Entry and Inefficiency in the Real Estate Brokerage Industry: Empirical Evidence and Policy Implications*

Lu Han
University of Toronto
lu.han@rotman.utoronto.ca

Seung-Hyun Hong
University of Illinois
hyunhong@uiuc.edu

July 8, 2008

Abstract

This paper studies entry in local real estate brokerage to investigate sources of potential cost inefficiency under free entry. Building upon recent empirical work on games of incomplete information, we construct a structural entry model and estimate the model using nested pseudo likelihood algorithms. Using data from 5% PUMS in 2000, we find direct evidence for two sources of cost inefficiency – wasteful non-price competition and loss of economies of scale. Using our estimates, we further investigate welfare implications of prohibiting rebates on commissions and the diffusion of the Internet. We find that rebate bans are welfare-reducing, not only because they discourage price competition from discount brokers, but also because they encourage excessive entry by traditional full-commission brokers. Removing these rebate bans would decrease the equilibrium number of realtors by 5.8% and reduce total variable costs by 3.9%. Welfare implications of the Internet diffusion are mixed, however. An increase in the number of Internet adopters alone could encourage traditional brokers' entry by helping them reach and match potential clients. However, a commensurate increase in online search intensity among the Internet adopters may discourage such entry by facilitating the development of alternative brokerage models, such as discount brokerage and FSBOs.

*We thank seminar participants at the Minneapolis Fed, University of California at Berkeley, University of Toronto, University of Wisconsin, 2008 IIOC conference and summer real estate symposium. All errors are our own. Max Rempel provided excellent research assistance in data processing. Contact information: Lu Han: Rotman School of Management, University of Toronto, 105 St. George St., Toronto, Ontario, Canada M5S 3E6. Seung-Hyun Hong: Department of Economics, University of Illinois at Urbana-Champaign, 428 David Kinley Hall, 1407 West Gregory Drive, Urbana, IL 61821.

1 Introduction

It is well known that entry can lead to social inefficiency under certain conditions. While a large theoretical literature on entry and social efficiency exists,[1] there are few empirical studies testing its direct implications. Testing inefficiency under free entry is difficult, not only because data on the relevant cost and benefit measures are difficult to obtain, but also because entry decisions are endogenous.[2] This paper therefore estimates a structural entry model and contributes to the literature by providing direct empirical evidence for inefficiency under free entry for the U.S. residential real estate brokerage industry. By efficiency, we mean cost efficiency in term of minimizing total costs of real estate broker services.

The U.S. residential real estate brokerage industry provides a particularly important setting for studying entry decisions. According to the Department of Justice, real estate agents earned $93 billion in commissions in 2006, which accounted for about 1% of GDP. Membership in the National Association of Realtors (NAR) nearly doubled between 1997 and 2006. Given the size and growth of the real estate brokerage industry, it is not surprising that the entry and competition in this industry have been recurrently featured in news reports and policy debates (see, e.g., White (2006)). Recently, the Department of Justice has brought several antitrust concerns to the forefront of economic debates about the industry.[3]

The nature of the real estate brokerage industry implies two sources of cost inefficiency under free entry. First, the industry does not have significant barriers to entry, if entry is perceived as gaining a license in order to practice (see DOJ and FTC Report (2007)). In 2004, there were 1.9 million active real estate brokers and agents (see GAO Report (2006)). Licensed agents and brokers tend to produce a standardized package of service, generating a "business stealing" effect. When average costs decline with output, a large number of entrants result in a significant loss of economies of scale, suggesting a first source of cost inefficiency. Second,

[1] See, e.g., Anderson, DePalma and Nesterov (1995), Chamberlain (1933), Dixit and Stiglitz (1977), Mankiw and Whinston (1986)and Sutton (1991).

[2] A notable exception is Berry and Waldfogel (1999), which estimates an entry model in the radio broadcasting industry and recovers the parameters for the demand function and the distribution of fixed costs. Although their estimates provide the strong evidence for social inefficiency under free entry, actual entry and exit are rare in the radio industry because of regulation by the Federal Communications Commission (Sweeting (2006)).

[3] Examples include an antitrust case against the Kentucky Real Estate Commissions for prohibiting agents from giving consumers a rebate and an antitrust suit against the National Association of Realtors for the Association's Virtual Office Website policy, both filed by the DOJ in 2005.

unlike most other industries, the real estate brokerage industry is characterized by relatively inflexible commission rates. Empirical evidence shows that real estate brokers often charge a commission of between 5% and 6% with little variance among firms, across housing markets or over time (Hsieh and Moretti (2003)).[4] In addition, several states banned agents from offering commission rebates to their clients, which explicitly prohibited agents from cutting prices. Therefore, it is unlikely that consumers would benefit from entry through price competition. Moreover, given that agents cannot compete on prices, free entry may lead to wasteful non-price competition in term of increasing realtor visibility and fighting for potential clients (Yinger (1981)), suggesting a second source of cost inefficiency.[5]

To test these two sources of cost inefficiency under free entry, this paper builds a simple structural model to estimate the entry decisions by real estate agents and brokers. Structural estimation of entry decision is useful for three reasons. First, as noted by Hsieh and Moretti (2003), the main difficulty in evaluating the inefficiency in the real estate brokerage industry is the lack of the data on costs. Structural estimation allows us to recover various components of the cost function and therefore to directly test different sources of inefficiency under free entry. Second, structural estimation allows us to overcome the endogeneity associated with entry decisions and therefore to evaluate the equilibrium effect of entry. Third, structural estimation allows us to explore a range of issues related to allegations of antitrust violations in the real estate brokerage industry – such as the state laws on prohibiting commission rebates – and to compute the associated welfare loss through counterfactual policy experiments.

Our empirical methodology relies on a basic insight of the traditional discrete-choice approach: the observed entry decisions are an indicator of the underlying profitability. That is,

[4]There are many explanations for uniform commission rates in the real estate brokerage industry. One possible explanation is that, while there are large number of agents in each city, there are relatively few large brokerage companies in most areas. Since brokers are allowed to control fee policies among their agents, sustaining collusive behavior is easier than if each agent set her own commission rate (Yavas (2001)). Such collusion is further facilitated by the fact that each home transaction involves both a seller's agent and a buyer's agent (Levitt and Syverson (2007)). In addition, the constrained variations on commission rates can also be attributed to the reluctance of homesellers to list their home at a commission rate below the prevailing rate. Homeowners must compete among themselves for the sales effort of agents and they may be unwilling to risk offering a lower commission rate (Goolsby and Childs (1988)). Finally, Williams (1998) provides a competitive equilibrium model in which a fixed commission rate across all brokers and clients minimizes their agency problems, to explain the constant 6% broker contract in U.S. residential housing markets.

[5]It is worth noting that many underwriters on initial public offerings (IPOs) in the U.S. charge exactly 7 percent (Chen and Ritter (2002)). Unlike the real estate brokerage industry, the entry barrier in the IPO underwriting industry is substantially higher. Therefore the two sources of cost inefficiency are unlikely to be present in the IPO underwriting industry.

potential agents decide to enter the market if and only if the expected variable profit exceeds the fixed cost plus reservation wages. To capture the empirical features of the real estate brokerage industry, we extend the standard entry model in two ways. First, given the large number of potential entrants, it is practically impossible for each potential agent to observe fixed costs and reservation wages of all other potential agents. The entry model in this paper is therefore characterized by an incomplete information setting in which fixed cost and reservation wage of each potential agent have a private component, which is observed to the agent herself but not to other potential agents.[6] Each agent thus forms a conjecture on other agents' entry strategies. Given this conjecture and private information, each potential agent simultaneously decides whether to enter. Further exploiting the fact that there are a large number of potential entrants in each market, we assume that each potential agent cares only about the average entry probability. Thus the resulting Bayesian Nash equilibrium is reduced to fixed points in entry probability in that agents' conjectures on average entry probability should coincide with the entry probability predicted from the model.

A second feature of our model is that we allow for a rich cost function that nests the two potential sources of inefficiency under free entry. First, by choosing a flexible functional form for the average variable cost function, we allow for the possibility that there may be economies or diseconomies of scale in the real estate brokerage market. If entry results in a loss of economies of scale, then we should expect that average costs of providing realtor service decline with the number of home transactions that a realtor facilitates. Second, given the lack of price competition in the real estate brokerage industry, our average variable cost function includes not only the cost of executing the transaction but also the cost of resources absorbed in obtaining potential clients. If entry results in an inefficiently large commitment of resources in promoting brokerage services, then we expect that average costs of wasteful non-price competition increase with the number of entrants.

Using the the 5 percent sample of the *Census of Population and Housing* in 2000, we estimate the model by employing a nested pseudo likelihood (NPL) algorithm (Aguirregabiria and Mira (2002, 2007)) in which the outer algorithm iterates on the choice probability to solve

[6] In this sense, we build the model upon the particularly recent empirical work on games with incomplete information (e.g. Augereau, et al. (2006); Bajari, et al. (2006); Seim (2006); Sweeting (2004)) and modify it to the case with many potential agents.

the fixed point problem, while the inner algorithm maximizes a pseudo likelihood function of individuals' entry decision given the fixed choice probability. This approach is attractive in our setting, because the agent's conjecture about market entry rate enters both the revenue and cost functions nonlinearly, which complicates the use of a standard fixed point algorithm.

Our first empirical result confirms the hypothesis that entry leads to a loss of economies of scale. In particular, the estimated average cost function in the current markets is downward-sloping, suggesting excessive entry in producing residential realtor services. Our second main empirical result is a confirmation of the hypothesis that entry leads to more intensified non-price competition in fighting for potential clients. In particular, the cost of competing with other agents for clients significantly increases with the number of realtors. Putting these two results together, we find that entry leads to inefficiently large expenses not only on producing, but also on marketing real estate brokerage services. In an average MSA in 2000, a 10% increase in the number of realtors increases the average variable cost of each transaction by 4.8% from $2,392 to $2,508 for a typical transaction, driven mostly by the wasteful non-price competition. These cost estimates are robust to various specification checks.

Finally, using our estimates, we further perform counterfactual experiments to investigate the welfare impact of anti-rebate rules which have often been criticized for discouraging price competition. We find that rebate bans are welfare-reducing, not only because they suppress price competition from discount brokers, but also because they encourage excessive entry by full-commission brokers. Removing these rebate bans would reduce realtor revenues, thereby decreasing the equilibrium number of realtors by 5.8% and reducing total variable costs by 3.9% in an average metropolitan area. Using the same methodology, we also examine the effect of the Internet diffusion on agents' entry decisions, but we find mixed welfare implications. Specifically, we find that an increase in the Internet adoption rate, if not accompanied by an increase in online search intensity, can lead to excessive entry by traditional brokers, likely because it helps them reach and match potential clients. A commensurate increase in the online search intensity, however, discourages entry by traditional brokers, presumably by facilitating the alternative real estate brokerage business models, such as discount brokers and FSBOs.

One limitation of our model is that it examines only cost inefficiency under free entry. From a social perspective, entry is optimal when at the margin, the costs of the resources absorbed are equal to the benefits measured in terms of the value of enhanced functioning in the real estate market. Given the lack of data on each individual transaction and each realtor's market share, we cannot fully specify the relevant benefits of entry. However, we provide robustness checks for whether entry by traditional full-commission realtors leads to dramatic benefits for consumers either by lowering commission fees or by improving quality of the service. First, using the commission data from the 1988-2002 *Consumer Expenditure Survey* and from the *Real Trends Brokerage Performance Reports*, we find that an increase in the number of full-commission realtors does not necessarily reduce the market commission rates. Second, using the data from the 2001 National Association Realtor's *Home Buyer and Seller Survey*, we find that competition among traditional realtors does not help home buyers reduce searching time or help home sellers reduce time on the market, nor does it improve the consumer satisfaction with their realtors. Thus, our welfare result based on the cost side analysis, while limited, is suggestive of the total social welfare inefficiency in the real estate brokerage industry.

There are at least two policy lessons that can be drawn from this exercise. First, with inflexible commission rates, realtors inefficiently engage in wasteful non-price competition rather than price competition. Improving market efficiency therefore requires public policy interventions that encourage price competition. Second, restricting excessive entry by traditional realtors may yield nontrivial cost savings by realizing economies of scale. Given the central role played by the multiple listing mechanisms, one step would be to remove the MLS barrier and allow participation by discount brokers. By the end of 2007, three out of thirteen states have abolished the anti-rebate rules. More recently, in May, 2008, the Department of Justice and the National Association of Realtors reached a major antitrust settlement that freed online discount brokers from using the multiple listing services. Given the findings in this paper, we believe that these public policy interventions are judicious and could have a marked impact on lowering costs of the real estate brokerage service.

This paper contributes to the literature in two respects. First, it provides a first *direct* empirical test of cost inefficiency under free entry in the real estate brokerage industry. It is well known that non-price competition, combined with free entry, could result in inefficiently large resources allocated to brokerage service and hence a reduction in welfare (Crockett (1982), Miceli (1992)). Despite the importance of the brokerage industry and the potentially large welfare loss, little empirical evidence has been provided. A seminar study by Hsieh and Moretti (2003) provides a first empirical test of the theory. Using the Census data in 1980 and 1990, they find that, in cities with higher house prices (1) there are more real estate agents relative to the city's workforce, (2) these agents are less productive (measured by sales per agent or sales per hours worked), and (3) wages for agents are not higher than wages in cities with lower housing prices. From these findings, they infer that free entry leads to an inefficient increase in costs, which dissipates any economic profit and generates a significant social loss. This conclusion serves as a catalyst to an investigation of competition in the real estate brokerage industry by the Federal Trade Commission and the Department of Justice Antitrust Division.[7] While intuitively appealing, their approach is *indirect*, as the effect of entry on social cost is not directly estimated, but rather derived from an estimated relationship between house price and productivity. Interpreting indirect evidence requires imposing potentially arbitrary assumptions.[8] In addition, average productivity itself does not reveal the sources of inefficiency. In contrast, we employ a structural approach, which allows us to directly test whether average costs of producing realtor service decline with output and whether average costs of non-price competition increase with entry. In addition to providing direct evidence on cost inefficiency under free entry, this approach also allows us to quantify different sources of inefficiency and to evaluate the welfare consequence of some anti-competitive rules in the brokerage industry.

