The Costs of Free Entry: An Empirical Study of Real Estate Agents in Greater Boston*

Panle Jia Barwick[†] Parag A. Pathak[‡]

First verion: August 2011
This version: March 2012

Abstract

This paper studies the consequences of fixed commissions and low entry barriers in Greater Boston's real estate brokerage industry from 1998-2007, a period with substantial agent turnover. We find that entry is not associated with increased sales probabilities or reduced sales time. Instead, it decreases the market share of experienced agents and reduces average service quality. We develop a dynamic empirical model motivated by these patterns to study how agents respond to different incentives. To accommodate a large state space, we approximate the value function using sieves and impose the Bellman equation as an equilibrium constraint. If commissions are cut in half, there would be 40% fewer agents implying social savings of 23% of industry revenue, and the average agent would facilitate 73% more transactions. House price appreciation of 50% during our sample period accounts for a 24% increase in the number of agents and a 31% decline in average agent productivity. Finally, improving information about past agent performance can increase productivity and generate significant social savings.

*We thank Alan Genz for helpful suggestions in the early stage of this project. We thank Paul Asquith, Lanier Benkard, Steve Berry, Xiaohong Chen, JP Dube, Hanming Fang, Phil Haile, Lu Han, Jerry Hausmann, Brad Larsen, Ariel Pakes, Michael Powell, Tavneet Suri, Maisy Wong, Juanjuan Zhang, and seminar participants at Columbia, Harvard, John Hopkins, Michigan, Olin Business School, Toronto, Toulouse, UCL, Wharton, and Yale for their comments. All errors are our own. Comments are welcome.

[†]Department of Economics, MIT and NBER. Email: pjia@mit.edu. Barwick gratefully acknowledges the research support of the National Science Foundation via grant SES-1061970.

[‡]Department of Economics, MIT and NBER. Email: ppathak@mit.edu.

1 Introduction

The current structure of the U.S. real estate brokerage industry puzzles economists, antitrust authorities, and the broader public. On one hand, most agents earn the same commission rate, even for properties with large differences in size, price, ease of sale, and other important dimensions.[1] On the other hand, substantial entry in the past decade and other competitive pressures make the absence of commission competition counterintuitive. Levitt and Syverson (2008a) show that fixed commissions imply conflicts of interest between agents and their clients, Hsieh and Moretti (2003) document that agents chasing after high house prices result in fewer houses sold per agent in more expensive cities, Han and Hong (2011) argue that more agents lead to higher brokerage costs, and Bernheim and Meer (2008) question whether brokers add any value to sellers.[2]

Theoretical arguments suggest that free entry can be socially inefficient if there are fixed costs (e.g., Mankiw and Whinston (1986)). Quantitative claims about brokerage market inefficiency, however, rest on assumptions about agent heterogeneity, the nature of agent competition, and the magnitude of agent fixed costs. This paper uses rich agent-level data from Greater Boston during 1998-2007 to investigate these assumptions and study sources of inefficiency. First, we find that agent productivity differs substantially by experience. Listings by agents with one year of experience are 13% less likely to sell than those by agents with nine or more years of experience. The typical agent intermediates 7.8 properties per year, but this distribution is highly skewed with the 75th percentile five times as large as the 25th percentile.

Second, agent entry and exit decisions respond to aggregate housing market conditions. The number of agents in Greater Boston almost doubles from 1998 to 2004 when house prices reach their peak. When prices fall in 2007, the number of entrants drops by about one-half. Despite significant entry and turnover, there is little evidence that consumers benefit in ways we can measure. In towns with intense agent competition, properties are neither more likely to sell nor do they sell more quickly relative to the overall Boston-wide trend. New entrants mostly compete for listings, taking the largest share from the bottom tier of incumbent agents. Unlike earlier work, our longitudinal data allows us to incorporate agent heterogeneity and their behavioral responses to quantify inefficiencies.

These descriptive patterns motivate our development of an empirical model focusing on agents' decisions to work as brokers to identify the cost of free entry.[3] Agent payoffs depend on the number of properties listed and sold, the price of these properties, and the commission rate. Together with entry and exit decisions, observed commission revenues identify the per-period cost of working as an agent. Knowing agents' per-period costs allows us to measure how market structure changes

[1] While a real estate broker usually supervises an agent, often as the owner of a firm, we use the terms agent, broker, and salesperson, interchangeably.

[2] Other recent work on real estate agents includes Rutherford, Springer, and Yavas (2005), Kadiyali, Prince, and Simon (2009), Hendel, Nevo, and Ortalo-Magné (2009), and Jia and Pathak (2010).

[3] Our approach follows the dynamic discrete choice literature, recently surveyed in Aguirregabiria and Nevo (2010). See also Bajari, Benkard, and Levin (2007), Collard-Wexler (2008), Dunne, Klimek, Roberts, and Xu (2011), Ryan (2011), and Xu (2008).

in response to variation in agent payoffs. The per-period cost also serves as a measure of social inefficiency for each additional agent since entry mostly dilutes the business of incumbent agents without benefiting consumers. Our preferred estimate of the per-period cost is about 80% of agents' observed revenue.

We use the empirical model to investigate several counterfactuals, taking into account agent heterogeneity and behavioral responses. These counterfactuals provide a benchmark for measuring inefficiency. The first exercise keeps commissions fixed, but at a lower level. It simulates alternative market structure given assumptions about what appropriate agent compensation ought to be. Alternatively, the exercise may be interpreted as the impact of an across-the-board reduction in commissions, mandated by a regulatory agency.[4] If the commission rate is cut in half, entry decreases by a third and there would be 40% fewer agents, generating a social gain equal to 23% of total commissions. Productivity substantially increases: each agent sells 73% more houses and the average sales likelihood is 2% higher.

Rigid commissions generate a wedge between an agent's effort and his earnings for handling properties. Under free entry, if commissions were flexible, agents would be compensated at a level that reflects their costs of transacting properties. Since these costs are not directly observed, we benchmark them using agent commissions in 1998. This is a conservative upper bound because real house prices are lowest in 1998, technological improvements such as the internet have reduced intermediation costs, and agents are unlikely to work at a loss. When agents are compensated by the 1998 average commission, there would be 24% fewer agents, total commissions paid by households would reduce from \$4.15 billion to \$3.04 billion, and the social savings is equal to 13% of total commissions. This scenario also illustrates the business stealing effect: if there were no house price appreciation during our sample period, the average commission per agent would be \$59,700; by comparison, when house prices rose 1.5 times, it is only \$63,300.[5]

Under a fixed commission rate, it is difficult to distinguish good agents from mediocre ones based only on the prices they charge. Our last counterfactual investigates the implications of providing consumers additional information on agents' past records. This would allow them to be more responsive to agent performance when selecting a broker, which follows the FTC's recommendation for consumers (FTC 2006). Given the records in the Multiple Listing Service platform, implementing a system which evaluates agents or provides feedback on their performance would be relatively easy. According to our analysis, more information reduces incentives for inexperienced agents to enter, thereby shifting business towards experienced incumbent agents and yielding productivity gains and social savings.

It is worth noting that to incorporate realistic dynamics in our framework, it is necessary to push forward methodological boundaries associated with large state spaces in dynamic models. Another contribution of our paper, therefore, is to illustrate how dynamic discrete choice methods

[4] The FTC and the Department of Justice have had several investigations of rigid commissions in the brokerage market.

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

may be useful for analyzing the housing market, where a rich state space is natural.[6] The common approach in these models involves discretizing the state space, which faces the well-known "curse of dimensionality" problem when the state space is large.[7] Instead, we treat the state space as continuous, approximate the value function using basis functions, and cast the Bellman equation as a model constraint following Su and Judd (2008). A similar procedure makes the computation of counterfactuals relatively straightforward. An independent study by Michelangeli (2010) is the only other paper we are aware of that estimates a dynamic model with value function approximation, though with only one state variable. Our estimator falls into the class of sieve estimators surveyed by Chen (2007), and Barwick, Kristensen, and Schjerning (2012) derive its asymptotic properties. More broadly, our method for estimation and computing counterfactuals may be applicable in problems where avoiding discretization is advantageous. It is also less computationally demanding than existing approaches for our problem.

The remainder of the paper is structured as follows: Section 2 provides industry background and describes our data sources. Section 3 presents a descriptive analysis of Greater Boston's real estate brokerage industry. Section 4 develops our model and Section 5 outlines the estimation approach. Section 6 describes our empirical results, while Section 7 presents our counterfactual analyses. The last section states our conclusions. Barwick and Pathak (2012) (hereafter BP2) contains 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 states require real estate brokers and agents to be licensed. The requirements by the Massachusetts Board of Registration of Real Estate Brokers and Salespersons in 2007 appear minimal: 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,

[6] Bayer, McMillan, Murphy, and Timmins (2011) is another application of these methods for the housing market.

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

completion of thirty classroom hours of instruction, passing a written exam, and paying a surety bond of five thousand dollars. 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 ($93 for salespersons and $127 for brokers).[8] Given the general perception that these requirements do not create significant barriers, it is unsurprising that entrants account for a large share (about 13%) of active agents each year in our dataset.

2.2 Data

The data for this study come from the Multiple Listing Service (MLS) network for Greater Boston, a centralized platform containing information on property listing and sales. We collect information on all listed non-rental residential properties for all towns within a fifteen-mile radius of downtown Boston. There are a total of 18,857 agents and 290,738 observations.[9] The list of 31 markets are shown in Figure 1, where we group together some smaller towns and cities with few agents (details in BP2). 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, and 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 construct a measure 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. While our data does not contain commissions paid to listing agents, it does contain the commission rate 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 a 2007 survey conducted by the National Association of Realtors, most agents are compensated under a revenue sharing arrangement, with the median agent keeping 60% of his commissions and submitting 40% to his firm (Bishop, Barlett, and Lautz 2007). We subsequently discuss how the assumption of a 60%-40% split impacts our analysis.

[8] MA Division of Professional Licensure, Board of Real Estate Brokers and Sales Persons. Available at http://license.reg.state.ma.us/public/dpl_fees/dpl_fees_results.asp?board_code=RE, last accessed in August 2011.

[9] To verify MLS's coverage of transactions in our cities, 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 led us to exclude the city of Boston from the empirical 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, including those who simply buy and sell their own properties, we only keep agents with an average of at least 1.5 listings and purchases per year. This sample restriction leaves us with 10,088 agents listing 257,923 properties, about 90% of the original records. BP2 provides more details on the sample construction.

3 Descriptive Analysis

3.1 Summary statistics

Greater Boston's housing market exhibits significant time-series variation in the number of properties listed, the likelihood of sale, and sales prices during our sample period. Table 1 shows that the number of listings varies from 20,000 to 23,000 in the late 1990s and early 2000s, but increases to 32,500 in 2005. There is a sharp decline in the number of listed houses after the onset of the decline in aggregate economic activity in 2007. Housing market weakness in the latter part of the sample also appears in the fraction of properties sold: before 2005, 75-80% of listed properties are sold; in 2007, only 50% are sold. The average real sales price of homes is $385,900 in 1999 and $529,200 in 2004, and it falls to $489,800 in 2007. The amount of time it takes for properties to sell leads 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.

Since an agent's expected revenue depends on the number of properties listed, sales likelihood, and price, it is unsurprising that entry and exit patterns of agents follow these market-wide trends. Table 2A shows that the number of incumbent agents increases 50% from around 3,800 in 1998 to a peak of more than 5,700 in 2005. The number of agents who leave the industry is around 400-500 each year during the early period, but rises to 700-800 when housing market conditions deteriorate with fewer transactions and lower prices.

