An Empirical Two-sided Equilibrium Search Model of the Real Estate Market

Paul E. Carrillo

October 2005

Abstract

Two-sided search is an important feature of the Real Estate market (REM). Existing empirical equilibrium search models assume that only one side of the market searches and, thus, can not adequately describe this market. We specify a two-sided equilibrium search model of the REM that incorporates five very important characteristics of the REM: a) buyers' and sellers' search behavior, b) heterogeneity in agents' motivation to trade, c) transaction costs, d) a trading mechanism with posting prices and bargaining, and e) the availability of an online advertising technology. The model is estimated using Maximum Likelihood methods and Multiple Listing Services data. We use the estimated model to analyze how information displayed in real estate ads on the internet affect prices and agent's search behavior. Our estimates suggest that, on average, only 3% of the relevant information that home-buyers collect before making a purchase decision is obtained through online ads. Furthermore, we find that improvements in online information displayed by Real Estate ads decrease equilibrium prices but increase the time that a property stays on the market.

Keywords: Advertising, Equilibrium Search Models, Internet, Real Estate Market.
JEL Codes: C51, D58, D80, D83.

* University of Virginia Department of Economics and Banco Central del Ecuador. Email: pcarrillo@virginia.edu. I owe thanks to my dissertation advisors, Steven Stern and Maxim Engers, for their invaluable help and guidance. I also gratefully acknowledge financial support from the University of Virginia's Bankard Fund for Political Economy and from the Banco Central del Ecuador.
1 Introduction

Equilibrium search models have been widely used to explain many important features of labor markets. In these type of models, workers search sequentially, drawing price offers from a known distribution. In the other side of the market, firms react optimally to individual’s behavior and price dispersion arises as an outcome of the equilibrium. One may use this framework to model other markets where search costs are high. However, to the best of my knowledge, every empirical model assumes that only one side of the market searches. This assumption is needed a) to make these models tractable and b) for identification because, usually, only data on the seller’s side are available.

Two-sided search is an important feature of many markets. For example, in the real estate market, both buyers and sellers actively participate in the search process to buy or sell a home and both incur in expensive costs during this process. For sellers, it is costly to keep their homes on the market, while buyers incur in high pecuniary and non-pecuniary expenses when visiting properties.

The existing empirical equilibrium search models can not adequately describe a market where both buyers and sellers search simultaneously. Thus, to answer any meaningful question about the real estate market, one ought to specify and estimate a model with two-sided search. In this research paper, we specify such a model and show that it is feasible to estimate with seller’s data only.

The theoretical model modifies the framework of existing equilibrium search models in the labor literature to capture the unique nature of the real estate market. To our knowledge, it is the first attempt in the literature to model in an equilibrium context five very important characteristics of the real estate market: a) buyers’ and sellers’ search behavior, b) heterogeneity in agents’ motivation to trade, c) transaction costs, d) a trading mechanism with posting prices and bargaining, and e) the availability of an online advertising technology.

In our model, both buyers and sellers are infinitely lived agents who search simultaneously for a potential trading partner. Buyers and sellers are heterogeneous, as some individuals are

---

1 See, for example, Eckstein and Wolpin (1990), Albrecht and Axell (1984), and Mortensen (1990).
2 A revision of these literature is presented in the next chapter.
more motivated to trade than others. Every seller uses the internet to advertise her product (housing unit) displaying online its posting price and some of its features. Each period, buyers use the internet to sample price postings (ads), learn the characteristics of the unit, and decide whether to visit a seller. There is an idiosyncratic random buyer-home value that can be learned fully only when a potential buyer visits a home. Part of this value, however, is revealed to the buyer when she looks at the ad on the internet. Sellers wait for a potential buyer to visit their property. When two potential trading partners meet, they play a well defined bargaining game, and trade may or may not occur. If a buyer and a seller engage in trade, they leave the market forever; otherwise, they return to the market and search for a trading opportunity next period. Sellers choose a posting price and a reservation value (price floor) given the characteristics of their home and their motivation to trade. Buyers, on the other hand, decide when to visit a property and when to purchase a home given their motivation to trade. An endogenous equilibrium in these optimal buyers’ and sellers’ strategies defines the solution to the theoretical model.

To estimate our theoretical model, we follow the growing literature on estimation of equilibrium search models and use maximum likelihood methods. The data used for estimation consists of Multiple Listing Services data (posting prices, transaction prices, time that a property stays on the market, and a detailed set of home characteristics) for real estate transactions in Charlottesville City and Albemarle County (VA) during the years 2000 through 2002.

