

The costs of free entry: an empirical study of real estate agents in Greater Boston

Panle Jia Barwick*

and

Parag A. Pathak**

This article studies the consequences of fixed commissions and low entry barriers in Greater Boston's real estate brokerage industry from 1998–2007. We find that agent entry reduces average service quality and use a dynamic empirical model to study the inefficiency in the current market structure. To accommodate a large state space, we approximate the value function using sieves and impose the Bellman equation as an equilibrium constraint. Our results suggest that a 50% cut in commissions would result in 40% fewer agents, social savings that amount to 23% of industry revenue, and 73% more transactions for the average agent.

1. Introduction

■ The structure of the US real estate brokerage industry puzzles economists, antitrust authorities, and the broader public. Most agents earn the same commission rate, even for properties with large differences in size, price, ease of sale, and other important aspects.¹ The absence of commission competition is counterintuitive in light of substantial entry and other competitive pressures in the past decade. Fixed commissions also adversely affect market performance. For example, Levitt and Syverson (2008a) illustrate how they lead to a conflict of interest between agents and clients, and Hsieh and Moretti (2003) document that agents chasing after high house prices result in fewer houses sold per agent in more expensive cities. Moreover, Han and Hong

*Cornell and NBER; panle.barwick@cornell.edu.

**MIT and NBER; ppathak@mit.edu.

We thank Alan Genz for helpful suggestions in the early stage of this project. Paul Asquith, Lanier Benkard, Steve Berry, Xiaohong Chen, JP Dube, Hanming Fang, Phil Haile, Lu Han, Jerry Hausman, Brad Larsen, Ariel Pakes, Michael Powell, Tavneet Suri, Maisy Wong, Juanjuan Zhang, and seminar participants at Columbia, Cornell, Duke, Georgetown, Harvard, Haas, John Hopkins, LSE, Michigan, Olin Business School, Princeton, Stanford GSB, Toronto, Toulouse, UCL, Wharton, and Yale provided helpful feedback. Constructive comments of the editor and referees improved the article. Barwick gratefully acknowledges the research support of the National Science Foundation via grant no. SES-1061970. All errors are our own.

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

(2011) argue that more agents lead to higher brokerage costs, and Bernheim and Meer (2012) 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 article provides model-based estimates of inefficiencies like Han and Hong (2011), but also investigates agent heterogeneity and the nature of agent competition using rich agent-level data from Greater Boston during 1998 – 2007. First, we find that agent productivity differs substantially depending on level of 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 equal to five times 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. Unlike earlier work, our longitudinal data allow us to incorporate agent heterogeneity and behavioral responses in quantifying market inefficiencies.

These descriptive patterns motivate our development of an empirical model that identifies the cost of free entry using the decisions of agents to work as brokers.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 costs of working as an agent, which includes costs associated with transacting properties as well as foregone labor income from working in an alternative profession. Knowing agents' per-period costs allows us to measure how market structure changes in response to variation in agent payoffs. The per-period cost also serves as a measure of social inefficiency for each additional agent, as 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.

To provide a benchmark for inefficiency, we use our empirical model to investigate several counterfactuals, which take agent heterogeneity and behavioral responses into account. The first exercise keeps the commission rate fixed, but at lower levels. The results measure the potential impact if a regulatory agency such as the Federal Trade Commission (FTC) or Department of Justice were to implement an across-the-board reduction in commissions. If the commission rate is cut in half, entry decreases by one third and there would be 40% fewer agents, generating a social gain equal to 23% of total commissions. Productivity increases substantially: each agent sells 73% more houses and the average sales likelihood is 2% higher.

Rigid commissions drive a wedge between agent effort and earnings. Under free entry, if commissions are flexible, agents would be compensated at a level that reflects their costs of transacting properties. The second counterfactual exercise quantifies inefficiency relative to a market structure where agent compensation is commensurate with the costs of buying and selling properties. As these costs are not directly observed, we use agent commissions in 1998 as a benchmark. This level is conservative because agents are unlikely to work at commission levels below their costs for an extended period, and that technological improvements such as the Internet have reduced their intermediation costs. When agents are compensated based on the 1998 average commission throughout our sample period, there would be 24% fewer agents, total commissions

² Other recent work on real estate agents includes Rutherford, Springer, and Yavas (2005); Kadiyali, Prince, and Simon (2014); 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 (2013); Dunne, Klimek, Roberts, and Xu (2011); and Ryan (2012).

paid by households decline from \$4.15 billion to \$3.04 billion, and social savings are equal to 13% of total commissions. This scenario also illustrates the “business stealing” effect: without 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 \$63,300.⁴

Under a fixed commission rate, it is difficult to distinguish good agents from mediocre ones based only on the prices they charge. The FTC recommends that consumers should investigate past agent performance when selecting a broker (FTC, 2006). Our last counterfactual investigates the implications of providing consumers with additional information on agents’ past records. Given the records in the Multiple Listing Service (MLS) platform, creating a system that evaluates agents or provides feedback on their performance would be relatively easy. According to our analysis, more information reduces the incentive for inexperienced agents to enter, thereby shifting business toward experienced incumbent agents and yielding productivity gains and social savings.

It is worth noting that to incorporate realistic dynamics in our framework, we push forward the methodological boundaries associated with large state spaces in estimating dynamic models. Another contribution of our article, therefore, is to illustrate how dynamic discrete choice methods may be useful for analyzing the housing market, where a rich state space is natural.⁵ The common approach in these models involves discretizing the state space, which leads to the well-known “curse of dimensionality” problem. We instead follow the tradition of value function approximation.⁶ Specifically, we treat the state space as continuous, approximate the value function using basis functions, and cast the Bellman equation as a model constraint using mathematical programming with equilibrium constraints (MPEC) following Su and Judd (2012).

One advantage of this procedure, besides avoiding the computational burden of solving for the value function, is that counterfactual analysis becomes relatively straightforward, even with a large state space. More broadly, this method for estimation and computing counterfactuals can be used in problems where avoiding discretization is advantageous. It is also less computationally demanding than existing approaches to our problem.⁷ A closely related study that adopts a similar procedure is an independent study by Michelangeli (2010), though with only one state variable. The estimator adopted in this article falls into the class of sieve estimators surveyed by Chen (2007). Arcidiacono, Bayer, Bugni, and James (2012) and Kristensen and Schjerning (2012) derive its asymptotic properties.

The remainder of the article 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, and then Section 7 presents our counterfactual analyses. The last section states our conclusions. Barwick and Pathak (2014) (hereafter, BP2) contains supplementary material on the sample construction, computational details, Monte Carlo simulations, and additional results not reported here.

2. Industry background and data

□ **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,

⁴ All dollar values in this article are in terms of 2007 dollars, deflated using the urban Consumer Price Index (CPI) (series CUUR0100SA0).

⁵ Bayer, McMillan, Murphy, and Timmins (2011) is another application of these methods for the housing market.

⁶ For examples of value function approximation, see Keane and Wolpin (1994); Ericson and Pakes (1995); Rust (1997); Judd (1998); Santos (1999); Farias, Saure, and Weintraub (2012); Fowlie, Reguant, and Ryan (2011); and many others.

⁷ For example, it avoids repeated interpolation (which is computationally demanding and difficult to achieve a desirable accuracy with high dimensions) and Monte Carlo simulations, as in Keane and Wolpin (1994).

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 modest: applicants for a salesperson license must take 24 hours of classroom instruction and pass a written exam. The qualifications for a broker's license involve a few additional requirements: one year of residence in Massachusetts, one year of active association with a real estate broker, completion of 30 classroom hours of instruction, passing a written exam, and paying a surety bond of \$5000. 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 6 to 12 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 data set.

□ **Data.** The data for this study come from the MLS network for Greater Boston, a centralized platform containing information on property listing and sales. We collect information on all listed nonrental residential properties for all towns within a 15-mile radius of downtown Boston from 1998 to 2007. There are a total of 18,857 agents and 290,738 observations.⁹ Figure 1 shows our 31 markets, which are created by grouping small towns and cities with neighboring ones (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.

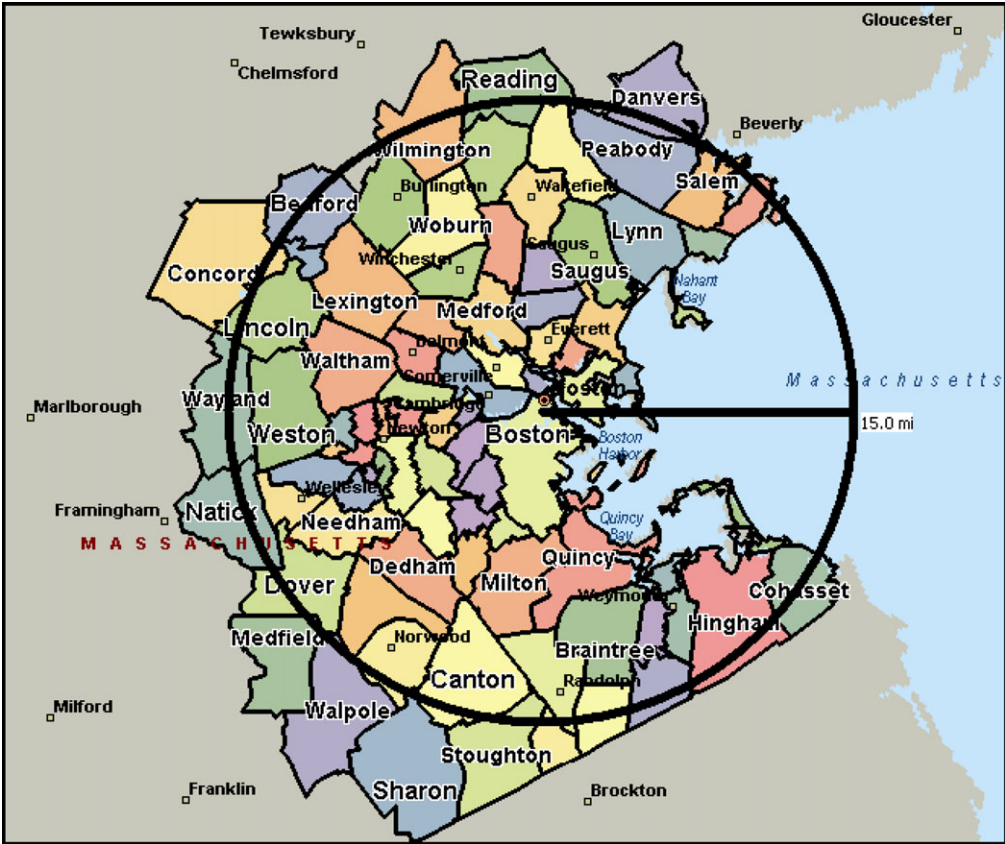
Information on commissions charged by real estate agents is difficult to obtain. Although our data does not contain commissions paid to listing agents, it does contain the commission rate paid to buyer's agents. Jia and Pathak (2010) report that the buyer's agent commission is 2.0% or 2.5% for 85% of listings in the sample. As 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 (NAR), 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.

The MLS data set 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.

⁸ MA Division of Professional Licensure, Board of Real Estate Brokers and Sales Persons. Available at 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 data set 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.

