Supplemental Online Appendix to Han and Hong, "Understanding In-House Transactions in the Real Estate Brokerage Industry"

Appendix A: An Agent-Intermediated Search Model

Our motivating theoretical framework follows closely Hagiu and Jullien (2011), who provide an inspiring economics analysis of search diversion in an online shopping setting. We apply their search diversion theory to the real estate brokerage industry and show that agents can misguide homebuyers by introducing noise in the home search process. Unlike their model that assumes an intermediary receives a fixed amount of revenues from each store visit by buyers regardless of actual sales, we make a different assumption to reflect the key compensation feature of the real estate brokerage industry. That is, agents receive a fixed percentage of *realized* sales revenues and this percentage is *larger* when a transaction occurs within the same brokerage firm. As shown later, such compensation feature is the driving source of agents' strategic promotion.

To simplify the analysis of the search process in the housing market, let us consider a setup where there are two types of buyers (buyer 1 and buyer 2), two types of houses (house 1 and house 2), and one cooperating agent.

Buyers: Buyers differ along two dimensions: preferences for houses and search costs. Along the first dimension, there are two types of buyers: type 1 buyers make up a fraction α of the population and derive net utilities u^H from visiting house 1 and u^L from visiting house 2; type 2 buyers make up a fraction $1-\alpha$ of the population and derive net utilities u^H from visiting house 2 and u^L from visiting house 1. We assume that $u^H > u^L$, which implies that *ex ante* type 1 buyers prefer house 1 over house 2, and that type 2 buyers prefer house 2 over house 1 in the sense that will be described below. Along the second dimension, buyers are differentiated by the search cost c they incur each time they visit a house. We use $F(c)$ to denote the cumulative distribution of c. They can only visit at most two houses sequentially.

More specifically, take buyer 1 as an example. Her valuation of a specific house h, v_h^1, is unknown prior to the visit but is learnt upon inspection of the house, so that the expected utility prior to visiting the house is $u_h^1 = \int_{p_h}^{k} (v_h^1 - p_h) dG(v_h^1)$, where $G(v)$ denote the cumulative distribution of v,[1] $k = H$ if $h = 1$ and $k = L$ if $h = 2$. Assuming that $0 < L < H$, it follows that $u_1^1 > u_2^1$, and that $u^H = u_1^1$ and $u^L = u_2^1$. In other words, *ex ante* house 1 is a better match for buyer 1 than house 2. Note that u_h^1 should be interpreted as encompassing the utility of just "looking around" the house plus the expected utility of actually buying the house, net of the price paid. Upon visiting a house, a buyer observes the realized value of being matched with a specific house, v_h^1, and then decides to whether to buy the house.

Houses: Houses also differ along two dimensions: matching quality and the listing brokerage firm. Along the first dimension, as described above, type 1 house stands for houses that *ex ante* match the buyer 1's preference best, whereas house 2 stands for houses that *ex ante* match the buyer 2's preference best. Along the second dimension, house 1 is listed by a firm that is different from the cooperating brokerage firm, whereas house 2 is listed by an agent affiliated with the same brokerage firm.

For simplicity, we assume that prices of houses are exogenously given at p_1 and p_2. This is because house prices are typically determined by general market conditions, which is much broader than the choice of intermediaries. In addition, the listing price of a house is publicly advertised before the cooperating agents and their buyers are engaged in the search process. To the extent that the sales and listing prices are highly positively correlated, the exogeneity assumption is justified.

Cooperating Agent: The cooperating agent observes each buyer's type (1 or 2) but not her search cost c. As the agent is assumed to have superior information about houses available in the market, he immediately knows which house *ex ante* fits the buyer's preference best. Following Hagiu

[1] For any given buyer, v_1^1 and v_2^1 are assumed to be independently distributed.

and Jullien (2011), we denote by q_1 the probability that the agent takes buyer 1 to house 1 for the *first* visit. If the cooperating agent always optimize the match process between buyers and houses, then we should expect $q_1 = 1$. In contrast, we say that the cooperating agent "strategically" promotes his own firm's listings (i.e., house 2) whenever $q_1 < 1$.