Second, this paper also emphasizes several methodological issues that are of particular importance in studying real estate brokerage markets. One difficulty of a direct approach in testing the inefficiency of the real estate brokerage industry stems from the lack of the detailed data on individual agents' market shares and costs, as well as the endogeneity of entry decisions. This paper overcomes these difficulties, first by exploiting the observed agents' revenue and entry decisions to make inferences about the underlying transaction volume and costs, and second, by imposing equilibrium conditions in estimating entry decisions. In this respect, the paper bears a resemblance to Berry and Waldfogel (1999), which recovers the distribution of fixed costs in radio industry. Because of a large number of potential agents, however, we use

[7]The Federal Trade Commission (FTC) and the Department of Justice Antitrust Division (DOJ) held a public workshop in October 2005 to address issues affecting competition in the residential real estate brokerage industry. For details, see DOJ and FTC Report (2007).

[8]As noted by Hsieh and Moretti (2003), "The main difficulty is that we do not observe costs, and we need to rely on assumptions that are necessarily arbitrary."

an incomplete information framework as in Seim (2006), and build our model more closely related to social interactions models in Brock and Durlauf (2001).

The paper is organized as follows. Section 2 constructs our equilibrium model and discusses our estimation method. Section 3 describes our data and variables. Section 4 presents the estimation results and discusses their implications, including the results from counterfactual experiments. Section 5 examines two simplifying assumptions in our model and provides suggestive evidence as to potential benefits from free entry. Section 6 concludes the paper.

2 Model

2.1 Set-up

We begin with a simple version of the model in which house price and the number of transactions in a market are exogenously given. There are M different markets. In each market m, S_m potential entrants simultaneously decide whether to enter the market or not. To examine a potential realtor's entry decision, consider a two-stage model: in the first stage, potential entrants simultaneously decide whether to enter the market; in the second stage, realtors in the market compete with each other and realize their profits. If potential agent i enters market m, then her post-entry profit is given by

$$\pi_{im} = R_{im}(N_m) - C_m(q_{im}, N_m) - F_{im}, \tag{1}$$

where $R_{im}(\cdot)$ denotes the revenue for agent i in market m which depends on the number of realtors N_m.[9] For each real estate transaction, realtors provide their clients with various services related to selling or buying houses, and q_{im} denotes the number of transactions carried out by each realtor i. The cost function consists of variable cost $C_m(\cdot)$ and fixed costs F_{im}. Note that we allow the market level heterogeneity in the variable cost function and the individual level heterogeneity in the fixed cost function. This is because realtors tend to provide a standardized package of service, the costs of which are likely to be market specific. In contrast, fixed costs could differ across agents. Though costs associated with obtaining licenses might be similar

[9] Ideally, the revenue function $R_{im}(\cdot)$ should depend on the number of transactions q_{im} as well. However, the Census data report R_{im} but not q_{im}. We therefore infer q_{im} from observed R_{im} and include it in the variable cost function $C_m(\cdot)$ only. Section 2.4. provides details on how to infer q_{im}.

in the same market, agents may incur additional startup costs to learn about neighborhoods, local real estate markets, related tax laws, and information on financing.

Potential agent i enters the market as long as her expected post-entry profit is positive. Given the simultaneous setting, the model does not explicitly distinguish between incumbent realtors and new entrants. In reality, existing incumbents may have better reputation and larger network. In addition, they do not need to incur additional sunk costs associated with entry. The difference between existing incumbents and new entrants is captured by allowing both observed and unobserved heterogeneity in R_{im}, q_{im}, and F_{im}.

The low entry barriers in the real estate brokerage industry imply that each market contains a large number of potential entrants. This has two implications for our model. First, a large number of agents suggest that our setting is more similar to "social interactions" models as in Brock and Durlauf (2001), rather than common settings in industrial organization in which the number of firms is limited. Because it is practically impossible for each potential agent to know the fixed costs of all other individuals in the market, we assume that there is private information component in fixed costs. In other words, each agent observes her own fixed costs and knows about other agents' fixed costs only up to the aggregate distribution.

A second implication of a large number of agents is that the common approach in entry literature (e.g. Berry (1992); Bresnahan and Reiss (1991); Berry and Waldfogel (1999)) of exploiting a free entry equilibrium condition – $\pi_{im}(N_m^*) \geq 0$ and $\pi_{im}(N_m^* + 1) < 0$, where N_m^* is the equilibrium number of firms – will be problematic for two reasons. First, N_m tends to be large in each market, suggesting that the value of the profit under N_m is likely to be very close to that under $N_m + 1$. Second, the 5 percent sample of Census of Population and Housing does not allow us to observe the exact number of realtors out of the whole population. Though the estimate for N_m from the Census is unbiased, it is hard to distinguish between N_m and $N_m + 1$. For this reason, we consider slightly different but still related equilibrium conditions as described in Section 2.4.

2.2 Unobservables and Conjectures

We consider two sources of unobservables. The first is private information on fixed costs. Though some components of fixed costs can be captured by observed variables, the idiosyncratic component of fixed costs, which we denote by ω_{im}, is unobserved both to econometricians and to other agents in the market. We assume that ω_{im} is an independently and identically distributed draw from a distribution known to all agents. When making entry decisions, each agent faces uncertainty about other agents' idiosyncratic fixed costs, which leads to uncertainty about other agents' entry decisions. The second source of unobservable, denoted by η_{im}, is agents' uncertainty about the demand shock to the revenue realized in the second stage, such as unexpected housing booms or slumps.[10]

Both sources of unobservables imply that potential agents need to form their expectations of variable profits in order to decide whether to enter in the first stage. In other words, potential agents enter as long as $E(V_i) \geq F_i$, where $V_i = R_i - C(q_i)$, suppressing the subscript m, and the expectation is taken over all other players' choices and agent uncertainty. The first type of unobservable, however, entails further implications for equilibrium conditions. Given uncertainty about other agents' fixed costs, agent i forms her subjective belief on other agents' actions and chooses her entry strategy based on her private information about F_i, which implies the choice probability of entry. Namely, $\Pr(d_i = 1) = \Pr(E(V_i) \geq F_i)$, where d_i is an indicator for agent i's entry. To the extent that agents' conjectures are rational, the equilibrium requires that the choice probability of entry of all agents should coincide with agents' subjective beliefs on other agents' entry.[11]

Note that V_i is a function of other agents' choices, denoted by d_{-i}, so that agents' conjectures, denoted by $\sigma(d_{-i})$, are contained in $E(V_i(d_{-i}))$. However, $V_i(d_{-i})$ may be a nonlinear function, in which case the computation of $E(V_i(d_{-i}))$ is likely to be demanding. Most literature on games with incomplete information (e.g. Aradillas-Lopez (2005); Augereau, et al. (2006); Bajari, et al. (2006); Seim (2006); Sweeting (2004)) therefore considers a payoff function in the reduced form and assumes that the number of entrants enters the payoff function linearly, thereby directly obtaining $E(V_i(d_{-i})) = V_i(E(d_{-i})) = V_i(\sigma(d_{-i}))$. We need to exploit similar simplifications, especially because the number of potential entrants in our setting is too large to allow for any tractable computation of the expectation. The assumption of a linear

[10] One difference between these two unobservables is that η_{im} is likely to be known to agents in the second stage, whereas ω_{im} is unknown to other agents even in the second stage.

[11] See Brock and Durlauf (2001) for similar definition of equilibrium. Most literature on games with incomplete information (e.g. Aradillas-Lopez (2005), Bajari, et al. (2006), Seim (2006) and Sweeting (2004)) use a similar Bayesian Nash equilibrium defined as fixed points in probability space.

profit function is very restrictive in our setting, however, since it prevents us from examining various sources of cost inefficiency. Consequently, we consider nonlinear functional forms for variable profits discussed in the subsequent sections, but impose the following assumptions on agent conjectures.

Specifically, potential agent i in market m first anticipates the average probability of agent entry, denoted by σ_m. Based on σ_m and information available in the first stage, agent i then predicts her expected revenue and quantity in the second stage. The assumption is that instead of making complicated calculation of expectation of variable profit, agent i simply conjectures the average number of realtors in the market[12] and then treats σ_m as known information in computing predicted variable profits. Note that under this assumption, if we do not allow for η_{im}, then agent i should immediately know her future revenue given σ_m, since there is no other uncertainty. For this reason, we introduce η_{im} to reflect other types of uncertainty that allow for the difference between predicted revenue and actual revenue in the second stage.

Given this assumption on agent conjectures, we can model the choice probability of agent entry. The following subsections describe details about the profit function in our model, and construct the equilibrium in terms of fixed points in entry probability.

2.3 Revenue and Costs

2.3.1 Revenue Function

In the second stage, agent i earns the revenue from commission fees. That is, $R_{im} = \sum_{k=1}^{q_{im}} c_{ikm} \times P_{ikm}$, where P_{ikm} is the price of house k sold by agent i in market m, and c_{ikm} is the commission rate for each transaction. Since we do not observe c_{ikm} and P_{ikm} on each individual transaction, and q_{im} by each individual realtor, we cannot construct a fully structural model for the second stage competition. Instead, we specify the revenue function for agent i in market m in the following reduced form

$$\log(R_{im}) = \gamma_0 + \gamma_1 Q_m + \gamma_2 N_m + \gamma_3 S_m + f(P_m) + Z_m^r \delta^r + X_{im}^r \beta^r + \eta_{im}, \tag{2}$$

where Q_m denotes total number of transactions in market m, N_m is total number of realtors in market m, S_m is total labor force in market m, Z_m^r is a vector of market characteristics,

[12] Given the expected realtor share σ_m and total number of potential agents S_m, the expected number of entrants N_m is $\sigma_m \times S_m$.

X^r_{im} is a vector of agent i's characteristics, and γ, δ^r, and β^r are parameters. The superscript r indicates that the variable or the parameter determines revenue. P_m denotes a vector of houses prices in market m, and $f(P_m)$ is a function of the distribution of housing prices. A simple example of this function is $f(P_{1m}, P_{2m}, \ldots, P_{Jm}) = \gamma_4 \bar{P}_m$, where $\bar{P}_m = \frac{1}{J}\sum_{j=1}^{J} P_{jm}$. An error term η_{im} reflects components in revenue unknown to agent i in the first stage.

2.3.2 Fixed Cost Function

On the cost side, we assume real estate agents' costs are characterized by a fixed cost, F_{im}, and a variable cost, $C(q_{im})$. Fixed costs measure the costs of becoming a real estate agent. They are assumed to vary both across markets and across agents within the same market. Using the superscript f to indicate that the variable or the parameter determines fixed cost, we write the fixed costs as

$$F_{im} = Z^f_m \delta^f + X^f_{im} \beta^f + \omega_{im}, \tag{3}$$

where β^f and δ^f are parameters to be estimated, and Z^f_m is a vector of market characteristics, including information on each market's licensing requirements. A vector of individual characteristics for agent i in market m is denoted by X^f_{im} which reflects the reservation wages that a potential agent could earn if she worked outside of the real estate brokerage sector within the same market. Different people may have different reservation wages, depending on their age, education, experience and skill set. The unobservable component of an individual's fixed costs, ω_{im}, follows an i.i.d standard normal distribution. It is observed to the individual herself, but not to other potential entrants or econometricians.

Note that we cannot distinguish reservation wages from standard fixed costs associated with starting real estate brokerage services. As a result, our estimates for fixed costs would include reservation wages and thus overestimate standard fixed costs. For this reason, this paper does not focus on cost inefficiency stemming from high fixed costs.

2.3.3 Average Variable Cost Function

Variable costs measure the costs involved in facilitating each house transaction. They are assumed to vary across markets and change with the number of housing transactions that agents facilitate. Heterogeneity in housing quality, however, can affect the quality of real estate brokerage services, thereby affecting the variable costs as well. This implies that variable costs might vary across different transactions even within the same market. In that case, our variable cost functions are presumed to measure the mean values of individual variable costs.

In market m, given the number of transaction q_{im}, the number of entrants N_m, and the number of total potential entrants S_m, the variable costs are defined as:

$$C(q_{im}, N_m) = (\theta_1 + \theta_2 q_{im} + \theta_3 q_{im}^2 + \alpha N_m + Z_m^v \delta^v + \mu S_m) \times q_{im}, \tag{4}$$

where δ^v and θ are parameters to be estimated, and Z_m^v is a vector of market level characteristics, such as average building ages, gas prices, and housing density, which presumably affect a realtor's marginal cost in each transaction. The average variable cost function, as shown inside the parenthesis in (4), is different from the usual average cost function in two respects.

First, we include the term αN_m to capture an important possible externality from the presence of other agents due to wasteful competition. Unlike most other markets where price competition is allowed, the real estate brokerage market is characterized by relatively inflexible commission rates. Since agents cannot directly compete on prices, increasing number of entrants must intensify competition along other dimensions, such as prospecting potential clients. That is, to compete for each sale, real estate agents have to spend additional amount of effort involving a wide range of activity, including marketing their own services to potential clients. As noted by Hsieh and Moretti (2003), such marketing activities include "paid advertisements in television, radio, print, or online media; informal networking to meet potential buyers and sellers, and giving away pumpkins at Halloween." The costs of these marketing activities include not only direct monetary costs of prospecting but also opportunity costs of time spent by realtors on these prospecting activities. Unlike the costs involved in selling or buying a house, most of these marketing expenses do not necessarily generate benefit enough to offset the resources committed to promoting. For this reason, we consider this part of variable costs as a cost of "wasteful non-price competition."

Second, we include linear and quadratic terms in the number of transactions in the marginal cost function. By choosing a flexible functional form, we allow for the possibility that there may be economies, diseconomies or both in the real estate brokerage market. In particular, our cost function specifications allow us to test whether the average cost function decreases or increases with individual output.

$$AC_{im} = (\theta_1 + Z_m^v \delta^v + \mu S_m + \alpha N_m) + \theta_2 q_{im} + \theta_3 q_{im}^2 + \frac{F_m}{q_{im}}. \tag{5}$$

One source of economies of scale is characterized by fixed costs. In an extreme case where $\theta_2 = \theta_3 = 0$, a sufficient amount of F_m would yield a natural monopoly as the optimal market structure. Another source of economies of scale comes from the shape and position of the average variable cost function. For example, when fixed costs are zero, the ranges of economies or diseconomies in the current market depend on whether $2\theta_3 q_{im} + \theta_2 < 0$.

