adjusted	0.93	0.77	0.96	0.96	0.59

Note: '*' significant at 10% level, '**' significant at 5% level, and '***' significant at 1% level. 'HP' is the product of the aggregate number of house listings and the average housing price index, 'Inv' is the sales-inventory ratio, 'L' is the listing share inclusive value, 'B' is the buying share inclusive value, and 'Skill' is agent i's number of transactions in the previous year. 'L05' and 'Ge05' are indicators for year<2005 and year>=2005, respectively. GMM-IV refers to the Arellano-Bond estimator. 'R^2 adjusted' for MLE and GMM-IV is pseudo adjusted R^2. Entrants as well as agents with 0 shares are excluded in Table 5A and column (5) of Table 5B.

Table 6A: Opportunity Cost Estimates

	δ=0.90 (Main Specification)		δ=0		δ=0.85		δ=0.95		Years of Experience		Two Cost-Parameters Per Market			
	C	std(C)	C	std(C)	C	std(C)	C	std(C)	C	std(C)	$C_{t<2005}$	std($C_{t<2005}$)	$C_{t\geq2005}$	std($C_{t\geq2005}$)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
ARLINGTON	0.42***	(0.05)	0.07***	(0.02)	0.37***	(0.05)	0.49***	(0.06)	0.53***	(0.01)	0.44***	(0.07)	0.37***	(0.08)
BROOKLINE	0.65***	(0.01)	0.21***	(0.02)	0.59***	(0.01)	0.71***	(0.01)	0.74***	(0.01)	0.66***	(0.04)	0.64***	(0.06)
CAMBRIDGE	0.69***	(0.02)	0.19***	(0.02)	0.61***	(0.02)	0.77***	(0.02)	0.74***	(0.02)	0.83***	(0.02)	0.40***	(0.05)
CONCORD	0.83***	(0.01)	0.24***	(0.02)	0.75***	(0.01)	0.95***	(0.02)	0.68***	(0.01)	0.77***	(0.04)	0.94***	(0.06)
DANVERS	0.31***	(0.05)	-0.05***	(0.02)	0.25***	(0.05)	0.40***	(0.05)	0.28***	(0.01)	0.30***	(0.06)	0.21**	(0.08)
DEDHAM	0.46***	(0.05)	0.08***	(0.02)	0.40***	(0.04)	0.54***	(0.05)	0.44***	(0.02)	0.41***	(0.06)	0.43***	(0.07)
HINGHAM	0.56***	(0.01)	0.13***	(0.02)	0.50***	(0.01)	0.62***	(0.01)	0.49***	(0.01)	0.54***	(0.04)	0.58***	(0.06)
LEXINGTON	0.60***	(0.01)	0.13***	(0.02)	0.53***	(0.01)	0.69***	(0.01)	0.58***	(0.01)	0.61***	(0.04)	0.61***	(0.05)
LYNN	0.38***	(0.03)	0.01	(0.01)	0.33***	(0.02)	0.43***	(0.03)	0.32***	(0.01)	0.34***	(0.04)	0.47***	(0.03)
MALDEN	0.40***	(0.02)	0.02*	(0.01)	0.35***	(0.02)	0.47***	(0.03)	0.35***	(0.01)	0.49***	(0.04)	0.37***	(0.04)
MARBLEHEAD	0.44***	(0.05)	0.10***	(0.02)	0.40***	(0.05)	0.50***	(0.06)	0.59***	(0.02)	0.44***	(0.06)	0.45***	(0.08)
MEDFORD	0.49***	(0.05)	0.06***	(0.02)	0.42***	(0.05)	0.58***	(0.06)	0.49***	(0.02)	0.39***	(0.08)	0.51***	(0.07)