FIGURE 1
GEOGRAPHIC AREAS IN GREATOR BOSTON



An agent sometimes facilitates transactions in multiple markets. We assign an agent to the market that accounts for the largest fraction of his total number of transactions during the entire sample period. If there is a tie, we use the market with the largest commission. An agent belongs to one market throughout our sample period. The top market accounts for 63% of an agent’s total transactions and 65% of his total commissions during the sample period. Once we assign agents to a particular market, the total number of transactions in each market is defined as the sum of the transactions facilitated by all agents who belong to that market. BP2 provides more details on the sample construction.

3. Descriptive analysis

□ **Summary statistics.** Greater Boston’s housing market exhibits significant variation over time 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

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) (7)
	Listed (1)	Sold (2)	Mean (3)	Standard Deviation (4)	Mean (5)	Standard Deviation (6)	
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 the urban Consumer Price Index (CPI) compiled by Bureau of Labor Statistics (BLS). Days on market is winsorized at 365.

TABLE 2A Real Estate Agent Listings and Sales By Year

Year	Entrants (1)	Incumbent Agents (2)	Exiting Agents (3)	Number of Properties Sold (4)	Sales per Listing Agent			Purchases per Buyer's Agent		
					Mean (5)	25th (6)	75th (7)	Mean (8)	25th (9)	75th (10)
1998	0	3840	0	18,256	4.75	1	6	3.76	1	5
1999	602	4054	388	18,094	4.46	1	6	4.43	1	6
2000	462	4013	503	17,235	4.29	1	6	4.15	1	6
2001	483	4052	444	17,645	4.35	1	6	3.94	1	6
2002	696	4344	404	17,872	4.11	1	5	3.91	1	6
2003	883	4791	436	19,418	4.05	1	5	3.72	1	5
2004	1005	5328	468	21,432	4.02	1	5	3.70	1	5
2005	1002	5763	567	21,078	3.66	1	5	3.38	1	5
2006	691	5671	783	17,198	3.03	0	4	2.75	1	4
2007	424	5227	868	13,648	2.61	0	3	2.90	1	4
All	6248	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.

in 2005, and by the end of the sample, a listed property requires about two months longer to sell than in 1998.

Agent entry and exit as well as performance follow market-wide trends. Table 2A shows that the number of incumbent agents increases 50% from around 3800 in 1998 to a peak of more than 5700 in 2005. The number of agents who exit is around 400 – 500 each year during the early period, but rises to 700 – 800 when housing market conditions deteriorate in the latter period. During the early 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. There is wide variation in agent performance within years as well. These facts suggest that most agents are not capacity constrained (especially in later years) and could intermediate more properties. There is also considerable agent heterogeneity, which is apparent in the skewness of the distribution of agents' transactions: both the number of listings sold per agent and the number of houses bought per agent at the 75th percentile is equal to four to six times that of the 25th percentile.

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	7459	2.93	2.73
CONCORD	925.47	67	174	91	2,581	2.68	2.46
NEWTON	746.73	215	434	195	8779	3.94	3.93
LEXINGTON	711.98	141	268	113	4814	3.27	3.27
HINGHAM	701.88	142	261	132	3715	2.78	2.75
WINCHESTER	694.65	76	161	90	2980	3.48	3.36
NEEDHAM	692.15	82	175	71	3347	3.48	2.97
BROOKLINE	616.98	129	244	104	6346	4.94	4.60
CAMBRIDGE	582.23	262	417	159	10,763	5.18	5.28
MARBLEHEAD	550.26	107	238	109	5769	4.33	4.38
WATERTOWN	528.37	157	259	106	5229	4.10	3.98
DEDHAM	516.62	110	207	103	3689	3.55	3.12
ARLINGTON	454.32	103	196	85	5230	4.96	4.86
WALPOLE	446.14	218	369	193	5496	3.36	2.73
SOMERVILLE	444.83	229	303	152	4762	3.87	3.89
READING	430.76	128	244	124	4918	3.95	3.45
WALTHAM	405.42	146	228	108	4823	4.58	4.10
WILMINGTON	399.37	148	250	150	3745	3.52	2.69
PEABODY	390.43	191	317	151	5529	3.73	3.28
STOUGHTON	386.12	272	453	235	7234	3.48	2.99
MEDFORD	385.21	113	191	86	4826	5.20	4.10
WAKEFIELD	381.57	243	403	208	7919	4.19	3.84
QUINCY	379.84	472	677	321	10,757	3.68	3.59
DANVERS	357.96	97	203	110	2771	2.89	2.59
MALDEN	347.58	404	495	215	7136	3.70	4.10
WOBURN	347.09	109	179	103	2918	3.80	3.16
REVERE	325.41	408	520	216	8454	4.04	4.10
WEYMOUTH	324.14	470	652	329	9938	3.52	3.04
SALEM	303.79	173	268	134	5103	4.18	3.75
LYNN	299.47	470	605	286	11,048	4.37	4.18
RANDOLPH	290.57	127	192	102	3798	4.81	3.60
All	495.38	6248	10,088	4861	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 BLS's urban CPI.

Home sales, agent entry and exit, and agent performance also vary across markets within Greater Boston, as shown in Table 2B. 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 agent turnover: it has the most entry and the second largest number of exits. Cambridge has about the same number of properties, but fewer agents and much less turnover. This translates into a higher number of properties sold and bought per agent in Cambridge than in Quincy: 10.46 versus 7.27. In general, agents in higher - priced towns are involved in fewer transactions, and the correlation coefficient between the average house price and the average number of transactions is around -0.43 across all markets.

Experience accounts for a large share of performance differences among agents. Panel A of Table 3 reports the average annual commissions of agents based on the number of years they

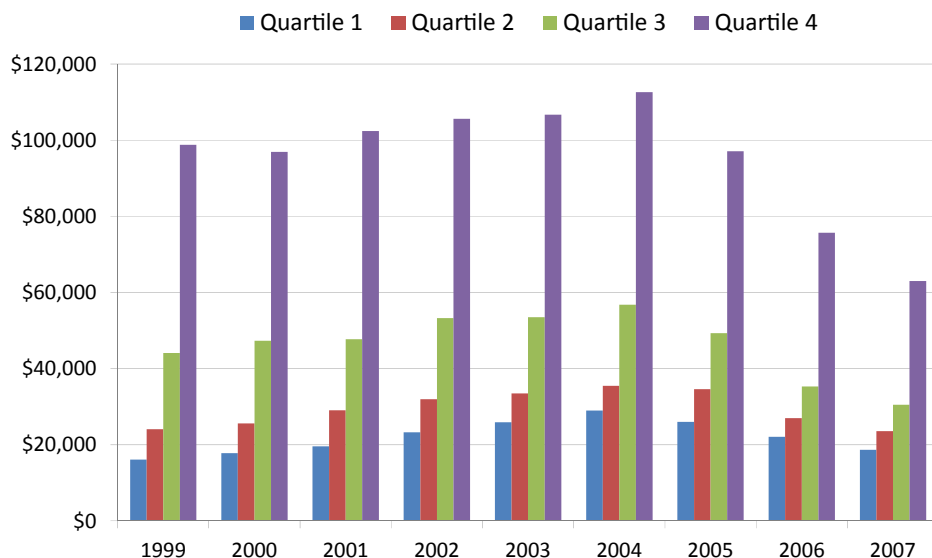
TABLE 3 Days on Market, Sales Probability, and Commissions By Agent Experience and Skill

Experience	N (1)	Sales Probability		Days on Market		Commissions (\$1000s)		Listing Commissions (\$1000s)		Sales Commissions (\$1000s)	
		Mean (2)	Median (3)	Mean (4)	Median (5)	Mean (6)	Median (7)	Mean (8)	Median (9)	Mean (10)	Median (11)
<i>A. Agents By Years of Experience</i>											
1	6,729	0.61	0.67	73.1	54.4	19.9	13.4	7.7	4.2	12.2	8.3
2	6635	0.63	0.67	71.9	56.0	35.8	24.3	14.0	7.6	21.8	15.0
3	5237	0.64	0.67	74.6	58.0	40.8	27.7	17.4	9.7	23.4	15.5
4	4184	0.64	0.68	72.2	55.7	45.1	30.5	21.0	12.0	24.1	16.1
5	3366	0.67	0.75	70.9	55.7	50.2	33.3	24.4	14.8	25.8	17.4
6	2657	0.70	0.76	68.9	53.7	55.1	37.4	28.0	17.2	27.1	18.5
7	2138	0.70	0.78	69.1	54.8	59.6	39.9	31.4	19.2	28.1	18.8
8	1788	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
<i>B. Agents By Deciles of Skill</i>											
Entrants	7421	0.62	0.67	71.6	53.7	20.3	13.6	8.0	4.3	12.3	8.3
<10%	3966	0.67	0.75	73.6	55.0	24.7	17.4	10.7	6.1	13.9	9.8
10-20%	3966	0.67	0.75	72.8	55.0	26.8	18.4	11.8	6.8	15.0	10.0
20-30%	3966	0.67	0.75	74.7	55.3	28.6	20.0	13.3	7.9	15.3	10.0
30-40%	3966	0.68	0.75	75.1	57.0	34.4	24.5	16.1	10.1	18.3	12.7
40-50%	3967	0.69	0.75	72.7	55.5	39.8	30.1	19.1	12.7	20.7	15.0
50-60%	3966	0.70	0.75	71.2	56.0	47.0	35.6	23.4	16.5	23.6	17.3
60-70%	3966	0.71	0.75	71.4	57.0	59.7	47.2	30.2	22.5	29.5	22.3
70-80%	3966	0.71	0.75	71.2	57.8	73.5	60.0	38.0	28.6	35.5	27.5
80-90%	3966	0.73	0.78	68.6	58.0	97.0	79.7	52.4	41.2	44.5	35.3
90%+	3967	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 BLS's urban CPI. Skill is proxied by the number of transactions in the previous year. Days on market is winsorized at 365.

FIGURE 2

COMMISSIONS BY QUARTILE FOR THE 1998 COHORT (2007 \$)



have worked as a broker.¹⁰ 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 nine or more years of 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 their listings faster, although the difference is modest.¹¹

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. As this measure is 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 2.¹² The top quartile agents consistently earn \$100,000 or more for most years, whereas the bottom quartile agents barely earn \$30,000, even during the housing boom. 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.