2.4 Equilibrium

In equilibrium, agent i enters market m if the predicted R_{im} exceeds the predicted costs $C_m + F_{im}$. To predict the revenue, agent i first conjectures the fraction of total number of realtors, which is denoted by σ_m. Given that the aggregate distribution of ω_{im} is i.i.d. and known to all agents, σ_m is a common conjecture made by all agents in market m. Next, agent i replaces N_m in the revenue function with $S_m \sigma_m$. Using the information on observed P_m, Q_m, and other market and individual characteristics, agent i determines the predicted revenue $\widehat{R}_{im}$ based on (2).[13]

One difficulty in predicting variable costs is that we do not observe q_{im}. However, the observed individual realtor's earnings allow us to infer the predicted q_{im} by making following simplifying assumption:

$$\widehat{q}_{im} \equiv \widehat{q}_{im}(X_{im}^r, Z_m^r, P_m, Q_m, \sigma_m, S_m) = \frac{\widehat{R}_{im}}{R_m} \times Q_m, \tag{6}$$

where R_m is total revenue for all realtors in market m. Equation (6) implies that agent i's predicted market share in transactions equals to the predicted market share in revenues (i.e. $\frac{q_{im}}{Q_m} = \frac{\widehat{R}_{im}}{R_m}$). This assumption is not too restrictive in the real estate brokerage industry. For example, equation (6) holds if, ex ante, agent i uses the average commission rate, $\bar{c}_m$, and the average housing prices, $\bar{P}_m$ to predict the second stage revenue:

$$\widehat{R}_{im} \equiv \widehat{R}_{im}(X_{im}^r, Z_m^r, P_m, Q_m, \sigma_m, S_m) = \bar{c}_m \times \bar{P}_m \times \widehat{q}_{im}(X_{im}^r, Z_m^r, P_m, Q_m, \sigma_m, S_m).$$

[13] In our application, we assume that η_{im} follows the standard normal distribution with variance σ^2, and compute the predicted revenues by $\widehat{R}_{im} = \exp\{\widehat{\log}(R_{im}) + \widehat{\sigma}^2/2\}$, where $\widehat{\log}(R_{im})$ is the predicted value for $\log R_{im}$, and $\widehat{\sigma}^2$ is the predicted value for σ^2. However, our results change little even if we use simpler measures to approximate $\widehat{R_{im}}$ (e.g. use $\exp[\widehat{\log}(R_{im})]$ for $\widehat{R}_{im}$).

and if R_m, total revenue in market m, is equal to $\bar{c}_m \bar{P}_m Q_m$. Given $\widehat{q}_{im}$, agent i is assumed to predict her variable cost by plugging $\widehat{q}_{im}$ into the known variable cost function.

Given the predicted expected variable profit, potential agent i decides to enter the market as long as the predicted variable profit is greater than the fixed cost, which suggests the following choice probability of entry:

$$\begin{aligned} &\Pr\{d_{im} = 1 | X_{im}, Z_m, P_m, Q_m, \sigma_m, S_m\} = \Pr\{\widehat{R}_{im} - C_m(\widehat{q}_{im}) \geq F_{im}\} \\ = \quad &\Phi\left[\frac{\widehat{R}_{im}(X^r_{im}, Z^r_m, P_m, Q_m, \sigma_m, S_m) - C_m(\widehat{q}_{im}(X^v_{im}, Z^v_m, P_m, Q_m, \sigma_m, S_m)) - F_{im}(X^f_{im}, Z^f_m)}{\lambda_m}\right] \end{aligned} \tag{7}$$

where d_{im} is an indicator of whether agent i enters market m and $\Phi(\cdot)$ is the cumulative distribution function of standard normal. We assume that ω_{im} follows the normal distribution of $N(0, \lambda_m^2)$.

The main dependent variable in our model is N_m, and the probit model in (7) generates N_m based on a potential agent's profit, which in turn depends on σ_m, agent conjecture on the fraction of total number of realtors. For this belief to be rational, it should coincide with the probability of entry predicted from (7). Specifically, the equilibrium σ_m^* should satisfy the following condition:

$$\sigma_m^* = \int \Pr\left\{\widehat{R}_{im}(\sigma_m^*) - C_m\left(\frac{\widehat{R}_{im}(\sigma_m^*)}{R_m} Q_m, \sigma_m^*\right) \geq F_{im}\right\} dG(X_{im}). \tag{8}$$

2.5 Estimation

If the probit model (7) does not depend on σ_m, then we can estimate the parameters by simply using the maximum likelihood estimation for a standard probit, except that $\widehat{R}_{im}$ needs to be estimated before applying a probit. Because the model depends on σ_m, however, we need to impose the equilibrium condition in (8). Several empirical studies on games with incomplete information (e.g. Augereau, et al. (2006); Seim (2006)) consider similar equilibrium conditions in probability space and use the nested fixed point algorithm, in which the outer algorithm maximizes a likelihood function, while the inner algorithm solves for the fixed point given the fixed parameters. Applying the nested fixed point algorithm to our context is difficult, since σ_m enters the equation (8) nonlinearly.

In contrast, the approach proposed by Aguirregabiria and Mira (2002, 2007), which they call the nested pseudo likelihood (NPL) algorithm, is more straightforward to apply to our context. Because we can easily compute nonparametric estimates for σ_m by computing the fraction of realtors in each market,[14] we can use $\frac{N_m}{S_m}$ as an initial guess for σ_m. We then estimate $\widehat{R}_{im}$ and finally estimate the probit model. This completes the first iteration. Using the estimates from the first iteration, we predict $\widehat{\sigma}_m^1$. More specifically, we use the following equation to predict $\widehat{\sigma}_m^1$:

$$\widehat{\sigma}_m^{k+1} = \sum_{i=1}^{S_m} \Pr\left\{\widehat{R}_{im}(\widehat{\sigma}_m^k) - C_m\left(\frac{\widehat{R}_{im}(\widehat{\sigma}_m^k)}{R_m}Q_m, \widehat{\sigma}_m^k\right) \geq F_{im}\right\} \times \frac{\text{weight}_{im}}{\sum_{l=1}^{S_m}\text{weight}_{lm}}, \tag{9}$$

where we use $\frac{N_m}{S_m}$ for $\widehat{\sigma}_m^0$, and the weights are provided by the Census data. We then replace $\widehat{\sigma}_m^0$ with $\widehat{\sigma}_m^1$, and repeat the same probit estimation. This completes the second iteration. We therefore iterate this procedure until $\widehat{\sigma}_m^k$ converges. This approach is a simple application of the NPL algorithm, in which the standard nested fixed point algorithm is swapped in the sense that the outer algorithm iterates on the choice probability to solve the fixed point problem, while the inner algorithm maximizes a pseudo likelihood function given the fixed choice probability.

2.6 Identification

Our goal is to uncover the cost function, which requires us to distinguish between revenues and costs in estimating entry decisions. To do so, we rely on both functional form assumptions and exclusion restrictions. First, using the information on realtors' earnings in the census data,[15] we predict revenues and numbers of sales for each potential agent. Given the log revenue function in the first stage, the predicted revenue enters the second stage equation nonlinearly. Second, to identify the cost functions, we further exploit exogenous variations that shift revenues but not costs. Recall that the variable cost function in (4) depends on q_{im}, $\frac{N_m}{S_m}$ (or σ_m), and Z_m^v, where q_{im} is inferred from the observed and predicted revenues. To the extent that these exogenous shifters in revenues do not affect other components of the cost function but only affect q_{im}, we can trace down the variable cost as a function of q_{im} . For this purpose, we consider two sets of excluded variables that enter the revenue function only. The first set of instrumental variables are the fraction of immigrants and the fraction of out-migrants in the past 5 years, which vary across MSAs and change over years. Markets with higher inflow and

[14] A consistent estimator for σ_m is $\frac{N_m}{S_m}$, in that $\sigma_m = \text{plim}_{S_m\to\infty}\frac{N_m}{S_m}$.

[15] Note that realtors' earnings consist mostly of commission fees. For a real estate agent or broker's definition, see the occupation description on real estate brokers and sales agents in the *Dictionary of Occupational Titles* by the U.S. Department of Labor, which is currently replaced by the *O*NET* at http://online.onetcenter.org.

outflow rates have higher demand for real estate brokerage service and therefore agents in these markets are likely to predict higher revenue. However, there is no obvious reason that the inflow and outflow rates would affect an individual's entry decision other than through the revenue channel. In this sense, these instruments help provide exogenous variations for the predicted revenue in entry decisions. A second set of instrumental variable is the change in land prices. The change in land prices is a main driver for the local house prices and hence a key factor that affects agents' potential revenue. On the other hand, the change in land prices is not correlated with the quality of houses and hence the costs of brokerage service, thus providing an additional exogenous variation in identifying the cost function. Section 4.2 provides more detailed discussion on this additional instrument.

Another challenge in identifying the cost function is that unobserved market conditions may affect both predicted revenues and the entry decision. For example, higher investment demand for vacation and retirement homes in resort areas could lead to both an increase in the demand for using local realtors and and increase in local house prices. In this case, $q_{im}(R_{im})$ and σ_m in the variable cost function are shifted simultaneously, creating additional difficulty in tracing down the cost function. One way to address this concern is to exploit the panel data structure and include the MSA fixed effects. We do so by using both the PUMS 1990 and the PUMS 2000 in Section 4.2.

3 Data

3.1 Basic Description

The main datasets are the 5 percent sample of the *Census of Population and Housing* in 1990 and 2000, commonly referred to as the PUMS 1990 and the PUMS 2000. Ideally, both the PUMS 1990 and the PUMS 2000 should be used in estimation. However, the occupation codes are not comparable across years. In the PUMS 1990, occupational categories are based on the *Standard Occupational Classification Manual: 1980* (SOC 1980), in which real estate sale occupation (code 254) includes real estate appraiser, sale superintendent, building consultant, residence leasing agent, and real estate sales agent. In the PUMS 2000, occupational categories are based on the SOC 2000 which precisely defines real estate brokers and sales agents (code 41-9020). Given the inconsistency in occupational classification across the PUMS 1990 and the PUMS 2000, as well as imprecise classification of real estate brokers and agents in the PUMS 1990, we restrict our main empirical analysis to the PUMS 2000. To control for the MSA fixed effects, we include both the 1990 and 2000 PUMS in a separate estimation as one of the robustness checks in Section 4.2

We choose to model the entry decision at the real estate agent and broker level rather than the brokerage firm level. As noted by Hsieh and Moretti (2003), in the real estate market, brokerage firm is relatively unimportant and the important capital and goodwill belong to the salesperson. To control for the possible influence of brokerage firms on the agent competition, we use a secondary dataset, Metro Business Patterns 1990 and 2000, in another separate estimation in Section 4.2 to examine the possible effects of local brokerage firms on individual agents' decisions.

Markets for real estate services are local, owing to the nature of the service.[16] There is no single, agreed upon method for empirical market definitions, although it is clear that the markets should be self-contained in the sense that there is not relevant competition from outside the market. We thus follow Bresnahan and Reiss (1991) by focusing on geographically isolated markets as a way of minimizing the possibility of competition from outside the defined market. For our main estimation in 2000, we use free-standing metropolitan statistical areas, which are generally surrounded by non-metropolitan territory and therefore are not integrated with other metropolitan areas. For the panel estimation in the robustness checks, a total of 184 MSAs are identified and matched across the 1990 and 2000 census.[17] Table 1 presents the sample statistics for these 184 MSAs in 1990 and 2000.

Table 2 presents the differences between real estate agents and other occupations in 2000. On average, real estate agents and brokers tend to be older, more educated and more likely to be married. In addition, real estate agents and brokers tend to earn higher income than

[16] Competition among realtors is local because real estate is fixed in a geographic location, and buyers and sellers often want in-person interactions with agents with experience and expertise to that particular location.

[17] We reassign the MSA codes based on a new geographical unit variable that the Census created in April 2007 for the 1990 and 2000 Census: CONSPUMA. Unlike the old geographical unit variable, PUMAs, CONSPUMAs are fully comparable across years. We first assign a CONSPUMA code to each year's observations based on the composition of state and PUMAs. We then redefine the 2000 MSAs to match the 1990 boundaries based on each MSA's composition of CONSPUMAs in 1990. In reassigning these MSAs, we drop the MSAs which contain areas that are "contaminated" either by including non-metropolitan areas or by sitting across different MSA boundaries. In the end, 197 "clean" MSAs are identified and are fully comparable over time. Among these MSAs, we drop 13 MSAs whose key variables are not available. As a result, 184 MSAs are used in our estimation. The computer code for assignment of MSAs is available upon request.

non-realtors. However, their income has larger standard deviation, suggesting that the real estate agents are more heterogenous or that their job is exposed to more market risk compared with non-realtors. Large variations in realtors' demographics and earnings suggest that it is important to allow for heterogeneity not only at the market level but also at the individual level. As a comparison, Table 3 reports similar statistics for the PUMS 1990. Most numbers are similar between the 1990 Census and the 2000 Census, but earnings (or total observations) for "real estate sales occupation" in 1990 are much lower (or larger) than those for real estate agents in 2000, which could be due to the inconsistency in occupational classification for "real estate sales occupation" reported by the PUMS.

As one way to assess the reliability of the Census' measure of the number of real estate agents and brokers and their annual earnings, we use data from the Occupation Employment Statistics collected by the Bureau of Labor Statistics. The numbers are fairly consistent. This suggests that using self-reported occupation and self-reported income in the Census reflects fairly accurately the actual number of realtors and their actual earnings.

3.2 Market Structure

Table 4 presents the summary statistics across different markets in both 1990 and 2000. To measure the number of house transactions in a city, we use information on the year in which the household moved to the current house, along with information on whether the household owns the house in which it lives. In 2000, an average MSA has a sample of $19,670$ house transactions and $2,152$ realtors. The Census also asks homeowners about the value of their house, which is virtually identical to the average price of houses sold in the last year. The average value of all houses in 2000 is $\$141,789$, while the average value of houses sold is $\$151,199$. In this paper, we take the second measure as the measure of the price of housing.