Agent performance is also related to overall trends in the housing market. During the early part of the 2000s, the number of properties each agent intermediates is about eight per year. By 2007, the average agent conducts a little more than five transactions. In addition, 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 75th percentile is four to six times that of the 25th percentile. During the down markets of 2006 and 2007, a large fraction of real estate agents are especially hit hard: more than 25% of the listing agents did not sell any properties at all. The within-agent standard deviation in the number of transactions is 4.44, which is sizeable given that the median number of transactions per agent is only 6. This suggests that most agents are not capacity constrained and could intermediate more properties if the opportunity arises.

Home sales, agent entry and exit, and agent performance also vary across markets within Greater Boston, as shown in Table 2B. Markets differ considerably in size. The largest five markets have four times as many transactions as the smallest five. The most expensive market in our sample is Wellesley, where the average house sold for more than $1 million. On the other end, in Randolph, the average sales price is $290,000. Quincy, a city with over 10,000 housing transactions, has significant turnover of agents: it is home to the most entries and 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 average house prices and the number of transaction per agent is around -0.43 across all markets.

An important component of performance differences among agents is their experience. Panel A of Table 3 reports the average annual commissions of agents based on the number of years they have 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 are active at the beginning of our sample. Most agents in the other categories (one to eight years of experience) 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 is modest.[10]

Performance differences are also closely related to agent skill. We measure skill based on the number of transactions an agent conducts 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. More skilled agents have higher sales probabilities, earn higher commissions, and a larger portion of their income comes from listing properties.

To examine performance variation over time, we assign the 1998 cohort (agents who were present in 1998) 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 earn $100,000 or more for most years, while the bottom quartile agents barely earn $30,000 in commissions, even during years of peak house prices. Moreover, agents in the top quartile earn significantly more than those in the second or third quartile. The gap in earnings between second and third quartile agents is much smaller and also compresses in down markets.

Earning differences 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 leave by 2007. In contrast, about three quarters of the agents in the

[10] Barwick, Pathak, and Wong (2012) show that performance differences are robust to controls for property attributes.

bottom quartile exit at some point during the ten-year period. Figure 3 presents our identification argument in a nutshell: variation in the exit rates of agents earning different commissions identifies the per-period cost of being a broker.

3.2 Descriptive regressions

We now turn to measuring the impact of competition on agent performance, adjusting for housing market conditions. This descriptive evidence informs the modeling choices we make in the next section. We measure competition between agents by counting the number of real estate agents working in the same market and year.

To estimate how agent performance is related to the competition he faces, we report estimates of agent performance, y_{imt}, for agent i in market m in year t from the following equation:

$$y_{imt} = \rho_y \log(N_{mt}) + \alpha s_{it} + \lambda_m + \tau_t + \theta_1 \log(H_{mt}) + \theta_2 \log(P_{mt}) + \theta_3 \text{INV}_{mt} + \epsilon_{imt}. \qquad (1)$$

N_{mt} is the number of competing agents, s_{it} is agent skill, proxied by the number of properties an agent intermediates in year $t-1$, following Table 3B.[11] λ_m and τ_t are market and year fixed effects. The parameter ρ_y reveals the impact of a percentage increase in the number of competing agents on the agent's performance.

An agent's performance depends on the underlying state of the housing market. Therefore, we include three market level controls to isolate the impact of competition relative to the housing market's overall state. These controls also feature in the structural model described in the next section. The first two are the total number of listed properties H_{mt}, which counts all houses for sale in year t and market m, and the average house price P_{mt}, which is the equal-weighted price of all houses that are sold in year t and market m. The third is the the inventory-sales ratio, INV_{mt}, defined as follows: for each month in year t, 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 12 months, averaged over all months in the year. This state variable has the greatest predictive power of whether properties sell, and is often cited by the National Association of Realtors when describing the state of the housing market.[12]

To avoid confounding changes in the composition of agents, we estimate equation (1) using a fixed cohort of agents, defined as the set of agents who are active or have entered as of a given initial year, and we follow them over time. For instance, the 1998 agent cohort includes all agents active in 1998. Agents who enter in subsequent years are excluded from the regression, but they contribute to the competition variable $\log(N_{mt})$.[13] We only report estimates of ρ_y for the first seven agent

[11] The estimates are similar using the number of years an agent has worked as a broker to measure skill as in Table 3A.

[12] See the NAR's research reports available at http://www.realtor.org/research/research/ehsdata, last accessed in August 2011.

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

cohorts (from 1998-2004), since the sample sizes are small for later ones. The patterns in Figure 2 imply that agents in bottom quartiles react to competition in different ways from the top quartile. This consideration motivates a variation on equation (1), where we allow for competition effects to be different for agents who are more established. Specifically, we assign each agent in a cohort to four groups according to his commission in his cohort year, and estimate equation (1) interacting ρ_y with a group dummy. We examine two measures of agent performance: log commissions and log number of transactions. The estimates are reported in columns (1)-(5) and columns (6)-(10) of Table 4, respectively.

Since agent entry is cyclical and our competition measure increases during a booming market when we expect properties to sell more quickly and at higher prices, we anticipate that our estimates of ρ_y are biased towards zero. Nonetheless, they are significantly negative and sizeable for almost all regressions we estimate, suggesting 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 4.6% decrease in average commissions and a 5.2% reduction in the number of transactions for the 1998 cohort. For the 2004 cohort, the impact is much larger: it leads to a 7.7% decrease in commissions and a 8.4% reduction in transactions. Across agent quartiles, the estimate is largest for the bottom quartile, and next for the third quartile, implying that competitors steal more business from the bottom tier agents, and have a smaller impact on agents in the top quartile.

Having documented a strong competition effect, we now examine whether home sellers benefit from more competition among agents. Let h_{imt} be a measure of the home seller's sales experience (the likelihood of sale, days on the market, or sales price) for a property intermediated by agent i working in market m in year t. We estimate property-level regressions of the form:

$$h_{imt} = \rho_h \log(N_{mt}) + \alpha s_{it} + \gamma' X_{i_h t} + \lambda_m + \tau_t + \theta_1 \log(H_{mt}) + \theta_2 \log(P_{mt}) + \theta_3 \text{INV}_{mt} + \upsilon_{imt}, \quad (2)$$

where N_{mt}, s_{it}, λ_m, τ_t, H_{mt}, P_{mt}, and INV_{mt} are defined as in equation (1). $X_{i_h t}$ represents a vector of attributes of the property listed by agent i 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. Table 5 reports estimates of ρ_h.

When the number of competing agents in a market increases, the likelihood that a property sells decreases, contrary to our prior that ρ_h is biased upward because more agents are 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 0.6% reduction in the sales probability (the average sales probability is 69%). A possible explanation for this negative coefficient is that with more agents in a booming market, sellers may list their property at a higher price to 'fish' for a buyer. A higher list price may indicate a more patient seller, so we include it as a control in column (2). The negative impact reduces to 0.4%, though this estimate is no longer significant. One might argue that our estimate ρ_h is negative because the composition of properties changes with market conditions: properties that are harder to sell (due to unobserved attributes) are more likely to be listed in a booming market.

We examine this possibility in two ways. First, we add property fixed effects to equation (2) and report estimates in column (3). The sample size is smaller because only 30% properties are listed more than once. Next, we interact the competition measure with indicators for before or after 2005 in column (4). The negative coefficient remains: 0.5% with property fixed effects, about 0.3% before 2005 and 0.4% post 2005 though none of the estimate is significant. Both approaches suggest that the reduction in sales probability with more agents is not driven by changes in property composition.

The impact of more competition on days on the market for sold properties is negative, but insignificant. 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 shown in columns (6), and when we also control for property fixed effects in column (7). Using either estimate, a 10% increase in competition generates a 0.1% increase in the sales price, which translates to roughly $700 for a typical home. Since a higher sales price is a transfer from buyers to sellers and has a negligible impact on aggregate consumer surplus, in subsequent discussions we do not focus on the impact agents have on sales prices.

Increased competition does not make it more likely that a given property will be sold or that it will sell faster. However, it is possible that competition among realtors may generate benefits in ways we cannot measure. For instance, enhanced competition may motivate agents to work harder at satisfying client requests. On the other hand, it may also cause them to spend more time marketing their services to attract clients rather than exerting effort to sell their listed properties. On net, our estimates indicate that these effects do not translate into consumer surplus for sellers and buyers. In the following sections, we turn to the task of quantifying the magnitude of inefficiency of excess entry under the assumption that consumer benefits from agent competition are modest.

4 Modeling Competition Between Agents

The patterns in the previous section show that competition centers on attracting listings and increased competition 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 per-period costs of working as brokers. 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 over time, 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 for three main reasons. First, we do not have access to 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. Second, including property-specific features in the state space substantially increases its dimension and therefore creates formidable challenges for estimation and counterfactual analyses that require solving for a new equilibrium. Third, as we further discuss in Section 5.2, structural parameters are identified from observed agent revenues and our parsimonious representation of the housing market allows for a reasonable fit of the data, so the marginal benefit from additional state variables seems modest.

We assume that agents' commissions are determined by two sets of payoff-relevant variables: agents' individual characteristics and aggregate variables. Individual characteristics include an agent's gender, firm affiliation, the number of years he has worked as a broker, and a count of his past transactions. The aggregate variables are H_{mt} (the total number of houses listed on the market), P_{mt} (average house prices), and $\textsc{Inv}_{mt}$ (the ratio of inventory-sales ratio) that are described in the previous section. We assume that these aggregate state variables transition exogenously. That is, we do not model potential feedback from agent entry to the aggregate housing market since it seems unlikely that this accounts for a significant fraction of housing market variation. Two other aggregate state variables measure the intensity of competition among agents and are discussed next.

4.2 Agent payoffs

Realtors earn commissions either from sales (as listing agents) or purchases (as buyer's agents) of homes. We model these two components of agent payoffs 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 characteristics (we omit the market subscript m throughout this subsection). 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. Since not all aspects of agent attributes are observed, we include the variable ξ_{it}^L to represent his unobserved quality (observed by all agents, but unobserved by the econometrician), as is commonly done in discrete choice models (e.g, Berry, Levinsohn, and Pakes (1995)). We report estimates assuming that ξ_{it}^L is independent across agents and time. In Section 6.1, we present evidence that our controls for agent skill mitigate the concern about correlated unobserved state variables. This assumption is needed because of computational difficulties involved in 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" (e.g. Aguirregabiria and Nevo (2010)). It is an aggregate state variable that measures the level of competition agents face in obtaining listings. Given the large number of brokers per market ($\geq$ 100), we assume that agents behave optimally against the aggregate competition intensity L_t, rather than tracking all rivals' decisions. Melnikov (2000) and Hendel and Nevo (2006) make similar assumptions. Without this assumption, agent identities and attributes would become state variables (as in many oligopoly models) and make the model intractable.

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

$$\Pr_{it}^{Sell} = \frac{\exp(X_{it}^S \theta^S)}{1 + \exp(X_{it}^S \theta^S)},$$

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 $R_{it}^{Sell} = r * H_t * P_t * ShL_{it} * \Pr_{it}^{Sell}$, where r is the commission rate, H_t is the aggregate number of houses listed, and P_t is the average price index.