The estimated model may be used as a tool to perform a variety of counterfactual experiments. We focus on a particular counterfactual that shows how improvements in the amount of information displayed in real estate ads on the internet affect this market. This is a particular important question for several reasons. First, the last decade has witnessed an explosive growth in the amount of information transmitted online. The information transmitted to web users is limited mostly by technological constraints (speed connections and specialized software) which have been rapidly changing over the last five years, and, according to Vanston et. al (2005), will be significantly less restrictive in the next decade.

\[\text{See, for example, Eckstein and Wolpin (1990), Kiefer and Neumann (1994) and Bunzel et. al. (2001).}\]

\[\text{According to Vanston et. al (2005), by 2006, U.S. broadband penetration will likely be above 50\% and a shift to data rates of 24 Mbps to 100 Mbps will have begun. By 2010, 75\% broadband penetration is likely,}\]
These technological changes foreseen for the future should affect buyer’s and seller’s behavior. Second, housing units, in the vast majority of cases, are advertised through the internet in specialized web sites. These online housing sites have notably improved the amount of information they display by incorporating pictures and virtual tours to the existing Multiple Listing Services, and important improvements are forecasted for the near future. Third, the widespread use of the internet has drawn the attention of many researchers who, using micro-data, have attempted to explain how the Internet has affected prices, search behavior, and offline markets. One can think, however, of two elements that determine the importance of electronic markets relative to offline markets: a) the size of the online markets, which is proportional to number of consumers who have internet access, and b) the amount and quality of the information that internet marketplaces provide to consumers. Clearly, changes in each of these elements should affect markets differently, and we are not aware of any study that has attempted to separate these effects. While the previous studies identify the global effect of internet use in some specific markets, they do not assess how improved information technology may influence buyers’ and sellers’ behavior. Finally, classical economic theory cannot adequately evaluate the effects of improvements in online information technology in the housing market because some assumptions underlying the traditional supply and demand model are invalid in the presence of market imperfections such as transactions costs or asymmetric information. To appropriately assess these effects, it is important to specify a model that depicts individual behavior of buyers and sellers that is consistent with the nature of the imperfections of the real estate market.

Parameter estimates have a reasonable interpretation, and the model is able to replicate the pricing data remarkably well. For instance, posting and transaction prices are predicted almost with the same accuracy of a linear regression model despite the variety of assumptions and the complexity of our structural model.

Our estimates suggest that, on average, only 3% of the relevant information that home-
buyers collect before making a purchase decision is obtained through on-line ads. Furthermore, through the counterfactual experiment we find that additional online information decreases market prices. For instance, if buyers could obtain from an online ad all the relevant information they need to make a purchase decision, posting and transaction prices would decrease by 2.9% and 3.1%, respectively; on the other hand, the average time that a home stays on the market would increase in about four days.

In the next section, we compare our research to the existing literature. The third section presents the theoretical model. In the fourth section, we describe the data. Section 5 discusses the estimation methods. The sixth section presents our results and the last one concludes.

2 Literature review

There is a growing literature that seeks to explain how the use of the internet has affected online and offline markets. For example, Brown and Goolsbee (2002) examine a relative homogeneous product (term life insurance) and conclude that the use of the internet has reduced term life prices by 8 to 15 percent. In a different study, Goolsbee (2000) suggests that applying existing sales taxes to Internet commerce might reduce the number of online buyers by up to 24 percent. D’urso (2002) studies the impact of internet use on the duration of search in the housing market, and, using Instrumental Variable Quantile Regression methods she finds that it’s use increases the duration of home search relative to employing other conventional search methods. While these previous studies identify the global effect of internet use in some specific markets, they do not assess how improved information technology may influence buyers’ and sellers’ behavior in these online markets.

The theoretical model was motivated by the economic theory of search. Standard search models assume identical agents search sequentially, drawing price offers from a known stationary distribution. They show how the characteristics of the market -discount rates, distribution of prices, and transaction costs- affect search behavior, and have been widely used to describe posting and transaction prices in the REM.\textsuperscript{7} Many variants of this standard model

\textsuperscript{7}Yinger (1981) was one of the first to use a search model in a formal analysis of real estate markets. Yavas (1992) presents a single period matching model examining the impact of buyers’, sellers’, and brokers’ search
have been created, but these typically do not explain the origin of the price distribution and are inconsistent with rational conduct.\textsuperscript{8}

Using assumptions about the sellers’ optimal behavior and agents’ heterogeneity, Burdett and Judd (1983), Albrecht and Axell (1984), and Mortensen (1990) have introduced theoretical models that determine non-degenerate price distributions and search behavior as an equilibrium outcome.\textsuperscript{9} However, these models assume that only one side of the market searches, and cannot adequately describe a market where both buyers and sellers search simultaneously (for example, the REM).

Mortensen and Wright (2002) introduce a two-sided search model where buyers and sellers with heterogeneous preferences simultaneously search and bargain over the price of an indivisible object. Agents have perfect information about each other’s preferences at the time they bargain, and the reservation prices, the size of the market, and the price distribution arise as an outcome of the equilibrium. While their model successfully describes search behavior in a two-sided market, it does not explain two very important features of the REM: heterogeneity in the agents’ motivation to trade, and posting prices and bargaining.

We introduce heterogeneity in the agents’ motivation to trade by assuming that the buyer’s and seller’s intertemporal discount factors are random variables.\textsuperscript{10} To introduce

\textsuperscript{8}That is, if individuals are homogeneous and face the same price distribution, they will have the same reservation price. Rational price-posting sellers would then post only that price, leading to a degenerate price distribution. If there is a degenerate price distribution, agents will not search, but search is precisely what these models are trying to explain (Diamond 1971).