Note that one cannot assume that all these transactions reported in the Census are conducted through real estate agents and brokers. Given the development in the Internet technology, some home sellers choose to avoid paying the brokerage fees by selling their homes themselves. Hendel, Nevo and Ortalo-Magne (2007) provide evidence on the importance of for-sale-by-owners (FSBO). However, using the National Home Buyer and Seller Survey data (2005), we find that the number of FSBO transactions account for only 10% of total transactions. This number may be even lower in 2000.[18]

Table 5 shows the structure of the real estate brokerage market. In both years, as the number of realtors increases, average number of households, average house value, and average realtor earnings increase substantially. Following Hsieh and Moretti (2003), we compute two measures of average productivity of real estate agents and brokers: sales per agent and sales per hour. Both measures of average productivity decrease with the number and share of realtors in the local market. If everything else was constant, then one could interpret this pattern as an indicator of excessive entry: average cost per transaction increases with the number of realtors. However, this evidence is only suggestive rather than supportive. One may worry that the negative correlation between average productivity and the number of agents simply reflects some unobserved heterogeneity across markets. For example, since more educated and skilled people tend to live together, we may find less but more capable brokers in cities with better economic conditions and more job opportunities. Given that the descriptive evidence here is not conclusive, this paper provides a formal test for the cost inefficiency hypothesis.

One purpose of employing a structural model in this paper is to test that, given the lack of price competition, whether entry results in agents inefficiently expanding resources in marketing and advertising their brokerage service. The main difficulty is that we do not observe the amount of promoting expenses by real estate agents. To get a rough idea about the magnitude of these expenses, we construct Table 6 based on the data from the Real Trends 500 Brokerage Performance Report (2002-2006). The marketing and advertising expenses include spending on mailing campaigns, handouts, inserts, open housing materials along with other company promotions. To our knowledge, Real Trends is the only source that publishes such data. Its data are derived from a survey of the top 500 brokerage firms in the U.S. and a group of rising firms just below the top 500. Note that these data are self-estimates of expenses by brokerage firms rather than by agents. Moreover, the data are reported at the regional level for the post-2000 period only. Nevertheless, they can still be suggestive. For example, Table 6 presents a comparison of marketing and advertising expenses and annual real estate agent growth across regions and years. While the observations are too few to present a clear

[18] In addition, two out of five of these FSBO transactions are between closely related parties, such as friends and relatives. That is, 40% of FSBOs are not placed on the open market and therefor are less likely to have direct effect on demand for real estate brokerage service.

pattern, one can reasonably say that in the same year, regions that experienced the highest agent growth had relatively large expenses on promoting activities (measured by the fraction of the total revenues.) Similarly, for the same region, years that experienced the highest agent growth had relative large expenses on promoting activities. Of course, these expenses account for total costs rather than average costs involved in marketing and advertising. In addition, these regions are subject to different market conditions in different years. Without further investigation, one cannot take this as evidence for excessive entry.

3.3 Fixed Costs Shifters

Both fixed costs and variable costs are affected by exogenous cost shifters. In this subsection, we consider three measures of fixed costs. First, real estate agents and brokers need to pass a number of exams obtain the license. Our data on the licensing requirements come from the Department of Financial and Profession Regulation's website. We use the following variables as measures of agents' fixed costs: the number of hours required to take real estate transaction course, the requirement for license renewal, and exam fees. Table 8 provides the number of hours, the license fees and exams fees in 2000. There is significant variation across state requirements. At one extreme, Texas requires a minimum of 270 study hours before an individual could take the exam. At the other extreme, Alaska requires 20 study hours only.[19]

Second, our measures of fixed costs also include the reservation wages of working outside of the real estate brokerage section in the same MSA. We consider two alternative measures of reservation wages for realtors: average earnings in the local market and average earnings for non-realtors. These measures are used to capture the reservation wage that a realtor could earn in the same MSA if she were not a realtor. An economic boom that drives up the house price and hence realtor earnings in an MSA will also increase the average earnings from working in the other sectors of the same market.

Finally, in addition to the market level variations in reservation wages, we include a set of demographics at the individual level, such as age, age squared, education, education squared, marital status, race and sex. These variables are important to control for the individual level

[19]Brokerage entry appears to be more difficult than agent entry. At a minimum, an entrant that wants to establish a brokerage must hire or become a licensed broker. Additionally, an entering broker may require an agent workforce, office space, an office staff, and advertising of listings to establish name recognition.

heterogeneity in terms of reservation wages. Agents with higher abilities in real estate or in sales skills could have earned higher wages by engaging in other businesses, hence their opportunity costs are also higher.

3.4 Variable Costs Shifters

To capture the presence of wasteful competition involved in fighting for clients, we include the number of realtors, which changes across markets and over time. If entry leads to more inefficient use of resources in marketing activities, larger realtor share would lead to an increase in average variable cost. In addition, we include the following cost shifters at the marginal cost level: average local building age, average local house density and average local gas price, all of which change across markets and over time. In general, one would expect new houses to be more easily sold. Newer houses can be expected to incorporate features that are more in demand by the market. There is less risk in their purchase because newer houses often involve fewer hidden features that require time-consuming investigation. Thus, one would expect that agents spend less time in helping to sell or purchase a new house. In addition, higher gas prices imply higher transportation cost and hence increases the cost per house visit.

3.5 Revenue Shifters

Referring back to the first stage estimation, an individual agent's revenue is determined by demand shifters, such as demographic factors (age, age squared, education, education squared, marital status, whether stayed in the same market for 5 years), working conditions (whether both working, working weeks, working hours, whether full-time, whether self-employed), and market conditions (total number of transactions, total number of agents, average house price, the rate of Internet adoption in each MSA, whether the local government has no-rebate policies, market size and market fixed effects). Real estate agents differ in their reputation, network and ability to sell. While we do not directly observe the number of transactions at the individual level, much of the agent heterogeneity can be attributed to the observed demographics. For example, those who have stayed within the same market for more than 5 years and have been working for longer hours are more likely to earn higher income. In addition, we include the market level brokerage firm concentration ratio and its interactions with the individual level demographics to capture the possible agent heterogeneity from the impact of local brokerage firms. (See Section 4.2.) As noted before, in order to identify the cost function, we include two sets of excluded variables that affect the revenue but not the cost. These variables include the number of immigrants into (and emigrants out of) the local MSA in the past 5 years, and the change in land prices.[20]

4 Estimation Results

4.1 Parameter Estimates

We estimate the model using the PUMS 2000. The NPL algorithm converged after 5 iterations. Table 8 shows the convergence in σ_m^t for some sample MSAs. In several MSAs, the initial estimates for σ_m^0 are close to the converged σ_m. In some MSAs, however, they are quite different from the converged σ_m, suggesting the importance of imposing equilibrium conditions. The difference between the initial estimates and the converged estimates tends to be larger as the sample size becomes smaller, which suggests that even using the consistent nonparametric estimates[21] for σ_m might not be sufficient to impose the equilibrium condition particularly when the sample size is small.

Table 9 presents the results of our first stage regression from equation (2). We show the results from both the first iteration and from the converged iteration. The coefficient estimates are largely similar across the two columns. The coefficient for #realtors is negative but is not statistically different from zero. Note that realtors' revenues from commission fees are proportional to house value. High house prices induce realtors' entry, then high entry will be positively correlated with high revenue. New entrants, however, may also steal business from existing realtors, hence resulting in a lower revenue for an average individual realtor. The insignificant coefficient for #realtors suggests that these two effects cancel each other out.

Market conditions and individual demographics have expected signs. At the market level, high house prices translate into high commission income. Larger market size, represented by total labor force and land area, increases the realtor revenue. In addition, more immigration inflows represent higher demand for local brokerage service and hence increase the realtor

[20] Another set of variables that only enter the revenue equation include the Internet adoption rate, the Internet search intensity, and the indicator dummy for rebate bans. Note that we initially included these variables in the cost function, but we found that their coefficient estimates were small and statistically indistinguishable from zero. We thus exclude them from the cost function in the main estimation.

[21] The consistent nonparametric estimator for σ_m is simply the frequency estimator ($\equiv \frac{N_m}{S_m}$) in our context.

revenue. In contrast, more outflows have the opposite effect. At the individual level, being male, white, married and older all increase individual revenue significantly. While there is no obvious indicator for whether an agent is a star agent in the local market, we find that agents who have stayed with the same MSA for more than 5 years and who have been working for longer hours tend to have higher revenues. In particularly, full-time realtors earn more than part-time realtors and brokers earn more than agents.[22] The evidence here suggests that controlling for heterogeneity in effort, skill and experience is important when determining realtors' revenue.

Two sets of coefficients are of particular interest. The first is the coefficient for anti-rebate policy. In 2000, 13 states prohibited, by law or regulation, real estate agents from giving consumers rebates on commissions. These 13 states were Alabama, Alaska, Iowa, Kansas, Louisiana, Mississippi, Missouri, New Jersey, North Dakota, Oklahoma, Oregon, Tennessee and West Virginia. We therefore create a dummy variable that indicates whether an MSA adopted the anti-rebate policy in 2000. . Presumably, in MSAs with anti-rebate policies, agents cannot compete directly on commission, which limits the competition from discount brokers. This has a positive effect both on the commission rate and on the market demand for traditional full-commission brokers, which in turn increases the individual realtor's revenue. The positive and fairly significant coefficient on the anti-rebate policy dummy variable is consistent with this hypothesis.

The second set of coefficients of interest is related to the diffusion of the Internet. We use the *2000 Current Population Survey: Supplement for Internet and Computer Use* to construct two Internet-related variables. One is Internet adoption, which indicates the fraction of survey respondents in each MSA who reported to be an Internet user. The coefficient for Internet adoption rate is positive and significant, confirming the common wisdom that Internet has become a power resource that brings more business to traditional realtors. The diffusion of the Internet, such as the use of emails and websites, helps realtors not only market their service but also match potential buyers and sellers. As a result, realtor revenue is higher in areas with higher Internet adoption rates. However, the diffusion of the Internet also suggests that potential buyers and sellers find it easier to search or sell the house on their own (Hendel,

[22]We follow Hsieh and Moretti (2003)'s notion that, among realtors, those who report to be self-employed are brokers and those who report to be non self-employed are agents.

Nevo and Ortalo-Magne (2007)) or to use online discount brokers (Levitt and Syverson (2007)), thereby lowering the potential revenue for traditional realtors. For this reason, we include a second Internet variable, online search, which indicates the fraction of respondents in each MSA who reported to regularly use the Internet to search for information. The corresponding negative and fairly significant coefficient confirms this second hypothesis related to the online home searching behavior.

We now present the results from the second stage estimation of the entry model. Table 10 presents the estimates from the first iteration and the converged iteration, and Table 11 reports the converged estimates, marginal effects, and mean values of all variables. The endogeneity of entry decisions is not controlled in the first iteration but is accounted for in the converged iteration. Most coefficient estimates are similar across different iterations, except for the coefficients in the variable cost function.

The main parameter is the coefficient for $q\times$#realtors. It is negative in both iterations, and its magnitude is much larger in the converged iteration than in the initial iteration. The negative and statistically significant coefficient for $q\times$#realtors confirms potential inefficiency due to wasteful competition under free entry: high entry intensifies competition in fighting for clients, which leads to a substantial increase in the marketing resources involved in each transaction.[23]

The coefficients for q^2 and q^3 indicate that the current real estate brokerage market is still in the declining part of the average variable cost function. In other words, average variable costs decline with output. Therefore, even without considering fixed costs, these estimates suggest that entry leads to the loss of economies of scale, since fewer agents could have handled the total transactions in a more cost efficient way. This is different from the radio broadcasting industry studied by Berry and Waldfogel (1999), in which the duplication of fixed costs is the main source of the loss of economies of scale.

The results thus far suggest that entry leads to inefficiency not only in term of the loss of economy of scale but also in term of the wasteful non-price competition. To further quantify the magnitude of the inefficiency under free entry, we compute the effect of an increase in

[23]Because wasteful non-price competition might be intensified if rebates on commissions are prohibited, we examined this possibility by including an interaction term for $q\times$#realtors$\times$anti-rebate in the cost function. The coefficient estimate turns out to be negative in the converged iteration, but it is statistically indistinguishable from zero. As a result, we drop this interaction term in the estimations henceforth.

entry on average variable costs. We find that, in an average MSA, when the number of realtors increases by 10%, the mean of average variable costs increases from $2,392 to $2,508. A 10% increase in the number of realtors therefore increases the average variable cost on each home transaction by 4.8%, which is attributed to both wasteful non-price competition and the loss of economies of scale. As a comparison, Hsieh and Moretti (2003) finds that $\frac{\Delta\text{AVC}}{\text{AVC}}/\frac{\Delta P}{P} \approx 0.3$ and $\frac{\Delta\text{N}}{\text{N}}/\frac{\Delta P}{P} \approx 0.9$, which implies that $\frac{\Delta\text{AVC}}{\text{AVC}}/\frac{\Delta N}{N} \approx 0.033/0.1$. That is, a 10% increase in the number of realtors increases the average variable costs by 3.3%. Recall that our estimation is based on the 2000 Census data while Hsieh and Moretti (2003) uses the 1980 and 1990 Census data with a much broader definition for real estate agents and brokers. In addition, as Hsieh and Moretti (2003) mentioned, their measure captures "social waste that takes the form of time spent by brokers doing things of marginal social value", but does not capture monetary costs of prospecting. This might also explain why the implied effect of entry on cost in their work is smaller than ours.

In 2004, commissions paid to real estate agents and brokers in the U.S. totaled roughly $61 billion (Hagerty, 2005). Do brokers provide commensurate value? Our model allows us to roughly examine the implied markup for real estate services that realtors provide. Assuming that the commission rate is 6%, we find that the weighted mean of MSA-level commission fees per transaction is $8,846.[24] Our cost estimates imply that the weighted mean of MSA-level average variable costs are $2,392, accounting for about 27% of commission fees on average.[25] That is, without controlling for fixed costs, the implied markup for the real estate brokerage industry is as high as 73%. This is qualitatively consistent with the key finding in Bernheim and Meer (2008): a 6% percent sales commission is a steep price to pay for the value rendered.

