Testing Cost Inefficiency under Free Entry in the Real Estate Brokerage Industry*

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

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

December 23, 2010

Abstract

This article provides an empirical framework to study entry and cost inefficiency in the real estate brokerage industry. We develop a structural entry model that exploits individual level data on entry and earnings to estimate potential real estate agents' revenues and reservation wages, thereby recovering costs of providing brokerage service. Using the Census data, we estimate the model and find strong evidence for cost inefficiency under free entry, particularly attributable to wasteful non-price competition. We further use the estimated model to evaluate welfare implications of the rebate bans that currently persist in some U.S. states. We find that removing rebate bans would decrease the equilibrium number of real estate agents by 5.14% and reduce total brokerage costs by 8.87%.

Keywords: real estate brokerage, entry, cost inefficiency, structural estimation

JEL classification: C35, C51, L85, R31

*We thank seminar participants at Minneapolis Fed, IIOC, NBER Summer Institute, SITE, WFA Summer Real Estate Symposium, University of California at Berkeley, University of Chicago, University of Toronto, and University of Wisconsin. We are grateful to the editor, the associated editor, and two anonymous referees for their valuable comments and suggestions that significantly improved the manuscript. Max Rempel provided excellent research assistance in data processing. All errors are our own. 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, 112 David Kinley Hall, 1407 West Gregory Drive, Urbana, IL 61821.

1 Introduction

Though a large body of theoretical research has shown that free entry could lead to social inefficiency (e.g., Anderson, DePalma and Nesterov 1995; Chamberlain 1933; Dixit and Stiglitz 1977; Mankiw and Whinston 1986; Sutton 1991), few empirical studies have tested for social inefficiency under free entry. The main difficulty is that relevant data on cost and benefit measures are hardly available. To address this difficulty, we use the individual level data on entry and earnings in the U.S. real estate brokerage industry, and construct a structural entry model to estimate potential agents' revenues and reservation wages, thereby recovering the costs of providing brokerage services. Based on the cost estimates, we investigate the efficiency of free entry equilibria when agents' ability to compete on commissions is limited. We further provide an analysis of the welfare consequences of removing the anti-rebate policies that currently persist in some U.S. states.

The U.S. residential real estate brokerage industry provides an important and suitable setting for our study. Entry and efficiency in this industry have been recurrently featured in the news and policy debates (e.g., White 2006). In particular, the industry is characterized by two unique features that are conducive to examining cost inefficiency under free entry. First, empirical evidence shows that commission rates in this industry range between 5% and 6% with little variation across markets over time (Hsieh and Moretti 2003). Second, barriers to entry appear to be low in the industry (DOJ and FTC Report 2007), leading to a large number of real estate agents when house prices are high. In particular, membership in the National Association of Realtors nearly doubled between 1997 and 2006. Given that agents have limited ability to compete on commissions, free entry is likely to lead agents to spend more resources on marketing and prospecting. To the extent that the resulting benefits do not offset the committed resources, the non-price competition induced by free entry is socially inefficient.

To test cost inefficiency under free entry, we estimate a structural entry model for the real estate brokerage markets. The model builds on the insight from the entry literature (e.g., Berry and Reiss 2007): the observed entry decision is an indicator of the underlying profitability. Hence, using the information on individuals' entry decisions as well as agents' revenues and reservation wages, one can in principle recover the cost estimates and analyze the key determinants of brokerage costs. This approach entails two difficulties, however. First, revenues are observed only for real estate agents, while reservation wages are observed only for those who have chosen alternative occupations, suggesting potential self-section bias in the estimation of revenues and reservation wages. Second, entry decisions are interdependent, in that individual entry decisions depend on the number of real estate agents which in turn is determined by individual entry decisions.

Given these difficulties, we first apply the Type 5 Tobit model in Amemiya (1985) to address potential selection bias. We then account for the interdependence of entry decisions by imposing the rational expectation equilibrium as in Brock and Durlauf (2001). The equilibrium is then represented by fixed points in entry probabilities in that agents' beliefs about other agents' entry coincide with the entry probabilities predicted from the model. We estimate our equilibrium model by employing a nested pseudo likelihood (NPL) algorithm (Aguirregabiria and Mira 2002, 2007). The main empirical specification is estimated based on cross-sectional variations in observed market and individual characteristics. To address the potential endogeneity concern due to unobserved market heterogeneity, we also estimate two additional specifications: the finite mixture version of the NPL algorithm and the panel data estimation with market fixed effects.

Using the 5 percent sample of the 2000 *Census of Population and Housing*, we find strong evidence for cost inefficiency under free entry, particularly attributable to wasteful non-price competition. Specifically, a 10% increase in the number of real estate agents would on average increase individual costs by 5.8% from $13,951 to $14,760. At the market level, we find that in a typical metropolitan area, a 10% increase in the number of agents increases total brokerage costs by 12.4% or $6.74 million, and wasteful competition accounts for at least 34% of the increase in total costs. Our cost estimates remain robust even after accounting for the presence of unobserved market heterogeneity. Using these 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. In an average metropolitan area with anti-rebate policies, removing these rebate bans would reduce real estate agents' revenues, thereby decreasing the equilibrium number of agents by 5.14% and reducing total brokerage costs by 8.87%.

This article contributes to two strands of the literature. First, the idea that free entry and lack of price competition together could lead to excessive entry in real estate brokerage markets goes back to prior work by Crockett (1982), Miceli (1992), and Turnbull (1996). By examining the relationship between house price and agent productivity, Hsieh and Moretti (2003) provide the first empirical evidence suggesting that entry is socially excessive in the real estate brokerage industry. We complement their study by employing a structural approach to recover the cost estimates. This allows us to investigate the underlying sources of cost inefficiency and to further perform counterfactual experiments. Second, from the methodological point of view, this article illustrates the importance of accounting for endogeneity and sample selection bias in studying entry decisions. To do so, we extend the NPL algorithm (Aguirregabiria and Mira 2002, 2007) by incorporating the Type 5 Tobit model. Our estimation approach is also related to Berry and Waldfogel (1999) who study entry decisions in the radio broadcasting industry, along with the recent structural work on games with incomplete information (e.g., Bajari, et al. 2010; Seim 2006; Sweeting 2009).

The article is organized as follows. Section 2 provides a simple theoretical framework on free entry and cost inefficiency in the real estate brokerage industry. Section 3 develops our structural model and discusses the estimation approaches we use in this article. Section 4 describes the data. The main results are presented in Section 5. This section also reports the robustness checks. Section 6 concludes.

2 Theoretical Framework

The purpose of this section is to develop a simple theoretical model that examines whether free entry is socially inefficient in the real estate brokerage industry. Let us consider a stylized environment in which all houses are identical and all agents provide identical real estate brokerage service. In each market, the transaction price of each house is P, and the total number of transactions is Q. The agent's commission rate is fixed at τ. For simplicity, we assume that P, Q and τ are exogenously given. The brokerage market comprises N identical agents, where N is endogenous under the assumption of free entry and exit. Given the symmetry assumption, it follows that the number of transactions facilitated by each agent is $q = \frac{Q}{N}$. The total revenue to each agent completing q transactions is $\tau P q$.

Turning to the cost side specification, for an agent that competes with $N-1$ agents, the cost of providing q transactions is $F + C(q, N)$, where F denotes fixed costs and $C(q, N)$ is a continuous variable cost function that satisfies the following assumptions.

Assumption 1. $\frac{\partial C}{\partial q} > 0$; $\frac{\partial^2 C}{\partial q^2} > 0$.

Assumption 2. $\frac{\partial C}{\partial N} > 0$; $\frac{\partial^2 C}{\partial N^2} > 0$; $\frac{\partial^2 C}{\partial q \partial N} < 0$

Assumption 3. $\frac{\partial AC}{\partial q} < 0$, *where* $AC \equiv \frac{F+C(q,N)}{q}$.

Assumption 1 simply says that variable cost function is increasing and convex in q. This is a standard assumption. Assumption 2 says that variable cost function is increasing and convex in N. The rationale for $\frac{\partial C}{\partial N} > 0$ is that, given that agents cannot compete on price, more entrants will induce incumbent agents to spend more resources on competing for potential clients. The negative cross-derivative in $C(N, q)$ implies that $\frac{\partial^2 C}{\partial q \partial N} = \frac{\partial(\partial C/\partial N)}{\partial q} < 0$. This is consistent with the notion that a decrease in the number of transactions intensifies competition, forcing existing agents to spend more resources attracting potential clients when competing with additional agents. Finally, Assumption 3 requires that average costs decline with the output of real estate brokerage service. That is, economies of scale are present in producing real estate brokerage service. This can be driven either by the presence of fixed costs or by declining average variable costs.

We model the entry process as a two-stage game. In the first stage, potential entrants decide whether to become a real estate agent. For those who become agents, their profits are realized in the second stage. Given N, an agent's post-entry profit is given by $\pi(N) \equiv \tau P q - C(q, N) - F$. Under free entry, a potential agent enters as long as her profit is larger than her reservation wage, w. Hence, the equilibrium number of agents, N^e, satisfies the following condition:

$$\pi(N^e) = \tau P q^e - C(q^e, N^e) - F = w, \tag{1}$$

where $q^e = \frac{Q}{N^e}$. To examine whether free entry is efficient in the real estate brokerage industry, we compare N^e with N^*, where N^* denotes the socially optimal number of agents that solves

$$\max_N V(N) = CS(P, Q, \tau) + \tau PQ - NC(q, N) - NF \tag{2}$$

where CS is total consumer surplus. Taking P, Q, and τ as given, entry is socially efficient if and only if it minimizes the total costs, $NC(q, N) + NF$.