\textsuperscript{9}Burdett and Judd (1983) assume that identical agents sample a subset of price offers. Agents with more than one offer are able to bargain for a price below their reservation price. This induces the sellers to set their prices using a mixed strategy that generates an endogenous distribution of prices in the market. Albrecht and Axell (1984) introduce heterogeneity in workers and firms (i.e. buyers and sellers). They assume that firms have different levels of productivity and workers differ in their valuation of leisure. Workers look sequentially for a job, and as firms maximize their profits, it may be optimal for heterogeneous firms to have different wage policies. Mortensen (1990) models the labor market by assuming that workers search while employed and change jobs if they sample a wage realization greater that the current wage. Firms offering high wages are able to keep workers for a longer period of time than firms offering low wages, but their per-period surplus is smaller. Therefore, firms face a trade-off between offering short and long run profits. In equilibrium, different wage policies are optimal for each firm, giving rise to an endogenous distribution of wages.

\textsuperscript{10}The level of the agents’ motivation to trade is clearly one of the most important elements in the housing market. However, there is little literature that addresses this topic. Glower et al. (1998) explain how sellers’ motivation fits into the standard (one-sided) search model. Then they use a small sample of sellers to explain
posting prices in our equilibrium search model we assume (as in Chen and Rosenthal 1996) that the list price constitutes a price ceiling and a commitment device. Unlike Chen and Rosenthal (1996), however, we use a simpler bargaining game that allows us to solve for the seller’s list price and reservation value analytically.

Because equilibrium search models provide a natural interpretation of interesting market phenomena, the estimation of such models has received considerable attention. Eckstein and Wolpin (1990) estimate a generalization of the Albrecht and Axell model.\textsuperscript{11} They use assumptions about the distribution of preferences and technology to identify the parameters of their model using workers’ data only. Their estimated model fails to conform to the data, and measurement error accounts for almost all of the dispersion in wages. Kiefer and Neumann (1994) estimate a version of the model of Mortensen (1990). Bunzel et. al. (2001) estimate a version of Mortensen (1990) with several variations that fit the data better (measurement error and heterogeneity in firms productivity). The empirical implementation of these models is still developing. Future research needs to address identification issues as well as generate models that fit the data more closely. The estimation of our structural model adds to the growing literature related to estimation of equilibrium search models.

3 Theoretical Model

3.1 The market

There exists a market with risk-neutral, infinitely lived agents. The agents are households who either are actively searching for a home (buyers), or who have a home for sale (sellers). Agents are alike except for how motivated they are to trade.

To model this heterogeneity, we define $\beta_b$ and $\beta_s$ as the buyers’ and sellers’ value of an opportunity to trade in the next period relative to the same opportunity in the current period and assume that they are random variables with constant distributions $K_b$ and $K_s$, respectively. Formally, each of these discount factors are formed by two components; that the role of their “motivation” in determining selling time, list price and sale price and use their empirical results to test the theoretical hypothesis. They conclude that motivation affects the expected time in the market and the sale price, but not the posting price. Our research addresses the same issue using a general equilibrium approach.

\textsuperscript{11}Canals and Stern (2001) provide a comprehensive survey of empirical search models.
is, $\beta_b = \beta^o \beta^*_b$ and $\beta_s = \beta^o \beta^*_s$. The first component $\beta^o = \frac{1}{1+r}$ discounts the future using a discount rate $r$ that is common to all buyers and sellers. On the other hand, the second components, $\beta^*_b$ and $\beta^*_s$, are random variables that capture idiosyncratic differences in buyers’ and sellers’ motivation when buying or selling their home. Note that the lower a household’s $\beta$ the more motivated and eager it is to engage in trade.

A home is considered to be an indivisible good that can be described fully by a vector of characteristics $X$ from which both buyers and sellers derive utility. Define $s_f$ as the per-period utility flow that sellers obtain by owning this good, and let $s = \frac{s_f}{1-\beta^o}$ be the level of lifetime utility that sellers obtain by owning this good; furthermore, let $s = X\gamma$, where $\gamma$ is a vector of parameters. Notice that $s$ represents the quality of the home and that properties with different features $X$ may provide sellers with the same level of utility $s$. To model heterogeneity in homes’ characteristics, we assume that $s$ is distributed according to an exogenous distribution $\Psi$, which is common knowledge to every agent in the market.