4.2 Robustness Checks for Cost Estimates

The preceding results provide evidence for potential cost inefficiency due to free entry, mostly driven by wasteful non-price competition. Our results so far are based on cross-section vari-

[24]The MSA-level average commission fee is the 6 percent of prices for houses sold. We then computed the weighted mean of these MSA-level average commission fees, using the number of labor force as weights.

[25]In a recent working paper, Bernheim and Meer (2008) divide real estate brokers' services into 6 components: promotional service (e.g. prepare a house for sale or recommend a house to a buyer), negotiation, matching, providing MLS access, providing market information and assisting with paper work. They infer the market value of the forth, fifth and sixth components is roughly $1,400. Our estimate of $2,392 is substantially higher, as it refers to the total costs of the bundled realtors' service in which all six components are included.

ations in the data. We are therefore concerned that unobserved characteristics, such as city-specific distribution of housing quality, may bias our cost estimates. In some areas, for example, the average quality of houses may be high, which drives up the overall house prices. In the same areas, the true quality of each individual house may also vary considerably and may be difficult to identify. As a result, selling each house requires more time and efforts, implying higher brokerage costs in these areas. Though there might be more realtors in this area because of high house prices, high costs do not result from wasteful competition among realtors. Because we do not observe the distribution of housing quality that results in high housing prices and more realtors in this area, we could overestimate wasteful costs by incorrectly attributing high costs to wasteful competition.

Note that the variable costs in our model are assumed to measure the mean values of individual variable costs that might vary across different transactions, and that we also allow for some market-level characteristics (e.g. building ages) to enter the variable costs. Nevertheless, our estimates for the mean values of variable costs can still be biased if unobserved characteristics are important. This section therefore provides some robustness checks for our cost estimates, and particularly the measure for wasteful competition.

To address this concern, we first consider a variable that affects housing prices but is not correlated with unobserved housing quality. Similarly to Hsieh and Moretti (2003), we use the change in the housing price index compiled by the Office of Federal Housing Enterprise Oversight (OFHEO). The OFHEO index is supposed to capture changes in the price of land, but not changes in housing quality.[26] Consequently, the change in the OFHEO index would change housing prices (and thus, realtor's revenues) but is unlikely to be correlated with housing quality and the quality (and the cost) of brokerage services. We re-estimate our model including the change in OFHEO index, and the estimates for the main coefficients are reported in Table 12.[27] As expected, the coefficient estimate for the OFHEO price growth is positive and fairly significant. This is consistent with the view that house price growth is one of the

[26] See Hsieh and Moretti (2003) for more discussion on the OFHEO index.

[27] In this specification, we also include some measures for market concentration. One may argue that the competition in the real estate brokerage industry is not between individual agents, but between brokerage firms. To examine how this possibility affects our estimates, we include the ratio of real estate firms with more than 100 employees, among total number of real estate firms. We obtain this ratio for each MSA from the *Metro Business Patterns*. However, the results show that they do not affect most of our estimates.

most important factors in realtors' entry decisions. Nevertheless, including this variable barely affects the cost estimates.

As alternative checks, we additionally use the logarithm of the number of realtor, as well as the fraction of the number of realtors among total labor force, because these measures are likely to be robust to the variations in market size which tends to be correlated with market-level unobservables. The results are also reported in Table 12. Because the log values and the values for the fraction are small, the coefficient estimates for $q\times$#realtors are larger than the baseline estimate. Nonetheless, the implied AVC and wasteful part of the AVC are reasonably consistent across different specifications.

More powerful way to control unobserved characteristics is to use panel data and difference out market fixed effects. As discussed in Section 3.1, the occupational category for realtors in PUMS 1990 is broad and imprecise. To combine PUMS 1990 and PUMS 2000, we thus redefine realtors in PUMS 2000 to be partially consistent with PUMS 1990.[28] Because MSA definition has changed between 1990 and 2000, we also redefine MSAs in 2000 to be consistent with the 1990 data. Because several variables (e.g. internet variables) are only defined over the original MSAs in 2000, we have to drop them. We then estimate the model including MSA fixed effects to control for market-level unobservables such as the distribution of housing quality.

Table 12 presents the results. The qualitative results for cost estimates are the same as those in the baseline. The coefficient estimate for $q\times$#realtors is negative and statistically significant. To compare these results with the results from PUMS 2000, we then compute the implied AVCs. The mean of the AVCs using the 1990-2000 data is larger than those using the 2000 data only. The percentage of wasteful AVC out of total AVC is somewhat lower than those using the 2000 data only. These differences could be due to the differences between the combined data and the 2000 data that we use in the main analysis. However, it is also possible that the measure for wasteful competition in the baseline (= \$1,120) may be even underestimated because the 1990-2000 data suggests that it is on average \$1,514.[29] Our robustness checks thus suggest that our cost estimates are fairly robust to different specifications, and that our measure for

[28] The occupational category for "real estate sale occupation" includes several other occupations including real estate appraiser. We can only identify real estate appraiser in PUMS 2000. Hence, redefined realtors include real estate brokers and sales agents, as well as appraisers.

[29] Moreover, higher values for the AVC in the 1990-2000 results may suggest that our cost estimates reflecting the loss of economies of scale might be underestimated in the baseline results.

wasteful competition is unlikely to be overestimated.

So far our model has taken the number of transactions, Q_m, and the average house price, P_m, as exogenously given. One might argue that both variables are likely to be correlated with entry decisions. A large number of realtors could lead to a different form of intensified non-price competition. In a more intensified competition, some realtors could spend substantial amount of effort to convert a potential transaction into a real transaction, which pushes up the number of total transactions in a given market. Additionally, when the competition intensifies, some realtors may have to increase the suggested listing prices (and hence potential transaction prices) in order to win the business from potential clients, which increases the mean of local house prices. If either of these hypotheses holds, the estimate of wasteful non-price competition in the cost function would be biased. To address these concerns, we could impose additional structure on the relationship between Q_m, P_m and N_m. However, without further complicating the model, our intuition suggests that the estimated cost of wasteful non-price competition, if biased, would be underestimated in magnitude, since the positive influence of entry decisions on Q_m and P_m could only bias down the coefficient on $N_m q_{im}$ in the cost function in magnitude. Given that the estimated coefficient is significantly negative (both statistically and economically), the evidence for wasteful non-price competition should be qualitatively consistent and robust, even in the presence of endogeneity of Q_m or P_m.

4.3 Counterfactual Analysis

The estimated model parameters allow us to evaluate the anti-rebate policy and further examine the effect of the Internet adoption. In this section, we calculate counterfactual revenues and costs using the parameter estimates. In computing counterfactual values, however, one needs to account for the equilibrium effect – a counterfactual experiment changes revenues or costs, thereby changing the individual's entry probability, which in turn changes the equilibrium number of realtors, hence further changing the equilibrium revenues and costs. Our model allows us to account for this equilibrium effect. Specifically, we solve for the fixed point in the entry probability under each counterfactual scenario below.

4.3.1 Anti-rebate Policy

Real estate agents can compete on price either by charging commission rates lower than 5 to 6 percent, or by offering rebates.[30] To the extent that 5 to 6 percent of the commission rate is an industry standard, rebates become an important form of price competition. In 2000, 13 states prohibited real estate agents from giving consumers rebates on commissions. Proponents of anti-rebate policy argue that under rebate bans, consumers are likely to choose agents based on the price of services rather than the quality of service, and also that the rebate bans protect consumers from false and misleading offers of rebates.[31] However, DOJ and FTC report (2007) argues that that there seems no evidence for harmful effects of rebates on consumer welfare. Instead, anti-rebate policy would harm consumers by preventing price competition. The explicit prohibition of price competition has attracted the notice of anti-trust authorities at the Justice Department who were planning to sue the National Association of Realtors over these policies. For example, in March 2005, the U.S. Department of Justice (DOJ) filed an antitrust suit against the Kentucky Real Estate Commission for prohibiting agents from giving consumers a rebate on some of the commissions they pay. By the end of 2007, three out of thirteen states have abolished the rebate bans. Despite public interest and antitrust implications, little empirical evidence on the implications of anti-rebate policy has been provided, in part because there is an absence of available data on the costs.

Given the estimates above, our model provides an appropriate setting to evaluate the potential benefit of removing anti-rebate bans. Specifically, we consider the MSAs where rebates were banned in 2000, and compute counterfactual results in the absence of anti-rebate policy. To do so, we first calculate predicted revenues and average variable costs for realtors. In equilibrium, turning off anti-rebate indicator variable changes an individual agent's predicted variable profits and therefore affects her entry probability. This in turn changes the market conjecture about the number of realtors, which again affects each individual's entry probability. We thus compute new equilibrium under counterfactual scenarios by solving for the fixed points

[30] In addition to rebates in the form of cash payments, agents can offer customers inducements such as gift certificates, coupons, vouchers, and discounted or free services relating to buying and selling a home. DOJ and FTC report (2007) defines both rebates and inducements as "rebates". See this report for more discussion on rebates and anti-rebate policy.

[31] Studies consistent with this view include Schroeter (1987) and Carroll (1989), which argue that fixed percentage commissions can promote welfare: clients who value brokerage services more highly offer to pay larger commissions and consequently receive more effort from the broker.

in entry probabilities.

Panel A of Table 13 presents the results. The positive coefficient estimate for anti-rebate dummy indicates that abolishing anti-rebate policy would decrease realtors' revenue, possibly because of price competition from both traditional full-commission and discount brokers. Reduced revenues, on the other hand, discourage potential entrants from entering the market, hence reducing the equilibrium number of traditional realtors. As a result, the competition is more intensified between traditional brokers and discount brokers but less intensified among traditional brokers. Reduced entry by traditional brokers lowers the variable costs both by realizing economies of scale and by saving on wasteful non-price competition. As shown in Table 13, for an average MSA with rebate bans, removing such bans would decrease the equilibrium number of realtors by 5.8% and reduce total variable costs by 3.9%.

Panel B of Table 13 presents the results from the converse counterfactuals, in which we consider MSAs without rebate bans and examine what if they prohibited rebates. The results show that adopting anti-rebate policy would increase the total and average variable costs and increase the number of realtors. For these MSAs, adopting anti-rebate policy would increase total variable costs by 5.9%. Therefore, an anti-rebate policy is harmful not only because it suppresses price competition, but also because it encourages excessive entry, hence resulting in wasteful competition.

Before we examine the counterfactual results related to the Internet, we note that the number of realtors in MSAs with rebate bans tends to be lower than that in MSAs without rebate bans, which explains relatively low wasteful AVC in MSAs with anti-rebate policy. As Table 14 shows, however, this difference seems to be due to the fact that MSAs without anti-rebate policy tend to be more densely populated, have more houses and transactions, and higher housing prices, than MSAs with anti-rebate policy.

4.3.2 Internet Diffusion

The Internet has become increasingly important in the real estate brokerage industry.[32] A few recent studies have revealed some of the economic benefits available to real estate service consumers who utilize the Internet. For example, Levitt and Syverson (2007) shows that

[32]See DOJ and FTC report (2007) for more discussion on the impact of the Internet on the real estate industry.

sellers who use online flat-fee agents could save an average of $5,000 compared to hiring traditional full-commission agents. In addition, Hendel, Nevo and Ortalo-Magne (2007) shows the importance of FSBOs accompanied by the Internet diffusion. However, the implication of the Internet for the traditional real estate service providers is far from being clear. On the one hand, higher Internet adoption rates increase the popularity of online discount brokers and online versions of FSBOs (for-sale-by-owner), both of which could steer the buyers and sellers away from traditional agents. On the other hand, as argued by the National Association of Realtors, the access to the Internet helps more traditional real estate agents to reach their potential clients, either by emails or by websites. How does the Internet affect entry decisions? Does the Internet promise real cost savings? In this section, we provide some rough answers to these questions by exploiting the estimated coefficients on the Internet adoption rate and online search intensity in our model.

According to the Current Population Survey, the Internet adoption rate, measured by the fraction of respondents who report to have adopted the Internet, increased by 10% between 2000 and 2003 in an average MSA. In addition, the NAR Profile of Home Buyers and Sellers has found that the Internet use in home searching, measured by the fraction of home buyers who use the Internet as information source in buying a home, has steadily increased from 71% to 80% between 2003 and 2006. We therefore perform a counterfactual experiment in which the Internet adoption rate in each MSA is increased by 10%, and the level of online search intensity is increased to full intensity so that all Internet adopters regularly use the Internet to search for information, including home information.[33] The results are presented in Panel A of Table 15. The combined effects of increasing Internet adoption rates and Internet search intensity are qualitatively similar to the effect of abolishing anti-rebate policy, though its magnitudes are rather modest.

As discussed in Section 4.1, a higher Internet adoption rate facilitates traditional realtors in reaching and matching their potential clients and therefore has a positive effect on revenue, whereas higher online search intensity facilitates the use of online discount brokers and FSBOs and there has a negative effect on revenue. To investigate each effect separately, we perform

[33] In the current sample, about 75% of Internet adopters in the MSAs included in our analysis report to regularly use the Internet to search for information. In the counterfactual experiment, we increase the online search intensity to full intensity so that 100% of Internet adopters, including existing and new adopters, all regularly use the Internet to search for information.

additional counterfactuals: (i) do not change the Internet adoption rate, but have full search intensity for the existing Internet adopters; (ii) increase the Internet adoption rate by 10%, but do not change the online search intensity. The results are respectively reported in Panel B and Panel C in Table 15. The results from the counterfactual (i) show that an increase in online search intensity lowers realtor's revenues, but it discourages excessive entry and reduces wasteful competition and total variable costs. In contrast, the results from the counterfactual (ii) show that an increase in the Internet adoption, if not accompanied by an increase in Internet search intensity, can lead to excessive entry by traditional brokers. These findings are consistent with the two opposite effects of the Internet diffusion discusses above. In our basic counterfactual setting, the effect of an increase in the Internet adoption is slightly dominated by the effect of a commensurate increase in online search intensity. The latter insures participation by online discount brokers and FSBOs and encourages the competition in commissions, thereby generating a parallel effect to removing the rebate bans.