The model for an agent's commissions from representing buyers is similar: $R_{it}^{Buy} = r * H_t^B * P_t * ShB_{it}$, where H_t^B is the total number of houses purchased by all home buyers, P_t is the average price index, and ShB_{it} is agent i's share of the buying market, with a similar expression as the listing share: $ShB_{it} = \frac{\exp(X_{it}^B \theta^B + \xi_{it}^B)}{\sum_k \exp(X_{kt}^B \theta^B + \xi_{kt}^B)}$. 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 buying side is $B_t \equiv \sum_k \exp(X_{kt}^B \theta^B + \xi_{kt}^B)$. Together, L_t and B_t are the aggregate state variables that measure the amount of competition agents face from rivals.

To reduce the number of state variables, we make the simplifying assumption that $H_t^B = 0.69 H_t$. In our sample, 0.69 is the average probability that houses are sold and the correlation between H_t^B and H_t is 0.94. 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 of a particular housing market.

Since agents earn commissions as both buyer's and seller's broker, agent i's revenue function is:

$$\begin{aligned} R(S_{it}) &= R^{Sell}(S_{it}) + R^{Buy}(S_{it}) \\ &= r * HP_t * (ShL_{it} * \mathrm{Pr}_{it}^{Sell} + ShB_{it} * 0.69). \end{aligned} \tag{5}$$

where $S_{it} = \{X_{it}^L, X_{it}^S, X_{it}^B, HP_t, \mathrm{INV}_t, L_t, B_t\}$. Despite this stylized representation of the housing market, the correlation between the model's predicted commission and the observed commission is 0.70. The model also captures the upward and downward trend of observed commissions. We provide details on the model's fitness in Section 6.5.

4.3 Transition process of state variables

When agents decide to enter or 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. In addition, most aggregate variables have a pronounced hump shape. We do not explicitly model agent's beliefs on how the aggregate state variables evolve. Following Aguirregabiria and Mira (2007), we adopt an AR(1) model but include a trend break before and after 2005, when house prices peaked in our sample, to capture the inverted U-shaped pattern.

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 S_{mt} is a vector of state variables, T_0 and T_1 are coefficient vectors of the trend break dummies, $1[\cdot]$ is an indicator function, T_2 is a matrix of autoregressive coefficients, α_m is the market fixed effect, and η_{mt} is a mean-zero multi-variate normal random variable. Market fixed effects in equation (6) are included to control for size differences across markets.

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 low, as we have only a few periods per market after 2005. Another alternative is to add 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. Finally, like the aggregate state variables, an agent's skill is also modeled as an AR(1) process, with a different constant before and after 2005.[14]

4.4 Entry and exit decisions

In the model, agents can make career adjustments each period: some incumbent agents continue to work as realtors, others leave the industry (exit), and new individuals become brokers (entry). At the beginning of a period, agents observe the exogenous state variables, their own characteristics,

[14] Market fixed effects are not included in the AR(1) process of agent skill. Adding them leads to little change in the autoregressive coefficient: 0.74 (with market fixed effects) vs. 0.75 (without market fixed effects).

as well as the 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 individuals become realtors and decrease when realtors quit and seek alternative careers. Agents observe their private idiosyncratic income shocks and simultaneously make entry and exit decisions.

Since agents start earning income as soon as they obtain listings, we assume that there is no delay between entry (becoming an agent) and earning commissions. This assumption contrasts with the literature on firm dynamics, which assumes that firms pay an entry cost at period t and start generating revenues in period $t+1$ after a delay from installing capital and building plants (e.g., Ericson and Pakes (1995)).

Let Z denote exogenous state variables and individual characteristics and Y denote the endogenous state variables L and B. An incumbent agent's decision making can be represented by the following Bellman equation:

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

where $E\left[R(Z_{it}, Y_t)|Z_{it}, Y_{t-1}\right]$ is his expected commission revenue conditional on observed state variables and δ is the discount factor. 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.[15] Since income shocks are private, agent i does not observe Y_t; it 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 per-period cost c captures agent i's costs of brokering house transactions. It includes the 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. More importantly, it includes his foregone labor income from working in an alternative profession, which captures his value of time. We assume that the cost of being a broker does not depend on the number of houses he handles. This is because the marginal *monetary* cost of handling more properties, such as gasoline expenses for showing properties and photocopying charges not covered by his firm, is likely swamped by the fixed costs. It is also necessary because the fixed costs and marginal costs cannot be separately identified, as we discuss in Section 4.5.[16] In all specifications, 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 agents in our sample. Relaxing this assumption would require estimating two value functions and substantially increase the complexity of the model.[17]

[15]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.

[16]In Section 7.5, we report counterfactual results under different assumptions on marginal cost.

[17]The estimation strategy would be similar, except that we need to use the exit choice probability to recast one

Private shocks $\tilde{\varepsilon}_0$ and $\tilde{\varepsilon}_1$ are 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 reduces to the usual log-sum form:

$$V(Z_{it}, Y_{t-1}) = \log\left[1 + \exp\left(\bar{R}(Z_{it}, Y_{t-1}, \beta) + \delta EV(Z_{it+1}, Y_t|Z_{it}, Y_{t-1})\right)\right], \tag{8}$$

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 simplify the notation. 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 exits at the end of period t is:

$$\Pr(\text{exit}_{it}|Z_{it}, Y_{t-1}, \beta) = 1 - \frac{\exp\left(\bar{R}(Z_{it}, Y_{t-1}, \beta) + \delta EV(Z_{it+1}, Y_t|Z_{it}, Y_{t-1})\right)}{1 + \exp\left(\bar{R}(Z_{it}, Y_{t-1}, \beta) + \delta EV(Z_{it+1}, Y_t|Z_{it}, Y_{t-1})\right)}. \tag{9}$$

The log likelihood for incumbent agents is:

$$LL(\beta) = \sum_{i,t} 1\left[\text{exit}_{it} = 1\right] * \log[\Pr(\text{exit}_{it}|\beta)] + \sum_{i,t} 1\left[\text{exit}_{it} = 0\right] * \log[1 - \Pr(\text{exit}_{it}|\beta)]. \tag{10}$$

Provided we can solve for EV and calculate the choice probability $\Pr(\text{exit}_{it}|\beta)$, we can estimate β by maximizing the sample log likelihood (10). In practice, solving EV with a large number of state variables is a difficult exercise. In Section 5, we explain in detail how we address this challenge.

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 (9), the probability of entry is:

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

---

of the choice-specific value functions as a fixed point of a Bellman equation as in Bajari, Chernozhukov, Hong, and Nekipelov (2009).

The log likelihood of observing N_t^E entrants out of a maximum of $\bar{N}^E$ potential entrants is:

$$LL^E(\beta) = \sum_{j \le \bar{N}^E,\, t} 1[\text{enter}_{jt} = 1] * \log[\Pr(\text{enter}_{jt}|\beta, \kappa)] + \sum_{j \le \bar{N}^E,\, t} 1[\text{enter}_{jt} = 0] * \log[1 - \Pr(\text{enter}_{jt}|\beta, \kappa)]. \tag{11}$$

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 (11) separately from equation (10). We report estimates of entry costs under three different assumptions on $\bar{N}^E$ in Section 6.3.

4.5 Discussion of modeling assumptions

The main parameter of interest is c, the average agent's per-period cost of working as a broker. The model does not allow c to depend on state variables because we only observe one action for each active agent (stay or exit). Therefore, we 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 a variable cost component which depends on the number of transactions because agent revenue is proportional to his total number of transactions.[18] 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 the National Association of Realtors 2007 agent survey, 79% of realtors report that real estate brokerage is their only source of income (Bishop, Barlett, and Lautz 2007). For 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 exiting the brokerage industry after his commission revenue reduces to \$30,000, then his foregone income is between \$30,000 and \$35,000 (ignoring the option value of future commissions), even though the value of his total working time is higher. Our estimate $\hat{c}$ correctly measures the average value of time that agents devote to being a broker.[19]

We do not endogenize the commission rate for a few reasons. First, we do not observe the commission rate paid to the listing agent. Second, 85% of buyer's agent commissions are either 2.0% or 2.5%. Third, commission rates are often determined by firms, with agents playing little, if any, role. Barwick, Pathak, and Wong (2012) document that the four largest firms intermediate roughly two-thirds of transactions and each sets its own policy on commissions.

Since we do not observe the actual contract terms between agents and their firms, we assume that agents keep 60% of total commissions. As mentioned in Section 2.2, a 2007 national NAR survey reports that the median split is 60%. Assuming that the buyer's agent and seller's agent evenly split the 5% commission, we fix the commission rate r in the revenue equation (5) at $1.5\% = 2.5\% * 60\%$. These assumptions affect our estimates proportionately: if the average commission is under-estimated by $\alpha\%$, then β_1 will be over-estimated by the same amount, and the per-period

[18] We report counterfactual results incorporating marginal costs in Section 7.5.

[19] We have tried to address part-time agents using a discrete mixture model that allows two types of agents with different per-period costs, but the likelihood is flat in a large region of parameter values.

cost $c = -\frac{\beta_2}{\beta_1}$ will be under-estimated by $\alpha\%$.

5 Solution Method

As explained in Section 4.4, the estimation of structural parameters β requires solving the unknown value function $V(\cdot)$ implicitly defined by the functional Bellman equation (8). The ability to quickly compute the value function is a crucial factor in most dynamic empirical models and in many cases is a determining factor in model specification. In our application, a reasonably realistic model of the housing market necessitates a rich set of state variables. Here we illustrate how we use sieve approximation combined with MPEC (mathematical programming with equilibrium constraints) to address challenges posed by a large number of state variables. BP2 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.

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. First, calculating the future value function $EV(S'|S)$, a high-dimensional integral of an unknown function using quadrature rules requires interpolation that is both slow and difficult to achieve a desirable accuracy. Second, the memory requirement of discretization increases exponentially.[20] Third, there are far fewer data points than the size of the state space. Discretizing the state space and solving the value function for the entire state space implies that most of the estimation time is spent solving value function $V(S)$ for states that are never observed in the data and hence not directly used in the estimation. Finally, both discretization and interpolation introduce approximation errors that grow with the number of state variables.

5.1 Sieve approximation of the value function

The alternative method we pursue approximates the value function $V(S)$ using sieves where unknown functions are approximated by parametric basis functions (e.g., Chen (2007)). 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 model constraint that has to be satisfied at the parameter estimates. This formulation reduces the computational burden significantly and makes it feasible to solve for the equilibrium of models with 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. There are two main downsides of our approach: a) the 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. BP2 presents Monte Carlo evidence showing that the method works well in our application: with a reasonable number of basis functions, the

[20] 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.

value function approximation error is small, the bias in parameter estimates is negligible, and the computation is very fast. We now present our solution algorithm.

Recall that our Bellman equation is:

$$V(S) = \log\left(1 + \exp\left[\bar{R}(S,\beta) + \delta EV(S'|S)\right]\right). \tag{12}$$

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.[21] This fact provides the theoretical foundation for using basis terms to approximate the value function $V(S)$.