The cooperating agent receives a fixed percentage of actual sales price as commission income when a transaction is completed. This income is then split with the agent's affiliated brokerage firm. In net, the cooperating agent obtains a fixed share of transaction price, τ_1 (or τ_2), from the sale of house 1 (or house 2). If an agent receives a bonus for promoting internal listings, then all else equal, $\tau_2 > \tau_1$. As a result, the cooperating agent for buyer 1 may sometimes find it profitable to recommend house 2 which generates the highest revenue, rather than house 1 which matches buyer 1's preference best. The incidence of $q_1 < 1$ captures precisely the inefficiency resulted from the commission structure described above.

Timing: The timing of decisions is as follows: (i) the agent publicly announces q_1; (ii) buyers observe q_1; (iii) buyers decide whether to follow agent's guidance, engage in the search process, and make their purchase decisions after visiting the house(s).

Solving the Model

Without loss of generality, let us focus our analysis on type 1 buyers. First, consider a type 1 buyer with high search costs, i.e., $c > u^H(p_1)$. In this case, the buyer would not visit any of the two houses, and as a result, the agent receives zero commission income.

Next, consider a type 1 buyer with low search costs, i.e., $c \leq u^L(p_2)$. She will visit both houses irrespective of where the agent directs her for her first visit. Upon visiting both houses, the buyer compares two houses and purchases the one that gives her the largest net realized utility, $\max\{v_1 - p_1, v_2 - p_2\}$.

Accordingly, the probability of buyer 1 purchasing house 2, ρ_2^1, is given by:

$$
\begin{aligned}
\rho_2^1 &\equiv \Pr(v_2^1 - p_2 > v_1^1 - p_1) \\
&= \int_{p_1}^{H} \int_{v_1^1 - p_1 + p_2}^{L} dG_L(v_2^1) dG_H(v_1^1) \\
&= \int_{p_1}^{H} \left(1 - G_L(v_1^1 - p_1 + p_2)\right) dG_H(v_1^1) \qquad (1)
\end{aligned}
$$

Thus the cooperating agent receives commission income $\tau_2 p_2$ with probability ρ_2^1 and $\tau_1 p_1$ with probability $1 - \rho_2^1$.

Finally, consider a type 1 buyer with intermediate search costs, i.e., $u^L(p_2) \leq c \leq u^H(p_1)$. In this case, if the buyer is first sent to house 1 (which happens with probability q_1), she would make a purchase and stop visiting another house if the net realized value from buying house 1, $(v_1 - p_1)$, is greater than the expected utility of continuing visiting house 2, $\max\{u^L(p_2) - c, 0\}$. Because $u^L(p_2) \leq c$, $\max\{u^L(p_2) - c, 0\} = 0$, so that she will not visit house 2 with probability 1. If she is first sent to house 2 (which happens with probability $1 - q_1$), she would stop searching if and only if the net realized utility, $(v_2 - p_2)$, is greater than the expected utility of continuing visiting house 1, that is, $\max\{u^H(p_1) - c, 0\} = u^H(p_1) - c$. In the event when buyer 1 visits both houses, she will purchase house 1 with probability $1 - \rho_2^1$ and house 2 with probability ρ_2^1.

Knowing the probability q_1, a type 1 buyer follows the agent's guidance if her search cost is above $u^L(p_2)$ and below some critical value u_1, where $u_1 = c$ is implicitly defined by

$$
q_1 u^H(p_1) + (1 - q_1) \int \max\left(v_2 - p_2, u^H(p_1) - c\right) g_L(v_2) dv_2 - c = 0
$$

Note that when $q_1 = 1$, we have $u_1 = u^H(p_1)$ and $\frac{du_1}{dq_1} = u^H(p_1) - u^L(p_2)$.

Turning to the agent's side, the revenue he derives from type 1 buyers is then:

$$
\begin{aligned}
\Pi_1 = {} & \left(\tau_1 p_1 (1 - \rho_2^1) + \tau_2 p_2 \rho_2^1\right) F(u_L) + q_1 \tau_1 p_1 \left(F(u_1) - F(u_L)\right) \\
& + (1 - q_1) \left[\left(\tau_1 p_1 (1 - \rho_2^1) + \tau_2 p_2 \rho_2^1\right) \int_{u^L}^{u_1} G_L(p_2 + u^H - c) f(c) dc \right. \\
& \left. + \tau_2 p_2 \int_{u^L}^{u_1} (1 - G_L(p_2 + u^H - c)) f(c) dc \right] \qquad (2)
\end{aligned}
$$