5 Robustness Checks for Key Assumptions

A full evaluation of real estate brokerage market requires assessing cost inefficiency against the benefit of reduced prices or improved quality of the service from entry. Ideally, we would like to have the data on market shares and commission rates in order to recover the demand function and measure the benefit from entry. Given the lack of the idea data, our model assumes that agents do not directly compete on prices and quality. For this reason, this section examines these simplifying assumptions and provides suggestive evidence as to potential benefits from free entry.

5.1 Does Entry Put Downward Pressure on Commission Rates?

Tables 16 and 17 indirectly show the relationship between entry and commission rates. Table 16 first shows a positive correlation between changes in log of housing prices and changes in log of the number (or fraction) of real estate brokers/agents in MSA. We compute the correlation coefficients using the MSAs whose geographical boundaries did not change between 1990 and 2000. This positive relationship is consistent with evidence presented by Hsieh and Moretti (2003). To examine the relationship between housing prices and commission rates, we then use the 1988-2002 Consumer Expenditure Surveys, because PUMS does not contain any information on commission fees. We follow Hsieh and Moretti (2003), and estimate the commission rate by dividing “total selling expenses” (including commission fees) by the price of housing. To reduce measurement errors, we drop observations with implausibly large or small estimated commission rate (less than 1 percent or more than 10 percent), again following Hsieh and Moretti (2003). The total number of observations then becomes 520, since only a handful of observations in the CEX reported selling their houses during the survey period. Table 17 reports the relationship between housing prices and commission rates. Though commission rates vary slightly across different price ranges, there is no clear positive or negative relationship between the increase in housing price and commission rates. Consequently, we find that the increase in the number of realtors in the market is positively correlated with the increase in housing prices, but housing prices do not appear to be correlated with commission rates. Therefore, entry is less likely to put downward pressure on commission rates.

To provide additional evidence on the changes in commission rates, Table 18 lists Real Trends national average commission rates and fees from 1991 through 2005.[34] As illustrated in Table 16, commission rates have fallen gradually over the 1990-2000 period, from 6.1 percent to 5.42 percent. The gradual reduction in commission rates is more likely to be due to rising house prices rather than the increasing entry.

To further explore the relationship between entry and commission rates, Table 19 presents the changes in real estate agent growth and the changes in average commission rates between 2002 and 2003 across 6 regions of the United States. The aggregate statistics are derived from the data based on the 2004 Real Trends Brokerage Performance Report. Using the percentage changes from 2002 to 2003, we control for any fixed region-specific factor that might affect both the commission rates and entry. In the regions which experienced most real estate agent entry (e.g. FW, MA and NE), the percentage changes in the average commission rates were 1.63%, 0% and −1.15% respectively. Given that the average commission rate was around 5.1% during that period, the magnitude of the changes in commission rates was almost zero. The number of observations and the level of geographic coverage prevent us from further investigation. However, it appears that entry does not have substantial effect on commission rate.

[34]Commission fees, measured in constant 2006 dollars, are based on median home prices so as to represent what a typical consumer would pay in real estate commissions to sell his or her home.

5.2 Does Entry Improve the Quality of Brokerage Service?

One may argue that the absence of price competition may imply a type of useful non-price competition, which is not considered in our model. That is, real estate agents increase the costs endogenously to improve the quality of their brokerage service. In this sense, the increase in average costs is not really wasteful because they help to generate better matches for consumers. A similar argument has proved to be wrong in Berry and Waldfogel (1999) using a simple theoretical framework. Even if entry indeed improved the average quality of products, they show that such improvements result in "stolen" business. In equilibrium, agents do not take into account that increases in own-quality reduces the demand for other agents, and therefore over-invest the resources.

In the rest of this subsection, we test whether entry by traditional brokers improves the quality of the service they receive. Unfortunately, the Census data do not provide valid information on the quality of home matching process. To estimate the effect of entry on the quality of real estate brokerage service, we use the 2001 Home Buyer and Seller Survey conducted by the National Association of Realtors. After excluding the MSAs with fewer than 10 observations and excluding transactions without the assistance of a real estate agent or broker, the sample consists of $3,145$ observations on home buyers and $1,785$ observations on home sellers. We then merge the sample with the PUMS 2000, which provides the total number of real estate agents and brokers in each local housing market at the MSA level.

The most common measures of quality of real estate brokerage service used in the literature are time on the market and the difference between sale price and listing (asking price). We therefore construct the following two quality measures. For home buyers, the quality of brokerage service is measured by the searching weeks for home buyers, divided by the buyer discount (measured by the percentage change between asking price and transaction price.) That is, for each percentage discount that a home buyer eventually obtains, how much time she waits on the market before the final match. For home sellers, the quality of brokerage service is measured by the time on the market for home sellers, divided the seller concession (measured by the percentage change between listing price and transaction price.)

To show the possible effect of entry on the quality of the brokerage service that home buyers receive, we first estimate a duration model for home buyers. The dependent variable is the searching time adjusted for buyers' discount while the independent variables include the fraction of realtors out of total labor force, original asking price, a set of buyers' demographics, a set of market and neighborhood conditions, a set of housing characteristics and a set set of searching conditions. The key coefficient on the fraction of realtors is -10 with a standard deviation of 34.33, which is statistically insignificant. We obtain similar results when we estimate a separate duration model for the home sellers. These estimates suggest that excessive entry does not provide significant benefit to consumers by reducing their time on the market.

While time on the market is informative, shorter time on the market does not necessarily indicate a better quality of matching process. For example, Levitt and Syverson (2005) find that homes owned by realtors sell for approximately 3.7% more and stay on the market about 10% longer than homes owned by non-realtors. To correct the possible bias by the using time on the market, we construct a more powerful measure of brokerage service quality: REA satisfaction. In particular, the NAR survey asks each respondent whether she would use the same agent again if there is another home transaction in the future. The dummy variable REA satisfaction equals to 1 if the answer is yes. We then run a probit model in which the dependent variable is REA satisfaction and the independent variables include the fraction of realtors out of total labor force in the local market, time on the market, buyers' discount (sellers' concessions for home sellers sample), housing characteristics, neighborhood characteristics, market conditions, search/selling process characteristics, home buyer/seller characteristics. As shown in Table 20, the key coefficients on the fraction of real estate agents in both the buyer and the seller samples are insignificant. This suggests that increasing number of realtors has a negligible impact on consumers' evaluation of the quality of the brokerage service. In addition, we find that home buyers' satisfaction with their realtors increases with the buyers' discount and decreases with the searching time. In contrast, home sellers are more concerned with the time on the market but do not care much about the concession to give.

In sum, entry by traditional realtors does not lead to obvious improvement in home matching quality. The evidence is consistent with the view that the traditional real estate brokerage industry tends to produce a standardized package of service. In this sense, our welfare result based on the cost estimates alone, while incomplete, is a strong indicator of the total welfare loss under free entry.

6 Concluding Remarks

In this paper, we study entry decisions in local real estate brokerage markets to investigate sources of potential cost inefficiency under free entry. We build upon recent empirical work on games of incomplete information, and construct a rational expectation equilibrium model. The model is estimated using recently developed pseudo likelihood algorithms. Using the data from the 5% 2000 PUMS, we find strong evidence for two sources of cost inefficiency – wasteful non-price competition and the loss of economies of scale.

Using our estimates, we further find that rebate bans are welfare-reducing, not only because they discourage price competition from discount brokers, but also because they encourage excessive entry by traditional full-commission brokers. Removing these rebate bans would decrease the equilibrium number of realtors by 5.8% and reduce total variable costs by 3.9%. Welfare implications of the Internet diffusion are mixed, however, in that an increase in the Internet adoption alone can lead to excessive entry by traditional brokers, whereas a commensurate increase in Internet search intensity can reduce excessive entry and wasteful costs.

References

Aguirregabiria, Victor and Pedro Mira "Swapping The Nest Fixed Point Algorithm: A Class of Estimators for Discrete Markov Decision Models." *Econometrica*, 2002, *70*, pp. 1519-1543.

Aguirregabiria, Victor and Pedro Mira "Sequential Estimation of Dynamic Discrete Games." *Econometrica*, 2007, *75*, pp. 1-53.

Anderson, Simon, Andre DePalma, and Yuru Nesterov "Oligopolistic Competition and the Optimal Provision of Products." *Econometrica*, 1995, *63*, pp. 1281-1302.

Aradillas-Lopez, Andres "Semiparametric Estimation of a Simultaneous Game with Incomplete Information." 2005, working paper.

Augereau, Angelique, Shane Greenstein, and Marc Rysman "Coordination vs. Differentiation in a Standards War: 56K Modems." *RAND Journal of Economics*, 2006, *37*, pp. 887-909.

Bajari, Patrick, Han Hong, John Krainer, and Denis Nekipelov "Estimating Static Models of Strategic Interactions." 2006, working paper.

Bernheim, Douglas and Jonathan Meer "How Much Value Do Real Estate Brokers Add? A Case Study." *NBER Working Paper No.13796*, 2008.

Berry, Steve "Estimation of a Model of Entry in the Airline Industry." *Econometrica*, 1992, *60*, pp. 889-917.

Berry, Steve and Joel Waldfogel "Free Entry and Social Inefficiency in Radio Broadcasting." *The RAND Journal of Economics*, 1999, *30*, pp. 397-420.

Bresnahan, Timothy and Peter Reiss "Entry and Competition in Concentrated Markets." *The Journal of Political Economy*, 1991, *99*, pp. 977-1099.

Brock, William. and Steven Durlauf. "Discrete Choice with Social Interactions." *The Review of Economic Studies*, 2001, *68*, pp. 235-260.

Carroll, Wayne "Fixed Percentage Commission and Moral Hazard in Real Estate Brokerage." *Journal of Real Estate Finance and Economics*, 1989, *2*, pp. 349-365.

Chamberlain, Edward "The Theory of Monopolistic Competition." *Cambridge, Mass.: Harvard University Press*, 1993.

Chen, Hsuan-chi and Jay Ritter "The Seven Percent Solution." *Journal of Finance*, 2000, *3*, pp. 1105-1131.

Crockett, John "Competition and Efficiency in Transacting: The Case of Residential Real Estate Brokerage." *American Real Estate and Urban Economics Association*, 1982, *10*, pp. 209-227.

Dixit, Avinash and Joseph Stiglitz "Monopolistic Competition and Optimum Product Diversity." *The American Economic Review*, 1977, *67*, pp. 297-308.

DOJ and FTC Report "Competition in the Real Estate Brokerage Industry." *U.S. Department of Justice and Federal Trade Commission*, 2007, April.

GAO Report "Real Estate Brokerage: Various Factors May Affect Price Competition." *United States Government Accountability Office*, 2006, July.

Goolsby, William and Barbara Childs "Brokerage Firm Competition in Real Estate Commission Rates." *The Journal of Real Estate Research*, 1988, *3*, pp. 79-85.

Hagerty, James "Discount Real Estate Brokers Spark A War Over Commissions." *The Wall Street Journal*, 2005, October.

Heckman, James and Burton Singer "A Method For Minimizing the Impact of Distributional Assumptions in Econometrica Models for Duration Data." *Econometrica*, 1984, *52*, pp. 271-320.

Hendel, Igal and Aviv Nevo and Francois Ortalo-Magne "The Relative Performance of Real Estate Marketing Platforms: MLS versus FSBOMadison.com." *Mimeo, University of Wisconsin*, 2007, April.

Hsieh, Chang-Tai and Enrico Moretti "Can Free Entry Be Inefficient? Fixed Commissions and Social Waste in the Real Estate Industry." *The Journal of Political Economy*, 2003, *111*, pp. 1076-1122.

Levitt, Steven and Chad Syverson "Market Distortions When Agents Are Better Informed: The Value of Information in Real Estate." *NBER Working Paper*, January, 2005, *11053*.

Levitt, Steven and Chad Syverson "Antitrust Implications of Home Seller Outcomes When Using Flat-Fee Real Estate Agents." *Mimeo, University of Chicago*, 2007.

Mankiw, Gregory and Michael Whinston "Free Entry and Social Inefficiency." *The RAND Journal of Economics*, 1986, *17*, pp. 48-58.

Miceli, Thomas "The Welfare Effects of Non-Price Competition Among Real Estate Brokers." *Journal of the American Real Estate and Urban Economics Association*, 1992, *20*, pp. 519532.

Schroeter, John "Competition and Value of Service Pricing in the Residential Real Estate Brokerage Market." *Quarterly Review of Economics and Business*, 1987, *27*, pp. 2940.

Seim, Katja "An Empirical Model of Firm Entry with Endogenous Product-Type Choices." *The RAND Journal of Economics*, 2006, *37*, pp. 619-640.

Sutton, John "Sunk Costs and Market Structure." *Cambridge, Mass.: MIT Press*, 1991.

Sweeting, Andrew "Coordination Games, Multiple Equilibria, and the Timing of Radio Commercials." 2004, working paper.

Sweeting, Andrew "The Costs of Product Repositioning: The Case of Format Switching in the Commercial Radio Industry." *Mimeo, Northwestern University*, 2006, November.

Williams, Joseph "Agency and Brokerage of Real Assets in Competitive Equilibrium." *Review of Financial Studies*, 1998, *11*, pp. 239-280.

White, Lawrence "The Residential Real Estate Brokerage Industry: What Would More Vigorous Competition Look Like?" *Law and Economics Research Paper Series*, 2006, *6-16*.

Yavas, Abdullah "Impossibility of a Competitive Equilibrium in the Real Estate Brokerage Industry." *Journal of Real Estate Research*, 2001, *3*, pp. 187-200.

Yinger, John "A Search Model of Real Estate Broker Behavior." *The American Economic Review*, 1981, *71*, pp. 591-605.

Table 1: Sample Statistics of Matched Data[a]

	2000	1990
# of MSAs	184	184
# of Realtors	18,178	24,566
# of Observations	3,286,662	3,284,417

[a]We matched 1990 PUMS and 2000 PUMS based on CONSPUMA. Note that the definition of PUMA changed over time, whereas the definition of CONSPUMA remained the same. We first obtained the list of 197 matched MSAs. Among these MSAs, we dropped 13 MSAs for which key variables are not available.