Earnings differences influence an agent's decision to work as a broker. Figure 3 follows agents in the 1998 cohort and reports the fraction of agents in each quartile who continue working as a broker each year. There are stark differences in the exit rates among the four groups. Only 25% of

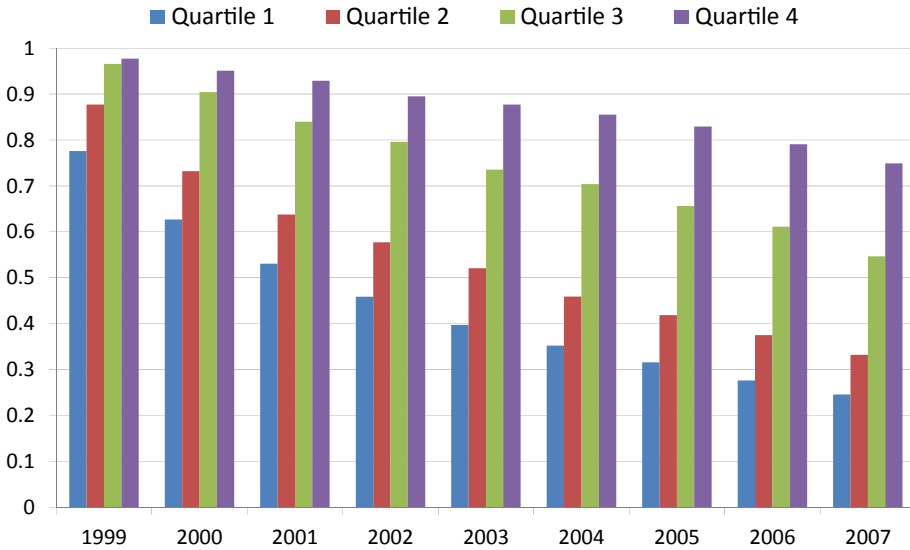
¹⁰ Note that the category of nine or more years of experience has 19,210 observations with a total of 3146 agents. These agents are active at the beginning of our sample. Most agents in the other categories (one to eight years of experience) entered during our sample period.

¹¹ Barwick, Pathak, and Wong (2014) show that performance differences are robust to controls for property attributes.

¹² Results using other cohorts are similar.

FIGURE 3

FRACTION OF REALTORS REMAINING BY COMMISSION QUARTILE FOR THE 1998 COHORT



the top quartile agents leave by 2007. In contrast, about three quarters of the agents in the bottom quartile exit at some point during the 10-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.

□ **Descriptive regressions.** Results from descriptive regressions inform the modelling choices we make in the next section. We focus on the impact of competition on agent performance as well as potential benefits home sellers derive from more competition among agents. We briefly describe our results here; more details are reported in Appendix A.

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

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

y_{imt} is either log commissions or log number of transactions, N_{mt} is the number of competing agents, and s_{it} is agent skill, proxied by the number of properties an agent intermediates in the previous year, following Table 3B.¹³ λ_m and τ_t are market and year fixed effects. H_{mt} , P_{mt} , and Inv_{mt} are three controls intended to capture market-year specific housing market patterns that are also used in the structural model described in the next section. H_{mt} is the total number of listed properties for sale in year t and market m , P_{mt} is the equal-weighted price of all houses that are sold in year t and market m , whereas Inv_{mt} is the inventory-sales ratio, 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. The parameter ρ_y reveals the impact of a percentage increase in the number of competing agents on performance.

Across all regressions that we estimate, ρ_y is significantly large and negative, 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 5.2% to 8.4% reduction in the number of transactions (and a similar reduction in commissions) for incumbent agents across cohorts (Table 4). Across agent quartiles, the reduction in transactions

¹³ Estimates are similar using the number of years an agent has worked as a broker to measure skill.

associated with increased competition is largest for the bottom quartile, implying that competitors attract more business from the bottom tier agents than from top ones.

Our second set of regressions examines whether home sellers benefit from more competition among agents. Let h_{imt} be a measure of the sales experience (the likelihood of sale, days on the market, or sales price) for a property intermediated by agent i working in market m in year t . We estimate property-level regressions of the form:

$$h_{imt} = \rho_h \log(N_{mt}) + \alpha s_{it} + \theta'_X X_{i(h)t} + \lambda_m + \tau_t + \theta_1 \log(H_{mt}) + \theta_2 \log(P_{mt}) + \theta_3 \text{Inv}_{mt} + v_{imt}, \quad (2)$$

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 (denoted by $i(h)$) 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 multifamily dwelling, and sometimes the log list price. Table 5 reports estimates of ρ_h .

For likelihood of sale, the ρ_h estimate is always negative across all the specifications, though it is sometimes insignificant. Hence, it does not appear that properties are more likely to sell when competition intensifies. For days on the market, the ρ_h estimate is negative and insignificant. Finally, for sales price, competition is associated with a small increase in the sales price of a property, and the effect shrinks when we control for the list price, as shown in columns (10) and (11) of Table 5. A 10% increase in competition generates a 0.1% increase in the sales price, which translates to roughly \$700 for a typical home. Because a higher sales price is a transfer from buyers to sellers and has a negligible impact on aggregate consumer surplus, in the rest of this article, we do not focus on the impact agents have on sales prices.

Our results suggest that 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 measurable benefits for sellers.¹⁴ We next turn to quantifying the magnitude of inefficient excess entry, assuming that consumer benefits from agent competition are modest.

4. Modelling 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 modelling 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. Before turning to the details of the model, we briefly describe why we develop a dynamic model.

□ **Motivating agent dynamics.** The longitudinal structure of the database naturally lends itself to considering dynamics in modelling entry and exit to recover agent's per-period costs. A static free entry model, which implies that agents' commission revenue are roughly equal to the fixed costs of being a realtor (as any deviation of commissions from fixed costs would be

¹⁴ It is possible that buyers gain from having more agents. For example, buyers might be more likely to be matched to their most desirable properties with more competition, which would result in higher sales prices provided that buyers and sellers divide the surplus generated from a better match. However, the limited price impact documented above suggest that this is unlikely to be significant.

TABLE 4 Impact of Competition on Agent Performance Across Cohorts

Agent Cohorts	All Agents (1)	Top Quartile (2)	Second Quartile (3)	Third Quartile (4)	Bottom Quartile (5)	All Agents (6)	Top Quartile (7)	Second Quartile (8)	Third Quartile (9)	Bottom Quartile (10)
I. Agent Commissions										
1998	-0.46*** (0.08)	-0.44*** (0.07)	-0.51*** (0.07)	-0.58*** (0.07)	-0.63*** (0.07)	-0.52*** (0.07)	-0.50*** (0.07)	-0.56*** (0.07)	-0.61*** (0.07)	-0.65*** (0.07)
1999	-0.45*** (0.08)	-0.39*** (0.07)	-0.47*** (0.07)	-0.53*** (0.07)	-0.59*** (0.07)	-0.52*** (0.07)	-0.47*** (0.07)	-0.52*** (0.07)	-0.57*** (0.07)	-0.63*** (0.07)
2000	-0.56*** (0.08)	-0.49*** (0.08)	-0.57*** (0.08)	-0.63*** (0.08)	-0.69*** (0.08)	-0.62*** (0.07)	-0.57*** (0.07)	-0.63*** (0.07)	-0.68*** (0.07)	-0.73*** (0.07)
2001	-0.70*** (0.09)	-0.59*** (0.08)	-0.67*** (0.08)	-0.74*** (0.08)	-0.80*** (0.08)	-0.72*** (0.08)	-0.63*** (0.08)	-0.69*** (0.08)	-0.75*** (0.08)	-0.80*** (0.08)
2002	-0.67*** (0.10)	-0.58*** (0.10)	-0.66*** (0.10)	-0.72*** (0.10)	-0.78*** (0.10)	-0.68*** (0.09)	-0.61*** (0.09)	-0.67*** (0.09)	-0.73*** (0.09)	-0.77*** (0.09)
2003	-0.73*** (0.12)	-0.57*** (0.11)	-0.65*** (0.11)	-0.73*** (0.11)	-0.77*** (0.11)	-0.75*** (0.11)	-0.62*** (0.10)	-0.69*** (0.10)	-0.75*** (0.10)	-0.79*** (0.10)
2004	-0.77*** (0.16)	-0.51*** (0.15)	-0.60*** (0.15)	-0.68*** (0.15)	-0.74*** (0.15)	-0.84*** (0.14)	-0.62*** (0.13)	-0.69*** (0.13)	-0.76*** (0.13)	-0.82*** (0.13)
II. Number of Transactions										
1998	-0.46*** (0.08)	-0.44*** (0.07)	-0.51*** (0.07)	-0.58*** (0.07)	-0.63*** (0.07)	-0.52*** (0.07)	-0.50*** (0.07)	-0.56*** (0.07)	-0.61*** (0.07)	-0.65*** (0.07)
1999	-0.45*** (0.08)	-0.39*** (0.07)	-0.47*** (0.07)	-0.53*** (0.07)	-0.59*** (0.07)	-0.52*** (0.07)	-0.47*** (0.07)	-0.52*** (0.07)	-0.57*** (0.07)	-0.63*** (0.07)
2000	-0.56*** (0.08)	-0.49*** (0.08)	-0.57*** (0.08)	-0.63*** (0.08)	-0.69*** (0.08)	-0.62*** (0.07)	-0.57*** (0.07)	-0.63*** (0.07)	-0.68*** (0.07)	-0.73*** (0.07)
2001	-0.70*** (0.09)	-0.59*** (0.08)	-0.67*** (0.08)	-0.74*** (0.08)	-0.80*** (0.08)	-0.72*** (0.08)	-0.63*** (0.08)	-0.69*** (0.08)	-0.75*** (0.08)	-0.80*** (0.08)
2002	-0.67*** (0.10)	-0.58*** (0.10)	-0.66*** (0.10)	-0.72*** (0.10)	-0.78*** (0.10)	-0.68*** (0.09)	-0.61*** (0.09)	-0.67*** (0.09)	-0.73*** (0.09)	-0.77*** (0.09)
2003	-0.73*** (0.12)	-0.57*** (0.11)	-0.65*** (0.11)	-0.73*** (0.11)	-0.77*** (0.11)	-0.75*** (0.11)	-0.62*** (0.10)	-0.69*** (0.10)	-0.75*** (0.10)	-0.79*** (0.10)
2004	-0.77*** (0.16)	-0.51*** (0.15)	-0.60*** (0.15)	-0.68*** (0.15)	-0.74*** (0.15)	-0.84*** (0.14)	-0.62*** (0.13)	-0.69*** (0.13)	-0.76*** (0.13)	-0.82*** (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 as the set of agents active or entering in the initial cohort year. The sample excludes 1631 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** (0.027)	-0.036 (0.026)	-0.051 (0.064)		-0.028 (0.093)	-0.046 (0.094)	-0.044 (0.181)		0.118*** (0.027)	0.012*** (0.003)	0.014*** (0.005)	
Log(Nmt) before 2005				-0.025 (0.030)				-0.022 (0.092)				0.013*** (0.003)
Log(Nmt) after 2005				-0.040 (0.025)				-0.055 (0.091)				0.012*** (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. Columns (7) and (11) use properties that are sold at least twice. Robust standard errors clustered by market.

eliminated by entry or exit), is inconsistent with some important features of the data. During our sample period, fixed costs such as the direct expenses related to buying and selling properties and the foregone labor income from alternative professions seem unlikely to vary considerably. For example, the real earnings in service-related professions in the greater Boston area have remained essentially flat from the late 1990s to the late 2000s, increasing less than 1% a year on average. In addition, technological innovations most likely have decreased transaction-related expenses during the same period.

Although fixed costs have probably stayed constant, agents' commission revenues exhibit strong cyclical patterns. For instance, the average annual commission is \$75,000 in 2004, but only \$46,000 in 2007. The cyclical pattern of commissions revenue is present for different agent cohorts and across different experience levels. For example, Figure 2 shows that top quartile agents in the 1998 cohort earn \$113,000 in 2004, but these same agents earn only \$63,000 in 2007, a reduction of 44%. There is a similar pattern for the other three quartiles of the 1998 cohort. A static model would rationalize these patterns by predicting that the costs of being an agent also vary significantly over time, which seems implausible.