The lifetime utility that properties provide to buyers varies for each buyer-home combination. To model this assumption, we let $\tilde{b}$ be a random buyer-home match value that captures the lifetime utility a specific buyer derives from owning one particular property. Furthermore, let $\tilde{b}$ depend on the home quality and two independent (from each other and for any home-and-buyer combination) mean zero random errors $b^o$ and $b^u$. That is, $\tilde{b} = \delta + b^o + b^u$, where $\delta$ is a scalar parameter and $b^o$ and $b^u$ are random variables with exogenous cumulative distributions $G_o$ and $G_u$, respectively. Notice that the parameter $\delta$ captures average percent differences in the properties’ valuations between buyers and sellers, and realizations of $\tilde{b}$ correspond to the specific value that one buyer assigns to one particular property.\footnote{The reader may notice that, for a given set of housing characteristics $X$, every seller receives the same lifetime utility $s = X\gamma$ from owning this particular house. On the other hand, we assume that the lifetime utility $b$ that buyers receive from owning the same good varies for each buyer-home combination. We are aware that, to the extent that today’s buyers may become sellers in the future, this assumption may not be realistic. However, we choose to use this assumption since it simplifies the analytical solution to the model. A natural way to relax this assumption consists of including a random component in $s$, which does not change the nature nor the main results of our model.}

A seller joins the housing market by placing an ad that informs all the other agents: a) that her home is for sale, b) the posting price and, c) the home characteristics. As in Horowitz (1992), and Chen and Rosenthal (1996a and 1996b), we assume that the posting price constitutes a price ceiling and a commitment device; that is, if a potential buyer wants
to buy the product at the posting price, the seller is obligated to engage in trade.\footnote{It is worth acknowledging, however, that houses do sometimes sell above the posting price. In certain cases, sellers are surprised by high demand and in others, sellers deliberately set low asking prices to foster competition among buyers. These cases, however, are a) infrequent (1.5\% in our sample), b) almost always regarded as a seller who is “bargaining in bad faith”, and, c) in some locations, ruled out by the existence of legal contracts that give the real estate broker the right to damages in the event that a seller does not agree to trade at the list price (Chen and Rosenthal 1996a). In any event, the assumptions of our paper rule out the possibility that the transaction price is above the posting price.}

Buyers search “home for sale” ads sequentially. Every period, a buyer samples an ad that provides her information about the home’s characteristics $X$, its posting price $p_s$, and $b^o$ (a fraction of her total random buyer-home match value). When a buyer observes an ad, she decides whether or not to visit the home. If she decides to visit the home, the buyer tours the house and $b^o$ is revealed to her.

After the buyer has visited the home, she meets the seller and both bargain over the transaction price. For simplicity, we adopt a reduced form representation of the bargaining process. With probability $\theta$, the seller is not willing to accept counter-offers, and the posting price $p_s$ constitutes a take-it-or-leave-it offer to the buyer. With probability $(1 - \theta)$, the buyer has the option to make a counter take-it-or-leave-it offer $p_b$ to the seller. It is assumed that, once a buyer has visited a property, she has perfect information about the seller’s preferences. That is, if she makes a counter take-it-or-leave-it offer, she will bid the seller’s reservation value $R_s$ (the minimum price at which she is willing to sell her property). During the meeting, the buyer decides whether she buys the home (paying either $p_s$ or $p_b$), or searches again for a new ad (posting price) next period. When a buyer (seller) buys (sells) a property, she exits the market forever.\footnote{Instead, we could use a Rubenstein (1982)-type bargaining model where the surplus from trade is shared in fixed proportions between the buyer and the seller (such as in Chen and Rosenthal (1996a) or in Mortensen and Wright (2002), for example). Using this approach, however, we could not find analytical solutions for the optimal seller’s posting and reservation prices.}

### 3.2 The seller’s problem

From a seller’s point of view, trade occurs only if a buyer visits her property and is willing to trade, either at the posting price $p_s$ or at her reservation value $R_s$. Let $q(p_s|s)$ (to be determined endogenously) be the rate at which buyers visit a particular seller who owns a type $s$ property and has posted a price $p_s$; also define $\gamma_b(p_s|s)$ ($\gamma_s(p_s|s)$) as the probability
that a buyer is willing to buy this property given that she has visited the home and did (did not) have the opportunity to make a counter offer (both to be determined endogenously).

Using these assumptions we define the seller’s expected gain from trade and searching as

\[ \Pi_s^e = q \theta [\gamma_s(p_s - s) + (1 - \gamma_s)W_s] + \]
\[ q(1 - \theta) [\gamma_b(R_s - s) + (1 - \gamma_b)W_s] + [1 - q]W_s \]
\[ = q [\theta \gamma_s(p_s - s) + (1 - \theta) \gamma_b(R_s - s)] + [1 - q(\theta \gamma_s + (1 - \theta) \gamma_b)]W_s, \]

and

\[ W_s = \beta_s E[\max\{p - s, W_s\}], \]

where \( \Pi_s^e \) is the seller’s expected profit from trade, \( W_s \) is her expected gain from searching (her value of search), and \( p \) is the random (from the seller’s point of view) transaction price.

Equation (1) states that, in every period, there is \( q \theta \gamma_s \) probability that a seller sells her home for the posting price and obtains \( p_s - s \) profit when trading; with probability \( q(1 - \theta) \gamma_b \), trade occurs at the seller’s reservation value, in which case her gains from trade are \( R_s - s \); finally, if trade does not happen, she returns to the market and keeps her value of search \( W_s \).

As presented in equation (2), the seller’s value of search is the discounted expected value of having an opportunity to trade next period, and it represents the utility that a seller obtains by staying in the market.