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

$$V(S) \simeq \sum_{j=1}^{J} b_j u_j(S), \tag{13}$$

with unknown coefficients $\{b_j\}_{j=1}^{J}$. Substituting equation (13) into equation (12), we have:

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

This equation should hold at all states observed in the data. Our approach is to choose $\{b_j\}_{j=1}^{J}$ to best-fit this non-linear equation in "least-squared-residuals":

$$\left\{\hat{b}_j\right\}_{j=1}^{J} = \arg\min_{\{b_j\}} \left\| \sum_{j=1}^{J} b_j u_j(S_{(k)}) - \log\left(1 + \exp\left[\bar{R}(S_{(k)},\beta) + \delta \sum_{j=1}^{J} b_j Eu_j(S'|S_{(k)})\right]\right) \right\|_2 \tag{14}$$

where $\left\{S_{(k)}\right\}_{k=1}^{K}$ denotes state values observed in the data and $\|\cdot\|_2$ is the L^2 norm. Essentially, $\{b_j\}_{j=1}^{J}$ are solutions to a system of first-order conditions that characterize how changes in $\{b_j\}_{j=1}^{J}$ 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. In general, the best basis function is application specific and 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, which is the discounted sum of future revenues. Note that if the revenue function $\bar{R}(S)$ increases in S, and a large S is more likely to lead to a large state next period, then the value function $V(S)$ increases in S.[22] This property motivates using basis functions that fit the revenue function $\bar{R}(S)$ for our choice of $u_j(S)$. Since these basis

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

[22] The formal argument follows from the Contraction Mapping Theorem and is in the appendix.

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. Ideally, we want an adaptable procedure to economize on the number of terms to reduce numerical errors and parameter variance. We adopt the 'Multivariate Adaptive Regression Spline' (MARS) method popularized by Friedman (1991,1993) to find spline terms that approximate the revenue function to a desired degree.[23] Once we obtain a set of spline basis terms that best fit our revenue function $\bar{R}(S)$, we substitute them for $\{u_j(S)\}$ in equation (14).

To further simplify the computational burden of the estimation, we follow Su and Judd (2008), Dube, Fox, and Su (2009), as well as other applications using MPEC (mathematical programming with equilibrium constraints). Instead of solving $\{\hat{b}_j\}_{j=1}^{J}$ explicitly in each iteration of the estimation procedure, we impose equation (14) as a constraint to be satisfied by parameter estimates that maximize the sample log likelihood.

The number of spline terms J is an important component of estimation. We propose a data dependent 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 the non-parametric bootstrap).

5.2 Identification

Identification of β_1 and β_2 follows from the argument for a standard entry model: different exit rates at different revenue levels pin down per-period costs. If there is substantial exit following a moderate reduction in revenue, then the coefficient measuring sensitivity to revenue, β_1, is large. On the other hand, if exit varies little when revenue falls, then β_1 is small. The coefficient β_2 is identified from the level of revenue when exit occurs. Following Rust (1994), we cannot separately non-parametrically identify δ, so we plug in a range of values in estimation.

Identification of the value function and spline coefficients, b, follows from Hotz and Miller (1993), which shows 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 normalized to 0. Choice probabilities, therefore, directly identify the value function and the spline coefficients b.

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

6 Estimates

We first examine estimates of the revenue function and state variables' transition process, and then present per-period 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_{mt}, L_{mt}, and B_{mt} are standardized with zero mean and unit standard deviation; the skill variable s_{it} is standardized with zero mean and 0.5 standard deviation, because it is more skewed (as seen in Table 2). BP2 includes additional details and alternative specifications not presented below.

6.1 First-step estimates: Agent payoffs

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:

$$\ln ShL_{imt} - \overline{\ln ShL_{\cdot mt}} = (X^L_{imt} - \overline{X^L_{\cdot mt}})\theta^L + (\xi^L_{imt} - \overline{\xi^L_{\cdot mt}}) = (X^L_{imt} - \overline{X^L_{\cdot mt}})\theta^L + \tilde{\xi}^L_{imt}, \quad (15)$$

where $\overline{\ln ShL_{\cdot mt}} = \frac{1}{N}\sum_{i=1}^{N} \ln ShL_{imt}$. The other two averages are defined similarly.

We estimate equation (15) 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 exclude observations with 0 shares, or entrants and second-year agents since their s_{it} is either undefined or biased downward, and end up with 32,237 agent-year observations.[24]

Since agents with many past transactions are more likely to receive referrals and attract new customers, the number of transactions an agent intermediates in the previous year is an important predictor of listing shares. When s_{it} is the sole regressor, the R^2 of the listing-share regression shown in Table 6A is 0.44, a high value given the extent of 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 60%. 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 6A, which only uses s_{it} as a regressor; alternative specifications are reported in BP2.

Given that our proxy s_{it} cannot fully capture all aspects of an agent's skill, residuals $\tilde{\xi}^L_{imt}$ may 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 regress the residual estimate $\widehat{\tilde{\xi}}^L_{imt}$ on its lags. 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 suggests mean reversion.

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

We repeat the analysis with the Arellano-Bond estimator that accommodates agent fixed effects. The coefficient of lagged $\widehat{\tilde{\xi}}_{imt}^{L}$ is slightly larger in the absolute value but again with a negative sign: -0.15. This indicates the possibility of a "luck" component in agent performance: a good year is often followed by a bad year, rather than unobservable persistent attributes that induce positive serial correlations. Overall, these results suggest that our controls mitigate the issue of persistent unobserved attributes.[25]

Once we estimate the listing share equation, we compute the state variable

$$\hat{L}_{mt} = \sum_i \exp(X_{imt}^L \hat{\theta}^L + \widehat{\tilde{\xi}}_{imt}^L),$$

for all markets and periods. Since we cannot estimate $\tilde{\xi}_{imt}^L$ for agents with $ShL_{imt} = 0$, we replace missing $\tilde{\xi}_{imt}^L$ with the average $\tilde{\xi}_{imt}^L$ among agents with the same experience.[26] Results of the purchasing share are similar, with a 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:

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

where X_{imt}^S includes both aggregate state variables and agent attributes. Assuming whether listed properties are sold are independent events conditional on X_{imt}^S, the probability that agent i sells T_{imt} properties out of a total of L_{imt} listings is:

$$\Pr(T_{imt}|L_{imt}) = \binom{L_{imt}}{T_{imt}} \Pr(\text{sell}_{imt})^{T_{imt}} (1 - \Pr(\text{sell}_{imt}))^{L_{imt} - T_{imt}}.$$

We control for market fixed effects and report MLE estimates of θ^S in column (3) of Table 6A. A linear probability model delivers similar results. A standard deviation increase in the inventory-sales ratio reduces the probability of sales by 11%, while a standard deviation increase in s_{it} improves the probability of sales by 3%.

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

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

where S_{imt} denotes state variables $\{HP_{mt}, \text{INV}_{mt}, L_{mt}, B_{mt}, s_{it}, \text{ whether } t < 2005\}$. 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 the distributions we estimate in Section 6.2.

[25] Allowing for persistent unobserved attributes is an important and difficult question. See Norets (2009) and Imai, Jain, and Ching (2009) for some recent progress.

[26] Replacing missing $\tilde{\xi}_{imt}^L$ with zero leads to nearly identical estimates of L_{mt}.

6.2 First-step: Transition of state variables

The transition process of the four aggregate state variables $HP, \textsc{Inv}, L$, and B is described by equation (6). We estimate it using the Arellano-Bond estimator and include market fixed effects 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 forecast future state variables. Our estimate of market fixed effects is the average residual within each market during the ten-year sample period.

We add the lag of HP in Inv's autoregression because a large number of listings in the previous year is likely to generate 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 6B, 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.

The fifth state variable is agent i's skill s_{it}. We estimate various 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 6B) includes the lag of skill as well as trend-break dummies as regressors.

6.3 Second-step estimates: Structural parameters

As described in Section 5.1, we approximate the value function $V(S_{imt})$ by $\sum_j b_j u_j(S_{imt})$ and impose the Bellman equation as an equilibrium constraint to estimate the structural parameters. Let $\beta = \{\beta_1, \beta_{2m}\}_{m=1}^M$, so that the per-period cost $c = \frac{\beta_{2m}}{\beta_1}$ differs across markets as discussed in Section 4.4, and $b = \{b_j\}_{j=1}^J$. The constrained log-likelihood maximization is:

$$\max_{\beta,b} LL(S;\beta,b) \text{ such that}$$

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

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 (10).[27]

[27] 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

Since we use a data-dependent approach to determine the number of spline basis functions that approximate the value function, we estimate model parameters multiple times, with an increasing number of spline terms. The standard errors of these estimates are computed using 100 non-parametric bootstraps.[28] We start with 24 spline terms and add three terms at a time until parameter estimates stabilize, where the element by element difference between two adjacent sets of parameters $\left\{\hat{\beta}^{k}, \hat{\beta}^{k-1}\right\}$ is smaller than half of their standard deviation:

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

Our parameters stabilize when the number of spline terms increases to 39. We continue this process for several additional terms and verify that the difference between two adjacent sets of parameters continues to be less than half of the standard deviation, though it is often much smaller. Then we take the ratio of $\hat{\beta}_{2m}$ to $\hat{\beta}_1$ to compute per-period costs: $\hat{c}_m = -\frac{\hat{\beta}_{2m}}{\hat{\beta}_1}$. The standard errors of $\hat{c}_m$ are calculated from the empirical sample of the bootstrap estimates. We report $\hat{c}_m$, standard errors, the number of observations, and the number of spline terms in the first two columns in Table 7. $\left\{\hat{\beta}_1^k, \hat{\beta}_{2m}^k\right\}$ for different sets of spline terms are reported in BP2.

6.4 Results and robustness

There is a total of 41,856 agent-year observations. Each estimate in Table 7 is positive and significant at the 0.01 level. On average, the per-period 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. This variation is consistent with residents in richer towns having better outside options.

An important premise of our 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 column (3) of Table 7. For a third of the markets, the estimates are negative or insignificant; they average \$9,850 for the remaining markets. Compared to the Bureau of Economic Analysis (BEA) estimate of Massachusetts' per capita income of \$46,000 in 2006, these numbers appear to be too small and suggest that agents are not entirely myopic. Empirically, a myopic model has a difficult time explaining inertia in agents' decisions: they rarely exit as soon as they experience a negative income shock.

The model requires a number of strong assumptions which we now probe. First, the main estimates assume $\delta = 0.90$.[29] In columns (5)-(8), we examine how results change with different

nonlinear constraints, experiment with different starting values, and use 10^{-6} for all tolerance levels.

[28] In these bootstrap estimations, we hold estimates of the revenue function and state variables' transition process fixed, because re-estimating them in bootstrap samples for each set of spline basis terms requires recomputing all elements of the model and would take too long to compute.

[29] Other dynamic empirical papers using a discount factor of 0.9 include Ryan (2011) and Kennan and Walker (2011). In the 2007 NAR agent survey, the median age of a broker is 51 (Bishop, Barlett, and Lautz 2007).

discount factors. Everything else equal, a smaller $\delta = 0.85$ leads to a lower discounted stream of future income. To offset the change in δ, the model relies on smaller cost estimates, which could be interpreted as diminished payoff from an alternative career. The 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 cost is about \$56,300, or 15% larger than the original estimate.

Second, agent skill is measured by the number of transactions in the previous period 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 years of experience in columns (9)-(10). These estimates are similar to those in column (1), with an average of \$47,300. Despite the similarity in $\hat{c}$, this alternative measure of skill results in a much worse fit of the data. The sample log-likelihood is -14,645, compared with -12,883 in column (1). Using years of experience also reduces the model's fit of observed commissions considerably.