The first term represents the revenue that the agent receives from type 1 with low search costs, i.e., with $c \leq u^L(p_2)$. The second term represents the revenue that the agent receives from type 1 buyers who have intermediate search costs, i.e., with $u^L(p_2) \leq c \leq u_1$, and have been efficiently matched to house 1 on their first visit. The third term represents the revenue that the agent receives from type 1 buyers who have intermediate search costs but have been strategically directed to house 2 first. Note that the first integrant term is the probability that the buyer decides to continue searching conditional on having visited house 2 in the first round of search. In this case, the agent receives $\tau_1 p_1$ with probability $1 - \rho_2^1$ and $\tau_2 p_2$ with probability ρ_2^1.

Maximizing (2) over q_1 yields the following proposition, which contains our baseline results:

Proposition 1 *The cooperating agent "strategically" promotes in-house transactions (i.e., $q_1 < 1$) if and only if*

$$\frac{\tau_2 p_2}{\tau_1 p_1} > \frac{F(u^H) - F(u^L) - (1-\rho_2^1)\int_{u^L}^{u_1} G_L(p_2 + u^H - c)f(c)dc + f(u^H)(u^H - u^L)}{F(u^H) - F(u^L) - (1-\rho_2^1)\int_{u^L}^{u_1} G_L(p_2 + u^H - c)f(c)dc} \tag{3}$$

Proof: The cooperating agent maximizes (2) over q_1. Using the fact $u_1(q_1 = 1) = u^H(p_1)$ and $\frac{du_1}{dq_1}(q_1 = 1) = u^H(p_1) - u^L(p_2)$, it is straightforward to show that $\frac{\partial \Pi_1}{\partial q_1}(q_1 = 1) < 0$ if and only if (3) holds. ■

Strategic In-House Transactions

Condition (3) is central to understanding of agents' incentives to strategically promote in-house transactions. In particular, at $q_1 = 1$, all type 1 buyers with intermediate search costs will be first directed to houses that match their preference best, leading to an efficient matching outcome. By laying out conditions under which the cooperating agent lowers q_1 below 1, condition (3) immediately delivers several predictions of strategic in-house transactions that can be taken to the data.

Prediction 1: The commission structure matters. It is clear from condition (3) that the optimal amount of strategic promotion increases with the ratio $\frac{\tau_2 p_2}{\tau_1 p_1}$. If the prices of two houses are not too different from each other (which is not too unreasonable given that buyers usually specify a price range for houses they search for), the larger is the ratio $\frac{\tau_2}{\tau_1}$, the more likely condition (3) will hold, and the stronger is the agent's incentive to promote her own firm's listings. In the brokerage industry, agents need to split commission fees with their affiliated brokerage offices, in return for the brand value and for the supporting services that brokerage offices provide. In practice, full commission brokerage firms, such as ReMax, let the agents keep all commission fees but require a fixed amount of upfront fees each month. More split fees per transactions firms, such as Royal LePage, split commission fees with their agents on the per-transaction basis. Naturally, the revenues in the latter type of brokerage firms strictly increase with the number of either end of transactions. Therefore, these brokerage firms are more likely to reward their agents for selling internal listings, making $\tau_2 > \tau_1$.[2] Thus, we expect that the per-transaction split commission structure is associated with a stronger presence of strategic in-house transactions.

In addition to commission structure, commission rate also matters. Note that the commission rate for a cooperating agent is typically predetermined when the listing is posted on the MLS. Whereas the commission rate is usually set at 2.5%, some listing agents would offer a higher or lower rate to cooperating agents. Intuitively, by rewarding cooperating agents a greater proportion of the commission, an external listing agent can effectively increase τ_1 in condition (3), and this helps offset the promotion bonus that the cooperating agent receives from her own firm for promoting internal listings. Conversely, when the commission rate offered by a listing agent is low, the cooperating agent is more likely to respond to the financial incentives offered by the brokerage firm for promoting in-house transactions. The strategy of using substandard commission rates to artificially increase the frequency of dual-agency transactions is discussed in Yavas, et al. (forthcoming) and also evidenced by a recent industry report.[3] Thus, we expect that lower commission rates offered by listing agents are associated

[2]For example, Gardiner, Heisler, Kallberg and Liu (2007) find that many brokerage firms give a financial reward to agents who successfully match internal clients with internal listings. Similarly, a popular industry practice book, *Buying a Home: The Missing Manual*, reports that some agencies pay agents a bonus for selling in-house listings because the agency makes more money in such transactions. In addition, a recent report by the *Consumer Advocates in American Real Estate* explicitly points out that agents who avoid in-house transactions may bear with some financial consequences, such as a less favorable commission split with the brokerage firm.