Proposition 1 *Suppose that Assumptions 1-3 hold. Then,* $N^e > N^*$. *That is, the equilibrium number of real estate agents is socially excessive.*

The proof is contained in Web Appendix A. In equilibrium, the marginal entrant can cause social inefficiency through two channels. First, new entrants cause existing agents to reduce q. If average costs decline with output (Assumption 3), this would lead to an inefficient increase in the costs of producing real estate brokerage service. This is the standard source of inefficiency that has been modeled in the entry literature (e.g., Mankiw and Whinston 1986; Berry and Waldfogel 1999). Second, new entrants also force existing agents to compete more aggressively in marketing their service to potential clients, leading to an inefficient increase in the marketing costs (Assumption 2). Given the institutional fact that commission rates are fixed, this second source of inefficiency is unique to the real estate brokerage industry and is the key assumption to test in our empirical analysis.

Before we proceed, we should highlight that our model relies on two simplifying assumptions: lack of price competition and limited product differentiation. These assumptions are justified in two ways. Empirically, using the data collected from alternative sources, Web Appendix C provides evidence that supports the two simplifying assumptions. Conceptually, the assumptions are consistent with the industry practice, namely, competition on commissions or service variety is often impeded by tacit collusion among local real estate agents or by explicit laws or regulations in some states. Of course, one may argue that competition could make some agents become more specialized in a certain neighborhood, or work for a certain type of clients or houses, thereby generating potential benefits to consumers. To address this, Web Appendix B also provides a formal model that incorporates the potential benefits of entry through product differentiation. The model shows that as long as the degree of product differentiation is limited, the cost inefficiency effect would dominate the potential consumer gains.

3 Econometric Framework

3.1 The Model

To describe our empirical model, let us examine the entry decision of a potential real estate agent i in market m, $m = 1, \ldots, M$. Following Section 2, we consider a two-stage entry model. In the first stage, I_m potential agents in market m simultaneously decide whether to enter the market or not. We define $d_{i,m} = 1$ if potential agent i enters real estate brokerage market m; $d_{i,m} = 0$, otherwise. We assume that potential agents are those in the labor force who are eligible for obtaining real estate agent licenses. Accordingly, if $d_{i,m} = 1$, then an individual i becomes a real estate agent; otherwise, she chooses a different occupation and becomes a non-agent. In the second stage, if agent i has entered market m, her post-entry net profit is given by

$$\Pi_{i,m} = R_{i,m}(N_m, X^{\mathrm{R}}_{i,m}) - C_{i,m}(N_m, X^{\mathrm{C}}_{i,m}) - W_{i,m}(X^{\mathrm{W}}_{i,m}), \tag{3}$$

where $R_{i,m}$ denotes revenues; $C_{i,m}$ indicates costs which include both variable costs and fixed costs; $W_{i,m}$ denotes reservation wages; N_m is the number of real estate agents in market m; $X^{\mathrm{R}}_{i,m}$, $X^{\mathrm{C}}_{i,m}$, and $X^{\mathrm{W}}_{i,m}$ include both individual-specific variables and market-specific variables that respectively determine revenues, costs, and reservation wages. In what follows, we suppress market subscripts m to simplify the exposition.

Individual i decides to enter the market if and only if she expects to earn non-negative net profits in the second stage. The entry decision is then determined by the following threshold rule

$$d_i = 1 \text{ iff } E(\Pi_i|\Omega_i) \geq 0 \tag{4}$$

where the expectation is taken over the information set Ω_i available for individual i. The information set includes observables $X_i \equiv \{X^{\mathrm{R}}_i, X^{\mathrm{W}}_i, X^{\mathrm{C}}_i\}$, as well as private information that is observed only by individual i, but not by the econometrician, nor by other individuals. We make use of this threshold condition to construct the probability model of entry, and further exploit observed R_i and W_i to recover the remaining cost function in the net profit.

One potential concern with this approach is that R_i is observed only for real estate agents, while W_i is observed only for non real estate agents. The actual earnings observed for agents and non-agents are unlikely to be random samples of the population. Thus, if we impute the expected revenues and reservation wages for all samples without accounting for the selection issue, the imputed values are likely to be biased, so that our cost estimates would be inconsistent. To address potential selection bias, we apply the approach developed by Lee (1978) and Willis and Rosen (1979), which is termed as the Type 5 Tobit model in Amemiya (1985). In what follows, we provide our econometric specifications and present our approach in more detail.

3.1.1 Revenue and Reservation Wage

Those who become real estate agents would earn R_i in the second stage. Ideally, an agent's revenue should be equal to the sum of her commission incomes generated from each transaction that she has facilitated, which in turn depends on the number of transactions, house transaction price, and commission rate in each transaction. Since we do not observe individual transaction data, we cannot construct a fully structural model for the second stage competition. Instead, we specify revenues in the following reduced form

$$R_i = \alpha N + X_i^{\mathrm{R}} \delta^{\mathrm{R}} + \eta_i, \tag{5}$$

where α and δ^{R} are parameters to be estimated, and X_i^{R} is a vector of observed variables that are presumed to determine agent i's earnings in market m. According to the baseline model laid out in Section 2, all agents are symmetric so that $R_i = \tau P \times Q/N$. In light of this model, X_i^{R} includes the average house transaction price P and the aggregate number of transactions Q. The symmetric assumption can be tested by examining whether $\frac{\partial \log R_i}{\partial \log N}$ is equal to -1, where $\frac{\partial \log R_i}{\partial \log N} = \alpha \times \frac{N}{R_i}$. In practice, however, agents differ in their expertise, efforts, and skills, thus earning different commission incomes. We therefore go beyond the simple symmetric model and allow for agent heterogeneity by including a rich set of individual demographics and other market-specific variables.

We assume that the error term in (5) contains two components, such that $\eta_i = \epsilon_i^{\mathrm{R}} + u_i^{\mathrm{R}}$. The first term ϵ_i^{R} represents a revenue shock that is realized in the second stage but unknown to agent i in the first stage. Examples include an unexpected housing boom or slump realized in the second stage. Because ϵ_i^{R} is unknown to individual i in the first stage, it is unlikely to affect the entry decision in the first stage. We assume that $E(\epsilon_i^{\mathrm{R}} | X_i^{\mathrm{R}}) = 0$. The second term u_i^{R} reflects private information that is known to individual i but unobservable to other agents and the econometrician. For example, if a potential entrant is socially well-connected, she may expect a longer list of potential clients. The presence of such private information introduces the possible correlation between her expected revenues and her entry decision. In this respect, we also rewrite (5) as $R_i = \widetilde{R}_i + \eta_i$, where $\widetilde{R}_i \equiv E(R_i|X_i) = \alpha E(N|X_i) + X_i^{\mathrm{R}} \delta^{\mathrm{R}}$.

For those who choose not to become real estate agents, we observe W_i. Similar to the revenue equation, we specify reservation wages in the following reduced form

$$W_i = X_i^{\mathrm{W}} \delta^{\mathrm{W}} + \nu_i, \tag{6}$$

where δ^{W} is a vector of parameters to be estimated, and X_i^{W} is a vector of observed variables that determine earnings of non-agents. Similar to the revenue specification, the error term ν_i contains two components, such that $\nu_i = \epsilon_i^{\mathrm{W}} + u_i^{\mathrm{W}}$, where ϵ_i^{W} captures idiosyncratic shocks in reservation wages that are realized in the second stage, but unpredictable to individual i in the first stage;

u_i^{W} reflects private information; and $E(\epsilon_i^{\mathrm{W}}|X_i^{\mathrm{W}}) = 0$. We also rewrite (6) as $W_i = \widetilde{W}_i + \nu_i$, where $\widetilde{W}_i \equiv E(W_i|X_i) = X_i^{\mathrm{W}}\delta^{\mathrm{W}}$, and we use tilde to denote the expected value.

3.1.2 Cost

Once we obtain $\widetilde{R}_i$ and $\widetilde{W}_i$, the remaining part of the net profit in (3) is the cost. Given the lack of information on individual agents' transactions, we cannot separate variable costs from fixed cost. Therefore, we consider the cost in the following reduced form

$$C_i = \beta N + X_i^{\mathrm{C}}\delta^{\mathrm{C}} + \zeta_i, \tag{7}$$

where β and δ^{C} are parameters to be estimated, and X_i^{C} is a vector of observed variables that determine costs. We assume that X_i^{C} includes two sets of variables. The first set is market m's characteristics, such as average building age, housing density, and licensing requirements. The second set is individual i's demographics. Similar to revenues and reservation wages, we specify the error term as $\zeta_i = \epsilon_i^{\mathrm{C}} + u_i^{\mathrm{C}}$, where ϵ_i^{C} captures idiosyncratic shocks in costs that are realized in the second stage, but unpredictable to individual i in the first stage; u_i^{C} reflects private information such as ability and experiences in providing the real estate brokerage service; and $E(\epsilon_i^{\mathrm{C}}|X_i^{\mathrm{C}}) = 0$.