A second pattern that is difficult for a static model to capture is the presence of agent learning. Interviews with realtors indicate that there is significant on-the-job learning, including familiarizing oneself with professional regulations, property inventories and local amenities, and networking with potential clients. In our sample, more than half of agents earn less than \$26,000 at the beginning of their career and 90% earn less than \$72,000 in their first two years. As they gain more experience and develop a client network, their commission income increases accordingly: in five years, the median within-agent increase in commissions is 36%, and the average increase is almost two-fold. Another pattern that is difficult to reconcile with a static model is inertia in agents' exit decisions. Few agents exit immediately after a significant negative income shock (as would be predicted by a static model). For example, among agents whose commission income decreases by a third or more in a given year compared to the previous year (which happens to one in six agents), the majority stay and only 17% exit the industry.

A dynamic model allows us to address some of these challenges. For example, agents' entry and exit decisions are driven by their lifetime discounted earnings, and hence, are less sensitive to short-run fluctuations. Learning can be directly incorporated in the model by a skill variable that varies over time. Finally, the option value of remaining in the profession in anticipation of future higher compensation would help to rationalize agent inertia in the presence of negative shocks.

A final limitation of a static model is that entry costs cannot be separately identified from fixed costs. Static models typically impose equilibrium restrictions with no entry or exit taking place. Individuals or firms are either in or out of the market; hence, only one cost can be estimated from the equilibrium conditions of static models. However, the distinction between entry costs and fixed costs is crucial in our counterfactual analysis: agents only incur entry costs once upon entry, whereas they incur fixed costs annually. The inability to separate these two costs prevents us from an accurate estimate of the efficiency gains associated with our various counterfactual experiments. Given these considerations, we believe a dynamic model is better suited for our purpose.

□ **Housing attributes.** To model the evolution of the housing market and how it affects the entry and exit decisions of agents over time, it is necessary to represent the housing market in terms of state variables. As our data includes information on the attributes of each property that an agent intermediates, we could, in principle, incorporate property attributes and model how agents are matched to particular properties.

We do not pursue this rich representation and instead work with a more stylized version of the housing market, for several 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, adding property attributes does not significantly improve the model's fit of agents' revenues (or the fit for various elements

of the revenue function, like the sales probability).¹⁵ As we discuss further in Section 5, structural parameters are identified from agent revenues. Our parsimonious representation of the housing market allows for a reasonable fit of the observed revenues (the correlation between fitted and observed agent revenues is 0.7); hence, the marginal benefit from introducing property attributes seems modest. Third, 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.

One might be concerned that there could be significant sorting between agents and properties. For instance, experienced agents might be more likely to list expensive properties. To the extent that agents self-select into different markets – experienced agents attract high-end properties whereas less established ones compete for inexpensive houses, our assumption would exaggerate the impact that entry exerts on experienced agents. As a result, our estimate of social inefficiencies with excess entry would be overstated.

To examine this issue, we proceed in several steps. First, we find that the correlation between agent characteristics and property attributes is positive, but small. For instance, the correlation between years working as an agent and the number of bedrooms is only 0.005. Table A1 in the Appendix compares property attributes for agents with above- or below-median number of transactions, where the median is computed by market-year. The average house listed by below-median agents has 3.31 bedrooms, 1.94 bathrooms, 1.8 other rooms, is about 1850 square feet, is listed at \$371,000, and sold for \$346,000. The average property listed by above-median agents has 3.34 bedrooms, 2.01 bathrooms, 1.79 other rooms, is about 1940 square feet, is listed at \$403,000, and sold for \$384,000. These differences are statistically significant given our large sample size (with about a quarter of a million observations), but are small in magnitude compared to the standard deviation of the relevant variables. For example, the difference in the number of bedrooms and sales price is 0.03 and \$38,000, respectively, whereas the standard deviation of bedrooms and sales price is 1.5 and \$270,000, respectively. Results are very similar when we examine agents and buying agents separately, or divide the sample by property type (condos versus single family houses), by year, or by market. Replacing the “median number of transactions” with “five years of experience” generates similar patterns: agents with more or less than five years of experience buy or sell properties with similar attributes.

It’s possible that established agents might be particularly good at selecting desirable properties along unobserved dimensions. To investigate this possibility, we project hedonic residuals onto agent types. Specifically, we first create indicator variables for each of the following property attributes: (i) property type (condo, single family, and multifamily), (ii) number of bedrooms (1 bed, 2 beds, . . . , 9 beds, 10 and more beds), (iii) number of bathrooms (1 bath, 1.5 baths, . . . , 7 and more baths), and (iv) number of other rooms. Then we regress the listing price on these attribute indicators together with square footage, as well as the full market-year interaction fixed effects for a total of 314 regressors. Finally, we obtain the residual from this regression and project the residual on the agent type as defined above. On average, established agents select properties that are \$6450 more desirable, conditioning on physical attributes and market-year fixed effects.¹⁶ Although the estimate is significant, the magnitude of the selection effect is economically small: it is about 1.7% of the average listing price, and much smaller than the standard deviation of the listing price (which is around \$285,000). Our conclusion is that it’s reasonable to model agents as competing across all listings given the relatively modest specialization across properties, conditional on market and year.

¹⁵ In the linear probability regression of whether a property gets sold, once we control for market aggregate state variables, agent characteristics, and year fixed effects, property attributes – including indicator variables for each of the following (property type, the number of beds, the number of baths), and square footage – have a modest impact on R^2 : the R^2 only slightly increases from 0.08 to 0.09.

¹⁶ Repeating the exercise with the sales price delivers a similar result, where established agents select properties that are \$7100 more desirable, everything else equal.

□ **State variables.** 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 Inv_{mt} (the ratio of inventory-sales ratio) that are described in Section 3. 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 as it seems unlikely that this accounts for a significant fraction of housing market variation. Two other aggregate state variables (L_{mt} and B_{mt}) measure the intensity of competition among agents, and are discussed after we introduce agents’ revenue function.

□ **Agent payoffs.** Realtors earn commissions either from sales (as listing agents) or purchases (as buyer’s agents) of homes. 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. As the aggregate variables are the same for all agents in market m and year t , the listing share depends only 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:

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

The variables X_{it}^L include agent i ’s demographics, work experience, firm affiliation, and proxies for agent skill. As not all aspects of his 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).

Variables X_{it}^L are observed state variables that capture agent heterogeneity, whereas ξ_{it}^L is an unobserved, time-variant, and potentially persistent state variable. Incorporating correlated unobserved state variables in dynamic discrete choice models with an infinite horizon is the subject of ongoing research. In our application, ξ_{it}^L is the residual from the share equation (3) and is “observed” once parameters θ^L are estimated. In principle, it would be possible to incorporate this variable as another state variable in estimation. However, we do not because of our interest in the counterfactual analysis, where there is no satisfactory approach to solve the new equilibrium with an unobserved and persistent time-varying state variable. We report estimates assuming that ξ_{it}^L is independent across agents and time. In Section 6, we present evidence that our controls for agent skill mitigate the concern about correlated unobserved state variables.

The denominator in equation (3),

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

L is a state variable that measures the level of competition agents face in obtaining listings. Given the large number of brokers per market (≥ 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.¹⁷ 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 $H_t^B = 0.69H_t$. 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. As 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.

As agents earn commissions as brokers on both the buy and sell side, agent i 's revenue function is:

$$R(S_{it}) = R^{Sell}(S_{it}) + R^{Buy}(S_{it}) \\ = r * HP_t * (ShL_{it} * Pr_{it}^{Sell} + ShB_{it} * 0.69), \tag{5}$$

where $S_{it} = \{X_{it}^L, X_{it}^S, X_{it}^B, HP_t, Inv_t, L_t, B_t\}$. 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. Section 6 contains details on the model's fit.

□ **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 agents' 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:

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

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 multivariate normal random variable.¹⁸ Market fixed effects in equation (6) are included to adjust for size differences across markets.

We also investigated splitting the sample at year 2005 and estimating a separate transition process for each subsample, 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

¹⁷ We treat the sales price as exogenous. Hence, this formulation does not allow for a trade-off between the probability of sale and the sales price.

¹⁸ Our state variables include dummy $\{t > 2005\}$ and the value function explicitly depends on this dummy variable. Our assumption is that the housing market is stationary, conditioning on this state variable. This is a strong assumption, but is needed for computing the value function.

0.96), and that our panel is relatively short.¹⁹ Finally, like the aggregate state variables, an agent's skill is also modelled as an AR(1) process, with a different constant before and after 2005.²⁰

□ **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, 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.

As 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 (H , P , and Inv) 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:

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

where $E [R(Z_{it}, Y_t) | Z_{it}, Y_{t-1}]$ is his expected commission revenue conditional on observed state variables and δ is the discount factor, and $\delta E \tilde{V}(Z_{i,t+1}, Y_t | Z_{it}, Y_{t-1})$ is the expected future value of being an agent. Conditioning on state variables, the revenue function also depends on ξ_{it}^L and ξ_{it}^B , which we integrate out using their empirical distributions.^{21, 22} As 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 as 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 to make this assumption because empirically we have a difficult time separately estimating the fixed costs and marginal costs, as discussed in Section 4.²³ In all specifications, c differs across markets, but is the same for agents

¹⁹ Like many other empirical exercises, the formulation of the state variable transition process assumes that agents have perfect foresight and can predict the reverse of the housing market condition after 2005, which seems unlikely. To address this concern, we estimated the model using only observations between 1998 and 2005, and assumed that agents expected the housing boom to continue forever and eventually plateau without experiencing a substantial downturn. The cost estimates are somewhat higher than those reported in Table 7, which suggests that our counterfactual analysis could be somewhat conservative.

²⁰ 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) versus 0.75 (without market fixed effects).

²¹ Once we have estimated the listing share as defined in equation (3) and the buying share, we obtain ξ_{it}^L and ξ_{it}^B as the residual of these two equations. These residuals give us the empirical distribution of ξ_{it}^L and ξ_{it}^B .

²² 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 .

²³ In Section 7, we report counterfactual results under different assumptions on marginal cost.

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. Reentering 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.²⁴

Private shocks $\tilde{\varepsilon}_0$ and $\tilde{\varepsilon}_1$ are i.i.d. extreme value random variables with variance $\frac{\pi^2}{6\beta_1^2}$, where $\beta_1 > 0$. Multiplying both sides of equation (7) by β_1 , the original Bellman equation can be rewritten as:

$$V(Z_{it}, Y_{t-1}) = E_\varepsilon \max_{\varepsilon_{i0t}} \left\{ \beta_1 E [R(Z_{it}, Y_t)|Z_{it}, Y_{t-1}] - \beta_1 c + \varepsilon_{i1t} + \delta EV(Z_{it+1}, Y_t|Z_{it}, Y_{t-1}) \right.$$

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 , and denote $\beta_1 E [R(Z_{it}, Y_t)|Z_{it}, Y_{t-1}] - \beta_1 c$ as $\bar{R}(Z_{it}, Y_{t-1}, \beta)$, the Bellman equation reduces to the usual log-sum form:

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

where γ is the Euler constant. 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:

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

The log likelihood for incumbent agents is:

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

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 ε_{j1t} and ε_{j0t} . We adopt the standard assumption that ε_{j1t} and ε_{j0t} are i.i.d. across individuals and time. Although this assumption is dictated by feasibility considerations, it does rule out certain patterns of entry and exit. For example, it’s possible that individuals of a certain type are more likely to enter and be successful realtors. The panel feature of our data set provide information about the importance of selection effects without requiring estimation of a full model. The selection story, taken to the extreme, would indicate that individuals’ commissions do not change much over time (conditioning on aggregate market conditions): entrants of the good type should earn high commissions from the beginning, whereas entrants of the bad type should earn low commissions throughout. As discussed in Section 4, most entrants experience significant commission increases with more years of experience, a fact that appears more consistent with on-the-job learning rather than selection.