Table 2: Individual level Demographics in 2000

	realtors		other occupation	
	mean	std.dev.	mean	std.dev.
age	48.27	13.24	39.53	13.10
education	9.79	1.55	9.01	2.31
%married	0.66	0.47	0.56	0.50
work.hours	42.00	14.18	40.10	12.22
earning	59,269	75,632	37,158	44,686
%male	0.46	0.50	0.53	0.50
%white	0.87	0.33	0.74	0.44
%house.owner	0.83	0.38	0.67	0.47
house.value	287,725	223,416	186,748	159,849
observations	18,178		3,259,851	

Table 3: Individual level Demographics in 1990

	"real estate sales occupation"		other occupation	
	mean	std.dev.	mean	std.dev.
age	48.22	13.08	37.89	12.99
education	9.71	1.6	8.77	2.31
%married	0.69	0.46	0.58	0.49
work.hours	41.28	13.96	39.62	12.17
earning	45,275	50,359	33,040	34,145
%male	0.49	0.50	0.55	0.50
%white	0.92	0.27	0.81	0.39
%house.owner	0.82	0.38	0.67	0.47
house.value	261,565	154,600	177,151	129,865
observations	24,566		3,268,484	

Table 4: Descriptive Statistics Across Markets[a]

Values per MSA	2000 mean	2000 std.dev.	1990 mean	1990 std.dev.
population	991995.40	1425097.00	789344.80	1217266.00
labor.force	408234.30	631383.80	408305.40	632420.50
#realtors	2152.33	3485.08	n.a.	n.a.
%realtors.among.labor.force	0.46	0.22	n.a.[b]	n.a.
average.realtors.earnings	49067.14	15889.39	n.a.	n.a.
average.earnings	31006.01	5576.62	27989.87	4512.27
average.house.value	141789.40	66615.41	127393.20	65924.07
average.value.house.sold	151199.90	66401.59	132067.20	65256.50
total.housing.transaction	19670.06	26584.87	17059.03	22956.12
average.realtors.hours	42.22	4.03	n.a.	n.a.
unemployment.rate	0.061	0.021	0.061	0.018
inflow.past.5.years	117436.9	165832.8	251689	761161.2
outflow.past.5.years	96835.67	143423.7	68916.01	280177.4
number of MSAs		184		184

[a]Table reports weighted values. Earnings and house values in 1990 are adjusted for inflation (values in 2000 dollar).

[b]We do not report the number and fraction of realtors in 1990 because the occupation classification for realtors in 1990 is much broader than in 2000.

Table 5: Mean Values of Market Structure Statistics by # Realtors[a]

	0-499	500-999	1000-1999	2000-3999	4000-7999	8000+
			Year = 2000			
%.realtors	0.29	0.91	0.55	0.61	0.68	0.58
population	285927	1499293	742893	1456809	2146221	5041920
labor.force	90550	174444	291752	552571	897924	2312757
average.realtors.earnings	40768	52350	52122	51038	59304	61476
average.earnings	27309	30491	31263	34643	37461	38085
average.house.value	104086	356763	154487	159425	211637	205969
average.value.house.sold	110451	351647	165129	175359	223910	219081
total.housing.transaction	4229	15789	14793	27028	48920	100182
building.age	33	32	32	30	33	32.5
gas.price	1.36	1.36	1.34	1.36	1.33	1.34
house.density	72.4	101.6	276.55	136.9	290.60	496.33
sales.per.agent	18.16	20.13	10.09	9.34	8.41	8.19
sales.per.hours	0.45	0.48	0.24	0.22	0.20	0.19
number of MSAs	69	38	28	20	17	12

[a]The number of realtors is the weighted sum of realtors in each MSA. Table reports the mean of weighted values.

Table 6: Marketing/Adversiting Expenses and Agent Growth: 2002-2006[a]

		2002	2003	2004	2005	2006
NE	Expenses	4.33%	3.94%	5.47%	4.82%	4.76%
	Agent Growth	11.7%	12.0%	5.2%	29.5%	0.7%
MA	Expenses	4.18%	3.38%	4.26%	4.70%	3.35%
	Agent Growth	10.6%	11.1%	12.4%	8.7%	5.9%
SE	Expenses	4.19%	3.97%	4.7%	4.24%	3.92%
	Agent Growth	8.3%	9.5%	7.2%	12.7%	1.6%
MW	Expenses	4.0%	3.77%	4.65%	4.26%	3.51%
	Agent Growth	1.33%	1.08%	1.69%	1.53%	1.12%
SW/MTN	Expenses	4.64%	2.5%	3.08%	2.53%	2.8%
	Agent Growth	17.0%	8.5%	10.0%	8.7%	11.0%
FW	Expenses	2.72%	2.15%	2.48%	3.46%	2.63%
	Agent Growth	10.9%	11.1%	7.0%	13.4%	4.0%

[a]Sources: The data are from the Real Trends 500 Brokerage Performance Report (2002-2006). "Expenses" refer to the marketing and advertising expenses, which include spending on mailing campaigns, handouts, inserts, open housing materials along with other company promotions. They are expressed as the fraction out of total brokerage firm revenues. "Agent Growth" refers to the percentage increase in the number of agents in a given region from the previous year. NE indicates Northeast region; MA indicates Mid-Atlantic region; SE indicates Southeast region; MW indicates Midwest region; SW/MTN indicates SW/Mountain region and FW indicates Far West region.

Table 7: Some Cost Variables[a]

	Study Hours	Renewal Interval (# years)	Hours of Continuous Education (per year)	Gas Price	Housing Density (density per square mile)	Average House Years Built
Mean	81.1	2.18	8.19	1.19	189.96	1967.54
Min	20	1	0	0.95	2.5	1954.41
Max	270	5	15	1.63	5153.8	1984.30
Std. Dev.	38.35	0.75	3.3	0.11	354.79	6.40

[a]Data Sources:
1. USA Real Estate Licensing Network;
2. C2ER Cost of Living Index, Council for Community and Economic Research;
3. PUMS 2000.

Table 8: Sample Evidence for NPL Convergence using 2000 Data

MSA	# obs	$\sigma_m^0 \equiv \frac{N_m}{S_m}$	σ_m^1	σ_m^2	σ_m^3	σ_m^4	σ_m^5
5120	56128	0.00471	0.00471	0.00467	0.00467	0.00467	0.00467
6280	47445	0.00329	0.00329	0.00324	0.00322	0.00322	0.00322
1520	32899	0.00560	0.00560	0.00537	0.00538	0.00538	0.00538
160	19932	0.00328	0.00328	0.00293	0.00293	0.00293	0.00293
3320	19275	0.00468	0.00468	0.00438	0.00438	0.00438	0.00438
1000	16141	0.00583	0.00583	0.00561	0.00561	0.00561	0.00561
200	14907	0.00729	0.00729	0.00675	0.00674	0.00674	0.00674
240	13032	0.00216	0.00216	0.00271	0.00274	0.00273	0.00273
1720	12638	0.00696	0.00696	0.00664	0.00664	0.00664	0.00664
1560	9723	0.00274	0.00274	0.00524	0.00526	0.00525	0.00526
2700	9579	0.01292	0.01292	0.00921	0.00918	0.00918	0.00918
3290	9046	0.00231	0.00231	0.00415	0.00414	0.00414	0.00414
9200	5745	0.00994	0.00994	0.00643	0.00642	0.00642	0.00642
2900	5686	0.00316	0.00316	0.00696	0.00698	0.00698	0.00698
1240	5503	0.00507	0.00507	0.00345	0.00344	0.00344	0.00344
480	5476	0.00614	0.00614	0.00537	0.00535	0.00535	0.00535
870	4191	0.00389	0.00389	0.00345	0.00344	0.00344	0.00344
5330	4083	0.01135	0.01135	0.00657	0.00657	0.00657	0.00657
4940	4048	0.00156	0.00156	0.00214	0.00214	0.00214	0.00214
3350	1693	0.00099	0.00099	0.00018	0.00017	0.00017	0.00017

Table 9: 2000 Results for 1st Stage Regression of Revenue for Realtors[a]

	1st iteration			5th iteration		
variable	estimate	s.e.	p-value	estimate	s.e.	p-value
total.transaction	-5.7E-08	2.0E-06	0.98	-6.1E-07	2.1E-06	0.77
total.labor.force	1.8E-07	9.9E-08	0.07	1.9E-07	1.0E-07	0.06
#realtors[b]	-3.2E-05	2.2E-05	0.14	-2.1E-05	2.5E-05	0.40
mean.value.house.sold	2.7E-06	3.8E-07	0.00	2.7E-06	3.8E-07	0.00
anti-rebate	0.0560	0.0311	0.07	0.0539	0.0312	0.08
internet.adoption[c]	0.5111	0.2223	0.02	0.4888	0.2226	0.03
internet.search[d]	-0.4215	0.2167	0.05	-0.3868	0.2149	0.07
inflow	1.2E-06	3.5E-07	0.00	1.0E-06	3.5E-07	0.00
outflow	-1.6E-06	4.1E-07	0.00	-1.6E-06	4.1E-07	0.00
land.area	3.0E-06	2.3E-06	0.19	3.6E-06	2.2E-06	0.10
male	0.2166	0.0206	0.00	0.2167	0.0206	0.00
age	0.0598	0.0044	0.00	0.0597	0.0044	0.00
age^2	-0.0006	0.0000	0.00	-0.0006	0.0000	0.00
white	0.2132	0.0344	0.00	0.2130	0.0344	0.00
education	-0.0342	0.0360	0.34	-0.0340	0.0360	0.35
education^2	0.0040	0.0017	0.02	0.0040	0.0017	0.02
live.same.msa.5yr	0.1072	0.0212	0.00	0.1074	0.0212	0.00
married	0.1139	0.0211	0.00	0.1142	0.0211	0.00
both.working	0.2843	0.0557	0.00	0.2840	0.0557	0.00
work.weeks	0.0312	0.0015	0.00	0.0312	0.0015	0.00
work.hours	0.0168	0.0012	0.00	0.0168	0.0012	0.00
full.time[e]	0.2942	0.0372	0.00	0.2942	0.0372	0.00
self.employed	0.0420	0.0203	0.04	0.0418	0.0203	0.04
constant	5.1911	0.2338	0.00	5.2025	0.2338	0.00

[a]Robust standard errors (adjusted for correlations within MSAs) are reported. The dependent variable is log(revenue). The sample includes 160 free-standing MSAs for which the Internet adoption rates can be computed from the CPS. The number of observations is 10,855.

[b]The 1st iteration uses the actual number of realtors, while the subsequent iterations use the expected number of realtors predicted from the previous iteration.

[c]Internet is the rate of Internet adoption in each MSA, computed from the 2000 CPS supplements for Internet and computer use.

[d]Internet.search is the proportion of respondents in each MSA who reported to use regularly the Internet to search for information. It is also computed from the 2000 CPS.

[e]Full.time is equal to 1 if the respondent works for more than 35 hours.

Table 10: 2000 Results for 2nd Stage Probit[a]

variable	1st iteration estimate	s.e.	p-value	5th iteration estimate	s.e.	p-value
$\widehat{R}$ (in $100,000)	0.9435	0.0739	0.00	0.9703	0.0752	0.00
q	0.0688	0.1901	0.72	-0.0829	0.1900	0.66
q^2	0.0003	0.0001	0.00	0.0003	0.0001	0.00
q^3	-1.1E-06	4.0E-07	0.01	-1.1E-06	3.7E-07	0.01
$q\times$#realtors	-1.3E-06	6.6E-07	0.04	-3.1E-06	7.2E-07	0.00
$q\times$mean.year.built	-4.2E-05	9.4E-05	0.66	3.4E-05	9.4E-05	0.72
$q\times$mean.gas.price	-7.3E-03	8.6E-03	0.40	-6.5E-03	8.2E-03	0.43
$q\times$mean.house.density	1.9E-05	1.2E-05	0.11	1.8E-05	1.3E-05	0.17
$q\times$total.labor.force	-1.2E-09	4.5E-09	0.78	8.9E-09	3.9E-09	0.02
land.area	3.5E-06	1.3E-06	0.01	3.8E-06	1.3E-06	0.00
live.same.msa.5yr	0.1082	0.0090	0.00	0.1077	0.0090	0.00
male	-0.1908	0.0125	0.00	-0.1925	0.0125	0.00
age	-0.0013	0.0021	0.55	-0.0016	0.0021	0.46
age^2	0.0002	0.0000	0.00	0.0002	0.0000	0.00
white	0.2245	0.0194	0.00	0.2241	0.0194	0.00
education	0.4801	0.0322	0.00	0.4808	0.0322	0.00
education^2	-2.0E-02	1.4E-03	0.00	-2.0E-02	1.4E-03	0.00
married	3.1E-02	9.0E-03	0.00	3.1E-02	9.0E-03	0.00
both.work	9.5E-02	2.2E-02	0.00	9.4E-02	2.2E-02	0.00
mean.earning.other.job	-1.9E-05	9.1E-06	0.04	-1.8E-05	8.6E-06	0.03
mean.family.income	7.5E-06	4.4E-06	0.09	7.0E-06	4.1E-06	0.09
mean.value.other.house[b]	-1.3E-07	7.5E-07	0.87	-1.2E-07	7.0E-07	0.87
unemployment.rate	-0.3745	0.7639	0.62	-0.6187	0.7493	0.41
population	2.3E-08	7.7E-08	0.76	1.9E-08	7.1E-08	0.79
total.houses	-1.7E-09	1.9E-07	0.99	1.1E-08	1.8E-07	0.95

[a]Robust standard errors (adjusted for correlations within MSAs) are reported. The dependent variable is whether the observation is a realtor. The regression includes state fixed effects whose coefficient estimates are suppressed. The 1st iteration uses the actual number of realtors, while the subsequent iterations use the expected number of realtors predicted from the previous iteration. The sample includes 160 free-standing MSAs for which the Internet adoption rates can be computed from the CPS. The number of observations is 2,115,499.

[b]The mean value of all other houses in the MSA, excluding houses sold.