Third, there is a shift in the aggregate economy at the end of our sample, and it is possible that agent outside options are impacted by this change. In columns (11)-(14), we estimate two per-period costs per market, $c_{m,t\geq 2005}$ and $c_{m,t<2005}$.[30] All but one parameter is significant at the 0.01 level. Interestingly, per-period costs (which include foregone labor income from an alternative career) are generally higher prior to 2005, although the differences are significant for only a few markets. Mechanically, lower cost estimates after 2005 are driven by the fact that conditional on observed commissions, exit rates were actually *smaller* than those prior to 2005 indicating worse outside options. Results from our preferred specification (column (1)) are in general between $\hat{c}_{m,t<2005}$ and $\hat{c}_{m,t\geq 2005}$, and closer to $\hat{c}_{m,t<2005}$ on average.

Aside from the specifications reported in Table 7, we have estimated a large number of alternative models. For instance, we experiment with a common c across markets. The estimate is $\$41,300$, but the fit as measured by log-likelihood is considerably worse (-14,088 compared with -12,883 when c varies across market), and the difference between the observed and fitted exit probabilities is greater than 0.02 for more than half of the markets. We also experiment with revenue functions that control for both agent experience and skill. Our per-period cost estimates display consistent patterns across these specifications. Therefore they seem to be driven by entry and exit patterns observed in our data and are not sensitive to these particular choices involving the revenue functions. Our preferred specification is column (1), and is the basis of the discussions below.

We report entry cost estimates in Table 8 using three different assumptions about the maximum number of entrants $\bar{N}_m^E$. The first assumption is that $\bar{N}_m^E$ is equal to the largest number of entrants ever observed, $\max(N_{mt}^E)$, which has been used in other studies (e.g., Seim (2006)). The second assumes that $\bar{N}_m^E = 2 * \max(N_{mt}^E)$. Since markets with more listed houses attract more realtors, this motivates the third assumption that $\bar{N}_m^E$ is proportional to the average number of listings $\bar{N}_m^E = \frac{H_m}{25}$, where $H_m = \frac{1}{T}\Sigma_{t=1}^T H_{mt}$.[31]

[30] 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.

[31] We also experiment with several other measures, including $2 * \text{mean}(N_{mt}^E)$, $\frac{H_m}{10}$, and $\frac{H_m}{20}$. These results are

The entry cost κ, its standard deviation, and the probability of entry (defined as $\frac{N^E_{m,t}}{\bar{N}^E_m}$) are reported for each market for each assumption. Entry costs increase mechanically with the assumed number of potential entrants. Assuming that $\bar{N}^E_m = \max(N^E_{mt})$, the average entry cost is $18,000, which is roughly the commission for transacting 2.5 properties.[32] The other two assumptions, $\bar{N}^E_m = 2 * \max(N^E_{mt})$ and $\bar{N}^E_m = \frac{H_m}{25}$, lead to an estimate of $79,000 and $26,800, respectively. These numbers might seem high given the general perception of low entry barriers of the realtor brokerage industry. Therefore, we use the most conservative estimate of entry cost in the counterfactuals and report social savings with and without entry costs.

6.5 Model's fit

To judge the fit of the model and its suitability for counterfactuals, we now compare the model's predictions to information directly used in estimation and information that is not used in estimation. Since agents' commissions are the main force driving entry and exit, we begin by comparing observed and fitted revenues. These may differ because the model does not incorporate data on properties' physical attributes and is only based on measures of the aggregate housing market. Moreover, observed commissions are realized *ex post*, while fitted commissions are the *ex ante* revenue that agents expect to earn. Despite our simplifications, the correlation coefficient between R and $E(R)$ across agent-years is 0.70. The first two columns of Table 9 Panel A tabulate observed and predicted commissions by year. The model replicates both the upward and the downward trend in revenues. The average observed commission is $63,300, while the average fitted expected commission is $63,900. In Panel B, we report observed vs. predicted commissions by market; the differences are small for most markets.[33] These results indicate that our state variables are predictive of observed revenues.

Since exit choices identify agents' per-period costs, we next examine how the model fits exit probabilities. Columns (3) and (4) in Table 9 compare the observed vs. fitted probability of exit by year (Panel A) and by market (Panel B), respectively. Across years, we observe that 12% of incumbents exit, a rate which climbs to 15% in down markets and drops to 10% in up markets. The difference between predicted and observed exit rates is smaller than 0.01 for all years except 2005 when it is 0.02. The model also captures the U-shaped pattern of exit probabilities during our time period. Across markets, we observe greater variation in exit rates, ranging from 8% to 16%. Panel B of Table 9 shows that the model closely approximates market-level exit rates, with the difference between the model's prediction and the observed exit rate smaller than 0.01 except for a few markets. This tight relation is likely driven by the market-specific cost parameter, $\hat{c}_m$, though the fit is notable given the model's nonlinear structure and equilibrium constraints.

Finally, to benchmark our estimate of costs $\hat{c}_m$, we compare it with measures that are not in our

available upon request.

[32]Three markets have negative entry costs, but these are necessary to justify the high entry rates observed in these markets.

[33]The largest gaps are $10,000 for Arlington and $7,000 for Revere. The discrepancy between the observed and predicted L and B for these two markets contributes to the large gaps.

dataset. By construction, part of the per-period cost consists of what agents would have earned in an alternative profession, a counterfactual concept that is never observed empirically. The closest measure we find is each city's 2007 median household income from http://www.city-data.com/.[34] Figure 4 plots the estimated cost $\hat{c}_m$, from the smallest to the largest, together with the median household income for each market in our sample. Costs are lower in poor cities and higher in rich ones. The correlation coefficient between $\hat{c}_m$ and the median household income is 0.74. On average, an agent's per-period cost is slightly higher than half the median household income. The high correlation coefficient and comparable magnitudes imply that the cost estimates are sensible and are a suitable ingredient for counterfactuals.

7 Measuring Inefficiency

So far, we have focused on measuring an agent's per-period cost of working as a broker. To quantify the extent of inefficiency in the current brokerage market, it is necessary to specify the alternative. Our estimate of the per-period cost factors into this counterfactual analysis because it impacts the decisions of agents to enter or exit when their payoffs change.

Social inefficiencies in the brokerage market are partly driven by the peculiar feature that agents earn a fixed fraction of the sales price regardless of the amount of effort involved in selling properties. We consider three alternatives for measuring inefficiency. The first approach takes the fixed commission rate as given, but lowers the level for all agents. This approach allows us to benchmark the extent of inefficiency based on alternative assumptions on appropriate commission rates. The second approach examines the consequences of commission competition where agent compensation reflects the costs of conducting transactions. Finally, we analyze changes in the existing system when consumers have more information about the past performance of agents.

Before presenting the results, we describe how we solve for the counterfactual equilibrium. When payoffs change, agents respond immediately since there are no adjustment costs in the model. As a result, we do not consider out-of-equilibrium dynamics. Readers only interested in results can proceed directly to the next subsection.

7.1 Methodology

The main issue in simulating counterfactuals involves finding the new transition process of L and B. These two endogenous state variables are determined by all agents' joint entry and exit decisions. In estimation, we obtain their transition process directly from data; in a counterfactual, we need to solve for the equilibrium transition process 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 beliefs about the distribution of L' and B' based on current state variables, agents individually solve the new Bellman equation and choose an optimal decision.

[34]This resource aggregates information on population and income of U.S. cities and provides an income estimate for years between those of the U.S. decennial census. Last accessed in August 2011.

These decisions then jointly determine the distribution of L' and B'. Given the large number of agents (≥100 per market) and the assumption of i.i.d. private random shocks $\{\varepsilon_{i0,}\varepsilon_{i1}\}$, the central limit theorem justifies our approximation of the distribution of L' and B' (conditional on current state S) by a normal random variable with two parameters: the mean and the variance.

Agent beliefs are consistent when they correspond to 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 mean determined by the following equation:

$$\begin{aligned} E(L'(S)) = & \sum_i E\left[1\left\{\tilde{R}(S_i,\beta) + \delta E_{L',B'}\left[V(S_i')|S_i\right] + \varepsilon_{i1} > \varepsilon_{i0}\right\}\right]\exp(\overline{X_i^L\theta^L}) \\ & + \sum_{j\le \bar{N}^E} E\left[1\left\{-\kappa + \tilde{R}(S_j,\beta) + \delta E_{L',B'}\left[V(S_j')|S_j\right] + \varepsilon_{j1} > \varepsilon_{j0}\right\}\right]\exp(\overline{X_j^L\theta^L}), \end{aligned} \tag{16}$$

where the first and second term sums over incumbents and potential entrants, respectively.[35] $\tilde{R}(S_i,\beta)$ 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 L', B', and other exogenous state variables, and $\exp(\overline{X_i^L\theta^L}) \equiv E_\xi\left[\exp(X_i^L\theta^L + \xi_i)\right]$. Equation (16) is similar to equation (4), except that the former takes expectation over agents' entry and exit decisions.

Replacing $E\left[1\left\{\cdot\right\}\right]$ with choice probabilities and omitting the dependence of L' on S, we have:

$$E\left(L'\right) = \sum_i \Pr(\text{active}_i; L', B')\exp(\overline{X_i^L\theta^L}), \tag{17}$$

where we write $\Pr(\text{active}; L', B')$ to emphasize that an agent's optimal decision depends on his belief about future competition intensity L' and B'. The variance of L' is:

$$\text{Var}(L') = \sum_i \Pr(\text{active}_i; L', B')(1 - \Pr(\text{active}_i; L', B'))\exp^2(\overline{X_i^L\theta^L}). \tag{18}$$

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 (16) has a unique fixed point when L and B are normally distributed. To solve the new equilibrium for each market m, we use the MPEC framework to cast

[35] The transition process of exogenous state variables remains unchanged throughout the counterfactual exercises.

the counterfactual analysis as another problem of constrained optimization:

$$\min_{\substack{E(L'),\text{Var}(L'),\\ E(B'),\text{Var}(B')}} \left\| \begin{array}{c} E(L'_1) - \sum_i \Pr(\text{active}_{i1}; L', B') \exp(\overline{X^L_{it}\theta^L}) \\ \text{Var}(L'_1) - \{\sum_i \Pr(\text{active}_{i1}; L', B')(1 - \Pr(\text{active}_{i1}; L', B')) \\ \exp^2(\overline{X^L_{i1}\theta^L})\} \\ \vdots \\ E(B'_T) - \sum_i \Pr(\text{active}_{iT}; L', B') \exp(\overline{X^B_{iT}\theta^B}) \\ \text{Var}(B'_T) - \{\sum_i \Pr(\text{active}_{iT}; L', B')(1 - \Pr(\text{active}_{iT}; L', B')) \\ \exp^2(\overline{X^B_{iT}\theta^B})\} \end{array} \right\|$$

$$\text{such that} \quad \{b_j\}_{j=1}^J = \arg\min \left\| \begin{array}{c} \sum_{j=1}^J b_j u_j(S) - \\ \log\left(1 + e^{\tilde{R}(S,\hat{\beta}) + \delta \sum_{j=1}^J b_j E_{L',B'}\{u_j(S'|S)\}}\right) \end{array} \right\|_2 . \qquad (19)$$

All standard errors of the counterfactuals are calculated using 100 nonparametric bootstrap simulations.