Testing whether entry induces wasteful non-price competition is formally equivalent to testing whether $\beta > 0$. Since real estate agents cannot directly compete on prices, an increase in the number of entrants intensifies competition along other dimensions. In particular, to compete for each sale, real estate agents have to spend additional amount of effort involving a wide range of activities, 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 the direct monetary costs of prospecting, but also the opportunity costs associate with the time spent by agents on these prospecting activities. Unlike the costs involved in selling or buying a house, most of these marketing expenses do not necessarily generate enough benefit to offset the resources committed to promoting, and are therefore considered "wasteful."

3.1.3 Entry and Selection

Having specified the underlying processes for revenues, reservation wages, and costs, we can combine them and construct individual i's expected net profit as

$$\widetilde{\Pi}_i = \widetilde{R}_i - \widetilde{W}_i - [\beta E(N|X_i, u_i) + X_i^{\mathrm{C}}\delta^{\mathrm{C}}] - u_i, \tag{8}$$

where $\widetilde{\Pi}_i \equiv E(\Pi_i|X_i, u_i)$, and $u_i = u_i^{\mathrm{C}} + u_i^{\mathrm{W}} - u_i^{\mathrm{R}}$. Following (4), the expectation is taken over Ω_i which includes both X_i and private information u_i. Note that u_i contains u_i^{R} and u_i^{W}, which introduces correlations between the entry decision, revenues, and reservation wages. Given that R_i is only observed for real estate agents and that W_i is only observed for non-agents, we need to explicitly account for potential selection bias when estimating the revenue equation and the reservation wage equation. To this end, we assume that $\{\eta_i, \nu_i, u_i\}$ are i.i.d. drawings from a trivariate normal distribution with zero means and variance-covariance matrix given by

$$\begin{pmatrix} \sigma_\eta^2 & \sigma_{\eta\nu} & \sigma_{\eta u} \\ \sigma_{\eta\nu} & \sigma_\nu^2 & \sigma_{\nu u} \\ \sigma_{\eta u} & \sigma_{\nu u} & \sigma_u^2 \end{pmatrix}.$$

Given this assumption on the error structure, we estimate revenues in (5) and reservation wages in (6), using the Type 5 Tobit model in Amemiya (1985). The selection, i.e., the entry decision, however, is determined by (8), which in turn depends on $\widetilde{R}_i$ and $\widetilde{W}_i$. To estimate revenues and reservation wages, we thus begin with the following reduced form profit

$$\Pi_i^* = \gamma_1 E(N|X_i, u_i) + X_i\gamma_2 - u_i, \tag{9}$$

where we use Π_i^* to distinguish the reduced form selection equation in (9) from the structural entry equation in (8). We then write the selection rule as

$$\Pr(d_i = 1) = \Pr(\Pi_i^* \geq 0).$$

Thus, when we estimate (5) using the real estate agent samples, we consider $E(R_i|X_i; d_i = 1)$, instead of $E(R_i|X_i)$. Similarly, we consider $E(W_i|X_i; d_i = 0)$, when we estimate (6) using the non-agent samples. The assumption of a trivariate normal distribution for $\{\eta_i, \nu_i, u_i\}$ then implies that the expected revenue can be written as

$$E(R_i|X_i; \Pi_i^* \geq 0) = \alpha E(N|X_i) + X_i^{\mathrm{R}}\delta^{\mathrm{R}} + \frac{\sigma_{\eta u}}{\sigma_u}\lambda_a, \tag{10}$$

where $E(\eta_i|X_i;\Pi_i^* \geq 0) = \frac{\sigma_{\eta u}}{\sigma_u}\lambda_a$, and λ_a is given by

$$\lambda_a = -\phi\left((\gamma_1 E(N|X_i) + X_i\gamma_2)/\sigma_u\right)/\Phi\left((\gamma_1 E(N|X_i) + X_i\gamma_2)/\sigma_u\right),$$

where $\phi(\cdot)$ is the standard normal probability density function, and $\Phi(\cdot)$ is the standard normal cumulative distribution function of a standard normal random variable. Similarly, the expected reservation wage can be written as

$$E(W_i|X_i;\Pi_i^* < 0) = X_i^{\mathrm{W}}\delta^{\mathrm{W}} + \frac{\sigma_{\nu u}}{\sigma_u}\lambda_b, \tag{11}$$

where $E(\nu_i|X_i;\Pi_i^* < 0) = \frac{\sigma_{\nu u}}{\sigma_u}\lambda_b$, and λ_b is given by

$$\lambda_b = \phi\left((\gamma_1 E(N|X_i) + X_i\gamma_2)/\sigma_u\right)/\left[1 - \Phi\left((\gamma_1 E(N|X_i) + X_i\gamma_2)/\sigma_u\right)\right].$$

Once we estimate the predicted revenues and reservation wages as above, we consider the expected profit in (8), and estimate the cost parameters using the entry model given by

$$\begin{aligned}
\Pr(d_i = 1|X_i) &= \Pr(\widetilde{R}_i - \widetilde{W}_i - [\beta E(N|X_i,u_i) + X_i^{\mathrm{C}}\delta^{\mathrm{C}}] - u_i \geq 0) \\
&= \Phi\left(\frac{\widetilde{R}_i - \widetilde{W}_i - \beta E(N|X_i,u_i) - X_i^{\mathrm{C}}\delta^{\mathrm{C}}}{\sigma_u}\right).
\end{aligned} \tag{12}$$

Due to the presence of $E(N|X_i,u_i)$, however, we cannot directly estimate (12). In the next section, we address this issue by imposing the rational expectation equilibrium condition.

3.1.4 Equilibrium

Potential agent i enters the market as long as $\widetilde{\Pi}_i \geq 0$, but $\widetilde{\Pi}_i$ depends on $E(N|X_i,u_i)$ which is agent i's belief about other agents' entry decisions conditional on her own entry, that is,

$$E(N|X_i,u_i) = 1 + \sum_{j \neq i} E(d_j|X_i,u_i).$$

To the extent that agents' beliefs are rational, the rational expectation equilibrium requires potential agent i's beliefs about other agents' entry to be correct, in that they coincide with the entry probabilities of other agents. In other words, $E(d_j|X_i,u_i) = \Pr(d_j = 1|X_j), \forall j \neq i$. Thus,

$$E(N|X_i,u_i) = 1 + \sum_{j \neq i}^{I} \Pr(d_j = 1|X_j).$$

Considering that there are a large number of potential entrants in each market, it is plausible to assume that the right hand side is the same for all potential agents in the same market. Therefore,

$E(N|X_i, u_i) = \widehat{N}$ for each $i = 1, \ldots, I_m$, where $\widehat{N}$ denotes the expected number of real estate agents in market m. The equilibrium condition is then written as the fixed points in $\widehat{N}$ given by

$$\widehat{N} = \sum_{i=1}^{I} \Pr(\widetilde{R}_i(\widehat{N}) - \widetilde{W}_i - \beta\widehat{N} - X_i^{\mathrm{C}}\delta^{\mathrm{C}} \geq u_i), \quad (13)$$

where we note that $\widetilde{R}_i$ also depends on $\widehat{N}$, and (13) should hold for each market $m = 1, \ldots, M$. This equilibrium is related to the rational expectation equilibrium in Brock and Durlauf (2001), and we impose the equilibrium condition (13) in our estimation.

3.2 Estimation

For our main results, we estimate the model using the nested pseudo likelihood (NPL) algorithm (Aguirregabiria and Mira 2002, 2007) augmented with the three-step estimation procedure described below. To address potential endogeneity concern due to unobserved market heterogeneity, we further consider the finite mixture version of the NPL algorithm as a robustness check, and our estimation procedure is discussed in Section 3.2.2.

3.2.1 Modified NPL with Three-Step Estimation

We consider the following three-step estimation approach. In the first step, we use the reduced form selection equation in (9) and estimate a probit model, which allows us to compute the inverse Mill's ratios $\widehat{\lambda}_a$ and $\widehat{\lambda}_b$. In the second step, we estimate (10) using observations with $d_i = 1$, and impute $\widehat{R}_i$ for all samples. Similarly, we estimate (11) using observations with $d_i = 0$, and impute $\widehat{W}_i$ for all samples. In the third step, we estimate the structural entry model in (12). For our estimation, we use a heteroskedastic probit model and account for potential heteroskedasticity in u_i. Specifically, we follow Harvey (1976) and model the variance as a multiplicative function of Z_i, where Z_i denotes a subset of variables that enter the entry equation and are likely to affect the variance of u_i.

If our model does not depend on $E(N|X_i, u_i)$, we can complete our estimation using the three-step estimation described above. Because the model depends on $E(N|X_i, u_i)$, however, we need to impose the equilibrium condition in (13). Several empirical studies on games with incomplete information (see, e.g. Augereau, et al. 2006; Seim 2006; Sweeting 2008) 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 $\widehat{N}_t$ enters the equation (13) through $\widehat{R}_i$ as well.