NEEDHAM	0.63***	(0.01)	0.14***	(0.02)	0.56***	(0.01)	0.72***	(0.01)	0.57***	(0.01)	0.71***	(0.04)	0.52***	(0.07)
NEWTON	0.62***	(0.02)	0.23***	(0.01)	0.59***	(0.02)	0.64***	(0.03)	0.74***	(0.01)	0.70***	(0.05)	0.56***	(0.03)
PEABODY	0.37***	(0.04)	0.01	(0.02)	0.32***	(0.03)	0.44***	(0.04)	0.36***	(0.01)	0.38***	(0.05)	0.34***	(0.06)
QUINCY	0.32***	(0.03)	0.01	(0.01)	0.28***	(0.02)	0.36***	(0.03)	0.33***	(0.01)	0.42***	(0.03)	0.31***	(0.03)
RANDOLPH	0.47***	(0.06)	0.01	(0.02)	0.40***	(0.05)	0.57***	(0.06)	0.36***	(0.02)	0.42***	(0.07)	0.43***	(0.07)
READING	0.41***	(0.04)	0.03**	(0.02)	0.35***	(0.04)	0.49***	(0.05)	0.40***	(0.01)	0.34***	(0.06)	0.43***	(0.07)
REVERE	0.30***	(0.03)	0.03**	(0.01)	0.27***	(0.03)	0.34***	(0.03)	0.37***	(0.01)	0.52***	(0.04)	0.34***	(0.04)
SALEM	0.37***	(0.04)	-0.03	(0.02)	0.30***	(0.04)	0.45***	(0.05)	0.30***	(0.01)	0.35***	(0.06)	0.36***	(0.06)
SOMERVILLE	0.59***	(0.05)	0.11***	(0.02)	0.51***	(0.04)	0.67***	(0.05)	0.47***	(0.01)	0.50***	(0.08)	0.61***	(0.06)
STOUGHTON	0.37***	(0.04)	0.00	(0.01)	0.31***	(0.03)	0.44***	(0.04)	0.32***	(0.01)	0.37***	(0.03)	0.31***	(0.05)
WAKEFIELD	0.44***	(0.03)	0.03**	(0.01)	0.38***	(0.03)	0.52***	(0.04)	0.36***	(0.01)	0.45***	(0.04)	0.36***	(0.05)
WALPOLE	0.42***	(0.03)	0.03**	(0.01)	0.36***	(0.03)	0.50***	(0.03)	0.39***	(0.01)	0.39***	(0.04)	0.39***	(0.06)
WALTHAM	0.44***	(0.05)	0.05***	(0.02)	0.37***	(0.04)	0.51***	(0.05)	0.43***	(0.01)	0.44***	(0.07)	0.41***	(0.07)
WATERTOWN	0.50***	(0.01)	0.10***	(0.02)	0.45***	(0.01)	0.57***	(0.01)	0.55***	(0.01)	0.53***	(0.04)	0.48***	(0.06)
WELLESLEY	0.87***	(0.01)	0.36***	(0.01)	0.81***	(0.01)	0.92***	(0.01)	0.79***	(0.01)	0.92***	(0.02)	0.79***	(0.05)
WEYMOUTH	0.34***	(0.03)	-0.03***	(0.01)	0.29***	(0.03)	0.41***	(0.03)	0.27***	(0.01)	0.35***	(0.03)	0.29***	(0.04)
WILMINGTON	0.41***	(0.05)	0.04***	(0.02)	0.35***	(0.05)	0.48***	(0.06)	0.41***	(0.01)	0.47***	(0.05)	0.35***	(0.08)
WINCHESTER	0.69***	(0.01)	0.20***	(0.02)	0.62***	(0.01)	0.78***	(0.01)	0.68***	(0.01)	0.61***	(0.04)	0.82***	(0.05)
WOBURN	0.38***	(0.07)	0.03	(0.02)	0.32***	(0.06)	0.46***	(0.08)	0.34***	(0.03)	0.45***	(0.08)	0.38***	(0.07)
Log-Likelihood	-12883		-12819		-12892		-12875		-14645		-12779			
Number of Observations	41856		41856		41856		41856		41856		41856			
Number of Splines	39		NA		39		39		39		39			