7.2 Inefficiency relative to lower commissions

Several recent developments in the brokerage industry imply downward pressure on the current commission rate. For instance, there has been an increasing interest in using non-traditional methods to buy and sell properties (e.g., Levitt and Syverson (2008b)). 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: how much inefficiency does the current market structure create relative to an alternative benchmark with lower commissions? Alternatively, this exercise examines the implications of a regulator mandating a lower commission rate. The long history of antitrust investigations for coordination of realtor fees motivates our analysis of the consequences of this price regulation. For example, 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.[36] We simulate the brokerage industry using ten different commission rates varying from 2.5% to 4.75% and report results in Table 10.

Lower commissions lead to fewer agents and higher productivity per agent. Currently, the average annual number of entrants and agents across markets is 23 and 154, respectively. The average number of transactions per agent is 7.78 and a typical agent earns $63,300 per year. When the commission rate is cut in half, the number of entrants declines by 31% and the number of incumbents drops to 91, which is only 59% of what is observed in the sample. Agent productivity increases by 73%, with a typical agent conducting 13.5 transactions annually. A 50% reduction in

[36] 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.

commission rate only leads to a moderate change in the revenue per agent – $54,400 vs. $63,300 – since fewer rivals partially offset the decline in the commission rate. The average sales probability increases by 2% because the remaining agents are more experienced and better at striking a deal.

The most noteworthy finding is that the magnitude of social savings relative to the current market structure is substantial. Savings in per-period cost amounts to $863 million, or 22% of total commissions paid by households during the same period. Using our most conservative estimate of entry costs, savings in entry cost are sizeable at about $36 million. The last column of Table 10 presents the reduction in commissions paid by households. Since commissions are transfers from households to realtors, these figures do not constitute increases in social surplus. However, if we care about distributional effects, these benefits are roughly twice the size of the social cost of excessive entry. For example, when the commission rate is reduced by half, households would save about $2 billion in commissions. Results for other commission rates shown in Table 9 are similar, though of smaller magnitudes.

What do our estimates imply for inefficiency in the nationwide brokerage industry? Total commissions in the mid 2000s in the United States are $100 billion annually according to the Bureau of Economic Analysis (NIPA 2008). A rough back-of-the-envelope calculation suggests that a one-half reduction in the commission rate nationally decreases per-period costs and entry costs by as much as $23 billion. Despite the many caveats with extrapolating our results to the national level, these magnitudes imply that policies encouraging lower commissions could generate substantial social savings.

While our simulation rests on the assumption that agents are not capacity constrained, we believe that this assumption is defendable for the situations we have 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 the NAR nearly doubled from 1998 to 2006, the number of national home sales only increased by 30% (Bishop, Barlett, and Lautz 2007). These patterns suggest that most agents are not capacity constrained and that a 40% reduction in the number of agents is unlikely to 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 agent 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 five or more listings at the same time and a property takes 12 weeks to sell, then a broker could conduct at least 20 transactions per year. At that level, the optimal fixed commission rate would be considerably lower than 2.5%.

7.3 Inefficiency relative to competitive commissions

The next issue we consider is what the market structure would be if there were more competition on commission rates, so that agent compensation reflects the costs of intermediating properties rather than a fixed rate. To measure the market structure under this scenario, it is necessary to have an estimate of the cost of intermediating properties.

Since these costs are not directly observed, we benchmark them using agent commissions per property in 1998, which is \$11,580, the lowest across our sample years. This estimate is conservative for two main reasons. First, while house prices experience significant appreciation during our sample period, the cost of selling houses has likely decreased with new technologies. For example, in a recent survey of California Home Buyers (CAR 2008), 78% used the internet as an important part of their home-buying process. Second, if agents do not work at a loss, then their costs are bounded above by their compensation.

The counterfactual results quantify social inefficiency of the existing system relative to this conservative representation of competitive commission rates. Panel A of Table 11 shows that if agents had been compensated at the 1998 level from 1999 to 2007, there would be 24% fewer agents, each facilitating 31% more transactions. Total commissions paid by households would fall by \$1.1 billion, and the social savings in per-period costs and entry costs would be \$525 million. The magnitude of social savings is 13% of industry revenue or 21.6% of overall agent earnings.

Since commissions in this counterfactual do not increase with housing prices, another way to interpret our exercise is that it measures the impact of the housing boom on the brokerage industry. Our results suggest that realtors profit little from house price appreciation, because of business stealing from new entrants: if there were no house price appreciation during our sample period, the average commission per agent would be \$59,700, a modest reduction from \$63,300 when house prices have risen 1.5 times.

7.4 Reducing inefficiency by providing information on agent performance

So far our exercises are mainly concerned with quantifying inefficiency relative to different assumptions about commissions. Now we turn to the following question: within the existing structure, are there simple policies to improve the efficiency of the brokerage market? With fixed commissions, households cannot use commissions rates to distinguish good agents from mediocre ones. Many rely on referrals, which are often subjective and can be 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). This motivates considering an agent experience rating program, which makes past agent performance public or uses this information to certify agents with different skill levels. With more information, consumers may be more responsive to agent quality. To investigate the implications, we simulate the model raising the skill coefficient by increments of 20% to 100%. A larger coefficient represents higher premiums to an agent's skill and shifts more business to skilled agents. Doubling the skill coefficient, for example, increases the revenue of the 75th percentile agent by 52% while reduces the revenue of the 25th percentile agent by 25%.

Panel B of Table 11 shows that when the skill coefficient doubles, entry declines by 34% and exit decrease from 18 to 15. On net, the number of active agents drops from 154 to 127, a 17%

reduction. Agent productivity increases by 23%, the average sales probability improves by 2%, and the average commission income rises from $63,300 to $77,200. Large skill coefficients diminish entry, dampen the business stealing effect, and raise revenue for skilled agents. There is no direct benefit to consumers (except for a higher sales probability), but total cost savings are still a sizeable $372 million. The other rows report results when we increase the skill coefficient from 20% to 80%. The economic forces push in the same direction, though the magnitudes are smaller.

Relative to regulations on commission rates, this proposal has the advantage of easy implementation using the information that is directly available in the Multiple Listing System. Moreover, our estimates suggest that it may have the support of incumbent agents whose average commissions would increase substantially.

7.5 Sensitivity of counterfactuals

Our methodology makes it relatively straightforward to see how sensitive the counterfactuals are to our modeling choices. We have repeated Table 10 and 11 using the other specifications discussed in Section 6.4. When agents are less forward-looking ($\delta = 0.85$), cost savings are roughly 10-12% less, while if agents are more forward-looking ($\delta = 0.95$), cost savings are 13-15% more. If cost estimates differ before and after 2005, total cost savings are $880 million when the commission rate is cut in half, compared to $899 million in Table 10.

One concern with our estimates of social cost savings is that the marginal monetary cost of intermediating housing transactions is assumed to be zero. 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. Incorporating these variable costs leads to lower social cost savings. On the other hand, with these costs factored in, the net earnings of agents decrease. Lower earnings lead to more exits, which translates into higher cost savings.

To address this issue, we simulate our model assuming the marginal monetary cost is $500 per transaction, which includes expenses such as gasoline and office supplies not paid for by the firm.[37] Using agent i's observed number of transactions in our MLS data, we first calculate his fixed cost (FC_{imt}) by 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 and T_{imt} is the number of transactions by agent i. Then we recompute the counterfactuals where agents incur both a fixed cost and a variable cost of buying or selling properties.

The two countervailing forces largely cancel each other out. For example, when the commission rate is reduced by half, there are on average 86 active agents with a $500 marginal cost, instead of 91 agents as reported in Table 10. Total cost savings are similar: $902 million (with marginal costs) vs. $899 million (without marginal costs). Counterfactual results assuming marginal cost is

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

$100 display a similar modest impact on our estimates of cost savings.

8 Conclusion

In this paper we use a new dataset to document stylized facts of realtor entry and exit in Greater Boston. Traditional arguments suggest that if the production process involves fixed costs, free entry might be socially inefficient. However, inefficiencies could be outweighed by benefits to consumers if free entry brings more variety, better products, or lower prices. We find that agents are differentiated by experience. New entrants compete by taking listings from bottom tier incumbents, but there is no evidence that consumers benefit from enhanced competition associated with entry either on sales probability or time to sale. Furthermore, most agents do not appear capacity constrained, so that alternative market structures with fewer agents seem unlikely to impact the total number of properties brokered.

Although the model we develop is stylized, it is able to match key moments of the data and hence may be useful for predicting counterfactual market structures corresponding to changes in agent payoffs. Each of the three scenarios we investigate – lower commissions, commissions based on break-even costs, or improved information about agents' past performance – indicate large social costs with the current fixed percentage commission regime. Even though each of these counterfactuals suggests that fewer agents may be preferable, it is worth emphasizing that explicit entry restrictions may create adverse effects that are not captured by the model.

Another contribution of this paper is to build and extend a small, but growing literature using new computational and econometric tools to study industry dynamics. Assuming that agents are entirely myopic provides a poor fit of their entry and exit behavior and, as a result, quantitative statements of inefficiency should consider forward-looking behavior. Moreover, the richness of the housing market necessitates working with a large state space. Our approach of approximating the value function and treating the Bellman equation as an equilibrium constraint shows that it is possible to estimate parameters and compute counterfactuals in dynamic models with large state spaces.

There are other benefits associated with lower or flexible commissions that are not captured by the 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 topic for future work is understanding forces that could lead to more flexible commission rates.

References

Aguirregabiria, V., and P. Mira (2007): "Sequential Estimation of Dynamic Discrete Games," *Econometrica*, 75, 1–53.

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

Bajari, P., L. Benkard, and J. Levin (2007): "Estimating Dynamic Models of Imperfect Competition," *Econometrica*, 75(5), 1331–1370.

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

Barwick, P. J., D. Kristensen, and B. Schjerning (2012): "Maximum Likelihood Estimation of Discrete Markov Decision Models by Simulations and Sieves," Work in progress, UCL.

Barwick, P. J., and P. Pathak (2012): "Supplemental Material to The Cost of Free Entry: An Empirical Study of Real Estate Agents in Greater Boston," Working paper, MIT.

Barwick, P. J., P. Pathak, and M. Wong (2012): "Market Power and the Commission Puzzle," Working paper, MIT.

Bayer, P., R. McMillan, A. Murphy, and C. Timmins (2011): "A Dynamic Model of Demand for Houses and Neighborhoods," NBER Working paper, 17250.

Bernheim, D., and J. Meer (2008): "Do Real Estate Agents Add Value When Listing Services are Unbundled?," NBER Working paper, 13796.

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

Bishop, P. C., H. Barlett, and J. Lautz (2007): "National Association of Realtors, Member Profile," Available from National Association of Realtors.

CAR (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.

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," forthcoming, *Econometrica*.

Dunne, T., S. D. Klimek, M. J. Roberts, and D. Y. Xu (2011): "Entry, Exit and the Determinants of Market Structure," NBER Working paper 15313.

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," forthcoming, *Rand Journal of Economics*.

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

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, last accessed: August 2011.

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., and A. Nevo (2006): "Measuring the Implications of Sales and Consumer Inventory Behavior," *EMA*, 74, 1637–1674.

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.

Imai, S., N. Jain, and A. Ching (2009): "Bayesian Estimation of Dynamic Discrete Choice Models," *Econometrica*, 77, 1665–1682.

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

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

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

Kennan, J., and J. Walker (2011): "The Effect of Expected Income on Individual Migration Decisions," *Econometrica*, 79(1), 211–251.

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.

Mankiw, G., and M. Whinston (1986): "Free-Entry and Social Inefficiency," *Rand Journal of Economics*, 17(1), 48–58.