The seller’s problem consists of choosing the optimal reservation value \( R_s^* \) and posting price \( p_s^* \) that simultaneously maximize her expected profit and value of search.

First, let us solve the seller’s search problem. For any given \( p_s^* \), and using the assumptions of our bargaining game, we work out the expectation in equation (2) and obtain

\[ W_s = \beta_s \left\{ q \theta \gamma_s(p_s^* - s) + q(1 - \theta) \gamma_b(R_s - s) + [1 - q(\theta \gamma_s + (1 - \theta) \gamma_b)]W_s \right\}. \]

Furthermore, following the solution techniques of standard search models, notice that any optimal seller’s behavior necessarily implies that \( R_s^* = W_s^* + s \).\(^{15}\) We replace this optimality condition in the previous equation, rearrange, and solve for the optimal reservation value

\[ R_s^* = \frac{\beta_s \theta (1 - \phi(p_s^*|s)) p_s^* + (1 - \beta_s)s}{1 - \beta_s[1 - \theta(1 - \phi(p_s^*|s))]}, \]

\(^{15}\)See, for example, Lipmann and McCall (1976).
where, for notational simplicity, we have defined $1 - \phi(p_s|s)$ as the probability that, given that the posting price is a take-it-or-leave-it offer to the buyer, a type $s$ home sells for the posting price; that is: $1 - \phi(p_s|s) = q(p_s|s)\gamma_s(p_s|s)$.

The previous result allows us to find the seller’s optimal posting price $p^*_s$. To determine $p^*_s$, we substitute the optimality condition ($R^*_s = W^*_s + s$) in equation (1) and obtain that, for any $R^*_s$,

$$\Pi^*_s = \theta q\gamma_s(p_s - s) + (1 - \theta q\gamma_s)(R^*_s - s).$$

Differentiating this equation with respect to $p_s$, we derive that the optimal seller’s posting price $p^*_s$ solves

$$p^*_s - R^*_s = \frac{1 - \phi(p^*_s|s)}{\phi'(p^*_s|s)}.$$

**Theorem 1:** As long as (a) the hazard function $h(p^*_s|s) = \frac{\phi(p^*_s|s)}{1 - \phi(p^*_s|s)}$ is non-decreasing in $p^*_s$, and (b) $\phi$ is decreasing in $s$, the optimal seller’s posting price and reservation value are defined by the unique pair $\{p^*_s, R^*_s\}$ that solves equations (3) and (4) simultaneously. For any seller, and for any $\beta_s$ and $s$: $p^*_s(\beta_s, s) \geq R^*_s(\beta_s, s) \geq s$; in addition, these functions are increasing in both arguments. Proof in Appendix.

The two assumptions of Theorem 1 are not restrictive in any sense. Condition (a) is a commonly used standard assumption about the shape of the demand function that guarantees the existence of a unique solution in similar problems and is satisfied by many standard distributions, such as the normal, uniform, and exponential. However, notice that $\phi(p^*_s|s)$ will be determined endogenously in our model. Thus, to guarantee that this assumption is satisfied in equilibrium, we need to specify this function, and place certain restrictions in the exogenous distributions of our model. Condition (b) makes the reasonable conjecture that, everything else constant, the higher the quality of a home, the higher the probability it is sold.

The results of Theorem 1 are all intuitive. First, we expect that a rational seller would

---

16 Notice that $1 - \phi(p_s|s)$ resembles a traditional demand function.

17 This assumption is equivalent to $1 - \phi(p^*_s|s)$ being strictly log-concave, and is commonly used in the literature (for example, in Anderson and Renault 2004 and in Chen and Rosenthal 1996b). For a list of distributions satisfying this assumption as well as some of its other applications, see Bagnoli and Bergstrom (1989).

18 This condition says that an increase in $s$ increases $p^*_s$ in the sense of first-order stochastic dominance.
post a sale price that is at least as high as her reservation value; and that the minimum price at which she is willing to sell her product must be no less than her outside option (the utility that she gets by keeping the product). Second, the model predicts that both posting and reservation prices diminish as the seller’s motivation to trade increases, and these predictions are consistent with other findings (Glower et al. 1998). Third, higher quality properties sell for higher prices. Finally, the monotonicity of $p^*_s(\beta_s, s)$ and $R^*_s(\beta_s, s)$ should facilitate the derivation of $\phi(p^*_s|s)$.

3.3 The buyer’s search problem

The buyer’s optimal behavior can be described with a two stage search model. In the first stage, she samples an ad and uses the information contained in it to decide whether she should visit this home. In the second stage, given that she has visited this property, she obtains additional information about its value and the outcome of the bargaining game and decides between purchasing it or searching for a new ad next period. In this section, we formally describe the solution to such problem.