The Bellman equation for potential entrant j is:

$$V^E(Z_{jt}, Y_{t-1}) = E_\varepsilon \max_{\varepsilon_{j0t}} \left\{ -\kappa + \bar{R}(Z_{jt}, Y_{t-1}, \beta) + \varepsilon_{j1t} + \delta EV(Z_{jt+1}, Y_t|Z_{jt}, Y_{t-1}) \right.$$

$$= \log[1 + \exp(-\kappa + \bar{R}(Z_{jt}, Y_{t-1}, \beta) + \delta EV(Z_{jt+1}, Y_t|Z_{jt}, Y_{t-1}))] + \gamma.$$

²⁴ The estimation strategy would be similar, except that the exit choice probability is exploited to recast one of the choice-specific value functions as a fixed point of a Bellman equation as in Bajari, Chernozhukov, Hong, and Nekipelov (2009).

Just as in equation (9), the probability of entry is:

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

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

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

As 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. In our counterfactual analysis discussed in Section 7, we report results with and without entry costs.

□ **Discussion of other modelling issues.** The main parameter of interest is c , the average agent's per-period cost of working as a broker. The current specification does not allow c to depend on state variables because we observe only one action for each active agent (stay or exit). In our experience, it proved difficult to separately identify the impact of a state variable working through c versus its impact working through revenue R on agent actions. Likewise, we do not allow a variable cost component which depends on the number of transactions because agent revenue is proportional to his total number of transactions.²⁵ 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} corresponds to the average value of time that agents devote to being a broker.²⁶

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 (2014) document that the four largest firms intermediate roughly two thirds of transactions and each sets its own policy on commissions for all of its employees.

As 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, 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 underestimated by $\alpha\%$, then β_1 will be overestimated by the same amount, and the per-period cost $c = -\frac{\beta_2}{\beta_1}$ will be underestimated by $\alpha\%$.

²⁵ We report counterfactual results incorporating marginal costs in Section 7.

²⁶ 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.

5. Solution method

■ As explained in Section 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 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.²⁷ 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.

□ **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. On the other hand, value function approximation introduces bias into parameter estimates and the nonparametric approximation typically 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 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:

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

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.²⁸ 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 $u_j(S)$:

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

²⁷ We ran out of memory on a server with 32GB of RAM, using 20 grid points for each of the four state variables.

²⁸ We thank Alan Genz for suggesting this reference.

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

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

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

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

where $\{S_{(k)}\}_{k=1}^K$ denotes state values observed in the data and $\|\cdot\|_2$ is the L^2 norm.²⁹ 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.

As 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 .³⁰ This property motivates using basis functions that fit the revenue function $\bar{R}(S)$ for our choice of $u_j(S)$. As these basis functions are chosen to preserve the shape of $\bar{R}(S)$, they should also capture the shape of the value function.

Choosing basis terms in high-dimensional models is not a simple matter. 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.³¹ 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 (2012), Dubè, Fox, and Su (2012), as well as other applications using MPEC. 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 constraint is stated as an *argmin* problem. To implement this constraint, we replace it with the first-order conditions with respect to b_j . In other words, the constraint is replaced with K equalities that the partial derivative of the constraint with respect to each b_j is equal to zero. Once we obtain the parameter estimates, we verify that the second-order conditions of the *argmin* problem are satisfied.³²

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

²⁹ Note that this formulation only uses states that are observed in the data. Whether this formulation is optimal is left for future research.

³⁰ The formal argument follows from the contraction mapping theorem and is in Appendix B.

³¹ 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.

³² In practice, if the second-order conditions are violated, we repeat the estimation using different starting values until we find parameter estimates that satisfy the second-order conditions.

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 nonparametric bootstrap).

□ **Identification and consistency.** Identification of β_1 and β_2 follows from the usual 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. As it is difficult to estimate the discount factor, δ , empirically (e.g., Rust, 1994), 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 spline coefficients.

Kristensen and Schjerning (2012) prove the consistency and asymptotic normality property of sieve maximum likelihood estimators (MLE) for a class of models that are more general than the one presented in this article, and provides convergence rates for parameter estimates when different basis functions are employed in the estimation, including Chebyshev polynomials and splines. We refer interested readers to that article for the theoretical analysis of these estimators.

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).

□ **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. Demeaning the log of the listing share (3), we obtain:

$$\ln ShL_{imt} - \overline{\ln ShL_{.mt}} = \left(X_{imt}^L - \overline{X_{.mt}^L} \right) \theta^L + \left(\xi_{imt}^L - \overline{\xi_{.mt}^L} \right) = \left(X_{imt}^L - \overline{X_{.mt}^L} \right) \theta^L + \tilde{\xi}_{imt}^L, \quad (15)$$

where $\overline{\ln ShL_{.mt}} = \frac{1}{N} \sum_{i=1}^N \ln ShL_{imt}$. 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, as their s_{it} is either undefined or biased downward, and end up with 32,237 agent-year observations.³³

As 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 .

³³ Including first- or second-year agents only slightly reduces s_{it} coefficient.

TABLE 6A Revenue Function Regressions

	Listing Share (1)	Buying Share (2)	Sold Probability (3)
Skill	1.27*** (0.01)	0.90*** (0.01)	0.21*** (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. R² adjusted for MLE is pseudo adjusted R² (computed by authors). Entrants and agent's with 0 shares are excluded.

TABLE 6B State Variable Autoregressions

	HP (1)	Inv (2)	L (3)	B (4)	Skill (5)
lag(HP)	0.74*** (0.05)	0.21*** (0.06)	0.35*** (0.02)	0.35*** (0.02)	
lag(Inv)		0.65*** (0.05)	-0.13*** (0.02)	-0.13*** (0.02)	
lag(L)			0.79*** (0.02)		
lag(B)				0.76** (0.02)	
lag(Skill)					0.75*** (0.00)
Year < 2005	0.29*** (0.03)	-0.10** (0.05)	0.03* (0.02)	0.04** (0.02)	0.04** (0.00)
Year ≥ 2005	0.17*** (0.05)	0.62*** (0.07)	0.12*** (0.03)	0.09*** (0.03)	0.00 (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. R² adjusted for GMM-IV is pseudo adjusted R² (computed by authors). Entrants as well as agents with 0 shares are excluded in column (5).

Experience is also an important predictor of listing shares, but it has a limited explanatory power once s_{it} is included.³⁴ Our preferred specification is column (1) of Table 6A, which uses only s_{it} as a regressor.

Given that our proxy s_{it} cannot fully capture all aspects of an agent's skill, residuals $\tilde{\xi}_{int}^L$ may be positively serially correlated: a good agent consistently outperforms his peers with the same observed value of s_{it} . To investigate this issue, we regress the residual estimate $\widehat{\xi}_{int}^L$ on its lags.

³⁴ These results are reported in Table A2 of BP2.

These residuals exhibit little persistence over time. The R^2 of the OLS regression is 0.002, and the coefficient of lagged $\hat{\xi}_{imt}^L$ is small and negative (about -0.04), which suggests mean reversion. We repeat the analysis with the Arellano-Bond estimator that accommodates agent fixed effects. The coefficient of lagged $\hat{\xi}_{imt}^L$ 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.³⁵

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

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

for all markets and periods. As we cannot estimate $\hat{\xi}_{imt}^L$ for agents with $ShL_{imt} = 0$, we replace missing $\hat{\xi}_{imt}^L$ with the average $\hat{\xi}_{imt}^L$ among agents with the same experience.³⁶ 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:

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

where X_{imt}^S includes 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:

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

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%, whereas a standard deviation increase in s_{it} improves the probability of sales by 3%.

Once empirical we estimate payoff parameters $\theta = \{\theta^L, \theta^B, \theta^S\}$, 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 $\{HP_{mt}, Inv_{mt}, L_{mt}, B_{mt}, s_{it}, \text{ whether } t < 2005\}$. 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 empirical distributions that we estimate.

□ **First-step estimates: transition of state variables.** The transition process of the four aggregate state variables HP , 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 10-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

³⁵ Allowing for persistent and time-varying unobserved attributes in dynamic discrete choice models with an infinite horizon is an important and difficult question. See Norets (2009); Imai, Jain, and Ching (2009); and Arcidiacono and Miller (2011) for some recent progress.

³⁶ Replacing missing $\hat{\xi}_{imt}^L$ with zero leads to nearly identical estimates of L_{mt} .

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, whereas 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.³⁷

□ **Second-step estimates: structural parameters.** As described in Section 5, 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, 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).³⁸

As 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 nonparametric bootstraps.³⁹ 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 $\{\hat{\beta}^k, \hat{\beta}^{k-1}\}$ is smaller than half of their standard deviation:

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

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. $\{\hat{\beta}_1^k, \hat{\beta}_{2m}^k\}$ for different sets of spline terms are reported in BP2.

³⁷ Alternative specifications are reported in Tables A5 – A9 of BP2.

³⁸ To minimize potential issues with numerical computing, we use the KNITRO optimization procedure for all estimation (including bootstrap simulations), provide analytic gradients for the objective function and the nonlinear constraints, experiment with different starting values, and use 10^{-6} for all tolerance levels.

³⁹ In these bootstrap estimations, we hold estimates of the revenue function and state variables' transition process fixed, because reestimating 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.