Note: standard errors are estimated via 100 bootstrap simulations, except for column (4) where standard errors are derived using the delta method. '*' significant at 10% level, '**' significant at 5% level, and '***' significant at 1% level. All opportunity costs are in $100,000 2007 dollars. The last row is the number of spline terms used in approximating the value function.

Table 6B: Entry Cost Estimates

Market	max(N^E)			2*max(N^E)			H/25		
	K	std(K)	Prob. Entry	K	std(K)	Prob. Entry	K	std(K)	Prob. Entry
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
ARLINGTON	0.44***	(0.03)	0.50	0.95***	(0.03)	0.25	0.54***	(0.03)	0.44
BROOKLINE	0.04	(0.04)	0.65	0.69***	(0.03)	0.33	0.47***	(0.03)	0.43
CAMBRIDGE	0.17***	(0.01)	0.55	0.81***	(0.01)	0.27	0.31***	(0.01)	0.48
CONCORD	0.21***	(0.06)	0.53	0.75***	(0.06)	0.27	0.21***	(0.06)	0.53
DANVERS	0.28***	(0.04)	0.60	0.87***	(0.04)	0.30	0.28***	(0.04)	0.60
DEDHAM	0.34***	(0.03)	0.53	0.88***	(0.03)	0.27	0.34***	(0.03)	0.53
HINGHAM	0.24***	(0.03)	0.63	0.86***	(0.03)	0.32	0.24***	(0.03)	0.63
LEXINGTON	0.04	(0.04)	0.75	0.70***	(0.05)	0.37	0.23***	(0.04)	0.62
LYNN	0.06***	(0.01)	0.67	0.72***	(0.01)	0.33	0.06***	(0.01)	0.67
MALDEN	0.34***	(0.01)	0.52	0.88***	(0.01)	0.26	0.34***	(0.01)	0.52
MARBLEHEAD	0.21***	(0.04)	0.63	0.83***	(0.04)	0.31	0.70***	(0.04)	0.38
MEDFORD	0.30***	(0.03)	0.52	0.83***	(0.03)	0.26	0.39***	(0.03)	0.47
NEEDHAM	0.21***	(0.04)	0.61	0.81***	(0.04)	0.30	0.36***	(0.04)	0.53
NEWTON	0.02	(0.02)	0.66	0.67***	(0.02)	0.33	0.33***	(0.02)	0.50
PEABODY	0.24***	(0.02)	0.61	0.84***	(0.02)	0.30	0.24***	(0.02)	0.61
QUINCY	0.38***	(0.01)	0.55	0.93***	(0.01)	0.27	0.38***	(0.01)	0.55
RANDOLPH	0.29***	(0.02)	0.50	0.81***	(0.03)	0.25	0.29***	(0.02)	0.50
READING	0.07*	(0.04)	0.68	0.74***	(0.05)	0.34	0.30***	(0.04)	0.57
REVERE	0.42***	(0.01)	0.53	0.96***	(0.01)	0.27	0.42***	(0.01)	0.53
SALEM	0.42***	(0.02)	0.49	0.93***	(0.02)	0.25	0.42***	(0.02)	0.49
SOMERVILLE	0.07***	(0.02)	0.61	0.66***	(0.02)	0.30	0.07***	(0.02)	0.61
STOUGHTON	0.15***	(0.02)	0.64	0.78***	(0.02)	0.32	0.15***	(0.02)	0.64
WAKEFIELD	0.09***	(0.02)	0.64	0.72***	(0.02)	0.32	0.09***	(0.02)	0.64
WALPOLE	0.03	(0.03)	0.69	0.71***	(0.03)	0.35	0.03	(0.03)	0.69
WALTHAM	0.25***	(0.02)	0.58	0.82***	(0.02)	0.29	0.25***	(0.02)	0.58
WATERTOWN	0.18***	(0.02)	0.65	0.82***	(0.02)	0.32	0.26***	(0.02)	0.61
WELLESLEY	-0.17***	(0.02)	0.72	0.57***	(0.02)	0.36	-0.06***	(0.02)	0.67
WEYMOUTH	-0.05***	(0.01)	0.74	0.69***	(0.01)	0.37	-0.05***	(0.01)	0.74
WILMINGTON	0.22***	(0.03)	0.61	0.82***	(0.03)	0.30	0.22***	(0.03)	0.61
WINCHESTER	-0.14**	(0.06)	0.70	0.56***	(0.06)	0.35	0.19***	(0.06)	0.54
WOBURN	0.29***	(0.03)	0.58	0.86***	(0.03)	0.29	0.29***	(0.03)	0.58

Note: parameter standard errors are estimated via 100 bootstrap simulations. '*' significant at 10% level, '**' significant at 5% level, and '***' significant at 1% level. Max. num. of potential entrants equal to max. num. of observed entrants for columns 1-3, twice the max. num. of observed entrants for columns 4-6, and the average num. of listings divided by 25 for columns 7-9. Entry costs in $100,000 2007 dollars.