When a buyer picks an ad, she observes the home’s features $X$, its posting price $p_s$, and $b^o$, a component of her buyer-home match value. From her point of view, a pair $\{p_s, s\}$ is an independent realization from the joint distribution of posting prices and home characteristics $\Gamma(p_s, s)$, while a specific value of $b^o$ is an independent realization from the distribution $G_o$. Notice that $\Gamma(p_s, s)$ is well defined by the optimal sellers’ pricing strategies $p^*_s(\beta_s, s)$ as well as the exogenous distributions of discount factors $K_s$ and valuations $\Psi$.

When the buyer visits a property, she pays a known visiting cost $c_b$. During the visit, the other component of her buyer-home match value $b^w$ and the outcome of the bargaining game are revealed to her. At this stage, she chooses between buying the property and staying in the market for another period. We assume that the buyer chooses the optimal strategy that maximizes her expected value of search.

Before she decides to visit a home, her value of search $W_b$ is the discounted expected value.\footnote{Visiting costs include transportation and monetary opportunity costs (time costs), as well as non pecuniary (emotional) costs of touring a property.}
utility of the maximum between visiting a property and waiting for another ad next period

\[ W_b = \beta_b \int \int \int \max \{ U^c(p^*_s, s, b^o, W_b), W_b \} d\Gamma(p^*_s, s)dG_o(b^o). \] (5)

\( U^c(p^*_s, s, b^o, W_b) \) is the buyer’s expected value of having an opportunity to visit a property at the time she looks at the listing, before observing the realization of \( b^o \), and before knowing who gets to make the take-it-or-leave-it offer

\[ U^c(p^*_s, s, b^o, W_b) = \theta \int \max \{ b^o + b^o + \delta s - p^*_s, W_b \} dG_u(b^u) \]

\[ + (1 - \theta) \int \max \{ b^o + b^o + \delta s - p_b, W_b \} dG_u(b^u) - c_b. \] (6)

Because buyers know the optimal strategies \( p^*_s(\beta_s, s) \) and \( R^*_s(\beta_s, s) \), they are aware of a seller’s reservation value \( R^*_s(p^*_s, s) \) once they have observed \( p_s \) and \( s \) in an ad. Hence, if buyers have the opportunity to make a counter offer, it is optimal for them to ask \( p_b = R^*_s(p^*_s, s) \) which does not depend on \( b^o \). Using this result, we integrate the right hand side of equation (6) by parts and obtain\(^{20}\)

\[ U^c(p^*_s, s, b^o, W_b) = W_b + j(p^*_s, s, b^o, W_b), \] (7)

where

\[ j(p^*_s, s, b^o, W_b) = \theta \int_A [1 - G_u(b^u)] db^u + (1 - \theta) \int_B [1 - G_u(b^u)] db^u - c_b, \]

\[ A = \{ b^u : b^u \geq W_b + p^*_s - \delta s - b^o \}, \]

and

\[ B = \{ b^u : b^u \geq W_b + R^*_s(p^*_s, s) - \delta s - b^o \}. \]

Now, let us find the value of search \( W^*_b \) that solves equations (5) and (7). We substitute equation (7) into (5), rearrange, and find

\[ W^*_b = \frac{\beta_b}{1 - \beta_b} \int \int \int \max \{ j(p^*_s, s, b^o, W^*_b), 0 \} d\Gamma(p^*_s, s)dG_o(b^o). \] (8)

\(^{20}\)Details of the integration are provided in the appendix.
It is easy to see that there is a unique $W^*_b$ that solves equation (8).

It is straightforward to show that $U^e$ is decreasing with respect to the posting price. This fact implies that it is optimal for the buyer to follow a reservation strategy such that, given a particular set of housing characteristics $s$ and buyer-home match component $b^o$, she visits the property if and only if the posting price $p^*_s$ is below a reservation price $p^*_b(s,b^o)$. Hence, for any $s$ and $b^o$, the optimal buyer’s reservation price must be such that her value of having an opportunity to visit a property equals her value of search

$$U^e(p^*_b, s, b^o, W^*_b) = W^*_b.$$ To find the optimal $p^*_b(s, b^o)$, we replace this optimality condition in (7) and solve

$$\theta \int_{A(p^*_b, s, b^o)} [1 - G_u(b^u)]db^u + (1 - \theta) \int_{B(p^*_b, s, b^o)} [1 - G_u(b^u)]db^u = c_b, \quad (9)$$

where

$$A(p^*_b, s, b^o) = \{b^u : b^u \geq W^*_b + p^*_b - \delta s - b^o\},$$

and

$$B(p^*_b, s, b^o) = \{b^u : b^u \geq W^*_b + R^*_s(p^*_b, s) - \delta s - b^o\}.$$

**Theorem 2:** The solution to the buyer’s two-step search is defined by a unique value $W^*_b$ and a function $p^*_b(s, b^o)$ that solve equations (8) and (9) respectively, along with the optimal strategies: (a) visit a property if, given a particular realization of $s$ and $b^o$, an observed ad’s posting price $p_s \leq p^*_b(s, b^o)$; (b) if she has visited a home and does not have the opportunity to make a counter offer, she buys the property if and only if $\hat{b} - p_s > W^*_b$; and (c) if she has visited a home and has the opportunity to make a counter offer, she should make a take-it-or-leave-it-offer (which will always be accepted) of $p_b = R^*_s(p_s, s)$ if and only

---

21 The right hand side of equation (8) is no less than zero, and decreasing in $W^*_b$ (since $j(p^*_b, s, b^o, W^*_b)$ is clearly decreasing in $W^*_b$). On the other hand, the right hand side of equation (8) crosses the origin and has a positive slope. Thus, a unique solution $W^*_b$ exists.