[3]For example, a recent report by the *Consumer Advocates in American Real Estate* states that "offering less than the going rate in your area will decrease the financial attractiveness of your home [to cooperating agents] and increases the likelihood that your broker will collect a double commission" (see an article titled "Dual Agency Schemes" in

with a stronger presence of strategic in-house transactions.

Prediction 2: The extent to which cooperating agents can promote in-house transactions depends on the difference in the matching quality that a given buyer obtains from internal and external listings. As we can see from condition (3), the bigger is $(u^H - u^L)$ and/or ρ_2^1, the smaller is the likelihood of strategic promotion ($q_1 < 1$). Intuitively, if the best house that a buyer can find from external listings is far better than the one she can find from internal listings, either *ex ante* (reflected by $u^H - u^L$) or *ex post* (reflected by ρ_2^1), then it becomes difficult for her agent to promote internal listings. Empirically, we do not observe matching quality. However, matching in housing markets is typically characterized by increasing returns to scale (Ngai and Tereyro 2014; Genesove and Han 2012b). When a brokerage firm has a larger number of listings which a buyer can choose among, there should be less dispersion in the buyer's valuation of her most-preferred house from the market-wide pool and from the internal listings. As a result, the brokerage firm will find it easier to promote its own listings. Although the promoted listings may not match the buyer's preference best, the resulting efficiency loss should be smaller as these listings are closer to the buyer's preference.

Prediction 3: The brokerage firm's ability to strategically promote in-house transactions also depends on whether buyers are aware of agents' incentives to strategically promote. So far, the model has assumed that buyers faced with a known probability of q_1. If buyers are not aware of agents' strategic incentives, this would remove the dependence of $\frac{du_1}{dq_1}$ in deriving the derivative of Π_1 with respect to q_1. As a result, the right-hand-side of condition (3) is reduced to 1. In this case, the agent's incentive to promote in-house transactions would purely depend on the financial reward ($\frac{\tau_2}{\tau_1}$) and search cost. The quality difference would no longer matter, because buyers believe that agents always match them to their first best house and hence would not be sensitive to the difference between the first and second best houses ($u^H - u^L$). As discussed later, our sample covers a natural experimental opportunity permitted by a legislation that required real estate agents engaged in in-house transactions to disclose

http://www.caare.org/ForBuyers, accessed August 1, 2014).

the possibility of strategic promotion to both buyers and sellers. This provides an opportunity for us to empirically test this prediction.

Efficient In-House Transactions

In-house transactions could also occur for efficiency rather than incentive reasons. We define an in-house transaction as "efficient" if the buyer's net utility from internal listings is larger than her net utility from external listings, either *ex ante* or *ex post*. In our model, the probability of efficient in-house transactions is given by:

$$P = \left(\alpha\rho_2^1 + (1-\alpha)(1-\rho_1^2)\right) F(u_L) + (1-\alpha)(F(u_2) - F(u_L)) \tag{4}$$

The first term in (4) refers to the probability of in-house transactions by type 1 and type 2 buyers with low search costs. With probability ρ_2^1, a type 1 buyer purchases house 2 because house 2 delivers larger net realized utility than house 1. Similarly, with probability $1-\rho_1^2$, a type 2 buyer purchases house 2. In both cases, transactions occur within the same brokerage firm, and these in-house transactions represent the outcome of buyers' own choices rather than agents' promotional efforts. In particular, the low search cost removes the reliance of buyers on agents in looking for ideal homes, resulting in an efficient match between buyers and houses, regardless of whether the transaction occurs within the same brokerage firm or not.