TABLE 7 Per-Period Cost Estimates (in \$100,000 2007 Dollars)

	Variations on Main Specification															
	Main Specification				Assumptions on Forward-Looking Behavior				Years of Experience as Skill Measure				Two Cost Parameters per Market			
	$\delta = 0.90$	$\delta = 0$ (myopic)	$\delta = 0.85$	$\delta = 0.95$	C	std(C)	C	std(C)	C	std(C)	C	std(C)	C_{t-2005}	std(C_{t-2005})	C_{t-2005}	std(C_{t-2005})
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)	
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)	0.66***	(0.06)
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)	0.83***	(0.05)
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)	0.77***	(0.06)
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)	0.30***	(0.08)
DANVERS	0.31***	(0.05)	-0.05***	(0.02)	0.25***	(0.05)	0.40***	(0.05)	0.28***	(0.01)	0.30***	(0.06)	0.21**	(0.08)	0.41***	(0.07)
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)	0.58***	(0.06)
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)	0.61***	(0.05)
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)	0.47***	(0.03)
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)	0.37***	(0.04)
MALDEN	0.40***	(0.02)	0.02	(0.02)	0.35***	(0.02)	0.47***	(0.03)	0.35***	(0.01)	0.49***	(0.04)	0.37***	(0.04)	0.45***	(0.08)
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)	0.51***	(0.07)
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)	0.52***	(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)	0.50***	(0.03)
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)	0.50***	(0.03)	0.34***	(0.06)
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)	0.42***	(0.03)
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)	0.43***	(0.07)
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)		

(Continued)

TABLE 7 Continued

	Variations on Main Specification															
	Main Specification							Assumptions on Forward-Looking Behavior							Two Cost Parameters per Market	
	$\delta = 0.90$		$\delta = 0$ (myopic)		$\delta = 0.85$		$\delta = 0.95$		Years of Experience as Skill Measure		C_{t-2005} (11)	$std(C_{t-2005})$ (12)	C_{t-2005} (13)	$std(C_{t-2005})$ (14)		
	C	std(C)	C	std(C)	C	std(C)	C	std(C)	C	std(C)	C	std(C)	C	std(C)		
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	39	39	39	39	39	39	39	39		
Log-likelihood	-12883		-12819		-12892		-12875		-14645		-12779		-12779			
N	41856		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 = $H/25$		
	K (1)	Std(K) (2)	Entry Probability (3)	K (4)	Std(K) (5)	Entry Probability (6)	K (7)	Std(K) (8)	Entry Probability (9)
ARLINGTON	0.44***	(0.03)	0.50	0.95***	(0.03)	0.25	0.54***	(0.03)	0.44
BROOKLINE	0.04	(0.04)	0.65	0.69***	(0.03)	0.33	0.47***	(0.03)	0.43
CAMBRIDGE	0.17***	(0.01)	0.55	0.81***	(0.01)	0.27	0.31***	(0.01)	0.48
CONCORD	0.21***	(0.06)	0.53	0.75***	(0.06)	0.27	0.21***	(0.06)	0.53
DANVERS	0.28***	(0.04)	0.60	0.87***	(0.04)	0.30	0.28***	(0.04)	0.60
DEDHAM	0.34***	(0.03)	0.53	0.88***	(0.03)	0.27	0.34***	(0.03)	0.53
HINGHAM	0.24***	(0.03)	0.63	0.86***	(0.03)	0.32	0.24***	(0.03)	0.63
LEXINGTON	0.04	(0.04)	0.75	0.70***	(0.05)	0.37	0.23***	(0.04)	0.62
LYNN	0.06***	(0.01)	0.67	0.72***	(0.01)	0.33	0.06***	(0.01)	0.67
MALDEN	0.34***	(0.01)	0.52	0.88***	(0.01)	0.26	0.34***	(0.01)	0.52
MARBLEHEAD	0.21***	(0.04)	0.63	0.83***	(0.04)	0.31	0.70***	(0.04)	0.38
MEDFORD	0.30***	(0.03)	0.52	0.83***	(0.03)	0.26	0.39***	(0.03)	0.47
NEEDHAM	0.21***	(0.04)	0.61	0.81***	(0.04)	0.30	0.36***	(0.04)	0.53
NEWTON	0.02	(0.02)	0.66	0.67***	(0.02)	0.33	0.33***	(0.02)	0.50
PEABODY	0.24***	(0.02)	0.61	0.84***	(0.02)	0.30	0.24***	(0.02)	0.61
QUINCY	0.38***	(0.01)	0.55	0.93***	(0.01)	0.27	0.38***	(0.01)	0.55
RANDOLPH	0.29***	(0.02)	0.50	0.81***	(0.03)	0.25	0.29***	(0.02)	0.50
READING	0.07*	(0.04)	0.68	0.74***	(0.05)	0.34	0.30***	(0.04)	0.57
REVERE	0.42***	(0.01)	0.53	0.96***	(0.01)	0.27	0.42***	(0.01)	0.53
SALEM	0.42***	(0.02)	0.49	0.93***	(0.02)	0.25	0.42***	(0.02)	0.49
SOMERVILLE	0.07***	(0.02)	0.61	0.66***	(0.02)	0.30	0.07***	(0.02)	0.61
STOUGHTON	0.15***	(0.02)	0.64	0.78***	(0.02)	0.32	0.15***	(0.02)	0.64
WAKEFIELD	0.09***	(0.02)	0.64	0.72***	(0.02)	0.32	0.09***	(0.02)	0.64
WALPOLE	0.03	(0.03)	0.69	0.71***	(0.03)	0.35	0.03	(0.03)	0.69
WALTHAM	0.25***	(0.02)	0.58	0.82***	(0.02)	0.29	0.25***	(0.02)	0.58
WATERTOWN	0.18***	(0.02)	0.65	0.82***	(0.02)	0.32	0.26***	(0.02)	0.61
WELLESLEY	-0.17***	(0.02)	0.72	0.57***	(0.02)	0.36	-0.06***	(0.02)	0.67
WEYMOUTH	-0.05***	(0.01)	0.74	0.69***	(0.01)	0.37	-0.05***	(0.01)	0.74
WILMINGTON	0.22***	(0.03)	0.61	0.82***	(0.03)	0.30	0.22***	(0.03)	0.61
WINCHESTER	-0.14**	(0.06)	0.70	0.56***	(0.06)	0.35	0.19***	(0.06)	0.54
WOBURN	0.29***	(0.03)	0.58	0.86***	(0.03)	0.29	0.29***	(0.03)	0.58

Note: *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.

□ **Results and robustness.** As shown in Table 7, the per-period cost is \$49,000 on average 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 the role of forward-looking behavior, we estimate the model with discount factor δ equal to zero and report \hat{c} in column (3) of Table 7. For one third of the markets, the estimates are negative or insignificant; they average \$9850 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 $\delta = 0.90$.⁴⁰ In columns (5) – (8), we examine how results change with different 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 reestimate 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 $c_{m,t < 2005}$.⁴¹ 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 $\bar{N}_m^E = 2 * \max(N_{mt}^E)$. As 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 $H_m = \frac{1}{T} \sum_{t=1}^T H_{mt}$.⁴²

The entry cost κ , its standard deviation, and the probability of entry (defined as $\frac{N_{m,t}^E}{\bar{N}_m^E}$), are reported for each market for each assumption. Entry costs increase mechanically with the assumed number of potential entrants. Assuming that $\bar{N}_m^E = \max(N_{mt}^E)$, the average entry cost is \$18,000, which is roughly the commission for transacting 2.5 properties.⁴³ The other two assumptions, $\bar{N}_m^E = 2 * \max(N_{mt}^E)$ and $\bar{N}_m^E = \frac{H_m}{25}$, lead to an estimate of \$79,000 and \$26,800, respectively. These numbers might seem high given the general perception of low entry barriers

⁴⁰ Other dynamic empirical articles using a discount factor of 0.9 include Ryan (2012) and Kennan and Walker (2011).

⁴¹ We also estimated the model using $c_{t \geq 2006}$ and $c_{t < 2006}$. This introduces another state variable (a dummy $\{t < 2006\}$). Results are similar, but estimates of $c_{t \geq 2006}$ are less stable as our sample ends in 2007.

⁴² We experimented additional measures including $2 * \text{mean}(N_{mt}^E)$, $\frac{H_m}{10}$, and $\frac{H_m}{20}$. Results are available upon request.

⁴³ Three markets have negative entry costs, which are necessary to justify high entry rates observed in these markets.

of the realtor brokerage industry.⁴⁴ Therefore, we use the most conservative estimate of entry cost in the counterfactuals. We also report social savings with and without entry costs in all the counterfactual analysis.

□ **Model 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. As 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 based only on measures of the aggregate housing market. Moreover, observed commissions are realized *ex post*, whereas 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, whereas the average fitted expected commission is \$63,900. In Panel B, we report observed versus predicted commissions by market; the differences are small for most markets.⁴⁵ These results indicate that our state variables predict observed revenues.

As 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 versus 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%. 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 data set. 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 www.city-data.com/.⁴⁶ 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.

⁴⁴ The empirical literature generally finds very large switching costs. Geyer and Sieg (2013) estimate that the cost of moving from a rental property to public housing in Pittsburg is about 10 times low income households' average annual income; they interpret this finding as evidence of habit persistence. Kennan and Walker (2011) show that the disutility of migration for young males with a high-school education is often above \$100,000. Although the setting in our article is different from those just cited, we note that our estimate of the costs of switching careers and becoming a real estate broker is about a quarter of the average annual commission earnings; this finding seems small in magnitude when compared with the findings in these other article.

⁴⁵ The largest gaps are \$10,000 for Arlington and \$7000 for Revere. The discrepancy between the observed and predicted L and B for these two markets contributes to the large gaps.

⁴⁶ This resource aggregates information on population and income of US cities and provides an income estimate for years between those of the US decennial census. Last accessed in August 2011.

TABLE 9 Model Fit for Commissions and Exit Probabilities

	Commissions (\$100,000s)		Exit Probabilities	
	Observed (1)	Fit (2)	Observed (3)	Fit (4)
<i>A. By Year</i>				
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
<i>B. By Market</i>				
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

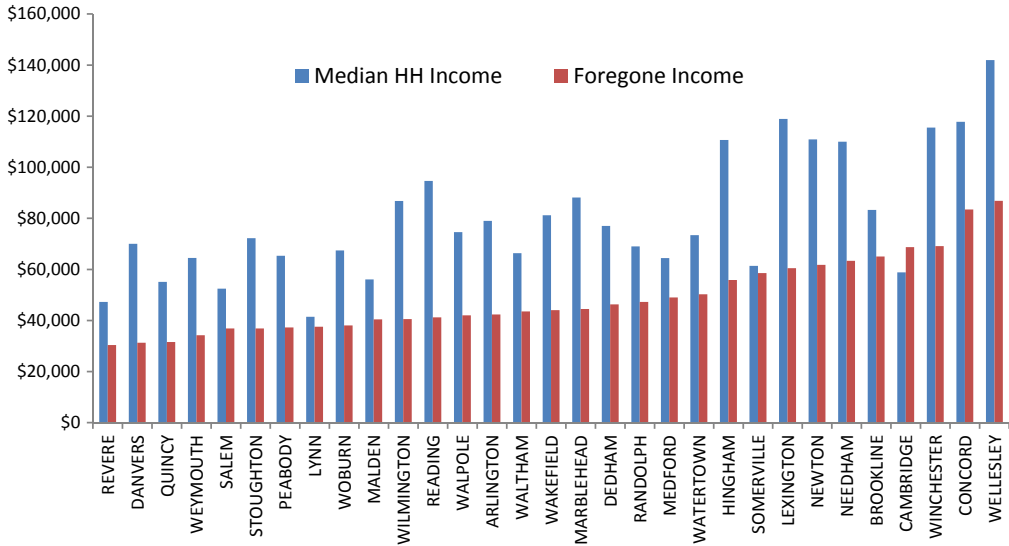
Notes: Commissions are in \$100,000 in 2007 dollars.

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 an 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.

FIGURE 4

FOREGONE INCOME VS. MEDIAN HOUSEHOLD INCOME (2007 \$)



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 because 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.