Table 11: 2000 Main Estimates[a]

variable	estimate	p-val.	marginal effect $\partial y/\partial x$	marginal effect p-val.	mean value
$\widehat{R}$ (in $100,000)	0.9703	0.00	0.0081	0.00	0.297784
q	-0.0829	0.66	0.0005	0.76	7.96
q^2	0.0003	0.00	0.0000	0.00	191.94
q^3	-1.1E-06	0.01	-9.1E-09	0.00	35306.90
$q\times$#realtors	-3.1E-06	0.00	-1.2E-08	0.03	23032.70
$q\times$mean.year.built	3.4E-05	0.72	-3.1E-07	0.70	15686.40
$q\times$mean.gas.price	-6.5E-03	0.43	-6.5E-05	0.37	9.48
$q\times$mean.house.density	1.8E-05	0.17	1.6E-07	0.11	1175.77
$q\times$total.labor.force	8.9E-09	0.02	-8.0E-12	0.83	4300000.00
land.area	3.8E-06	0.00	3.1E-08	0.01	3948.05
live.same.msa.5yr	0.1077	0.00	0.0010	0.00	0.32
male	-0.1925	0.00	-0.0017	0.00	0.54
age	-0.0016	0.46	0.0000	0.51	39.11
age^2	0.0002	0.00	0.0000	0.00	1701.83
white	0.2241	0.00	0.0016	0.00	0.80
education	0.4808	0.00	0.0041	0.00	10.44
education^2	-2.0E-02	0.00	-1.7E-04	0.00	115.64
married	3.1E-02	0.00	2.6E-04	0.00	0.57
both.work	9.4E-02	0.00	7.2E-04	0.00	0.93
mean.earning.other.job	-1.8E-05	0.03	-1.6E-07	0.04	31454.00
mean.family.income	7.0E-06	0.09	6.2E-08	0.09	63972.90
mean.value.other.house	-1.2E-07	0.87	-8.6E-10	0.89	134303.00
unemployment.rate	-0.6187	0.41	-0.0032	0.62	0.05
population	1.9E-08	0.79	1.8E-10	0.78	1300000.00
total.houses	1.1E-08	0.95	5.3E-11	0.97	534222.00

[a]The table reports the estimates from the 5th iteration in Table 11, and additionally reports marginal effects.

Table 12: Robustness Checks for Cost Estimates

	Baseline		Add Price Growth and Firm Ratio			log (#realtors) and log (total.labor.force)		$\frac{\text{#realtors}}{\text{total.labor.force}}$		1990-2000 data[a] (MSA fixed effects)	
	estimate	s.e.	estimate	s.e.	mean	estimate	s.e.	estimate	s.e.	estimate	s.e.
A. Estimates for the First Stage Regression of Realtor's Revenues											
#realtors	-2.1E-05	2.5E-05	-1.9E-05	2.6E-05						-9.6E-07	6.2E-06
$\log N$						0.080	0.037				
N/S								8.005	6.768		
total.labor.force	1.9E-07	1.0E-07	2.2E-07	1.0E-07						-3.3E-07	9.6E-08
$\log S$						-0.002	0.045				
mean.value.house.sold	2.7E-06	3.8E-07	2.4E-06	3.9E-07		2.4E-06	3.7E-07	2.4E-06	3.7E-07	1.7E-06	3.2E-07
anti-rebate	0.054	0.031	0.047	0.031		0.057	0.031	0.053	0.031		
internet	0.489	0.223	0.456	0.224		0.384	0.219	0.349	0.218		
internet.search	-0.387	0.215	-0.384	0.216		-0.388	0.209	-0.317	0.209		
OFHEO.price.growth[b]			0.96	0.52	0.039						
firm.ratio[c]			-6.16	8.16	0.003						
firm.ratio×age			0.12	0.16	0.124						
firm.ratio×live.same.msa			4.94	4.50	0.001						
B. Estimates for the Second Stage Probit											
$\widehat{R}$	0.970	0.075	0.958	0.073		0.975	0.078	0.965	0.112	0.517	0.066
q	-0.083	0.190	-0.125	0.191		0.337	0.209	0.096	0.196	-0.019	0.003
q^2	2.9E-04	7.1E-05	2.7E-04	6.6E-05		4.0E-04	1.0E-04	2.7E-04	7.4E-05	3.6E-04	8.4E-05
q^3	-1.1E-06	3.7E-07	-9.4E-07	3.4E-07		-2.1E-06	7.1E-07	-1.1E-06	3.8E-07	-1.5E-06	4.7E-07
q×#realtors	-3.1E-06	7.2E-07	-3.2E-06	7.3E-07						-9.1E-07	3.6E-07
$q \times \log N$						-3.8E-03	1.2E-03				
$q \times \frac{N}{S}$								-2.123	1.002		
q×total.labor.force	8.9E-09	3.9E-09	9.4E-09	3.8E-09						2.5E-09	1.9E-09
$q \times \log S$						1.1E-03	1.8E-03				
q×mean.year.built	3.4E-05	9.4E-05	5.5E-05	9.5E-05		-1.7E-04	1.1E-04	-5.3E-05	9.9E-05		
q×mean.gas.price	-6.5E-03	8.2E-03	-5.7E-03	8.4E-03		-1.4E-02	9.8E-03	-7.9E-03	9.6E-03		
q×mean.house.density	1.8E-05	1.3E-05	2.0E-05	1.4E-05		2.0E-05	1.2E-05	1.1E-05	1.2E-05		
C. Mean of Implied Average Variable Costs											
wasteful.AVC		1,120		1,171			1,463		1,092		1,526
AVC		2,419		2,394			2,631		2,438		4,137
%wasteful.AVC in AVC		0.46		0.49			0.56		0.45		0.37
# observations		2,115,499		2,056,738			2,115,499		2,115,499		6,571,079

[a] To combine PUMS 1990 and PUMS 2000, we redefine realtors in 2000 to be partially consistent with 1990 data. We also redefine MSAs in 2000 to be consistent with 1990 data, and recompute most of MSA-level variables accordingly, but several MSA-level variables defined over the original MSA (e.g. internet variables) cannot be recomputed. Note also that we use both free-standing MSAs and non-free-standing MSAs. MSA fixed effects are included.

[b] OFHEO.price.growth is the change in annual price index from the Office of Federal Housing Enterprise Oversight (OFHEO).

[c] Firm.ratio is equal to the ratio of the number of real estate firms with more than 100 employees among total number of real estate firms.

Table 13: Counterfactual Results for Anti-rebate Policy[a]

	Initial Equilibrium	Counterfactual Equilibrium
	A. MSAs with Anti-rebate Policy (counterfactual: anti-rebate = 0)	
#realtors	2,473	2,329
mean of realtor revenue	\$33,738	\$31,460
mean of AVC	\$2,141	\$2,081
mean of wasteful AVC	\$797	\$750
total realtor revenue	\$85.4 million	\$75.6 million
total variable cost	\$43.9 million	\$42.2 million
#MSAs		34
	B. MSAs without Anti-rebate Policy (counterfactual: anti-rebate = 1)	
#realtors	3,776	3,998
mean of realtor revenue	\$34,657	\$36,872
mean of AVC	\$2,444	\$2,530
mean of wasteful AVC	\$1,217	\$1,288
total realtor revenue	\$146 million	\$163 million
total variable cost	\$82.9 million	\$86.8 million
#MSAs		126

[a]The table reports weighted means of MSA-level values, using the total labor force as weights. The results for initial equilibrium are computed using the equilibrium entry probabilities estimated from the converged iteration. For the counterfactual results, we compute new equilibrium entry probabilities, and recalculate all the values.

Table 14: Comparison of MSA-level Characteristics for MSAs with or without Anti-rebate

	With Anti-rebate		Without Anti-rebate	
	mean	std.dev.	mean	std.dev.
mean.value.house.sold	134,047	19,532	150,205	41,504
total.transaction	28,947	22,626	38,592	36,773
#realtors	2,500	1,954	3,924	4,115
#labor.force	538,565	432,978	667,491	555,297
$\frac{\#\text{realtors}}{\#\text{labor.force}}$	0.0045	0.0012	0.0053	0.0023
converged σ_m	0.0042	0.0011	0.0051	0.0020
Internet	0.4244	0.0755	0.4306	0.0856
Internet.search	0.3226	0.0769	0.3202	0.0860
year.built	1969.27	4.00	1970.44	6.95
housing.unit	463,061	346,760	548,143	429,721
land.area	3,530	1,783	4,041	5,480
population.density	285	102	424	249
housing.density	119	42	174	101

Table 15: Counterfactual Results for Changes in Internet Adoption Rate and Internet Search Intensity[a]

	Initial Equilibrium	Counterfactual Equilibrium
	A. Increase internet by 10% and full search intensity[b] (internet.search = internet)	
#realtors	3,554	3,435
mean of realtor revenue	$34,500	$33,273
mean of AVC	$2,392	$2,349
mean of wasteful AVC	$1,145	$1,109
total realtor revenue	$135 million	$126 million
total variable cost	$76.3 million	$74.1 million
#MSAs		160
	B. Do not change internet but full search intensity (internet.search = internet)	
#realtors	3,554	3,399
mean of realtor revenue	$34,500	$32,883
mean of AVC	$2,392	$2,334
mean of wasteful AVC	$1,145	$1,098
total realtor revenue	$135 million	$124 million
total variable cost	$76.3 million	$73.5 million
#MSAs		160
	C. Increase internet by 10% but do not change internet.search	
#realtors	3,554	3,744
mean of realtor revenue	$34,500	$36,522
mean of AVC	$2,392	$2,469
mean of wasteful AVC	$1,145	$1,207
total realtor revenue	$135 million	$150 million
total variable cost	$76.3 million	$79.5 million
#MSAs		160

[a]For this counterfactual, we consider what if Internet adoption rate increased by 10% and all Internet users regularly use the Internet to search for information. The counterfactual results are computed using new equilibrium entry probabilities. The table reports weighted means of MSA-level values, using the total labor force as weights.

[b]According to the 2000 CPS, about 75% of Internet users in the MSAs included in our analysis report to regularly use the Internet to search for information. The counterfactual for full search intensity is that 100% of Internet users regularly use the Internet to search for information.

Table 16: Commission Rate from the CEX[a]

Year	Rate	House Price	Rate	House Price	Rate
1988	0.061	20000	0.053	200000	0.059
1989	0.064	30000	0.046	220000	0.049
1990	0.054	40000	0.060	230000	0.063
1991	0.055	50000	0.055	240000	0.083
1992	0.055	60000	0.054	250000	0.053
1993	0.051	70000	0.065	260000	0.060
1994	0.061	80000	0.062	270000	0.034
1995	0.052	90000	0.053	280000	0.069
1996	0.045	100000	0.066	300000	0.056
1997	0.065	110000	0.058	310000	0.065
1998	0.053	120000	0.052	320000	0.056
1999	0.059	130000	0.058	330000	0.042
2000	0.062	140000	0.051	360000	0.060
2001	0.053	150000	0.052	370000	0.042
2002	0.057	160000	0.054	380000	0.063
		170000	0.055	390000	0.073
		180000	0.047	400000+	0.056
		190000	0.054		

[a]Table reports the average commission rates by year, and by housing prices (at $10,000 intervals). The commission rate is computed from the 1988-2002 Consumer Expenditure Surveys.

Table 17: Correlation Between Entry and Housing Price Changes from 1990 to 2000

	changes in log(average price of houses)
changes in log(total realtors)	0.4564
changes in $\log(\frac{\text{total realtors}}{\text{total labor force}})$	0.5042

Table 18: Commission Rates and Real Commission Fees: 1991-2005[a]

		Median House Prices		Commission Fees	
Year	Rate	2006 Dollar	%Change	2006 Dollar	%Change
1991	6.10%	$153,925		$9,389	
1992	6.04%	$153,235	-0.45%	$9,255	-1.43%
1993	5.94%	$153,632	0.26%	$9,126	-1.40%
1994	5.88%	$155,145	0.98%	$9,123	-0.04%
1995	5.83%	$155,365	0.14%	$9,058	-0.71%
1996	5.75%	$158,029	1.71%	$9,087	0.32%
1997	5.64%	$162,168	2.62%	$9,146	0.66%
1998	5.48%	$167,881	3.52%	$9,200	0.59%
1999	5.44%	$171,031	1.88%	$9,304	1.13%
2000	5.42%	$172,427	0.82%	$9,346	0.45%
2001	5.12%	$177,939	3.20%	$9,110	-2.52%
2002	5.14%	$188,634	6.01%	$9,696	6.42%
2003	5.12%	$198,557	5.26%	$10,166	4.85%
2004	5.08%	$212,655	7.10%	$10,803	6.26%
2005	5.02%	$230,059	8.18%	$11,549	6.91%

[a]Sources: This table is taken from Report by FTC and DOJ released in April 2007. Commission rates are from REAL Trends 500; real median home prices are from U.S. Department of Housing and Urban Development, U.S. Housing Market Conditions, 4th Quarter 2006, Tables 6-9 (Feb. 2007), and are a weighted average of new and existing home prices, based on annual sales; median home prices are converted into 2006 dollar with consumer price index for all goods for all urban consumers (CPI-U) from Bureau of Labor Statistics (http://data.bls.gov/PDQ/servlet/SurveyOutputServlet); commission fees are calculated by multiplying commission rates by real median home prices.

Table 19: Agent Growth and Commission Rate Growth Across Regions 2002-2003[a]

Region	SW/MTN	MW	SE	FW	MA	NE
Agent Growth	8.50%	8.70%	9.50%	11.10%	11.10%	12.00%
Commission Rate Growth	1.94%	-4.27%	-2.18%	1.63%	0.00%	-1.15%

[a]Sources: This table is derived from the statistics based on the 2004 Real Trends 500 Brokerage Performance Report. The agent growth and commission rate growth are computed as the percentage changes in real estate agents and percentage changes in average commission rates from 2002 to 2003. NE indicates Northeast region; MA indicates Mid-Atlantic region; SE indicates Southeast region; MW indicates Midwest region; SW/MTN indicates SW/Mountain region and FW indicates Far West region.

Table 20: Effects of Realtor Competition on Realtor Satisfaction[a]

	Home Buyer Sample	Home Seller Sample
Dependent Variable	REA Satisfaction	REA Satisfaction
Agents/Labor Force	-0.74	2.93
	(12.83)	(16.95)
Number of Observations	3145	1785

[a]Standard errors in parentheses. The data source is the 2001 National Association of Realtors Surveys on Home Buyers and Sellers. Additional control variables include housing characteristics, neighborhood characteristics, market conditions, search/selling process characteristics, home buyer/seller characteristics.