Melnikov, O. (2000): "Demand for Differentiated Durable Products: The Case of the U.S. Computer Printer Market," manuscript, Department of Economics, Yale University.

Michelangeli, V. (2010): "Does it Pay to Get a Reverse Merger?," Working paper, Congressional Budget Office.

NIPA (2008): "National Income and Product Accounts Table 5.4.5. Private Fixed Investment in Structures by Type. 1929-2008.," Bureau of Economic Analysis, Available at: http://www.bea.gov/national/nipaweb/index.asp.

Norets, A. (2009): "Inference in Dynamic Discrete Choice Models with Serially Correlated Unobserved State Variables," *Econometrica*, 77, 1665–1682.

Rust, J. (1994): "Estimation of dynamic structural models, problems and prospects: discrete decision processes," in *Advances in Econometrics: Sixth World Congress*, ed. by C. Sims. Cambridge University Press.

Rutherford, R. C., T. M. Springer, and A. Yavas (2005): "Conflicts between Principals and Agents: Evidence from Residential Brokerage," *Journal of Financial Economics*, 76, 627–665.

Ryan, S. (2011): "The Costs of Environmental Regulation in a Concentrated Industry," forthcoming, *Econometrica.*

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.

Su, C.-L., and K. Judd (2008): "Constrained Optimization Approaches to Estimation of Structural Models," forthcoming, *Econometrica.*

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_{it} is a vector of agent i's characteristics, including demographics, past experience, firm affiliation; $\xi_{i,t}$ represents agent i's unobserved quality; and τ_{hit} is the i.i.d. 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 i.i.d. extreme value random variables, agent i's listing sharing is:

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

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 $S = \{HP, \textsc{Inv}, -L, -B, s\}$ 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:

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

and f is the density of S'. It is straightforward to show that the operator

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

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 $\Gamma\left[C'\right] \subseteq C''$, 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, $R(S)$ 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 $\Gamma\left[C'\right] \subseteq C''$, 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:

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

where f_ϵ are the density of ϵ. 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 (16) under normality

Given a vector of state variables S, we can approximate L' by a normal random variable:

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

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):

$$\begin{aligned} \mu &= E(L') = \sum_i \Pr(\text{active}_i)e^{\overline{X_i^L\theta^L}} \\ &= \sum_i \frac{e^{R_i(\mu)+\beta\int V(L')f(L';\mu)dL'}}{1+e^{R_i(\mu)+\beta\int V(L')f(L';\mu)dL'}}e^{\overline{X_i^L\theta^L}}. \end{aligned} \tag{20}$$

We first show that the right-hand-side of this equation is strictly monotonic in μ. Since expected profits reduce with more competition, $R_i(\mu)$ strictly 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 strictly decreases in μ. Hence, equation (20) has at most one fixed point. When μ approaches 0 (so that few agents are active), $\Pr(\text{active}_i)$ approaches 1, so the right-hand-side of equation (20) exceeds μ. When μ approaches ∞, the right-hand-side of equation (20) is bounded above by $\sum_i e^{\overline{X_i^L\theta^L}}$ and, hence, is smaller than μ. Therefore, there is a unique fixed point.

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

Year	Properties (1000s)		Sales Price ($1000s)		Days on Market		Total Commissions ($mil)
	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: Numbers include all properties in the Multiple Listing Service listed and sold by 10,088 agents in the Greater Boston Area. List of towns is in the supplemental material. All prices are 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	Entrants	Incumbent Agents	Exiting Agents	Number of Properties Sold	Sales per Listing Agent			Purchases 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 does not work in the previous year (either as a listing or a buyer's agent), an incumbent is one who works 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

Market	Average Sales Price ($1000s) (1)	Entering Agents (2)	Incumbent Agents (3)	Exiting Agents (4)	Number of Properties Sold (5)	Sales per Listing Agent (6)	Purchases 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 does not work in the previous year (either as a listing or a buyer's agent), an incumbent is one who works as an agent in the year, and an exiting agent does not work in subsequent years. All sales prices are in 2007 dollars, deflated using the BLS's urban CPI.

Table 3. Days on Market, Sales Probability, and Commissions by Agent Experience and Skill

		Sales Probability		Days on Market		Commissions ($1000s)		Listing Commissions ($1000s)		Sales Commissions ($1000s)	
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 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 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 from the 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.

Table 4. 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	-0.46***	-0.44***	-0.51***	-0.58***	-0.63***	-0.52***	-0.50***	-0.56***	-0.61***	-0.65***
	(0.08)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
1999	-0.45***	-0.39***	-0.47***	-0.53***	-0.59***	-0.52***	-0.47***	-0.52***	-0.57***	-0.63***
	(0.08)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
2000	-0.56***	-0.49***	-0.57***	-0.63***	-0.69***	-0.62***	-0.57***	-0.63***	-0.68***	-0.73***
	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
2001	-0.70***	-0.59***	-0.67***	-0.74***	-0.80***	-0.72***	-0.63***	-0.69***	-0.75***	-0.80***
	(0.09)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
2002	-0.67***	-0.58***	-0.66***	-0.72***	-0.78***	-0.68***	-0.61***	-0.67***	-0.73***	-0.77***
	(0.10)	(0.10)	(0.10)	(0.10)	(0.10)	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)
2003	-0.73***	-0.57***	-0.65***	-0.73***	-0.77***	-0.75***	-0.62***	-0.69***	-0.75***	-0.79***
	(0.12)	(0.11)	(0.11)	(0.11)	(0.11)	(0.11)	(0.10)	(0.10)	(0.10)	(0.10)
2004	-0.77***	-0.51***	-0.60***	-0.68***	-0.74***	-0.84***	-0.62***	-0.69***	-0.76***	-0.82***
	(0.16)	(0.15)	(0.15)	(0.15)	(0.15)	(0.14)	(0.13)	(0.13)	(0.13)	(0.13)

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, agent skill, log of the aggregate number of listed properties, log of house price index, inventory-sales ratio, and 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 row is estimated using a cohort, defined at the set of agents active or entering in the initial cohort year. The sample excludes 1,631 agent-years without transactions (out of 47,083, or 3%).

Table 5. Impact of Competition on Property Sales

	Sales Probability				log(Days on Market)				log(Sales Price)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Log(Nmt)	-0.059**	-0.036	-0.051		-0.028	-0.046	-0.044		0.118***	0.012***	0.014***	
	(0.027)	(0.026)	(0.064)		(0.093)	(0.094)	(0.181)		(0.027)	(0.003)	(0.005)	
Log(Nmt) Before 2005				-0.025				-0.022				0.013***
				(0.030)				(0.092)				(0.003)
Log(Nmt) After 2005				-0.040				-0.055				0.012***
				(0.025)				(0.091)				(0.003)
Property Fixed Effects	N	N	Y	N	N	N	Y	N	N	N	Y	N
Listing Price	N	Y	Y	Y	N	Y	Y	Y	N	Y	Y	Y
R^2	0.09	0.10	0.16	0.10	0.12	0.12	0.16	0.12	0.86	0.99	0.995	0.99
N	239462	239252	116707	239252	171314	171217	52812	171217	171228	171212	52842	171212

Note: * significant at 10% level, ** significant at 5% level, and *** significant at 1% level. Each cell reports coefficient 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, agent skill, and the housing state variables pf log of the aggregate number of listed properties, log of house price index, and inventory-sales ratio. Column (3) uses properties that are listed at least twice. Column (7) and (11) use properties that are sold at least twice. Robust standard errors clustered by market.

Table 6A. 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)
Year < 2005			1.27***
			(0.03)
Year >= 2005			0.83***
			(0.03)
Market FEs	No	No	Yes
Estimation Method	OLS	OLS	MLE
R^2 adjusted	0.44	0.30	0.18
N	32237	30986	32237

Note: * significant at 10% level, ** significant at 5% level, and *** significant at 1% level. Skill measures an agent number of transactions in the previous year. Inv is the inventory-sales ratio. Dependent variables in columns (1) and (2) are demeaned log shares as explained in text. R^2 adjusted for MLE is pseudo adjusted R^2 (computed by authors). Entrants and agents with 0 shares are excluded.

Table 6B. 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)
Year < 2005	0.29***	-0.10**	0.03*	0.04**	0.04**
	(0.03)	(0.05)	(0.02)	(0.02)	(0.00)
Year >= 2005	0.17***	0.62***	0.12***	0.09***	0.00
	(0.05)	(0.07)	(0.03)	(0.03)	(0.00)
Market FEs	Yes	Yes	Yes	Yes	No
Estimation Method	GMM-IV	GMM-IV	GMM-IV	GMM-IV	OLS
R^2 adjusted	0.93	0.77	0.96	0.96	0.59
N	279	279	279	279	30648

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 inventory-sales 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. GMM-IV refers to the Arellano-Bond estimator. R^2 adjusted for GMM-IV is pseudo adjusted R^2 (computed by authors). Entrants as well as agents with 0 shares are excluded in column (5).

Table 7. Per-Period Cost Estimates (in $100,000 2007 Dollars)

	Main Specification (δ=0.90)		Variations on Main Specification											
			Assumptions on Forward-Looking Behavior						Years of Experience as Skill Measure		Two Cost Parameters Per Market			
			δ=0 (myopic)		δ=0.85		δ=0.95							
	C	std(C)	C	std(C)	C	std(C)	C	std(C)	C	std(C)	$C_{t<2005}$	std($C_{t<2005}$)	$C_{t\geq 2005}$	std($C_{t\geq 2005}$)
	(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)
INV	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)
Spline Terms	39		NA		39		39		39		39			
Log-Likelihood	-12883		-12819		-12892		-12875		-14645		-12779			
N	41856		41856		41856		41856		41856		41856			

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

Table 8. Entry Cost Estimates (in $100,000 2007 Dollars)

	Potential entrants=max(N^E)			Potential entrants=2*max(N^E)			Potential entrants=H/25		
	κ	std(κ)	Entry Probability	κ	std(κ)	Entry Probability	κ	std(κ)	Entry Probability
	(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
INV	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: * significant at 10% level, ** significant at 5% level, and *** significant at 1% level. Parameter standard errors are estimated via 100 bootstrap simulations. Maximum number of potential entrants equal to maximum number of observed entrants for columns 1-3, twice the maximum number of observed entrants for columns 4-6, and the average number of listings divided by 25 for columns 7-9. Entry costs are in $100,000 2007 dollars.