The second term in (4) refers to the probability of in-house transactions by type 2 buyers with intermediate search costs. It is straightforward to show that with probability $q_2 = 1$, all type 2 buyers will be first directed to house 2.[4] Moreover, these buyers would end up purchasing house 2, because the expected utility of visiting house 1 is less than the search cost. In this case, the agents' incentive to promote their own listings is consistent with the buyers' interest, because these listings match the buyers' *ex ante* preference best. This type of in-house transactions, although promoted by cooperating

[4] To see this, note that for type 2 buyers, a similar condition as condition (3) can be obtained by changing only the left-hand side in (3) to $\frac{\tau_1 p_1}{\tau_2 p_2}$. Assuming that houses 1 and 2 are in the same price range, with the in-house promotion bonus, the left-hand side is less than 1, whereas the right-hand side is greater than 1. This implies that the threshold condition will never be met, and hence $q_2 = 1$.

agents, represents an efficient matching outcome.

Thus, the model predicts two types of efficient in-house transactions. Under the first type, a buyer receives the largest *ex post* utility from an internal listing through her own comparison of all available listings. Under the second type, a buyer is directed by an agent to an internal listing that matches her *ex ante* preference best. Because the first type of in-house transactions are not driven by agents, we focus our discussion on the second type of in-house transactions, which is driven by mutual interests of buyers and their agents. Note that our model takes the buyer's choice of the cooperating brokerage as given (reflected by an exogenous α), hence we are unable to explicitly model the sources of efficient in-house transactions. In practice, buyers' *ex ante* preference for an internal listing may agree with the agent's self-promotion interest for various reasons. For example, an in-house transaction may lower transaction costs and improve the efficiency in the bargaining and closing stage, making buyers more likely to favor transactions within the same brokerage house. Alternatively, a buyer may choose a cooperating agent simply because the agent's affiliated firm specializes in listing houses that fit the buyer's interest best. In both cases, in-house transactions are caused by a mixture of transaction efficiencies and information advantages.

Appendix B: Details on Three-Step Estimation Approach

To estimate our model, we follow and modify the estimation approach used by Bajari and Kahn (2005) which involves three steps. The first step estimates the hedonic price function using nonparametric methods, and recovers buyer-specific utility parameters β_i. In the second step, we estimate $V^1(\beta_i)$ and $V^0(\beta_i)$. In the third step, we estimate the distribution of γ_i and c_i. In what follows, we describe our approach in detail. We also provide discussion on identification where applicable.

Step 1: estimating the price function and recovering ξ_j and β_i

In the first step, we recover the slope of the price function in a local neighborhood of the characteristics of house j^* chosen by buyer i. To this end, we use a nonparametric estimation of the hedonic price function, and in particular, we use the local linear regression.[5]

Following Bajari and Kahn (2005), we consider a linear approximation of $\mathbf{p}_t(X_j, \xi_j)$ in a local neighborhood of house j^*'s observed characteristics. Specifically, we consider

$$\log\left(\mathbf{p}_t(X_j, \xi_j)\right) = \alpha_{j,0} + \sum_{k=1}^{m} \alpha_{j,k} x_{j,k} + \eta_t + \xi_j, \tag{5}$$

where $\alpha_j = (\alpha_{j,0}, \ldots, \alpha_{j,m})$ is a vector of the hedonic coefficients, η_t captures market fixed effects, and we use a logarithm of the price function instead of its level in order to improve the fitting of the price function. In our estimation, we first regress $\log(p_j)$ on X_j, MLS district fixed effects, and year×month fixed effects. We then use the demeaned prices, $\widetilde{\log(p_j)} \equiv \log(p_j) - \hat{\eta}_t$, and estimate α_{j^*} for each value of $j = 1, \ldots, J_t$ by using local fitting methods which solve

$$\min_{\alpha} \sum_{j=1}^{J_t} \left\{ \widetilde{\log(p_j)} - \alpha_0 - \sum_{k=1}^{m} \alpha_k x_{j,k} \right\}^2 K_B(X_j - X_{j^*}), \tag{6}$$

where $K_B(\mathbf{v})$ is the kernel function. Given $K_B(X_j - X_{j*})$, α_{j^*} can be estimated by weighted least squares for each j^*. As for $K_B(\mathbf{v})$, we use the product of univariate Gaussian kernel, following Bajari and Kahn (2005) who used $K_B(\mathbf{v}) = \prod_{k=1}^{m} \frac{1}{b} N(\frac{1}{b} \frac{v_k}{\hat{\sigma}_k})$, where b is a scalar bandwidth, $N(\cdot)$ is the univariate Gaussian kernel, and $\hat{\sigma}_k$ is the sample standard deviation of v_k.

