Web Appendix to

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

A Proof for Proposition 1

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

PROOF: Differentiating (2) with respect to N yields

$$\begin{aligned} V'(N) &= -C(q,N) - N\left(\frac{\partial C}{\partial q}\frac{\partial q}{\partial N} + \frac{\partial C}{\partial N}\right) - F \\ &= -C(q,N) + q\frac{\partial C}{\partial q} - N\frac{\partial C}{\partial N} - F \qquad (16) \end{aligned}$$

where (16) follows from the symmetric assumption $q = Q/N$. Adding and subtracting τPq and rearranging the terms yields

$$V'(N) = \pi(N) - \left(\tau P - \frac{\partial C}{\partial q}\right) q - N\frac{\partial C}{\partial N}$$

The socially optimal number of agents, N^*, satisfies $V'(N^*) = 0$.

At the free-entry equilibrium number of agents, we have

$$\begin{aligned} V'(N^e) &= \pi(N^e) - q^e\left(\tau P - \frac{\partial C}{\partial q}\right) - N^e\frac{\partial C}{\partial N} \\ &= \pi(N^e) - w - q^e\left(\frac{C(q,N^e)+F}{q} - \frac{\partial C}{\partial q}\right) - N^e\frac{\partial C}{\partial N} \\ &= \pi(N^e) - w \underbrace{-q\left(AC(q^e,N^e) - MC(q,N^e)\right)}_{\text{Loss of Economies of Scale}} \underbrace{-N^e\frac{\partial C}{\partial N}}_{\text{Wasteful Competition}} \qquad (17) \end{aligned}$$

where the second equality follows from (1), and the last equality follows from definitions that $AC \equiv \frac{C(q,N)+F}{q}$ and $MC \equiv \frac{\partial C}{\partial q}$. Note that the first term, $\pi(N^e) - w$, is equal to zero under free entry. The second term is negative under Assumption 3, as declining average costs mean that marginal cost is necessarily below average cost. The third term is negative under Assumption 2. Thus, $V'(N^e) < V'(N^*) = 0$.

We complete the proof by establishing the global concavity of $V(N)$. From (16), we have

$$V''(N) = -2\frac{\partial C}{\partial N} - N\frac{\partial^2 C}{\partial N^2} + 2q\frac{\partial^2 C}{\partial q \partial N} - \frac{q^2}{N}\frac{\partial^2 C}{\partial q^2},$$

which is negative under Assumptions 1-2. It follows that $N^e > N^*$. ∎

B Simple Model with Product Differentiation

In this section, we explicitly take into account product differentiation and examine how this would affect the direction of the entry bias identified in Proposition 1. Following Mankiw and Whinston (1986), we define the total consumer surplus in the following way:

$$G(Nf(q), P, \tau)$$

Note that total consumer surplus now depends not only the aggregate level of output, but also on the number of real estate agents. For given P and τ, the function $G(\cdot)$ satisfies the following assumptions.

ASSUMPTION 4. $f(0) = 0$, $f'(q) > 0$, and $f''(q) \leq 0$ for all $q \geq 0$.

ASSUMPTION 5. $G'(z) > 0$ and $G''(z) \leq 0$ for all $z \equiv Nf(q) \geq 0$.

Assumptions 4-5 imply that households prefer increasing variety of real estate brokerage service and that the types of brokerage service provided by different agents are substitutes.

In equilibrium, consumer maximization implies that the commission that a household pays to a real estate agent is exactly equal to its marginal utility, $G'(Nf(q))f'(q)$. Thus, the equilibrium level profit per agent is

$$\pi \equiv G'(Nf(q))f'(q)q - C(q, N) - F$$

The objective of the social planner is to select N^* to solve:

$$\max_N W(N) \equiv G(Nf(q)) - NC(q, N) - NF$$

To examine whether entry is socially efficient, we begin by differentiating $W(N)$ as follows

$$W'(N) = G'\left(f(q) + Nf'\frac{\partial q}{\partial N}\right) - N\left(\frac{\partial C}{\partial q}\frac{\partial q}{\partial N} + \frac{\partial C}{\partial N}\right) - C(q, N)$$

Adding and subtracting $G'f'q$ and rearranging terms give

$$W'(N) = \pi + N\frac{\partial q}{\partial N}\left(G'f' - \frac{\partial C}{\partial q}\right) - N\frac{\partial C}{\partial N} + G'\left(f - f'q\right)$$

The socially optimal number of agents, N^*, satisfies $W'(N^*) = 0$.