For this reason, we use the NPL algorithm developed by Aguirregabiria and Mira (2002, 2007), which is more straightforward to apply in our context. Note that a consistent nonparametric estimator for $\widehat{N}$ is simply the actual number of real estate agents. Thus, we use the actual number of real estate agents in each market as an initial guess for $\widehat{N}^0$. Because the Census data described in Section 4 are representative random samples of the U.S. population, we use the weighted sum of d_i to estimate $\widehat{N}^0$ in our application. That is, $\widehat{N}^0 = \sum_{i=1}^{T} d_i \times \mathsf{weight}_i$, where weight_i is the weight provided by the Census data, and T is the actual number of observations for market m in our data, so that $I = \sum_{i=1}^{T} \mathsf{weight}_i$. We then use the three-step approach described above. This completes the first iteration. Using the estimates from the first iteration, we predict $\widehat{N}^1$. More specifically, we predict $\widehat{N}^h$ at the h-th iteration by using

$$\widehat{N}^h = \sum_{i=1}^{T} \Phi \left(\frac{\widehat{R}_i(\widehat{N}^{h-1}) - \widehat{W}_i - \widehat{\beta}\widehat{N}^{h-1} - X_i^{\mathrm{C}}\widehat{\delta}^{\mathrm{C}}}{\widehat{\sigma}_{u_i}} \right) \times \mathsf{weight}_i, \tag{14}$$

where $\widehat{R}_i$, $\widehat{W}_i$, $\widehat{\beta}$, $\widehat{\delta}$, and $\widehat{\sigma}_{u_i} = \exp(Z_i\widehat{\mu})$ are estimated at the h-th iteration. Once we compute $\widehat{N}^1$ for all markets, we use $\widehat{N}^1$ and follow the three-step approach above to estimate the model parameters. We repeat this procedure until $\widehat{N}$ 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.

Because of the interdependence between the estimation of our entry equation and the estimation of revenues and reservation wages, we also check the convergence in $\widehat{R}_i$ and $\widehat{W}_i$. Note that the NPL algorithm iterates over N, but in each iteration, we also compute new estimates for $\widehat{R}_i$ and $\widehat{W}_i$. Hence, the NPL algorithm additionally allows us to check the convergence in $\widehat{R}_i$ and $\widehat{W}_i$, which ensures an internal consistency in our model. One more comment on our estimation is that we replace $\widetilde{R}_i$ and $\widetilde{W}_i$ in (8) with their predicted values $\widehat{R}_i$ and $\widehat{W}_i$. The expected net profit we use for our estimation is then rewritten as

$$\widetilde{\Pi}_i = \widehat{R}_i - \widehat{W}_i - [\beta E(N|X_i) + X_i^{\mathrm{C}}\delta^{\mathrm{C}}] - u_i^*,$$

where $u_i^* = u_i + [\widetilde{R}_i - \widehat{R}_i] - [\widetilde{W}_i - \widehat{W}_i]$, so that u_i^* contains not only private information but also prediction errors. To account for these prediction errors, we use the bootstrap method to estimate the standard errors.

3.2.2 NPL with Finite Mixture

The error term in our entry model includes private information which is known to each agent but is not known to other agents. However, there may be market-specific unobservables that are commonly observed by agents in the same market, but not observed to the econometrician. Let ξ_m denote unobserved market heterogeneity for each market m. We then include ξ_m in the expected net profit as follows

$$\widetilde{\Pi}_i = \widetilde{R}_i - \widetilde{W}_i - [\beta E(N|X_i, u_i) + X_i^{\mathrm{C}}\delta^{\mathrm{C}}] - \xi - u_i,$$

where the subscript m is suppressed. The presence of ξ suggests potential endogeneity concern, in that ξ may be correlated with market-level variables including the expected number of real estate agents in particular. Ideally, we would like to use panel data to different out market-fixed effects, but as we discussed in Section 4, the precise information on the entry decision d_i is included only in one year of the data. As a compromise, we use two approaches to address the concern due to unobserved market heterogeneity. First, using the panel data with imprecise information on d_i, we estimate the model by including market fixed effects. Second, using the cross-sectional data with more precise information on d_i, we assume the finite mixture distribution for ξ and implement a finite mixture version of the NPL algorithm as described in Web Appendix D. In this case, our identification relies mainly on functional form given the lack of the panel data.

3.3 Identification

The identification of our model mainly relies on the fact that we observe individuals' entry decisions, revenues, and reservation wages. More specifically, an individual's entry decision is an indicator of her expected net profit, namely, the difference between the expected revenues and the expected reservation wages and brokerage costs. Thus, after controlling for the effect of N on agents' revenues, the remaining effect of N on the entry decision should be associated with costs.

In order for this identification strategy to work, we need to deal with three additional issues. The first issue is the potential selection bias, given that revenues are observed only for real estate agents and reservation wages are observed only for non-agents. As discussed in Section 3.1.3, we address this issue by combining the empirical entry model with the Type 5 Tobit approach developed by Lee (1978) and Willis and Rosen (1979). The second issue stems from the presence of unobserved market heterogeneity. As described in Section 3.2.2, we provide two additional specifications to address this concern: a finite mixture version of the NPL algorithm that relies on our main cross-sectional data, and a market fixed effect model that relies on panel data. The results from both specifications are presented as robustness checks in Section 5.2.1.

The final identification issue concerns how to separate cost estimates from other components of the profit function. To this end, we provide two sets of exclusion restrictions that provide further identification of the cost function. First, to separate the expected revenues and costs from the reservation wages, we consider a dummy variable that indicates whether real estate brokerage commission rebates are prohibited. Intuitively, this variable captures the degree to which agents can compete on commissions, which should affect an individual's expected revenues and costs as a real estate agent. But there is no evident reason that the anti-rebate policy would affect reservation wages that the individual could earn from other professions.

Second, to further separate the expected revenues from costs, we include the net inflows of immigrants to a certain MSA in the past 5 years in the revenue and reservation wage equations but not in the cost function. Intuitively, larger inflows of immigrants represent higher demand for the local real estate brokerage service. Thus, individuals in markets with larger net inflows are likely to predict higher revenues from the real estate brokerage business. In addition, larger net inflows are also associated with higher regional growth, leading to higher reservation wages in this market as well. However, there is no obvious reason that the net inflow would affect the costs of the real estate brokerage services after controlling for the market level cost drivers.

4 Data

The primary source of our data is the 5 percent sample of the 2000 *Census of Population and Housing* Public Use Microdata Series, commonly referred to as the 2000 PUMS. We additionally use the 1990 PUMS data. However, the occupation codes are not comparable across different years. In the 1990 PUMS, 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 2000 PUMS, 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 between these two years' data, as well as the imprecise classification of real estate brokers and agents in the 1990 PUMS, we restrict our main empirical analysis to the 2000 PUMS. For our robustness check in Section 5.2.1, we use both PUMS data to control for market fixed effects.

Markets for real estate services are local, owing to the nature of the service given that real estate is fixed in a geographic location. There is no single, agreed upon method for empirical market definitions, but it is clear that the markets should be self-contained in the sense that there is little 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. Specifically, we use free-standing metropolitan statistical areas (MSAs), which are generally surrounded by non-metropolitan territory and therefore are not integrated with other metropolitan areas.

In the real estate market, brokerage firms are relatively unimportant while the important capital and goodwill belong to the salesperson (Hsieh and Moretti 2003). Therefore, we model the entry decision at the real estate agent and broker level, rather than the brokerage firm level. In principle, anyone can become a real estate agent as long as he or she obtains real estate agent licenses. For this reason, we assume that potential real estate agents are those in the labor force who are eligible for obtaining real estate agent licenses. Specifically, we consider all individuals in our data who worked in 2000 and were at least 18 years old and high school graduates. Table 1 presents the summary statistics of individual level demographic variables. The table shows the differences between real estate agents and other occupations. On average, real estate agents and brokers tend to be older, slightly more educated, and more likely to be married. In addition, real estate agents tend to earn higher income than non-agents. The differences shown in the table suggest that real estate agents are unlikely to be randomly selected from the population, suggesting the potential importance of accounting for selection bias in estimating agents' entry decisions.

Because we observe the number of transactions at the market level rather than at the individual level, much of the agent heterogeneity explored in our empirical analysis will be attributed to the observed agent demographics and revenues. The former includes a set of individual level characteristics, such as agent, education, marital status, race, gender, whether to work full time, and whether being self-employed. The agents' revenues are measured by their self-reported earnings. Implicit in this measure are two assumptions: (1) agents' earnings come exclusively from sales of residential homes; and (2) agents report their revenues from commissions as their earnings. Although we cannot directly test these two assumptions, we provide two ways to assess the reliability of the revenue measure.

First, we regress the log of the 6 percent of total home sales in each market on the log of total self-reported earnings for real estate agents. The coefficient is 1.048 (s.e. of 0.002). Thus, the reported agents' earnings in the Census appear to closely reflect revenues from housing sales, consistent with Hsieh and Moretti (2003). Second, we compare earning data reported by the Census from those reported by the Occupation Employment Statistics (OES). The distributions of real estate agents' earnings reported by these two sources are fairly consistent, providing further support for our measures of real estate agents' revenues.

Table 2 presents the summary statistics for MSA-level variables. To measure the number of house transactions in each market, 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 currently lives. The table shows that an average MSA has a sample of 36,985 house transactions and 3,549 real estate agents. The Census also asks homeowners about the values of their houses. Using this information, we construct the mean value of houses in each city and the average value of houses sold in each city in the previous year. In an average MSA, the mean value of all houses is \$135,739, while the mean value of houses sold is \$147,472. In this article, we use the mean value of houses sold as the measure of the house price.

Table 3 shows the structure of the real estate brokerage market. As the number of real estate agents increases, the average house values, the average agent earnings, and the density of houses and population increase substantially. Following Hsieh and Moretti (2003), we compute two measures of average productivity of real estate agents: sales per agent and sales per hour. Both measures of average productivity decrease with the number and the share of real estate agents in the local market. One may consider this pattern as an indicator of excessive entry: the average cost per transaction increases with the number of real estate agents. However, the cities with a large number of real estate agents tend to have more expensive houses, and it may be more difficult and costly to provide real estate brokerage services for transactions of expensive houses. The preceding descriptive statistics thus provide only suggestive evidence.