Once we estimate α_{j^*}, we can recover an estimate of ξ_{j^*}. Following Bajari and Benkard (2005) and Bajari and Kahn (2005), we recover an estimate of ξ_{j^*} from the residual in (5), which yields

$$\xi_{j^*} = \log(p_{j^*}) - \alpha_{j^*,0} - \sum_{k=1}^{m} \alpha_{j^*,k} x_{j^*,k} - \eta_t.$$

We then use (**??**) to recover $\beta_{i,k}$ as follows.

$$\hat{\beta}_{i,k} = \frac{\partial \mathbf{p}_t(X_{j^*}, \xi_{j^*})}{\partial x_{j,k}} = \frac{\partial \log(\mathbf{p}_t(X_{j^*}, \xi_{j^*}))}{\partial x_{j,k}} \times p_{j^*} = \hat{\alpha}_{j^*,k} \times p_{j^*}, \quad \forall k = 1, \ldots, m.$$

To recover $\beta_{i,0}$, the coefficient on ξ_j in (**??**), we use a similar equation as above. Because $\frac{\partial \log(\mathbf{p}_t(X_{j^*},\xi_{j^*}))}{\partial \xi_j} = 1$ in (5), we can easily recover $\beta_{i,0}$ by $\hat{\beta}_{i,0} = \frac{\partial \mathbf{p}_t(X_{j^*},\xi_{j^*})}{\partial \xi_j} = \frac{\partial \log(\mathbf{p}_t(X_{j^*},\xi_{j^*}))}{\partial x_{j,k}} \times p_{j^*} = p_{j^*}$.

[5] Fan and Gijbels (1996) provide detailed treatment on local linear (or polynomial) regression. We could instead use other nonparametric methods such as a kernel estimator (e.g. Nadaraya-Watson estimator or Gasser-Müeller estimator) or a series estimator. However, Bajari and Benkard (2005) found that a local linear kernel estimator as in Fan and Gijbels (1996) worked best. For this reason, we also use the local linear regression.

Step 2: estimating $V^1(\beta_i)$ and $V^0(\beta_i)$

For in-house transactions, we compute $V^1(\beta_i)$ by plugging the recovered β_i and ξ_j into (**??**). Similarly, we compute $V^0(\beta_i)$ for cross-house transactions. To estimate $V^0(\hat{\beta}_i)$ for buyer i with $d_{j^*} = 1$ and $V^1(\hat{\beta}_i)$ for buyer i with $d_{j^*} = 0$, we need to compute the weighted mean of $U_s(\beta_i)$ by putting higher weights on houses with similar characteristics as house j^*, while putting lower or no weights on houses with different characteristics. For this reason, we use a local linear matching method[6] to estimate $E[U_s(\beta_i)|s \in D_i^0]$ for buyer i with $d_{j^*} = 1$ and $E[U_s(\beta_i)|s \in D_i^1]$ for buyer i with $d_{j^*} = 0$. Specifically, the local linear weighted mean is given by the intercept μ_0 in the minimization problem

$$\min_{\mu_0,\mu_1} \sum_{s \in D_i^{1-d_{j^*}}} \left\{U_s(\beta_i) - \mu_0 - (X_s - X_{j^*})'\mu_1\right\}^2 K_B(X_s - X_{j^*}) \times K_b(\xi_s - \xi_{j^*}),$$

where $K_B(\mathbf{v})$ is defined above, $K_b(v) = \frac{1}{b}N(\frac{1}{b}\frac{v}{\hat{\sigma}_k})$, and D_i^1 (or D_i^0) denotes a set of internal (or external) listings in the same market, so that if $d_{j^*} = 1$, we compute the local linear weighted mean by using houses in $D_i^{1-d_{j^*}} = D_i^0$.