22 We use equation (7) -and Leibnitz rule- to show that $U^e$ is monotone

$$\frac{\partial U^e}{\partial p^*_b} = -\theta(1 - G_u(V_p)) - (1 - \theta)\frac{\partial R^*_s}{\partial p_s}(1 - G_u(V_r)) \leq 0,$$

where $V_p = W_b + p^*_s - \delta s - b^o$, and $V_r = W_b + R^*_s(p^*_s | s) - \delta s - b^o$. 

13
if $\tilde{b} - R_s^*(p_s, s) > W_b^*$. In addition, $W_b^*$ is increasing in $\beta_b$, while $p_b^*$ is decreasing in $\beta_b$. Proof in appendix.

It is useful to analyze the case when the distributions of $s$ and $b^o$ are degenerate. When this is the case, we are able to find analytical solutions for $W_b^*$ and $p_b^*$ and provide an intuitive interpretation of the buyer’s optimal decisions.

**Theorem 2a:** When the distributions of $s$ and $b^o$ are degenerate, the solution to the buyer’s two-step search model is defined by the unique pair $\{W_b^*, p_b^*\}$ that solves equations (10) and (11) simultaneously, along with the optimal strategies described in (a), (b), and (c) in Theorem 2.

\[
\frac{\beta_b}{1-\beta_b} \int \Gamma(p_s^*) \left(-\frac{\partial U^s}{\partial p_s^*}\right) dp_s^* = W_b^* 
\]

\[
\theta \int_{W_b^* + p_b^* - \delta_s} W_b^* \Gamma(p_s^*) \left[1 - G_u(b^o)\right] db^o + (1 - \theta) \int_{W_b^* + R_s^*(p_b^*) - \delta_s} \left[1 - G_u(b^o)\right] db^o = c_b
\]

Proof in appendix.

Equation (10) states that the expected benefits from sampling a posting price lower than the buyer’s reservation price $p_b^*$ (the left hand side) should be the same as the per-period expected return of staying in the market (since $\frac{1-\beta_b}{\beta_b}$ equals the per-period discount rate). Equation (11) implies that the optimal $W_b^*$ and $p_b^*$ must be such that, the buyer’s expected benefit from visiting a property equals her visiting cost.

### 3.4 Equilibrium

From the sellers’ point of view, there is a distribution of heterogenous buyers in the market, each one of them with a different value of search. Because sellers are rational individuals who know the optimal buyers’ strategies $p_b^*(\beta_b, s, b^o)$, $W_b^*(\beta_b)$, and the relevant exogenous cumulative distribution functions, each is able to determine the probability that a buyer visits her property and is willing to trade.

The probability that a buyer visits a seller who has posted a price $p_s$ and owns a type $s$ home is $q^*(p_s|s) = \Pr\{p_b^*(\beta_b, s, b^o) > p_s\}$. In addition, $\gamma_s^*(p_s|s) = \Pr\{\tilde{b} - p_s >$
$W^*_b(\beta_b) | p^*_b(\beta_b, s, b') > p_s$ is the probability that the buyer is willing to buy the property given that she has visited the home and did not have the opportunity to make a counteroffer; and $\gamma^*_b(p_s | s) = \Pr\{\bar{b} - R^*_s(p_s) > W^*_b(\beta_b) | p^*_b(\beta_b, s, b') > p_s\}$ is the probability that the buyer buys the property given that she has visited the property and had the option to make a counteroffer.

Notice that $q^*(p_s | s), \gamma^*_s(p_s | s)$, and $\gamma^*_b(p_s | s)$ are well defined by the buyer’s optimal strategies and the exogenous cumulative distribution functions $G_o, G_u$ and $K_b$. Furthermore, we show in the appendix that $1 - \phi^*(p_s | s) = q^*(p_s | s) \gamma^*_s(p_s | s)$ is a well defined decreasing function.

**Equilibrium conditions:** The equilibrium of the model is determined by a fixed point in the following probability distributions:

$$q(p_s | s) = q^*(p_s | s) ; \quad \gamma^*_s(p_s | s) = \gamma_s(p_s | s) ; \quad \gamma^*_b(p_s | s) = \gamma_b(p_s | s)$$

Thus, the solution to the model is defined by a Bayesian Nash Equilibrium, where, given every buyers’ and sellers’ beliefs and optimal strategies, no one has an incentive to deviate from them.

**Theorem 3:** There exists probability functions $\{\phi^*, \gamma^*_s, \gamma^*_b\}$ such that the equilibrium conditions are satisfied. Proof: A heuristic proof is presented in the appendix.