5 Results

5.1 Estimation Results

5.1.1 Revenues and Reservation Wages

Table 4 reports the converged estimates from the second-step regression of revenues. The table shows that the coefficient on the number of real estate agents is negative and statistically significant. This points to the presence of both business stealing effect and some form of price competition. To further examine the competition effect, we include an interaction term between the number of agents and the anti-rebate dummy variable, and the estimated coefficient is positive and significant. Intuitively, anti-rebate laws prohibit agents from giving rebates on their commissions, thereby limiting the degree of price competition, which in turn reduces the magnitude of the negative entry effect on revenues. In addition, the revenue elasticities computed from our estimates indicate that the symmetric assumption $\frac{\partial \log R_i}{\partial \log N} = -1$ is rejected, suggesting that the real estate brokerage market is characterized by substantial amount of agent heterogeneity and some degree of product differentiation. We control for agent heterogeneity by including a rich set of individual demographics in our empirical specifications. In Web Appendix C, we use a separate dataset to examine the effect of product differentiation and discuss its implications on our results.

Table 4 also shows that the coefficient on the Internet search variable is negative and fairly significant. The Internet search variable is imputed from the *2000 Current Population Survey: Supplement for Internet and Computer Use*, and it indicates the fraction of respondents in each MSA who reported regularly conducting searches on the Internet. Previous literature finds that the diffusion of the Internet makes it easier for sellers to sell their houses on their own (Hendel, Nevo and Ortalo-Magne 2009) or to use online discount brokers (Levitt and Syverson 2007). Intuitively, these outside options should create competitive pressure on traditional real estate agents and reduce their commission income, which is confirmed by the estimated coefficient.

Turning to the remaining coefficients in Table 4, we find that most variables have the expected signs. At the individual level, being male, white, married, older, and more educated, all increase individual revenues 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 worked for longer hours tend to have higher revenues. In particular, full-time agents earn significantly more than part-time agents. All together, these estimates demonstrate a significant degree of heterogeneity in efforts, skills and experiences at the agent level.

At the market level, consistent with the basic model laid out in Section 2, we find that, everything else being equal, higher local house prices and more transactions translate into higher commission income. In addition, larger markets, proxied by land area, tend to produce more business for real estate agents. Markets that attract more immigrants have higher commission incomes for real estate agents, presumably because new immigrants, in search of local housing in order to settle down, increase the demand for real estate brokerage services. Lastly, as shown in the bottom of Table 4, λ_a is significantly negative, suggesting that a model that does not account for the selection effect could lead to biased results.

Table 5 presents the converged estimates from the second-step regression of reservation wages. Similar to the revenue estimation results, we find that the individual demographics and market conditions have the expected signs. For example, at the individual level, being white, male, and married are positive indicators of higher wages. Moreover, older and more educated people tend to earn more. In addition to these life-cycle variables, working effort also matters. In particular, those who work longer hours are likely to earn more. At the market level, higher wages are positively correlated with higher local house prices and larger inflows of immigrants.

5.1.2 Cost Estimates

Table 6 presents the converged estimates from the third-step probit model, in which we fix the coefficient on $\widehat{R}_i - \widehat{W}_i$ to be 1, and estimate the cost parameters using the entry model in (12). Most of the cost shifter have expected signs. For example, areas with older houses have relatively lower brokerage costs, and this building age effect is marginally decreasing. Selling houses is also less costly in areas characterized by high building density. In addition, the adoption of the Internet has a negative effect on the brokerage costs, presumably because the Internet helps agents to reduce their cost of obtaining housing market information and reaching their potential clients. At the individual level, we find that older and more educated agents tend to have lower brokerage costs. We also include state dummy variables to capture differences in licensing requirements and other heterogeneity across states, but their coefficients are suppressed in the table.

The parameters of interest are coefficients on the number of real estate agents and its interaction with the anti-rebate dummy. If entry leads to more inefficient use of resources in marketing activities, a larger number of real estate agents would result in an increase in the cost of providing brokerage service, pointing to the presence of wasteful competition. Such effects would be more substantial in the states with anti-rebate policies. These hypotheses are confirmed by the negative and statistically significant coefficients on $\widehat{N}$ and on its interaction with the anti-rebate dummy.

Using these coefficient estimates, we further compute the predicted costs for individual agents and seek to quantify the wasteful competition effect. Given that we estimate the reduced form cost in (7), however, we cannot directly measure the magnitude of the wasteful competition effect, since we can only compute the total effect of entry on costs, that is, $\frac{\mathrm{d}C_i}{\mathrm{d}N}$. Nevertheless, information on $\frac{\mathrm{d}C_i}{\mathrm{d}N}$ is still useful because it provides a lower bound on the sheer effect of wasteful non-price competition. To see this, recall that the simple model presented in Section 2 shows that

$$\frac{\mathrm{d}C_i}{\mathrm{d}N} = \frac{\partial C_i}{\partial q_i}\frac{\partial q_i}{\partial N} + \frac{\partial C_i}{\partial N}. \tag{15}$$

The wasteful competition effect due to entry refers only to the second term $\frac{\partial C_i}{\partial N}$ in (15), which is the direct effect of entry on costs. Since we do not observe individual-level transactions q_i, we cannot recover the exact magnitude of $\frac{\partial C_i}{\partial N}$. However, one can make the following observations. First, Table 4 shows that the individual agent's revenue, R_i, decreases with N. To the extent that R_i depends positively on the number of transactions facilitated by an individual agent, one could infer that q_i is also negatively correlated with N. Second, $\frac{\partial C_i}{\partial q_i}$ represents the marginal cost. In a standard setting, costs cannot be strictly decreasing in q_i, and the real estate brokerage transaction is not a special exception. Hence, as long as costs are identified, the marginal cost should not be negative. These two observations establish the following inequality, $\frac{\partial C_i}{\partial q_i}\frac{\partial q_i}{\partial N} \leq 0$. Therefore, the imputed total effect of entry, $\frac{\mathrm{d}C_i}{\mathrm{d}N}$, provides a lower bound for the true magnitude of wasteful non-price competition effect, $\frac{\partial C_i}{\partial N}$.

We thus quantify the effect of entry on costs. To do so, we increase N by 10% while fixing all the other variables, and then compute the resulting change in the predicted cost, $\frac{\Delta \widehat{C}_i}{\widehat{C}_i}$, for each observation. Note that the mean of the predicted individual cost is \$13,951, and we find that the mean of $\frac{\Delta C_i}{C_i}$ is 0.058 with the 95% confidence interval given by $[0.054, 0.062]$. In other words, a 10% increase in the number of real estate agents increases individual costs on average by 5.8% from \$13,951 to \$14,760. To quantify the effect on market-level costs, we compute total costs for each MSA by summing up individual agents' costs. Table 7 presents the summary statistics of the imputed TC and other MSA-level imputed variables, where TC denotes total cost. We find that the mean of TC is \$46.3 million, and the mean of wasteful TC is \$19.2 million, where wasteful TC is the sum of individual costs attributed to $\widehat{N}$ and its interaction term with the anti-rebate dummy. We further calculate ΔTC, the increase in TC due to a 10% increase in N for each MSA. We find that in an average MSA, a 10% increase in N would increase TC by 12.4%, or \$6.74 million. We also find that the resulting increase in wasteful TC would on average account for 34% of ΔTC, which is equivalent to \$2.2 million. As discussed above, these estimates should be interpreted as lower bounds for the true effect of wasteful competition. Therefore, our results provide strong evidence for the presence of wasteful non-price competition in the real estate brokerage markets.

We have thus far established strong evidence for the first source of cost inefficiency – wasteful competition. As shown in Section 2, a second source of cost inefficiency is loss of economies of scale, which could be driven either by the presence of fixed costs or by declining average variable costs. Given that we do not observe individual transaction q_i, we cannot recover the curvature of the variable cost function, nor can we recover the fixed costs for each agent. However, based on the cost estimates above, we can provide suggestive evidence on fixed costs. To do so, we consider a simple linear specification for MSA-level total costs given by $TC_m = Q_m(\gamma_1 + X_m^{\text{VC}}\gamma_2) + X_m^{\text{F}}\gamma_3 + \varepsilon_m$, where X_m^{VC} is a vector of variables related to variable costs, including market level cost shifters in Table 6; and X_m^{F} is a vector of variables associated with fixed costs, including state-level licensing requirements such as the number of hours required to take real estate transaction courses, and the requirements for license renewal and exam fees.

Using 130 MSAs for which information on licensing requirements is available, we estimate the simple linear regression specified above. Two findings are worth mentioning. First, the coefficient on Q_m is 0.2 with a standard error of 0.03. The coefficients on its interaction terms are also positive and statistically significant. These results indicate that total market costs increase with total output, providing support for Assumption 1 in Section 2 at the aggregate level. Second, the coefficients on licensing variables are small and statistically insignificant, and the predicted values for $X_m^{\text{F}}\hat{\gamma}_3/N$ are negligible. This is consistent with the conventional wisdom that barriers to entry in the real estate brokerage industry are extremely low. Thus the evidence from this preliminary regression of TC_m does not lend support to cost inefficiency due to loss of economies of scales.