Step 3: estimating the distribution of γ_i and c_i

To obtain more information on the extent of strategic promotion, we need to estimate the distribution of γ_i and c_i. To this end, we use (**??**) and impose a parametric assumption on the distribution of $\delta_i = \gamma_i + c_i$. Hence, we do not attempt to fully separate c_i from γ_i, but instead focus on the marginal effect of strategic promotion by using exclusion restrictions and a natural experiment from a policy change. Let us begin by considering the following specifications for γ_i and c_i:

$$\begin{aligned} \gamma_i = &\quad \gamma_0 + W_{1,i}\lambda_1 + W_{2,i}\lambda_2 + \epsilon_i, \\ c_i = &\quad c_0 + Z_i\theta_1 + W_{2,i}\theta_2 + \omega_i, \end{aligned} \tag{7}$$

where γ_0 and c_0 are the intercepts, ϵ_i and ω_i are the error terms, and W_i is a vector of variables related to transaction costs, but $W_{1,i}$ is only related to transaction costs, whereas $W_{2,i}$ is related to

[6] See, e.g., Heckman, et al. (1997, 1998) and Hong (2013) for more details on a local linear matching method.

both transaction costs and strategic promotion. In (7), Z_i is a vector of variables related to strategic promotion but not related to transaction costs. Though we use excluded variables Z_i that only affect strategic promotion, we cannot separately identify γ_i and c_i, because we cannot distinguish γ_0 from c_0 without further restrictions, and $W_{2,i}$ enters both γ_i and c_i.

Therefore, our main approach for the step 3 considers $\delta_i = \gamma_i + c_i$ as follows:

$$\delta_i = \delta_0 + Z_i\theta_1 + W_{1,i}\delta_1 + W_{2,i}\delta_2 + \eta_i, \tag{8}$$

where $\delta_0 = \gamma_0 + c_0$, $\delta_1 = \lambda_1$, $\delta_2 = \lambda_2 + \theta_2$, and $\eta_i = \epsilon_i + \omega_i$. Hence, as long as we have excluded variables Z_i, we can identify and estimate the marginal effect of strategic promotion due to changes in Z_i. If we do not impose any assumption on η_i, we can obtain bounds on θ_1. To obtain point identification, we follow Bajari and Kahn (2005) and impose a parametric distribution on η_i. However, note that the identification of θ_1 does not rely on a particular parametric assumption for η_i. In our application, we assume a logistic distribution, simply because it is straightforward to estimate the model. We thus estimate the parameters using the following likelihood function based on (**??**):

$$\begin{aligned} L(\theta_1, \delta) &= \prod_i F\left(V^1(\beta_i) - V^0(\beta_i) + \delta_0 + Z_i\theta_1 + W_{1,i}\delta_1 + W_{2,i}\delta_2\right)^{d_j} \\ &\quad \times \left[1 - F\left(V^1(\beta_i) - V^0(\beta_i) + \delta_0 + Z_i\theta_1 + W_{1,i}\delta_1 + W_{2,i}\delta_2\right)\right]^{1-d_j}, \end{aligned}$$

where $F(\cdot) = \exp(\cdot)/(1 + \exp(\cdot))$.

Appendix C: Estimation of Preferences for Discrete Attributes from First Order Conditions

If a housing attribute, $x_{j,k}$, only takes integer values, the first order conditions will hold with inequality. Any estimated $\beta_{i,k}$ that satisfies the inequality should be consistent with the optimal choice, so that the true $\beta^*_{i,k}$ will be only partially identified. Suppose that $x_{j,1}$ is the number of bedrooms. If the optimal $x_{j,1}$ is, say, 2, then consumer maximization implies

$$U_j(x_{j,1} = 2, p_j(x_{j,1} = 2)|\beta_i) \geq U_j(x_{j,1} = n, p_j(x_{j,1} = n)|\beta_i), \quad \forall n \neq 2,$$

where we fix all $x_{j,k}$'s ($k \neq 1$) and they are not included above to simplify the exposition. Given our linear utility assumption, the inequality above can be written as

$$\beta_{i,1} \times 2 + \sum_{k=2}^{m} \beta_{i,k} x_{j,k} + \beta_{i,0}\xi_j - p_j(x_{j,1} = 2) \geq \beta_{i,1} \times n + \sum_{k=2}^{m} \beta_{i,k} x_{j,k} + \beta_{i,0}\xi_j - p_j(x_{j,1} = n), \quad \forall n \neq 2,$$

which can be rewritten as

$$\beta_{i,1} \times (2 - n) \geq p_j(x_{j,1} = 2) - p_j(x_{j,1} = n), \quad \forall n \neq 2,$$