Table 7: Model Fit, by Year

	Commissions		Probability of Stay	
	Observed	Fit	Observed	Fit
1999	0.60	0.58	0.90	0.90
2000	0.63	0.59	0.88	0.89
2001	0.66	0.67	0.89	0.89
2002	0.72	0.73	0.90	0.90
2003	0.73	0.74	0.90	0.90
2004	0.75	0.76	0.90	0.89
2005	0.67	0.71	0.89	0.87
2006	0.51	0.56	0.86	0.87
2007	0.46	0.46	0.85	0.86
All	0.63	0.64	0.88	0.88

Table 8: Model Fit, by Market

	Commissions		Probability of Stay	
	Observed	Fit	Observed	Fit
ARLINGTON	0.79	0.71	0.91	0.91
BROOKLINE	1.07	1.04	0.91	0.91
CAMBRIDGE	1.05	1.11	0.91	0.89
CONCORD	0.78	0.81	0.90	0.90
DANVERS	0.32	0.34	0.87	0.87
DEDHAM	0.60	0.62	0.89	0.88
HINGHAM	0.61	0.65	0.89	0.89
LEXINGTON	0.78	0.77	0.91	0.92
LYNN	0.47	0.48	0.87	0.88
MALDEN	0.50	0.52	0.87	0.88
MARBLEHEAD	0.76	0.79	0.91	0.91
MEDFORD	0.64	0.67	0.90	0.89
NEEDHAM	0.76	0.77	0.92	0.92
NEWTON	0.96	0.99	0.90	0.90
PEABODY	0.47	0.50	0.89	0.88
QUINCY	0.49	0.48	0.88	0.88
RANDOLPH	0.44	0.46	0.85	0.85
READING	0.54	0.55	0.89	0.89
REVERE	0.48	0.58	0.88	0.90
SALEM	0.45	0.43	0.88	0.88
SOMERVILLE	0.62	0.65	0.86	0.86
STOUGHTON	0.44	0.43	0.87	0.87
WAKEFIELD	0.53	0.51	0.88	0.88
WALPOLE	0.49	0.48	0.87	0.86
WALTHAM	0.65	0.59	0.88	0.89
WATERTOWN	0.73	0.73	0.91	0.91
WELLESLEY	1.03	0.99	0.88	0.88
WEYMOUTH	0.39	0.38	0.87	0.86
WILMINGTON	0.43	0.43	0.84	0.86
WINCHESTER	0.76	0.80	0.89	0.89
WOBURN	0.43	0.44	0.85	0.87

Note: commissions in Table 7 and 8 are in $100,000 2007 dollars.

Table 9: Market Structure with Different Commission Rates

	Ave. Number Transactions	Ave. Number of Entrants	Ave. Number of Active Agents	Ave. Number of Exiting Agents	Avg. Commission	Avg. Sales Probability	oppCst Savings ($mil)	Entry Cost Savings ($mil)	Commission Savings ($mil)
Actual (5%)	7.78	22.52	153.78	18.05	0.63	0.70			
Counterfactual									
4.75%	8.13	21.70	147.18	18.23	0.63	0.70	90.37	4.25	193.52
	(0.15)	(0.22)	(2.37)	(0.58)	(0.00)	(0.00)	(3.75)	(0.51)	
4.50%	8.50	20.93	140.81	18.40	0.62	0.71	177.25	8.26	387.03
	(0.16)	(0.22)	(2.35)	(0.57)	(0.00)	(0.00)	(4.05)	(0.68)	
4.25%	8.91	20.17	134.47	18.57	0.62	0.71	263.90	12.16	580.55
	(0.18)	(0.23)	(2.32)	(0.56)	(0.00)	(0.00)	(4.40)	(0.87)	
4.00%	9.36	19.44	128.15	18.75	0.61	0.71	350.31	15.95	774.06
	(0.19)	(0.23)	(2.29)	(0.55)	(0.00)	(0.00)	(4.79)	(1.05)	
3.75%	9.86	18.72	121.85	18.93	0.60	0.71	436.47	19.63	967.58
	(0.21)	(0.24)	(2.26)	(0.53)	(0.00)	(0.00)	(5.20)	(1.24)	
3.50%	10.42	18.03	115.57	19.11	0.59	0.71	522.36	23.20	1161.09
	(0.23)	(0.24)	(2.22)	(0.52)	(0.00)	(0.00)	(5.62)	(1.42)	
3.25%	11.04	17.35	109.32	19.29	0.58	0.71	607.95	26.66	1354.61
	(0.26)	(0.24)	(2.18)	(0.50)	(0.00)	(0.00)	(6.04)	(1.59)	
3.00%	11.74	16.70	103.09	19.47	0.57	0.71	693.23	30.01	1548.12
	(0.28)	(0.24)	(2.14)	(0.48)	(0.00)	(0.00)	(6.47)	(1.77)	
2.75%	12.54	16.06	96.89	19.66	0.56	0.72	778.19	33.25	1741.64
	(0.31)	(0.24)	(2.09)	(0.47)	(0.00)	(0.00)	(6.88)	(1.93)	
2.50%	13.46	15.45	90.70	19.84	0.54	0.72	862.82	36.39	1935.15
	(0.35)	(0.24)	(2.03)	(0.45)	(0.00)	(0.00)	(7.27)	(2.10)	

Note: average commissions are in $100,000 2007 dollars. In each of the counterfactual simulations, we reduce the commission rate to the denoted fraction of the original commission rate. Standard errors (in brackets) are derived from 100 bootstrap simulations.