5.2 Robustness Checks

5.2.1 Unobserved Market Heterogeneity

The main results in Section 5.1 are estimated using the cross-sectional variations in observed market and individual characteristics. As discussed in Section 3.3, one potential concern for these results is that the estimates of revenues and costs might be biased if unobserved market characteristics were important. In this section, we provide two additional specifications to address this concern. In the first specification, we assume the finite mixture distribution for unobserved market heterogeneity and implement the finite mixture version of the NPL algorithm as discussed in Section 3.2.2. In the second specification, we fit the baseline model to the panel data to difference out market fixed effects.

The results from the finite mixture estimation are displayed in Specification II of Table 8. For comparison purpose, Specification I presents the estimates from the baseline model reported in Tables 4 and 6. Panel A shows a sample of the coefficient estimates from the revenue regression. Panel B reports the coefficient estimates from the entry probit model, and Panel C presents the average values for several imputed variables including TC. Most estimates in Specification II are fairly close to those in Specification I. In particular, the coefficients on $\widehat{N}$ and its interaction with the anti-rebate dummy are similar in both specifications, and their absolute magnitudes are slightly larger in the finite mixture estimation than in the baseline estimation. Thus, the estimate of the wasteful competition effect in our main specification, if biased, is likely to be underestimated rather than overestimated. These results suggest that our main empirical findings in Section 5.1 are robust to some form of unobserved market heterogeneity.

To make the computation feasible, we choose five discrete points to implement the finite mixture algorithm above. This would address some representation of unobserved heterogeneity present in the sample, but may not be sufficient to capture the entire set of unobserved characteristics at the market level. For this reason, we also consider the second approach that relies on panel data to difference out market-level fixed effects. Because there are inconsistencies between the 1990 PUMS data and the 2000 PUMS data, we make three adjustments for the sample. First, as discussed in Section 4, the occupational category for real estate agents in the 1990 PUMS is broad and imprecise, in that it includes several other occupations. In the 2000 data, however, we cannot identify these other occupations, except for real estate appraiser. Hence, we redefine real estate agents in 2000 to include real estate brokers and sales agents, as well as appraisers, so that the indicator for real estate agents is partially consistent between the 1990 data and the 2000 data. Second, because the MSA definition has changed between 1990 and 2000, we also redefine MSAs in 2000 to be consistent with the 1990 data. Finally, we are forced to drop the anti-rebate variable, as we do not have information on which markets had anti-rebate policies in 1990.

The results from our panel data estimation are presented in Specification III of Table 8. As shown in Panel A, the coefficient on $\widehat{N}$ in the revenue estimation is still negative and statistically significant, but larger in magnitude than in Specifications I and II. The difference in magnitude is largely due to the fact that the sample for the panel estimation has been adjusted in three different ways in order to incorporate the 1990 PUMS data. Panel B shows that some of the cost estimates are also different between the baseline results and the panel data results. Nevertheless, the coefficients on $\widehat{N}$ and its interaction term with the anti-rebate policy are reasonably similar to those from the baseline estimation. To examine the implications of these estimates, we then compute the implied costs. The results reported in Panel C show that the implied costs are fairly consistent with those from the baseline model. Therefore, the results from the panel data estimation suggest that our main empirical results in Section 5.1 are still robust to more general form of unobserved market heterogeneity.

5.2.2 Endogeneity Concerns

Our model thus far has taken the number of transactions, Q, and the average house price, P, as exogenously given. One might argue that both variables are likely to be correlated with entry decisions. Intensified competition may force some agents to increase the amount of effort in order to convert a potential transaction into a real transaction, leading to an increase in the number of transactions in a given market. Alternatively, some agents may have to suggest higher listing prices in order to obtain the business from potential clients. Higher listing prices partly translate to higher transaction prices, leading to an increase in 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, P, and N. 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. To see this, note that, with a negative coefficient on the number of agents, the positive influence of entry decisions on Q and P could only bias down the coefficient estimate for N in magnitude. Given that the estimated coefficient is negative (both statistically and economically) across different specifications, the evidence for wasteful non-price competition should be qualitatively consistent and robust, even in the presence of endogeneity of Q or P.

5.3 Counterfactual Analysis: Anti-rebate Policy

In this section, we use our model to further perform counterfactual experiments and evaluate the effects of the anti-rebate policy. In computing counterfactual values, one needs to account for the equilibrium effect: a counterfactual experiment of abolishing the anti-rebate policy would change predicted revenues, thereby altering the entry probability; this in turn increases or decreases predicted revenues and costs, thus further changing the entry probability. Therefore, we compute counterfactual equilibrium by solving for the new fixed points in the entry probabilities. Note that the counterfactual interventions have no effect on house price or transaction volume, since P and Q are both taken as given in our model. In Web Appendix E, we provide a similar counterfactual experiment to evaluate the effects of the Internet diffusion in the real estate brokerage industry.

Real estate agents can compete on price either by charging lower commission rates, or by offering rebates which include cash payments and various inducements such as gift certificates and coupons. To the extent that a 5%-6% commission rate is an industry standard, rebates become a powerful tool for price competition among agents. However, several states prohibit real estate agents, by law or regulation, from giving consumers rebates on brokerage commissions. In 2000, rebates were banned in the following 15 states: Alabama, Alaska, Iowa, Kansas, Kentucky, Louisiana, Mississippi, Missouri, New Jersey, North Dakota, Oklahoma, Oregon, South Dakota, Tennessee, and West Virginia. Proponents of anti-rebate policies argue that under rebate bans, consumers are likely to choose agents based on the quality of services rather than the price of services, and that rebate bans protect consumers from false and misleading rebate offers. In contrast, DOJ and FTC report (2007) argues that there seems to be no evidence for harmful effects of rebates on consumer welfare, and that anti-rebate policies actually harm consumers by preventing price competition. Moreover, the explicit prohibition of price competition has raised antitrust concerns. In March 2005, for example, the U.S. Department of Justice filed an antitrust suit against the Kentucky Real Estate Commission for rebate bans. By the end of 2007, three out of 15 states had abolished the rebate bans.

Despite public interest and antitrust implications, little empirical evidence has been provided for the welfare implications of anti-rebate policies, partly due to lack of relevant data on costs. For this reason, we use our model estimates and quantify potential benefits of removing anti-rebate bans. To do so, we consider the MSAs that had the rebate bans in 2000 and compute counterfactual results in the absence of the anti-rebate policy. Note that removing anti-rebate bans affects agents' expected profits and their entry decisions in two ways. On the one hand, it decreases agents' revenues by putting downward pressure on their commission rates, hence discFouraging potential agents from entering the market. On the other hand, it reduces agents' marketing expenses by allowing them to compete along the price dimension. However, lowered costs could encourage more potential agents to enter the market. Therefore, it is not immediately clear what the resulting equilibrium outcomes will be.

Using our estimated model, we thus compute the new equilibrium under the counterfactual scenario. Panel A of Table 9 presents the summary statistics of key variables in the initial equilibrium and the new equilibrium. We find that abolishing rebate bans would decrease the number of real estate agents in the new equilibrium. In a typical MSA with the anti-rebate policy, the equilibrium number of real estate agents is decreased by 5.14%, and total brokerage costs are declined by 8.87%, thus generating substantial cost savings in the real estate brokerage industry.

Panel B of Table 9 presents the results from the converse counterfactuals, in which we consider MSAs without rebate bans and examine what would happen if they prohibited rebates. The table reports that, in an average MSA, adopting an anti-rebate policy would increase the number of real estate agents by 9.93% and total brokerage costs by 19.64%. Note that the counterfactual effects are larger in magnitude in Panel B than in Panel A. As Table 10 shows, this difference may be due to the fact that MSAs without anti-rebate policy tend to be more densely populated, have more houses and transactions, and higher house prices, than MSAs with anti-rebate policies. Lastly, the table also reports the interim results for the counterfactual scenario when the equilibrium effect is ignored. These interim results are considerably different from the final counterfactual results, suggesting the importance of accounting for the equilibrium effect when computing counterfactuals.

To conclude, our counterfactual results show that an anti-rebate policy is harmful not only because it has a negative impact on consumer welfare in terms of higher commission rates, but also because it leads to excessive entry and higher brokerage costs.

6 Conclusion

In this article, we provide an empirical framework to study entry and cost inefficiency in the real estate brokerage industry. We develop a structural entry model that exploits individual level data on entry and earnings to estimate potential real estate agents' revenues and reservation wages, thereby recovering the costs of providing brokerage service. Using the 2000 PUMS data, we find strong evidence for cost inefficiency under free entry. Unlike other industries that have been analyzed in the entry literature, our estimates show that the cost inefficiency in the real estate brokerage industry is largely due to agents' inability to compete on commissions, rather than loss of economies of scale from fixed costs.

Using our model estimates, we evaluate the welfare implications of the anti-rebate policies that still currently persist in some U.S states. In a typical MSA with rebate bans, removing such bans would decrease the equilibrium number of real estate agents by 5.14% and save total brokerage costs by 8.87%. Our results show that the anti-rebate policies are welfare-reducing, not only because they suppress price competition, but also because they encourage excessive entry.

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.

Amemiya, Takeshi *"Advanced Econometrics."* Harvard University Press, 1985, Princeton NJ.

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." *Journal of Business and Economic Statistics*, 2010, *28 (4)* pp. 469-482.

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 Reiss, Peter "Empirical Models of Entry and Market Structure." in *Handbook of Industrial Organization, Volume 3*, eds. M. Armstrong and R. H. Porter. Amsterdam: Elsevier, 1845-1886.

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." *American Economic Review*, 2007, *99*, pp.1878-98.