These inequalities can be reduced to

$$p_j(x_{j,1} = 2) - p_j(x_{j,1} = 1) \leq \beta_{i,1} \leq p_j(x_{j,1} = 3) - p_j(x_{j,1} = 2), \tag{9}$$

because $p_j(x_{j,1} = 3) - p_j(x_{j,1} = 2) \leq \frac{p_j(x_{j,1}=n) - p_j(x_{j,1}=2)}{n-2}$ for $n \geq 4$, as long as $p_j(x_{j,1})$ is convex or linear in $x_{j,1}$. In our application, we consider

$$\log\left(\mathbf{p}_t(X_j, \xi_j)\right) = \alpha_{j,0} + \sum_{k=1}^{m} \alpha_{j,k} x_{j,k} + \eta_t + \xi_j,$$

where $\alpha_j = (\alpha_{j,0}, \ldots, \alpha_{j,m})$ is a vector of random coefficients. Therefore, $p_j(x_{j,1}) = \exp(\alpha_{j,0} + \sum_{k=1}^{m} \alpha_{j,k} x_{j,k} + \eta_t + \xi_j)$, which is convex in $x_{j,1}$.

The discussion above suggests two results on identification of $\beta^*_{i,1}$. First, if the price function is linear in $x_{j,1}$ (e.g. $\mathbf{p}(X_j) = \alpha_{j,0} + \sum_{k=1}^{m} \alpha_{j,k} x_{j,k} + \xi_j,$), the inequality (9) becomes equality, because $p_j(x_{j,1} = 2) - p_j(x_{j,1} = 1) = p_j(x_{j,1} = 3) - p_j(x_{j,1} = 2)$. Thus, even though $x_{j,1}$ is not continuous, we can recover $\beta^*_{i,1}$ from the equality FOC. Second, if the price function is convex in $x_{j,1}$ as in our application, $\beta^*_{i,1}$ is partially identified by the inequality (9). Because $\frac{\partial p_j(x_{j,1}=2)}{\partial x_{j,1}}$ lies between $p_j(x_{j,1} = 2) - p_j(x_{j,1} = 1)$ and $p_j(x_{j,1} = 3) - p_j(x_{j,1} = 2)$, if we set $\beta_{i,1} = \frac{\partial p_j(x_{j,1}=2)}{\partial x_{j,1}}$, this is similar to using the mid point value for an interval variable. In this case, the issue is not that the buyer has chosen a particular attribute violating the equality FOC (this is the case of the suboptimal choice, which can be captured by the optimization error), but rather that the buyer has chosen an optimal amount of a given attribute, but this optimal choice can be rationalized by any values within the interval given by (9).

We further examine conditions under which our approach of setting $\hat{\beta}_{i,1} = \frac{\partial p_j}{\partial x_{j,1}}$ does not result in any bias. To this end, suppose that $\beta^*_{i,1}$ denotes the true value of $\beta_{i,1}$, and so the inequality (9) is satisfied for $\beta^*_{i,1}$. We can then consider a prediction error ϵ_i, so that $\hat{\beta}_{i,1} = \beta^*_{i,1} + \epsilon_i$. Note that our matching estimator ($V^1(\hat{\beta}_i)$ or $V^0(\hat{\beta}_i)$) is constructed from $U_j(\hat{\beta}_i) = \sum_{k=1}^{m} \hat{\beta}_{i,k} x_{j,k} + \hat{\beta}_{i,0}\xi_j - p_j$. Hence, for our matching estimator to be consistent, we need that $E(\beta_{i,1}x_{j,1})$ be equal to $\beta^*_{i,1}x_{j,1}$. This is satisfied when $E(\epsilon_i) = 0$ and $E(\epsilon_i|x_{j,1}) = 0$, because $E(\beta_{i,1}x_{j,1}) = E(\beta^*_{i,1}x_{j,1} + \epsilon_i x_{j,1}) = \beta^*_{i,1}x_{j,1} + E(\epsilon_i x_{j,1})$. We believe that these conditions are not too strong in our case.