Table 10: No Price Appreciation

	Ave. Number Transactions	Ave. Number of Entrants	Ave. Number of Active Agents	Ave. Number of Exiting Agents	Avg. Commission	Avg. Sales Probability	oppCst Savings ($mil)	Entry Cost Savings ($mil)	Commission Savings ($mil)
Actual	7.78	22.52	153.78	18.05	0.63	0.70			
Counterfactual									
No Price Appreciation	10.20	17.99	116.93	19.12	0.60	0.71	502	23	1105.06
	(0.22)	(0.24)	(2.23)	(0.52)	(0.00)	(0.00)	(5.50)	(1.42)	

Note: average commissions are in $100,000 2007 dollars. Standard errors (in brackets) are derived from 100 bootstrap simulations.

Table 11: Improved Information on Agent's Past Performance

	Ave. Number Transactions	Ave. Number of Entrants	Ave. Number of Active Agents	Ave. Number of Exiting Agents	Avg. Commission	Avg. Sales Probability	oppCst Savings ($mil)	Entry Cost Savings ($mil)
Actual	7.78	22.52	153.78	18.05	0.63	0.703		
Counterfactual								
Raise Skill Coef by 20%	8.09	20.83	148.33	17.17	0.66	0.71	62.80	8.90
	(0.14)	(0.19)	(2.33)	(0.54)	(0.00)	(0.00)	(3.93)	(0.64)
Raise Skill Coef by 40%	8.43	19.16	142.56	16.41	0.69	0.71	134.32	17.74
	(0.16)	(0.18)	(2.29)	(0.51)	(0.00)	(0.00)	(4.52)	(1.02)
Raise Skill Coef by 60%	8.76	17.69	137.60	15.74	0.72	0.71	193.71	25.57
	(0.17)	(0.17)	(2.28)	(0.48)	(0.00)	(0.00)	(5.08)	(1.38)
Raise Skill Coef by 80%	9.15	16.26	132.22	15.19	0.75	0.71	269.94	32.94
	(0.19)	(0.17)	(2.28)	(0.46)	(0.00)	(0.00)	(5.63)	(1.77)
Double Skill Coef	9.53	14.95	127.48	14.68	0.77	0.72	332.76	39.63
	(0.21)	(0.17)	(2.30)	(0.45)	(0.00)	(0.00)	(6.13)	(2.09)

Note: average commissions are in $100,000 2007 dollars. In each of the counterfactual simulations, we increase the skill coefficient by the denoted percentage. Standard errors (in brackets) are derived from 100 bootstrap simulations.

Figure 1: Markets in Greater Boston

Figure 2: Commissions by Experience Quartile for the 1998 Cohort (2007 $)