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.

Lee, Lung-Fei "Unionism and Wage Rates: A Simultaneous Equations Model with Qualitative and Limited Dependent Variables." *International Economic Review*, 1978, *19*. pp. 415-433.

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. 519-532.

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 "The Strategic Timing of Radio Commercials: An Empirical Analysis Using Multiple Equilibria." *The RAND Journal of Economics*, 2009, *40*, pp. 710-742.

Turnbull, Geoffrey "Real Estate Brokers, Nonprice Competition and the Housing Market." *Real Estate Economics*, 1996, *24*, pp. 293-216.

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

Willis, Robert and Sherwin Rosen "Education and Self-Selection." *Journal of Political Economy*, 1979, *87*, pp. 7-36.

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: Individual Level Demographics[a]

	Real Estate Agent		Other Occupation	
Variable	Mean	S.D.	Mean	S.D.
earnings	53,475	67,370	35,194	39,650
age	48.12	13.36	39.62	12.55
education	11.52	1.65	11.15	1.87
married	0.68		0.59	
male	0.46		0.53	
white	0.92		0.82	
lived.same.MSA	0.33		0.31	
full hour	0.35		0.20	
hours worked	42.31	14.15	40.66	11.75
weeks worked	47.98	9.13	47.16	10.41
house owner	0.83		0.67	
house value	287,725	223,416	186,748	159,849
observations		10,541		1,803,320

[a]The table reports the weighted mean of each variable in the 2000 PUMS, using the Census weights. The mean of the variable without standard deviation is the fraction of observations. The sample includes all observations who worked in 2000 and were eligible for obtaining real estate licenses, that is, those who were at least 18 years old and high school graduates. Education values greater than 8 indicate high school graduates or above. Lived.same.MSA is the indicator for whether the respondent has lived in the same MSA for the past 5 years. Full hour is equal to 1 if the respondent worked for more than 50 hours per week.

Table 2: Descriptive Statistics Across Markets[a]

Variable	Mean	S.D.	Min	Max
total house transaction	36,985	34,783	2,410	134,709
mean value of all houses	135,739	40,501	60,780	346,785
mean value of houses sold	147,472	38,793	67,383	339,427
mean value of houses unsold	134,303	40,623	59,841	347,652
# real estate agents	3,549	3,713	44	15,014
% real estate agents	0.0052	0.0022	0.0011	0.0204
population	1,306,038	1,040,811	112,249	4,112,198
total labor force	621,396	515,750	41,847	2,055,769
mean of real estate agent earnings	51,587	12,755	7,250	88,825
mean of other occupation earnings	35,060	3,716	24,981	44,587
total earnings of real estate agents	1.94E+08	2.23E+08	308,500	9.13E+08
past 5 years inflow/population	0.17	0.06	0.04	0.48
past 5 years outflow/population	0.13	0.04	0.02	0.31
mean of house ages	29.78	6.53	15.70	44.78
house density (per km^2)	164.73	95.33	13.40	526.90
population density (per km^2)	400.57	235.15	29.00	1,460.80
% using internet for search	0.32	0.08	0.03	0.58
transactions per agent	12.43	5.59	5.21	68.91
transactions per working hour	0.30	0.15	0.13	2.54
# MSAs				160

[a]The table reports the weighted means of MSA-level variables in the 2000 PUMS, weighted by total labor force. The sample includes 160 free-standing MSAs for which the Internet adoption rates can be computed from the 2000 CPS supplements for Internet and computer use. % using internet for search (or internet.search) is the proportion of respondents in each MSA who reported to use the Internet regularly to search for information.

Table 3: Mean Values of Market Structure Statistics by # Real Estate Agents[a]

	# Real Estate Agents						
Variable	0-199	200-399	400-599	600-999	1000-1499	1500-2499	2500+
sales.per.agent	30.18	21.57	15.26	13.38	10.96	10.75	10.51
sales.per.hour	0.75	0.54	0.36	0.33	0.27	0.26	0.25
total.transaction	4,133	6,567	7,664	10,772	13,834	21,676	67,050
mean.house.value	99,314	105,259	128,816	133,096	136,485	149,583	164,484
mean.revenue	35,801	40,168	47,959	46,034	50,625	50,466	56,953
mean.other.wage	29,797	30,136	32,697	32,598	32,863	35,083	37,605
% real.estate.agent	0.0020	0.0027	0.0037	0.0043	0.0051	0.0052	0.0062
total.labor.force	77,097	125,488	149,309	205,835	276,628	436,117	1,072,956
population	220,194	289,552	315,901	453,470	602,015	973,726	2,212,433
house.density	64.85	87.04	114.60	108.44	125.18	197.77	200.95
population.density	154.42	217.03	273.84	265.47	304.60	479.88	488.75
internet.search	0.27	0.27	0.29	0.30	0.28	0.30	0.36
# MSAs	20	20	21	35	19	22	23

[a]The table reports the weighted means of MSA-level variables in the 2000 PUMS, weighted by total labor force. Mean.house.value is the mean of houses sold in the MSA. Mean.revenue is the mean of earnings for real estate agents. Mean.other.wage is the mean of earnings for other occupation in the MSA.

Table 4: Results for the Revenue Regression[a]

Variable	Estimate	S.E.	p-value
$\widehat{N}$	-0.000013	0.000006	0.021
anti.rebate$\times\widehat{N}$	0.000034	0.000003	0.000
total.transaction (in 10,000)	0.013891	0.006219	0.026
mean.house.value (in $10,000)	0.030886	0.001001	0.000
internet.search	-0.363322	0.040899	0.000
net.inflow	0.499291	0.078605	0.000
vacant.units (in 10,000)	-0.011997	0.014305	0.402
land.area (km^2)	0.000007	0.000001	0.000
male	0.442072	0.010005	0.000
age	0.060716	0.001217	0.000
age.squared	-0.000694	0.000013	0.000
white	0.233930	0.011959	0.000
education	0.070484	0.001747	0.000
live.same.msa.5years	0.120239	0.007400	0.000
married	0.144645	0.005811	0.000
full.hour	0.471852	0.008619	0.000
self.employed	-0.214340	0.027315	0.000
$\widehat{\lambda}_a$	-0.419264	0.038284	0.000
constant	-0.899630	0.150737	0.000

[a]The table reports the converged estimates from the second-step regression in our modified NPL algorithm. Bootstrapped standard errors are reported. The dependent variable is the revenue for real estate agents. The regression is estimated by using observations of real estate agents from the 2000 PUMS data. The number of observations is 10,541. We compute the inverse Mill's ratio $\widehat{\lambda}_a$ by using the estimates from the first-step probit for the reduced form selection equation.

Table 5: Results for the Reservation Wage Regression [a]

Variable	Estimate	S.E.	p-value
internet.search	0.005064	0.001636	0.002
net.inflow	0.060077	0.002495	0.000
mean.house.value (in $10,000)	0.007298	0.000034	0.000
male	0.282832	0.000245	0.000
age	0.047761	0.000052	0.000
age.squared	-0.000473	0.000001	0.000
white	0.067460	0.000299	0.000
education	0.100629	0.000061	0.000
married	0.100149	0.000242	0.000
both.work	0.146228	0.000463	0.000
full.hour	0.329693	0.000299	0.000
self.employed	0.109289	0.000784	0.000
unemployment.rate	-0.349346	0.008075	0.000
$\widehat{\lambda}_b$	0.434181	0.011453	0.000
constant	-2.068058	0.001443	0.000

[a]The table reports the converged estimates from the second-step regression in our modified NPL algorithm. Bootstrapped standard errors are reported. The dependent variable is the earnings for non-agents. The regression is estimated by using observations of non-agents from the 2000 PUMS data. The number of observations is 1,803,320. We compute the inverse Mill's ratio $\widehat{\lambda}_b$ by using the estimates from the first-step probit for the reduced form selection equation.

Table 6: Results for the Entry Model Probit[a]

Variable	Estimate	S.E.	p-value
$\widehat{N}$	-0.000017	0.000001	0.000
anti.rebate$\times\widehat{N}$	-0.000004	0.000002	0.024
mean.house.ages	-0.178951	0.017688	0.000
mean.house.ages.squared	0.036199	0.002527	0.000
mean.unsold.house.value (in $10,000)	0.000262	0.000019	0.000
house.density	-0.000003	0.000000	0.000
land.area (km^2)	0.001355	0.000749	0.071
internet.search	0.425064	0.026695	0.000
vacant.units (in 10,000)	-0.340115	0.015541	0.000
population (in 100,000)	0.193832	0.013223	0.000
male	-0.159640	0.002698	0.000
age	0.001489	0.000312	0.000
age.squared	0.000167	0.000004	0.000
white	0.148959	0.003416	0.000
education	0.032918	0.000604	0.000
married	0.018797	0.001530	0.000
$\widehat{R}_i - \widehat{W}_i$ (in $50,000)	1.000000		

[a]The table reports the converged estimates from the third-step heteroskedastic probit estimation in our modified NPL algorithm. Bootstrapped standard errors are reported. The dependent variable is whether the observation is a real estate agent. The number of observations is 1,813,861. We model the variance of the heteroskedastic probit to be $\{\exp(Z_i\mu)\}^2$, where Z_i includes the predicted revenues and reservation wages. The coefficient on $\widehat{R}_i - \widehat{W}_i$ is fixed to be 1. The probit model includes state fixed effects whose coefficient estimates are suppressed.