□ **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. 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 its mean determined by the following equation:

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

where the first and second term sums over incumbents and potential entrants, respectively.⁴⁷ $\tilde{R}(S_i, \beta)$ is agent i 's expected revenue in the counterfactual, $E_{L',B'}[\tilde{V}(S'_i)|S_i]$ is the expectation of his counterfactual value function $\tilde{V}(S'_i)$ over the distribution of L' , B' , and other exogenous state variables, and $\exp(\overline{X_i^L \theta^L}) \equiv E_{\xi}[\exp(X_i^L \theta^L + \xi_i)]$.⁴⁸ Note that both $\tilde{R}(S_i, \beta)$ and $E_{L',B'}[\tilde{V}(S'_i)|S_i]$ depend on the distribution of L' and B' . Equation (16) is similar to equation (4), except that the former takes expectation over agents' entry and exit decisions.

Replacing $E[1\{\cdot\}]$ on the right-hand side of equation (16) with choice probabilities and omitting the dependence of L' on S , we have:

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

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

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

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' for each S , which are fixed points of the equation system defined in (17) and (18). To solve the new equilibrium, we use the MPEC framework to cast the counterfactual analysis as another problem of constrained optimization, where the constraint corresponds to the Bellman equation that is associated with the new equilibrium.⁴⁹ All standard errors of the counterfactuals are calculated using 100 nonparametric bootstrap simulations.

□ **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 nontraditional 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 *US V. National Association of Real Estate Boards* in 1950, which first prohibited coordination of realtor fees.⁵⁰ We simulate the brokerage

⁴⁷ The transition process of exogenous state variables remains unchanged throughout the counterfactual exercises.

⁴⁸ The set of splines for \tilde{V} is the same as that for V .

⁴⁹ One might be concerned about multiple equilibria (i.e., multiple fixed points associated with equation (16)). In all of our counterfactual analyses, despite a large number of starting values that we experimented, we never found more than one fixed point.

⁵⁰ Recent FTC and Department of Justice investigations have also examined Internet and virtual real estate offices. See www.ftc.gov/bc/realestate/index.htm (last accessed February 2011) for additional information on the FTC's investigations of the real estate industry.

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 commission rate only leads to a moderate change in the revenue per agent – \$54,400 versus \$63,300 – as 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 amount 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. As 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 10 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.

Although our simulation rests on the assumption that agents are not capacity constrained, we believe that this assumption is defensible 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, although 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, the socially optimal commission rate is one that minimizes agent idle capacity. As 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%.

□ **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.

As 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, although house prices experience significant appreciation

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)	Reduction in Commission (\$mil) (9)
Actual (5%)	7.78	22.52	153.78	18.05	0.63	0.70			
Counterfactual Commission Rates									
4.75%	8.13 (0.15)	21.70 (0.22)	147.18 (2.37)	18.23 (0.58)	0.63 (0.00)	0.70 (0.00)	90.37 (3.75)	4.25 (0.51)	193.52
4.50%	8.50 (0.16)	20.93 (0.22)	140.81 (2.35)	18.40 (0.57)	0.62 (0.00)	0.71 (0.00)	177.25 (4.05)	8.26 (0.68)	387.03
4.25%	8.91 (0.18)	20.17 (0.23)	134.47 (2.32)	18.57 (0.56)	0.62 (0.00)	0.71 (0.00)	263.90 (4.40)	12.16 (0.87)	580.55
4.00%	9.36 (0.19)	19.44 (0.23)	128.15 (2.29)	18.75 (0.55)	0.61 (0.00)	0.71 (0.00)	350.31 (4.79)	15.95 (1.05)	774.06
3.75%	9.86 (0.21)	18.72 (0.24)	121.85 (2.26)	18.93 (0.53)	0.60 (0.00)	0.71 (0.00)	436.47 (5.20)	19.63 (1.24)	967.58
3.50%	10.42 (0.23)	18.03 (0.24)	115.57 (2.22)	19.11 (0.52)	0.59 (0.00)	0.71 (0.00)	522.36 (5.62)	23.20 (1.42)	1161.09
3.25%	11.04 (0.26)	17.35 (0.24)	109.32 (2.18)	19.29 (0.50)	0.58 (0.00)	0.71 (0.00)	607.95 (6.04)	26.66 (1.59)	1354.61
3.00%	11.74 (0.28)	16.70 (0.24)	103.09 (2.14)	19.47 (0.48)	0.57 (0.00)	0.71 (0.00)	693.23 (6.47)	30.01 (1.77)	1548.12
2.75%	12.54 (0.31)	16.06 (0.24)	96.89 (2.09)	19.66 (0.47)	0.56 (0.00)	0.72 (0.00)	778.19 (6.88)	33.25 (1.93)	1741.64
2.50%	13.46 (0.35)	15.45 (0.24)	90.70 (2.03)	19.84 (0.45)	0.54 (0.00)	0.72 (0.00)	862.82 (7.27)	36.39 (2.10)	1935.15

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 parentheses) are computed from 100 bootstrap simulations.

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.

As 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.

□ **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 MLS. Moreover, our estimates suggest that it may have the support of incumbent agents whose average commissions would increase substantially.

□ **Sensitivity of counterfactuals.** Our methodology makes it relatively straightforward to see how sensitive the counterfactuals are to our modelling choices. We have repeated Table 10 and 11, using the other specifications discussed in Section 6. When agents are less forward-looking ($\delta = 0.85$), cost savings are roughly 10%–12% less, whereas if agents are more forward-looking

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 (0.22)	17.99 (0.24)	116.93 (2.23)	19.12 (0.52)	0.60 (0.00)	0.71 (0.00)	502.23 (5.50)	23.42 (1.42)
B. Counterfactual with Improved Information on Past Agent Performance								
Raise Skill Impact by 20%	8.09 (0.14)	20.83 (0.19)	148.33 (2.33)	17.17 (0.54)	0.66 (0.00)	0.71 (0.00)	62.80 (3.93)	8.90 (0.64)
Raise Skill Impact by 40%	8.43 (0.16)	19.16 (0.18)	142.56 (2.29)	16.41 (0.51)	0.69 (0.00)	0.71 (0.00)	134.32 (4.52)	17.74 (1.02)
Raise Skill Impact by 60%	8.76 (0.17)	17.69 (0.17)	137.60 (2.28)	15.74 (0.48)	0.72 (0.00)	0.71 (0.00)	193.71 (5.08)	25.57 (1.38)
Raise Skill Impact by 80%	9.15 (0.19)	16.26 (0.17)	132.22 (2.28)	15.19 (0.46)	0.75 (0.00)	0.71 (0.00)	269.94 (5.63)	32.94 (1.77)
Raise Skill Impact by 100%	9.53 (0.21)	14.95 (0.17)	127.48 (2.30)	14.68 (0.45)	0.77 (0.00)	0.72 (0.00)	332.76 (6.13)	39.63 (2.09)

Note: Average commissions are in \$100,000 2007 dollars. Standard errors (in parentheses) 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.

($\delta = 0.95$), 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 nonzero 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.⁵¹ 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 per market with a \$500 marginal cost, instead of 91 agents as reported in Table 10. Total cost savings are similar: \$902 million (with marginal costs) versus \$899 million (without marginal costs). Counterfactual results assuming marginal cost is \$100 display a similar modest impact on our estimates of cost savings.

⁵¹ As explained in Section 4, the marginal cost cannot be separately identified from the fixed cost.

8. Conclusion

■ In this article, we use a new data set 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. Entrants compete by taking listings from 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 article is to build and extend a 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. Although we take the absence of price competition as given throughout this article, an interesting topic for future work is understanding forces that could lead to more flexible commission rates.

Appendix A

In this section, we provide some details of the descriptive regressions discussed in Section 3. The first equation that we estimate is:

$$y_{imt} = \rho_y \log(N_{mt}) + \alpha s_{it} + \lambda_m + \tau_t + \theta_1 \log(H_{mt}) + \theta_2 \log(P_{mt}) + \theta_3 \text{Inv}_{mt} + \epsilon_{imt}. \quad (\text{A1})$$

To avoid confounding changes in the composition of agents, we estimate equation (A1) 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 $\log(N_{mt})$.⁵² We report only estimates of ρ_y for the first seven agent cohorts (from 1998 – 2004), because 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 (A1), 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 (A1) interacting ρ_y with a group dummy. Results for log commissions and log number of transactions are reported in columns (1) – (5) and columns (6) – (10) of Table 4, respectively.

As 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 toward zero. Nonetheless, they

⁵² Changes in the composition of cohorts do arise when agents exit. We also estimate equation (A1) 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.

TABLE A1 Property Attributes by Agent Type

Panel A: Across All Markets and Years				
	Inexperienced Agents		Established Agents	
	Mean	Standard Deviation	Mean	Standard Deviation
Sales Price (\$100k)	3.46	(2.52)	3.84	(2.94)
Listing Price (\$100k)	3.71	(2.85)	4.03	(3.25)
Number of Beds	3.31	(1.48)	3.34	(1.46)
Number of Baths	1.94	(0.89)	2.01	(0.94)
Sqft (1000ft ²)	1.85	(1.09)	1.94	(1.15)
Number of Other Rooms	1.80	(1.20)	1.79	(1.20)

Panel B: By Year				
	Inexperienced Agents		Established Agents	
	Sales Price (\$100k)	Number of Bedrooms	Sales Price (\$100k)	Number of Bedrooms
1999	2.58 (2.03)	3.43 (1.54)	3.21 (2.60)	3.36 (1.37)
2000	2.97 (2.54)	3.42 (1.53)	3.64 (2.98)	3.33 (1.33)
2001	3.30 (2.50)	3.31 (1.40)	3.59 (2.97)	3.39 (1.47)
2002	3.64 (2.61)	3.31 (1.43)	3.91 (2.93)	3.33 (1.41)
2003	3.73 (2.49)	3.28 (1.41)	3.99 (2.92)	3.33 (1.46)
2004	3.87 (2.55)	3.26 (1.46)	4.15 (3.03)	3.32 (1.47)
2005	3.90 (2.50)	3.27 (1.54)	4.07 (2.91)	3.35 (1.52)
2006	3.70 (2.52)	3.28 (1.51)	3.94 (2.96)	3.34 (1.53)
2007	3.64 (2.68)	3.26 (1.48)	3.84 (2.98)	3.32 (1.51)

Note: Inexperienced agents is defined as those with below median number of transactions in a given year and market, where as established agents are those with above median number of transactions in a given year and market. All prices are in 2007 dollars, deflated using urban CPI.

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.

Our second set of descriptive regressions examine whether home sellers benefit from more competition among agents:

$$h_{imt} = \rho_h \log(N_{mt}) + \alpha s_{it} + \theta'_X X_{i(h)t} + \lambda_m + \tau_t + \theta_1 \log(H_{mt}) + \theta_2 \log(P_{mt}) + \theta_3 \text{Inv}_{mt} + v_{imt}. \quad (\text{A2})$$

When the number of competing agents in a market increases, the likelihood that a property sells decreases, shown in Table 5, which is a somewhat surprising result given 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 (A2) and report estimates in column (3). The sample size is smaller because only 30% of 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 results suggest that the reduction in sales probability with more agents is not driven by changes in property composition.⁵³

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 (10), and when we also control for property fixed effects in column (11). 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. Overall, our results suggest that increased competition does not make it more likely that a given property will be sold or that it will sell faster.

Appendix B

This Appendix proves the following claim referenced in Section 5.