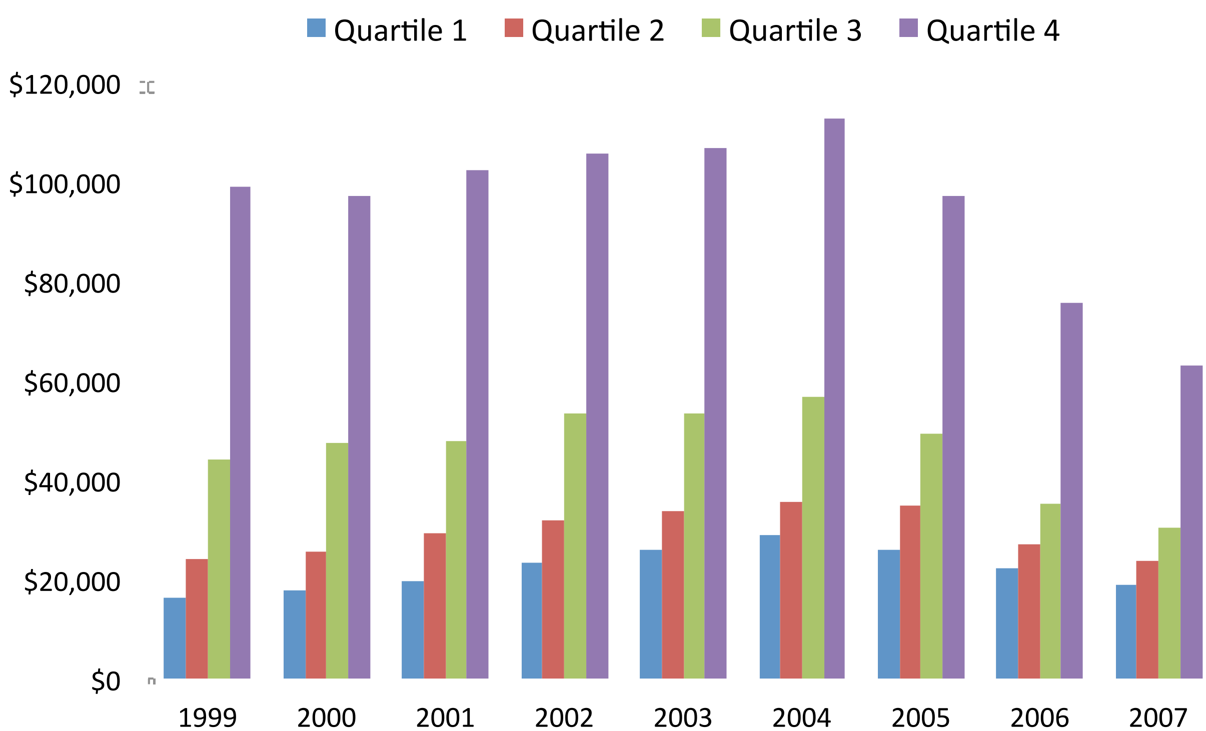

Figure 3: Fraction of Realtors Remaining by Commission Quartile for the 1998 Cohort

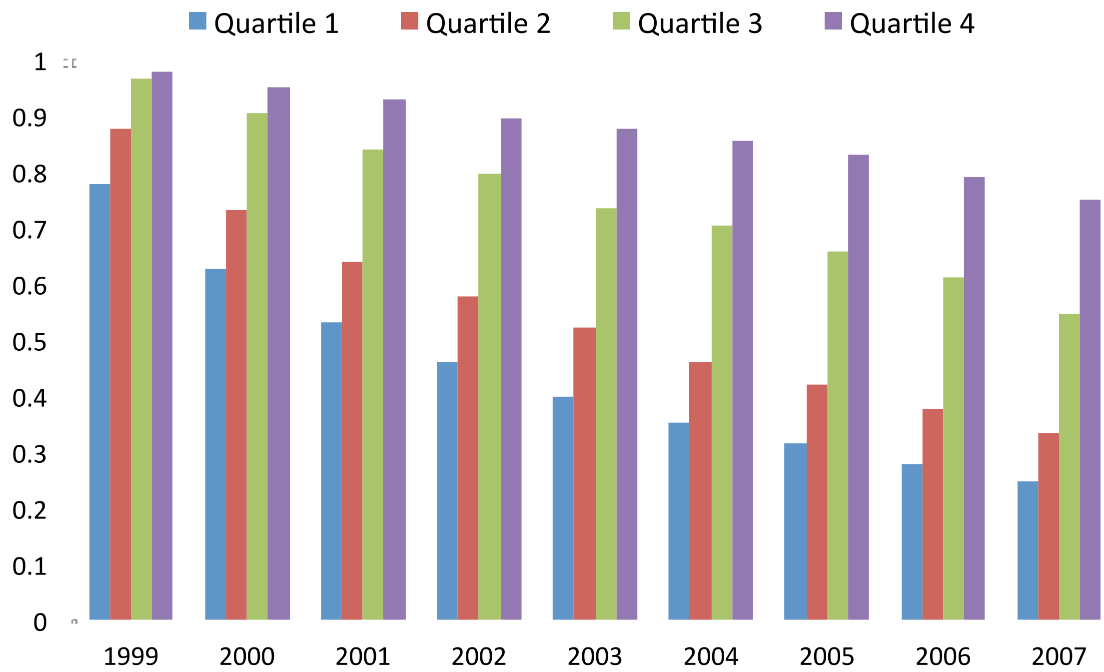

Figure 4: Foregone Income vs. Median Household Income (2007 $)