The market equilibrium requires that $G'f' = \tau P$ (consumer optimality) and that $\tau P = AC(q, N^e)$ (free entry condition). Thus the market equilibrium number of agents, N^e, satisfies

$$W'(N^e) = \pi(q, N^e) + \underbrace{N^e\frac{\partial q}{\partial N}\left(AC(q, N^e) - MC(q, N^e)\right)}_{\text{Loss of Economies of Scale}} \underbrace{-N^e\frac{\partial C}{\partial N}}_{\text{Wasteful Competition}} \underbrace{+G'\left(f - f'q\right)}_{\text{Product Differentiation}} \quad (18)$$

The equation above is a generalization of (17). The first three terms are identical to those in (17), and the sum of these components is negative under Assumptions 1-3, as shown earlier. The fourth term represents the effect of product differentiation, and Assumptions 4-5 imply that it is positive. More specifically, if households have a preference for variety of brokerage service, a marginal real estate agent increases consumer surplus by providing more variety, but does not capture this gain in profits. If this effect dominates the loss of economies of scale and wasteful competition, then entry will be socially insufficient. Nevertheless, it is plausible to conclude that entry is still socially excessive, to the extent that the product differentiation effect is limited to a certain degree.

C Evidence on Price Competition and Product Differentiation

The main empirical findings in this article have established evidence for cost inefficiency under free entry in the real estate brokerage market. As discussed in Section 2, entry could potentially benefit consumers either through price competition or through product differentiation. Given the lack of information on the demand side for the real estate brokerage service in the PUMS, we cannot recover the magnitude of these potential benefits. In this appendix, using separate datasets, we show that consumers' benefits from possible price competition and product differentiation are quite limited, thereby providing evidence for the two simplifying assumptions described at the end of Section 2.

C.1 Price Competition

It is well known that real estate brokerage is characterized by relatively fixed commission rates (e.g., FTC report, 1983; Hsieh and Moretti, 2003). One observation from the American Bankers Association is that "one would expect to see variations in brokerage commissions across geographical regions as the supply and demand varies dramatically across the United States, but little if any variation exists." In light of the motivation of this paper, it is reasonable to ask whether entry has helped bring down commission rates in recent years.

Tables 8 and 9 indirectly show the relationship between entry and commission rates. Table 8 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 (CEX), because the 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 9 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 number of real estate agents is positively correlated with housing prices, but housing prices do not seem to be correlated with commission rates. Therefore, entry is less likely to put downward pressure on commission rates.

To further explore the relationship between entry and commission rates, Table 10 resents 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 region-specific fixed effects that might affect both entry and commission rates. The percentage changes in the average commission rates were respectively 1.63%, 0%, and -1.15% in FW, MA, and NE, which are the regions with high growth in the number of real estate agents. Given that the average commission rate was around 5.1% during that period, the magnitude of the changes in commission rates was negligible. Though lack of more detailed information prevents us from further investigation, the descriptive evidence here suggests that entry does not have noticeable effects on commission rates.

C.2 Product Differentiation

Just like price competition, product differentiation in the real estate brokerage industry is impeded by implicit agreement or state legislation. Many states in the U.S. have imposed so-called minimum-service requirements on real estate agents. As the name suggests, these laws and regulations enumerate specific tasks that a real estate agent must perform for a client. According to the joint statement from the FTC and DOJ, minimum-service requirements limit the degree of differentiation in the real estate brokerage service and reduce the options available to consumers by prohibiting agents from offering a la carte services.

While agents are required to provide the standardized package of service to their clients, they may differentiate themselves by trying to improve the quality of their service. Even in this case, entry could still be socially excessive (e.g., Berry and Waldfogel, 1999). The reason is that such improvements result in "stolen" business. In equilibrium, agents do not take into account the fact that the increase in own-quality reduces the demand for other agents, and therefore over-invest in quality improvement.

In this section, we empirically examine whether entry benefits consumers by improving the quality of the brokerage service. Because the PUMS does not provide valid information on the quality of home matching process, we use the 2001 Home Buyer and Seller Survey conducted by the NAR. 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 merge this sample with the PUMS 2000 to obtain the MSA-level information on the number of real estate agents and brokers.