Table 7: Summary Statistics of MSA-level Imputed Variables[a]

Variable	Mean	S.D.	P10	P50	P90
mean of $\widehat{R}_i$	42,632	8,060	33,579	42,144	52,657
mean of $\widehat{W}_i$	34,899	3,039	31,279	34,855	38,789
mean of $\widehat{C}_i$	13,951	5,183	8,615	12,559	20,306
TC	46,288,585	64,559,029	1,958,049	18,607,494	175,664,544
wasteful TC	19,180,068	38,381,453	111,131	3,051,687	57,238,120
ΔTC	6,738,666	10,487,218	209,271	2,318,160	23,862,640
$\left(\frac{\Delta TC}{TC}\right)$	0.124	0.020	0.105	0.119	0.157
$\left(\frac{\text{wasteful } \Delta TC}{\Delta TC}\right)$	0.340	0.214	0.099	0.306	0.588
average markup	0.682	0.057	0.613	0.691	0.750

[a]The table reports the summary statistics of the MSA-level imputed variables for 160 MSAs. For the mean of $\widehat{R}_i$ (or $\widehat{W}_i$, $\widehat{C}_i$), we compute the weighted mean of $\widehat{R}_i$ for each MSA, and the table reports the weighted mean of these means. The mean reports the weighted mean, using total labor force as weights. P10 is the 10th percentile, and similarly for P50 and P90. TC is MSA-level total cost, computed by $TC = \sum_{i=1}^{T} \widehat{C}_i \times \widehat{p}_i \times \mathsf{weight}_i$, where $\widehat{p}_i$ is the predicted entry probability, and weight_i is the Census weight for observation i. ΔTC is the change in TC due to a 10% increase in N. Wasteful TC is computed by $\sum_{i=1}^{T}(\beta\widehat{N} + \delta^{\mathrm{A}} A \times \widehat{N}) \times \widehat{p}_i \times \mathsf{weight}_i$, where A is the indicator for anti-rebate policy. Average markup is computed by $\frac{\text{mean of } \widehat{R}_i - \text{mean of } \widehat{C}_i}{\text{mean of } \widehat{R}_i}$.

Table 8: Robustness Results[a]

	Specification I Baseline		Specification II Finite Mixture		Specification III Panel Data	
Variable	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
	A. Revenue Estimation Results					
$\widehat{N}$	-1.34E-05	5.84E-06	-1.38E-05	5.89E-06	-2.97E-05	2.31E-06
anti.rebate$\times\widehat{N}$	3.41E-05	2.59E-06	3.40E-05	2.60E-06		
total.transaction	0.01389	0.00622	0.01415	0.00624	0.02129	0.00236
house.value	0.03089	0.00100	0.03092	0.00100	0.03760	0.00061
	B. Cost Estimation Results					
$\widehat{N}$	-1.65E-05	1.45E-06	-2.01E-05	1.40E-06	-1.95E-05	3.02E-06
anti.rebate$\times\widehat{N}$	-3.54E-06	1.57E-06	-4.66E-06	1.55E-06		
house.ages	-0.17895	0.01769	-0.19034	0.01517	-0.35027	0.02601
house.ages.squared	0.03620	0.00253	0.03751	0.00249	-0.00895	0.00207
unsold.house.value	0.00026	0.00002	0.00159	0.00034	-0.02460	0.00270
house.density	-2.92E-06	3.59E-07	2.69E-04	1.75E-05	8.93E-04	1.10E-04
land.area	0.00135	0.00075	-2.95E-06	3.50E-07	3.76E-05	9.18E-06
internet.search	0.42506	0.02670	0.19612	0.01258		
vacant.units	-0.34012	0.01554	0.42610	0.00628	0.27789	0.02107
population	0.19383	0.01322	-0.33083	0.00488	-0.41768	0.02227
male	-0.15964	0.00270	-0.15973	0.00135	-0.09066	0.00218
age	0.00149	0.00031	0.00149	0.00030	0.02917	0.00050
age.squared	0.00017	0.00000	0.00017	0.00000	-0.00009	0.00000
white	0.14896	0.00342	0.14910	0.00246	0.64418	0.00674
education	0.03292	0.00060	0.03293	0.00035	0.07776	0.00075
married	0.01880	0.00153	0.01882	0.00149	0.04929	0.00207
# observations		1,813,861		1,813,861		2,493,202
	C. MSA-level Imputed Variables					
	Mean	S.D.	Mean	S.D.	Mean	S.D.
mean of $\widehat{C}_i$	13,951	5,183	13,954	5,183	13,468	6,788
TC	46.3 mil	64.6 mil	45.6 mil	63.7 mil	31.7 mil	49.2 mil
ΔTC	6.7 mil	10.5 mil	7.1 mil	11.2 mil	5.1 mil	8.2 mil
$\left(\frac{\Delta TC}{TC}\right)$	0.124	0.020	0.130	0.024	0.138	0.034
wasteful TC	19.2 mil	38.4 mil	23.1 mil	46.0 mil	17.4 mil	32.0 mil
wasteful ΔTC	4.0 mil	8.1 mil	4.8 mil	9.6 mil	3.7 mil	6.7 mil
$\left(\frac{\text{wasteful } \Delta TC}{\Delta TC}\right)$	0.340	0.214	0.394	0.238	0.451	0.294

[a] Specification I reports the results from the main specification using the 2000 PUMS data, which are presented in Tables 4 and 6. The imputed costs in Panel C report the weighted mean of each variable, using the total labor force as weights. AC is the average cost, which is equal to the total cost divided by N, and ΔAC is the change in AC due to a 10% increase in N. Specification II reports the results from the finite mixture estimation using the 2000 PUMS data. Specification III presents the results from panel data estimation using the 1990 and 2000 PUMS data, where we redefine real estate agents in the 2000 data to be partially consistent with the 1990 data. We also redefine MSAs in 2000 to be consistent with the 1990 data. Specifications I and II include state fixed effects, whereas Specification III includes MSA fixed effects.

Table 9: Counterfactual Results for Anti-rebate Policy[a]

Variable	Mean	S.D.	P10	P50	P90
	A. MSAs with Anti-rebate Policy (counterfactual: anti-rebate = 0)				
$\widehat{N}$ initial equilibrium	2,294	1,866	377	1,774	5,576
$\widehat{N}$ counterfactual	2,126	1,692	372	1,686	5,170
% change in $\widehat{N}$	-5.14	3.25	-0.95	-4.55	-10.86
mean of $\widehat{C}_i$ initial eqm.	14,053	3,868	9,653	12,701	19,363
mean of $\widehat{C}_i$ interim	13,734	3,659	9,598	12,575	18,561
mean of $\widehat{C}_i$ counterfactual	13,592	3,523	9,596	12,559	18,155
% change in mean of $\widehat{C}_i$	-2.84	1.84	-0.60	-2.46	-6.24
TC initial equilibrium	30,286,687	31,119,595	1,784,937	15,097,020	81,808,032
TC interim	29,086,961	29,663,159	1,771,212	14,634,407	78,235,248
TC counterfactual	25,978,848	25,836,360	1,754,261	13,780,815	69,261,736
% change in TC	-8.87	5.58	-1.51	-7.62	-17.69
#MSAs					33
	B. MSAs without Anti-rebate Policy (counterfactual: anti-rebate = 1)				
$\widehat{N}$ initial equilibrium	3,691	3,795	518	2,190	8,387
$\widehat{N}$ counterfactual	4,321	4,766	528	2,372	9,781
% change in $\widehat{N}$	9.93	7.87	1.93	7.91	24.43
mean of $\widehat{C}_i$ initial eqm.	13,927	5,444	8,557	12,559	22,970
mean of $\widehat{C}_i$ interim	14,581	5,848	8,722	13,378	24,193
mean of $\widehat{C}_i$ counterfactual	15,211	6,506	8,752	13,507	24,322
% change in mean of $\widehat{C}_i$	7.89	8.72	1.01	4.74	24.26
TC initial equilibrium	49,606,704	70,555,842	2,041,044	18,607,494	175,664,544
TC interim	54,698,804	77,995,752	2,073,242	19,502,946	187,949,360
TC counterfactual	71,403,157	114,569,787	2,128,698	20,946,310	232,676,800
% change in TC	19.64	18.45	3.29	14.18	37.62
#MSAs					127

[a]For the counterfactual results, we compute new equilibrium entry probabilities, and recalculate all the values. The interim costs are computed by changing the dummy for anti-rebate policy, while fixing N. In the table, the mean reports the weighted mean, using total labor force as weights. P10 is the 10th percentile, and similarly for P50 and P90. TC is MSA-level total cost. % change reports the percentage change between the initial equilibrium value and the counterfactual equilibrium value.

Table 10: Comparison of MSAs with or without Anti-rebate

	With Anti-rebate		Without Anti-rebate	
	mean	std.dev.	mean	std.dev.
mean.value.house.sold	134,517	18,956	149,633	42,046
total.transaction	26,487	20,319	38,023	36,260
#real estate agents	2,236	1,753	3,726	3,902
#labor.force	472,732	373,371	634,505	525,856
$\frac{\text{\#real estate agents}}{\text{\#labor.force}}$	0.0046	0.0013	0.0053	0.0023
internet.search	0.3203	0.0762	0.3206	0.0845
housing.unit	427,480	309,123	543,917	426,186
land.area (km^2)	3,290	1,653	4,004	5,388
population.density	289.49	105.96	420.89	250.30
housing.density	121.44	44.18	173.18	101.76