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

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

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

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

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

is a contraction mapping because $h \leq g$ implies $\Gamma(h) \leq \Gamma(g)$, and $\Gamma(V + a) \leq \Gamma(V) + \beta a$ for $a > 0, \beta \in (0, 1)$. According to Stokey, Lucas, and Prescott (1989), if the contraction mapping operator satisfies $\Gamma[C'] \subseteq C''$, where C' is the set of bounded, continuous, and nondecreasing functions, whereas 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 $\Gamma[C'] \subseteq C''$, we only need to show that if $V(S)$ is nondecreasing in S , then $\int V(S') f(S'|S) dS'$ is nondecreasing in S . Note that:

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

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 $f_\varepsilon \geq 0$.

Q.E.D.

References

- AGUIRREGABIRIA, V. AND MIRA, P. "Sequential Estimation of Dynamic Discrete Games." *Econometrica*, Vol. 75 (2007), pp. 1–53.
- AGUIRREGABIRIA, V. AND NEVO, A. "Recent Developments in IO: Dynamic Demand and Dynamic Games." *Econometric Society World Congress*, 2010.
- ARCIDIACONO, P., BAYER, P., BUGNI, F.A. AND JAMES, J. "Approximating High-Dimensional Dynamic Models: Sieve Value Function Iteration." NBER Working Paper, no. 17890, 2012.
- ARCIDIACONO, P. AND MILLER, R.A. "Conditional Choice Probability Estimation of Dynamic Discrete Choice Models with Unobserved Heterogeneity." *Econometrica*, Vol. 79(6) (2011), pp. 1823–1867.

⁵³ Note that we control for agent observed attributes (including skill) and use a cohort analysis that should alleviate the concern that these negative results are driven by entry of inexperienced agents. However, it is possible that matching between a seller's agent and a buyer's agent becomes harder with an overall increase of inexperienced agents.

- BAJARI, P., BENKARD, L., AND LEVIN, J. "Estimating Dynamic Models of Imperfect Competition." *Econometrica*, Vol. 75(5) (2007), pp. 1331–1370.
- BAJARI, P., CHERNOZHUKOV, V., HONG, H. AND NEKIPELOV, D. "Nonparametric and Semiparametric Analysis of a Dynamic Discrete Game." Working Paper, Stanford University, 2009.
- BARWICK, P.J. AND PATHAK, P. "Supplemental Material to the Cost of Free Entry: An Empirical Study of Real Estate Agents in Greater Boston." Working Paper, MIT, 2014.
- BARWICK, P.J., PATHAK, P., AND WONG, M. "Market Power and the Commission Puzzle." Working Paper, MIT, 2014.
- BAYER, P., McMILLAN, R., MURPHY, A., AND TIMMINS, C. "A Dynamic Model of Demand for Houses and Neighborhoods." NBER Working Paper, no. 17250, 2011.
- BERNHEIM, D. AND MEER, J. "Do Real Estate Agents Add Value When Listing Services Are Unbundled?" *Economic Inquiry*, Vol. 51(2) (2012) pp. 1166–1182.
- BERRY, S., LEVINSOHN, J., AND PAKES, A. "Automobile Prices in Market Equilibrium." *Econometrica*, Vol. 63(4) (1995) pp. 841–890.
- BISHOP, P.C., BARLETT, H., AND LAUTZ, J. "National Association of Realtors, Member Profile." Available from National Association of Realtors, 2007.
- CAR. "2008 Survey of California Home Buyers." California Association of Realtors, 2008. california.realestaterama.com/2008/09/19/car-2008-survey-of-california-home-buyers-ID0233.html. Assessed August 2011.
- CHEN, X. "Large Sample Sieve Estimation of Semi-Nonparametric Models, Chapter 76." In J.J. Heckman and E. Leamer, eds., *Handbook of Econometrics (Volume 6B)* Vol. 5. Amsterdam, The Netherlands: Elsevier, 2007.
- COLLARD-WEXLER, A. "Demand Fluctuations in the Ready-Mix Concrete Industry." *Econometrica*, Vol. 81(3) (2013), pp. 1003–1037.
- DUBÈ, J.-P., FOX, J., AND SU, C.-L. "Improving the Numerical Performance of BLP Static and Dynamic Discrete Choice Random Coefficients Demand Estimation." *Econometrica*, Vol. 80(5) (2012), 2231–2267.
- DUNNE, T., KLIMEK, S.D., ROBERTS, M.J., AND XU, D.Y. "Entry, Exit and the Determinants of Market Structure." NBER Working Paper no. 15313, 2011.
- ERICSON, R., AND PAKES, A. "Markov Perfect Industry Dynamics: A Framework for Empirical Work." *Review of Economic Studies*, Vol. 62(1) (1995) pp. 53–82.
- FARIAS, V., SAURE, D., AND WEINTRAUB, G. "An Approximate Dynamic Programming Approach to Solving Dynamic Oligopoly Models." *RAND Journal of Economics*, Vol. 43(2) (2012), pp. 253–282.
- FOWLIE, M., REGUANT, M., AND RYAN, S., "Pollution Permits and the Evolution of Market Structure." Working Paper, MIT, 2011.
- FRIEDMAN, J.H. "Multivariate Adaptive Regression Splines." *Annals of Statistics*, Vol. 19(1) (1991), pp. 1–67.
- . "Fast MARS." Technical Report, Department of Statistics, Stanford University, 1993.
- FTC. "Selling Your Home? Tips for Selecting a Real Estate Professional." FTC Facts for Consumers, Bureau of Competition, 2006. www.ftc.gov/bc/edu/pubs/consumer/homes/zrea01.pdf. Accessed August 2011.
- GEYER, J. AND SIEG, H. "Estimating a Model of Excess Demand for Public Housing." *Quantitative Economics*, Vol. 4(3) (2013) pp. 483–513.
- HAIJAVASSILOU, V. "Some Practical Issues in Maximum Simulated Likelihood." In R. Mariano, T. Schuermann, and M.J. Weeks, *Simulation-Based Inference in Econometrics: Methods and Applications* Vol. 1. New York, NY: Cambridge University Press, 2000.
- HAN, L. AND HONG, S.-H. "Testing Cost Inefficiency under Free Entry in the Real Estate Brokerage Industry." *Journal of Business and Economic Statistics*, Vol. 29(4) (2011), pp. 564–578.
- HENDEL, I. AND NEVO, A. "Measuring the Implications of Sales and Consumer Inventory Behavior." *Econometrica*, Vol. 74 (2006) pp. 1637–1674.
- HENDEL, I., NEVO, A. AND ORTALO-MAGNÉ, F. "The Relative Performance of Real Estate Marketing Platforms: MLS versus FSBOMadison.com." *American Economic Review*, Vol. 5 (2009) pp. 1878–1898.
- HOTZ, V.J. AND MILLER, R. "Conditional Choice Probabilities and the Estimation of Dynamic Models." *Review of Economic Studies*, Vol. 60 (1993) pp. 497–529.
- HSIEH, C.-T. AND MORETTI, E. "Can Free Entry Be Inefficient? Fixed Commission and Social Waste in the Real Estate Industry." *Journal of Political Economy*, Vol. 111(5) (2003) pp. 1076–1122.
- IMAI, S., JAIN, N., AND CHING, A. "Bayesian Estimation of Dynamic Discrete Choice Models." *Econometrica*, Vol. 77 (2009) pp. 1665–1682.
- JIA, P. AND PATHAK, P. "The Impact of Commissions on Home Sales." *American Economic Review, Papers & Proceedings*, pp. 475–479, 2010.
- JUDD, K. *Numerical Methods in Economics*. Cambridge, Mass.: MIT Press, 1998.
- KADIYALI, V., PRINCE, J., AND SIMON, D. "Is Dual Agency in Real Estate Transactions a Cause for Concern?" *Journal of Real Estate Finance and Economics*, Vol. 48(1) (2014), pp. 164–195.
- KEANE, M.P. AND WOLPIN, K.I. "The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation and Interpolation: Monte Carlo Evidence." *Review of Economics and Statistics*, Vol. 76 (1994) pp. 648–672.

- KENNAN, J., AND WALKER, J. "The Effect of Expected Income on Individual Migration Decisions." *Econometrica*, Vol. 79(1) (2011) pp. 211–251.
- KRISTENSEN, D. AND SCHJERNING, B. "Implementation and Estimation of Discrete Markov Decision Models By Sieve Approximation." Work in progress, UCL, 2012.
- KUMAR, S. AND SLOAN, I. "A New Collocation-Type Method for Hammerstein Integral Equations." *Math. Computation*, Vol. 48 (1987) pp. 585–593.
- LEVITT, S. AND SYVERSON, C. "Antitrust Implications of Home Seller Outcomes when using Flat-Fee Real Estate Agents." Brookings-Wharton Papers on Urban Affairs, 2008a.
- . "Market Distortions when Agents are Better Informed: The Value of Information in Real Estate Transactions." *Review of Economics and Statistics*, Vol. 90 (2008b) pp. 599–611.
- MANKIW, G. AND WHINSTON, M. "Free-Entry and Social Inefficiency." *Rand Journal of Economics*, Vol. 17(1) (1986) pp. 48–58.
- MELNIKOV, O. "Demand for Differentiated Durable Products: The Case of the U.S. Computer Printer Market." Manuscript, Department of Economics, Yale University, 2000.
- MICHELANGELI, V. "Does It Pay to Get a Reverse Merger?" Working Paper, Congressional Budget Office, 2010.
- NIPA. "National Income and Product Accounts Table 5.4.5. Private Fixed Investment in Structures By Type, 1929–2008." Bureau of Economic Analysis, 2008. www.bea.gov/national/nipaweb/index.asp. Assessed August 2011.
- NORETS, A. "Inference in Dynamic Discrete Choice Models with Serially Correlated Unobserved State Variables." *Econometrica*, Vol. 77 (2009) pp. 1665–1682.
- RUST, J. "Estimation of Dynamic Structural Models, Problems and Prospects: Discrete Decision Processes." In C. Sims, ed., *Advances in Econometrics: Sixth World Congress*. Cambridge, UK: Cambridge University Press, 1994.
- . "Using Randomization to Break the Curse of Dimensionality." *Econometrica*, Vol. 65(3) (1997) pp. 487–516.
- RUTHERFORD, R.C., SPRINGER, T.M. AND YAVAS, A. "Conflicts between Principals and Agents: Evidence from Residential Brokerage." *Journal of Financial Economics*, Vol. 76 (2005) pp. 627–665.
- RYAN, S. "The Costs of Environmental Regulation in a Concentrated Industry." *Econometrica*, Vol. 80(3) (2012), pp. 1019–1061.
- SANTOS, M.E. "Numerical Solution of Dynamic Economic Model." In J.B. Taylor and M. Woodford, eds., *Handbook of Macroeconomics (Volume 1A)* Amsterdam, The Netherlands: Elsevier, 1999.
- SEIM, K. "An Empirical Model of Firm Entry with Endogenous Product-Type Choices." *RAND Journal of Economics*, Vol. 37(3) (2006) pp. 619–640.
- STOKEY, N., LUCAS, R., AND PRESCOTT, E. *Recursive Methods in Economic Dynamics*. Cambridge, Mass.: Harvard University Press, 1989.
- SU, C.-L. AND JUDD, K. "Constrained Optimization Approaches to Estimation of Structural Models." *Econometrica*, Vol. 80(5) (2012), pp. 2213–2230.