To measure the quality of real estate brokerage service, we use the household's response to the following question in the NAR survey: "Would you use the same agent again for the future home transaction?" We define an indicator variable REA satisfaction to be equal to 1 if the answer is yes. We then estimate the probit model in which the dependent variable is REA satisfaction and the independent variables include the number of real estate agents 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 and home buyer/seller characteristics. As shown in Table 11 , the key coefficients on the number of real estate agents in both the buyer and the seller samples are insignificant. This suggests that increasing number of real estate agents has a negligible impact on consumers' evaluation of the quality of the brokerage service. Therefore, entry by traditional real estate agents does not lead to obvious improvement in home matching quality.

D Implementation of NPL with Finite Mixture

To implement a finite mixture version of the NPL algorithm, we assume that ξ has a discrete distribution with finite points of support. We then consider the following log-likelihood function at the h-th iteration of the NPL algorithm for given $\widehat{R}_i$ and $\widehat{W}_i$

$$\mathcal{L}^h(\theta,\varphi) = \sum_{m=1}^{M} \ln\left(\sum_{k=1}^{K} \varphi_k \times f_k(X_m;\theta,\widehat{N}_{m,k}^{h-1},\omega_k)\right), \tag{19}$$

where $\theta = \{\alpha, \beta, \delta^{\text{R}}, \delta^{\text{W}}, \delta^{\text{C}}\}$; X_m is a vector of market-level observables; $\widehat{N}_{m,k}^{h-1}$ is the number of real estate agents in market m for type k, estimated from the $(h-1)$-th NPL iteration; ξ takes values from $\{\omega_1, \ldots, \omega_K\}$, and ξ is i.i.d across markets with probability mass function $\varphi_k = \Pr(\xi = \omega_k)$. Because we allow for market-specific unobserved heterogeneity, we consider the market-level density function for type k given by

$$f_k(X; \theta, \widehat{N}_k^{h-1}, \omega_k) = \prod_{i=1}^{I} \Pr(d_i = 1 | X_i; \theta, \widehat{N}_k^{h-1}, \omega_k)^{d_i} (1 - \Pr(d_i = 1 | X_i; \theta, \widehat{N}_k^{h-1}, \omega_k))^{1-d_i}.$$

where we suppress market subscripts m.

The maximum likelihood estimation of the finite mixture model is usually difficult to estimate, because the log-likelihood function as in (19) does not allow for any additive separability. However, Arcidiacono and Jones (2003) develop an approach to reintroduce additive separability in the log-likelihood function. The key insight of their approach is that maximizing the log of the unconditionally-type-avearged likelihood ($\ln[\sum_k \varphi_k f_k(X_m; \theta)]$) is equivalent to maximizing the conditionally-type-averaged log-likelihood ($\sum_k \Pr(k|X_m) \ln[f_k(X_m)]$), where $\Pr(k|X_m)$ is the probability that market m is of type k, conditional on having observed X_m. We thus incorporate the Expectation-Maximization (EM) algorithm proposed by Arcidiacono and Jones (2003) as follows. At the beginning of iteration l, we use $\widehat{\theta}^{l-1}$ and $\widehat{\varphi}^{l-1}$ to update $\Pr(k|X_m; \theta, \varphi)$, where we suppress $\widehat{N}_{m,k}^{h-1}$ to simplify the notation, and from Bayes' theorem, we have $\Pr(k|X_m; \theta, \varphi) = \frac{\varphi_k f_k(X_m; \theta, \omega_k)}{\sum_{k=1}^{K} \varphi_k f_k(X_m; \theta, \omega_k)}$. This completes the "E" step. In the "M" step, we obtain the ML estimates $\widehat{\varphi}^l$ by $\widehat{\varphi}_k^l = \frac{1}{M} \sum_{m=1}^{M} \Pr(k|X_m; \widehat{\theta}^{l-1}, \widehat{\varphi}^{l-1})$, and the ML estimates $\widehat{\theta}^l$ by

$$\widehat{\theta}^l = \arg\max_{\theta} \sum_{m=1}^{M} \sum_{k=1}^{K} \Pr(k|X_m; \widehat{\theta}^{l-1}, \widehat{\varphi}^{l-1}) \times \ln[f_k(X_m; \theta, \omega_k)].$$

We repeat the "E" step and the "M" step until $\widehat{\theta}$ and $\widehat{\varphi}$ converge.