Table 9. Model Fit for Commissions and Exit Probabilities

	Commissions ($100,000s)		Exit Probabilities	
	Observed	Fit	Observed	Fit
	(1)	(2)	(3)	(4)
		A. By Year		
1999	0.60	0.58	0.10	0.10
2000	0.63	0.59	0.12	0.11
2001	0.66	0.67	0.11	0.11
2002	0.72	0.73	0.10	0.10
2003	0.73	0.74	0.10	0.10
2004	0.75	0.76	0.10	0.11
2005	0.67	0.71	0.11	0.13
2006	0.51	0.56	0.14	0.13
2007	0.46	0.46	0.15	0.14
All	0.63	0.64	0.12	0.12
		B. By Market		
ARLINGTON	0.79	0.71	0.09	0.09
BROOKLINE	1.07	1.04	0.09	0.09
CAMBRIDGE	1.05	1.11	0.09	0.11
CONCORD	0.78	0.81	0.10	0.10
DANVERS	0.32	0.34	0.13	0.13
DEDHAM	0.60	0.62	0.11	0.12
HINGHAM	0.61	0.65	0.11	0.11
LEXINGTON	0.78	0.77	0.09	0.08
LYNN	0.47	0.48	0.13	0.12
MALDEN	0.50	0.52	0.13	0.12
MARBLEHEAD	0.76	0.79	0.09	0.09
MEDFORD	0.64	0.67	0.10	0.11
NEEDHAM	0.76	0.77	0.08	0.08
NEWTON	0.96	0.99	0.10	0.10
PEABODY	0.47	0.50	0.11	0.12
QUINCY	0.49	0.48	0.12	0.12
RANDOLPH	0.44	0.46	0.15	0.15
READING	0.54	0.55	0.11	0.11
REVERE	0.48	0.58	0.12	0.10
SALEM	0.45	0.43	0.12	0.12
SOMERVILLE	0.62	0.65	0.14	0.14
STOUGHTON	0.44	0.43	0.13	0.13
WAKEFIELD	0.53	0.51	0.12	0.12
WALPOLE	0.49	0.48	0.13	0.14
WALTHAM	0.65	0.59	0.12	0.11
WATERTOWN	0.73	0.73	0.09	0.09
WELLESLEY	1.03	0.99	0.12	0.12
WEYMOUTH	0.39	0.38	0.13	0.14
WILMINGTON	0.43	0.43	0.16	0.14
WINCHESTER	0.76	0.80	0.11	0.11
WOBURN	0.43	0.44	0.15	0.13

Notes: Commissions are in $100,000 in 2007 dollars.

Table 10. Market Structure with Different Commission Rates

	Average Number of Transactions (1)	Average Number of Entrants (2)	Average Number of Active Agents (3)	Average Number of Exiting Agents (4)	Average Commissions ($100,000s) (5)	Average Sales Probability (6)	Per-period Cost Savings ($mil) (7)	Entry Cost Savings ($mil) (8)	Commission Savings ($mil) (9)
Actual (5%)	7.78	22.52	153.78	18.05	0.63	0.70			
Counterfactual Commission Rates									
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. Each row indicates the reduced commission rate. Entry cost savings use assumption that the number of potential entrants is the maximum number of distinct agents in market in our sample. Standard errors (in parenthesis) are computed from 100 bootstrap simulations.

Table 11. Counterfactual Market Structures and Social Savings

	Average Number of Transactions (1)	Average Number of Entrants (2)	Average Number of Active Agents (3)	Average Number of Exiting Agents (4)	Average Commissions ($100,000s) (5)	Average Sales Probability (6)	Per-period Cost Savings ($mil) (7)	Entry Cost Savings ($mil) (8)
Actual Market Structure	7.78	22.52	153.78	18.05	0.63	0.70		
A. Counterfactual with Compensation Based on Break-even Costs in 1998								
Cost-based compensation	10.20	17.99	116.93	19.12	0.60	0.71	502.23	23.42
	(0.22)	(0.24)	(2.23)	(0.52)	(0.00)	(0.00)	(5.50)	(1.42)
B. Counterfactual with Improved Information on Past Agent Performance								
Raise Skill Impact 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 Impact 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 Impact 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 Impact 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)
Raise Skill Impact by 100%	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. Standard errors (in brackets) are derived from 100 bootstrap simulations. In Panel A, we simulate the model assuming agents are compensated per property according to break-even levels in 1998. The total commission savings is $1.1 billion. In Panel B, we increase the coefficient on agent skill in the listing equation by the denoted percentage. Entry cost savings use assumption that the number of potential entrants is the maximum number of distinct agents in market in our sample.

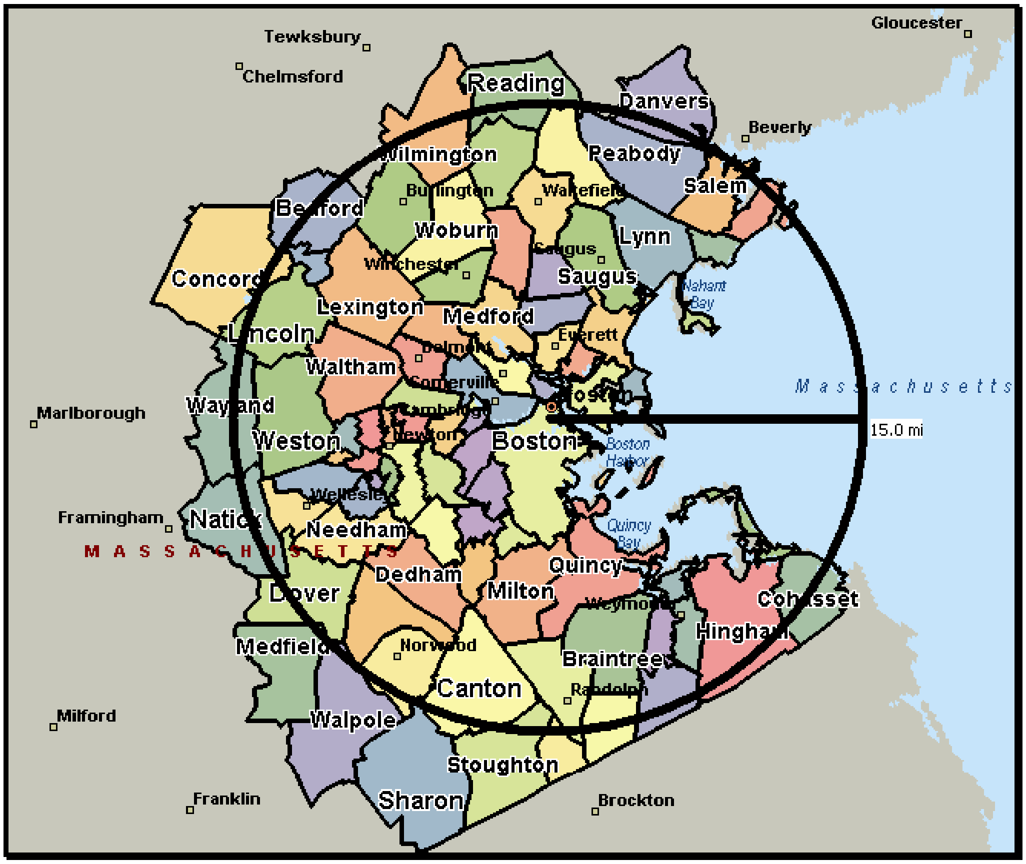

Figure 1: Geographic Areas in Greater Boston

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

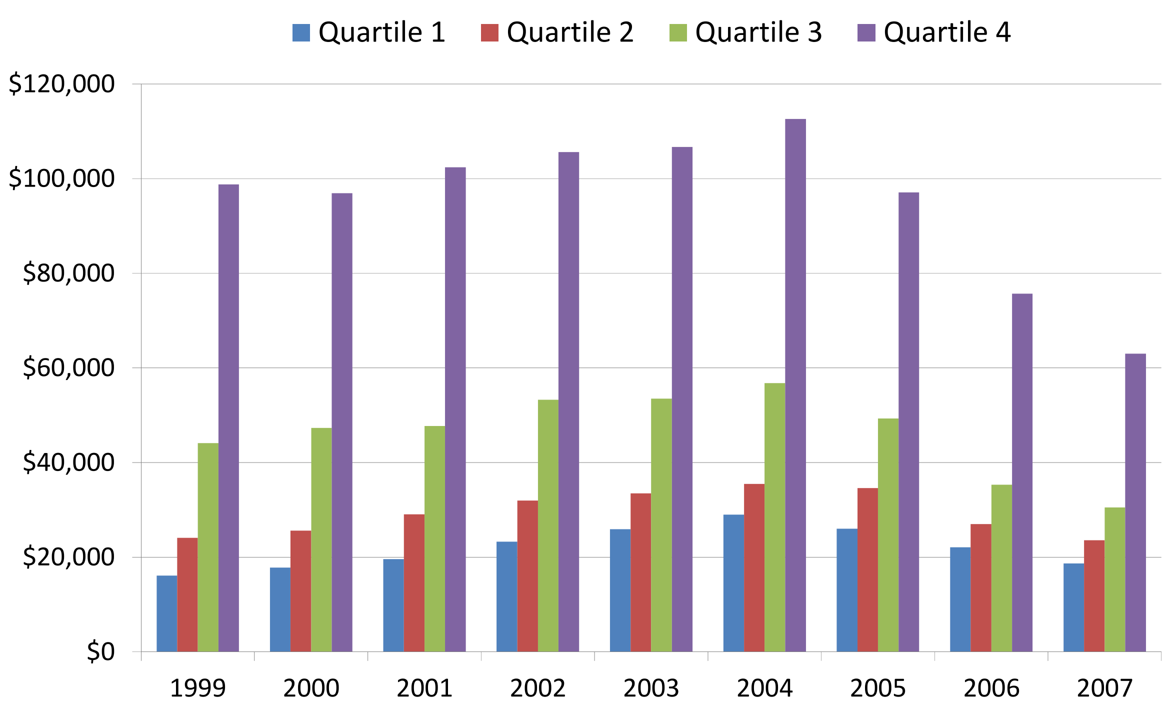

Figure 3: Fraction of Realtors Remaining by Commission Quartile for the 1998 Cohort

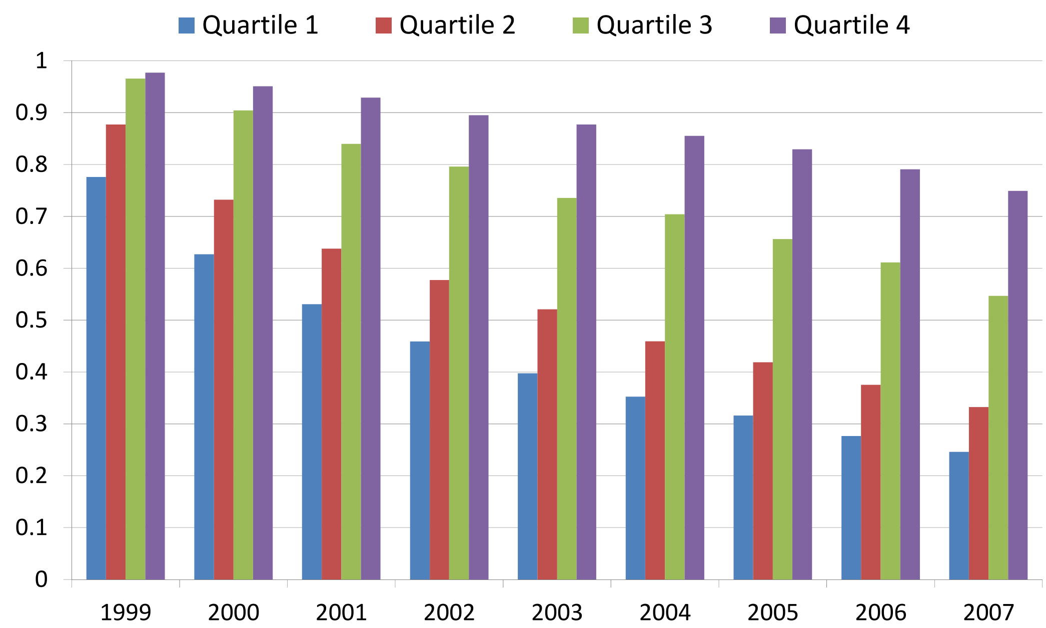

Figure 4: Foregone Income vs. Median Household Income (2007 $)