E Counterfactual Analysis: Internet Diffusion

The Internet has become increasingly important in the real estate brokerage industry. 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) show that home sellers could save an average of \$5,000 by using online flat-fee agents, instead of hiring traditional full-commission agents. In addition, Hendel, Nevo and Ortalo-Magne (2007) show 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. There are two opposing effects of the Internet. First, higher Internet adoption rates increase the popularity of online discount brokers and online versions of FSBOs, both of which could steer the buyers and sellers away from traditional agents. This effect discourages potential agents from entering the market. Second, as argued by the National Association of Realtors (NAR), the access to the Internet helps more traditional real estate agents to reach their potential clients, either by emails or by websites. This effect encourages potential agents to enter the market. Though both effects are plausible, which effect would dominate is an empirical question. Therefore, we use our model estimates to investigate this question and further quantify the effect of the Internet on cost savings.

According to the Current Population Survey, the Internet adoption rate 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, steadily increased from 71% to 80% between 2003 and 2006. In this respect, we perform a counterfactual experiment in which the Internet adoption rate is increased by 10%, and all Internet users regularly use the Internet to search for information. The results are presented in Table 12.

We find that both effects of the Internet are present. First, increasing the online search rate reduces the average agent revenues by 9.27%. As discussed above, this is likely due to the competition from the online discount brokers, which accompanies the increase in the online search intensity. Second, increasing the online search rate also reduces the average agent costs by 15.75%. This is consistent with the conjecture that the access to the Internet makes it easier for agents to reach their potential clients and to match potential buyers with sellers. However, we find that the second effect dominates the first effect, in that the resulting equilibrium number of real estate agents is increased by 2.22% in a typical MSA. This increase in entry could lead to wasteful competition, but we find that total brokerage costs decrease on average by 17.07%. Despite the potentially negative impact of the Internet on traditional real estate agents, we find that an increase in online search intensity would increase the number of agents in equilibrium and could significantly reduce total brokerage costs, suggesting that the beneficial effect of the Internet is likely to be dominant in this industry.

Table 8: Correlation Between Entry and Housing Price[a]

	changes in log(average price of houses)
changes in $\log(N)$	0.4564
changes in $\log(N/M)$	0.5042

[a]The table reports the correlation coefficients between the changes in log of the number of real estate agents and the changes in the average housing prices between 1990 and 2000 in the PUMS. In the table, N denotes the number of real estate agents, and M denotes total labor force in a market.

Table 9: 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 10: Agent Growth and Commission Rate Growth Across Regions 2002-2003[b]

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

[b]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 11: Effects of Competition on Consumer 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 and home buyer/seller characteristics.

Table 12: Counterfactual Results for Changes in Internet Search Intensity[a]

Variable	Mean	S.D.	P10	P50	P90
	Increase Internet adoption rate by 10% and full search intensity				
$\widehat{N}$ initial equilibrium	3,427	3,554	486	1,969	8,387
$\widehat{N}$ counterfactual	3,503	3,644	491	2,021	8,654
% change in $\widehat{N}$	2.22	1.02	1.07	2.24	3.53
mean of $\widehat{R}_i$ initial eqm.	42,632	8,060	33,579	42,144	52,657
mean of $\widehat{R}_i$ counterfactual	38,801	8,015	29,212	37,907	48,528
% change in mean of $\widehat{R}_i$	-9.27	3.07	-6.15	-8.90	-13.26
mean of $\widehat{C}_i$ initial eqm.	13,951	5,183	8,615	12,559	20,306
mean of $\widehat{C}_i$ counterfactual	11,993	5,188	6,857	10,755	17,802
% change in mean of $\widehat{C}_i$	-15.75	7.28	-8.34	-15.07	-25.89
TC initial equilibrium	46,288,585	64,559,029	1,958,049	18,607,494	175,664,544
TC counterfactual	41,427,847	59,505,869	1,611,195	14,994,424	163,628,656
% change in TC	-17.07	8.82	-7.59	-16.32	-28.50
#MSAs					160

[a]The counterfactual results are computed using new equilibrium entry probabilities. 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