The strategy to prove the existence of an equilibrium is to show that the space of functions $\{\phi^*, \gamma^*_s, \gamma^*_b\}$ maps continuously into itself. Once it is shown that the space spanned by these functions is a compact convex subset of a Banach space, existence follows from Schauder’s Fixed Point Theorem.

### 3.5 Changes in the information technology

Developments in the information technology enhances the content of a home listing and provides buyers with additional information at the time they look at an ad. This additional information should change the agents’ optimal strategies and the equilibrium of the market.

---

23The structure of Theorem 3’s proof follows the existence proof in Stern (1990).
In this section, we show how our theoretical model could be used to predict the effects of these improvements on the real estate market.

A natural way to evaluate the effects of better information technology consists of assuming that this new technology allows buyers to learn a greater portion of their buyer-home match value at the time they look at an ad. That is, we should explore how the solution to the equilibrium search model changes when the variance of $b^a$ decreases relative to the variance of $b^0$. These technological changes should affect the distribution of posting and transaction prices, the buyer’s visiting rate, and the average time that a seller stays in the market.

Because in our equilibrium model the variables are related in complicated nonlinear ways, we cannot give precise theoretical insights about the size of these effects. For example, as more information becomes available, buyer’s are able to make a more careful screening process before they decide to visit a property. Thus, the buyer’s value of search increases, and the visiting rate diminishes. It is not clear how improved information may affect the probability of agreement between a buyer and a seller given that the buyer has decided to visit a particular property. On the one hand, buyers visit only those properties with a relative high observed match value $b^0$; on the other hand, the value of their outside option (value of search) has increased. Therefore, the final shift of the sellers’ demand function $1 - \phi^*$ is uncertain. When buyers’ optimal strategies change, sellers’ optimal behavior is affected as well. It is straightforward to see that positive shifts in the demand function $1 - \phi^*$ makes sellers set higher posting prices and reservation values. However, we do not know the direction of the shift in demand and, thus, cannot make any theoretical predictions about the changes in optimal sellers’ behavior nor about the changes in equilibrium.

To answer our question of interest, we use numerical methods to solve our equilibrium model and perform comparative statics exercises. However, before we attempt to do these tasks, we need to estimate the model to have a reliable benchmark for our comparisons.
4 Data and reduced form analysis

4.1 Study Area and Data

The area of our study includes Charlottesville City and Albemarle County. These are two adjacent locations that are part of the Charlottesville, VA Metropolitan Statistical Area. The City of Charlottesville is located in Central Virginia, approximately 100 miles southwest of Washington, D.C. and 70 miles northwest of Richmond, Virginia. Albemarle County surrounds Charlottesville City, and its north border lies approximately 80 miles southwest of Washington, D.C. These areas occupy approximately 733 square miles (Charlottesville 10 and Albemarle 723). As one of the fastest growing areas in the state, the population increased by 16.7% between 1990 and 2003. According to the US Census, their combined population was 126,832 in 2003. In 2002, the total number of housing units in Charlottesville City and Albemarle County was 52,716, and, of those units, 58% were owner-occupied.

The Charlottesville and Albemarle Association of Realtors (CAAR) has provided us with Multiple Listing Services (MLS) data for all completed real estate transactions in Charlottesville City and Albemarle County during the years 2000 through 2002. The property data consist of 3,910 individual transaction records with information on posting prices, transaction prices, number of days in the market, and detailed property characteristics that include the home address.

To avoid biases in our analysis produced by outliers, we exclude from our database 160 observations corresponding to properties that were sold for less than $45,000 or more than $450,000. In addition, to be consistent with our theoretical model, we also exclude 58 transactions (1.5%) where the transaction price was above the posting price. Then, using the individual addresses, we were able to match 2,876 observations with the US Census Block Codes and construct a matched dataset with both housing and neighborhood characteristics.\footnote{To match our database with the US Census, we assigned a Census Block Code (CBC) to each of our records. However, in 816 cases, we were unable to link the reported addresses with the CBC. We dropped these unmatched observations from the sample.} We include five variables from the US Census that we believe are important to explain neighborhood desirability; these are population density, proportion of blacks, median age, household size, and household income. The first four variables were tabulated for each Cen-
The density has been evaluated at 100 equally-spaced points in the range of the data using a normal kernel function (with a bandwidth of 16.5).

sus Block while the variable “median household income” was obtained for each Census Block Group only. Descriptive statistics for this matched dataset are presented in Table 1.

Based on our 2,876 records, the average transaction price was $196,400, with a minimum of $50,000 and a maximum of $449,300. The posting price was, on average, $4,200 higher than the transaction price. Despite this fact, the distribution of both posting and transaction prices are quite similar. As shown in Figure 1, both distributions are unimodal and skewed to the right.

In this area, most homes sell relatively quickly. While the mean time that a home stays on the market is 43 days, twenty two percent of the properties sold in less than one week, and fifty percent sold in less than 26 days. On the other hand, a small number of homes (8.8%) stay on the market for more than four months. The density of the time that a home stays on the market is skewed to the right and unimodal (see Figure 2).

A typical home is about 24 years old, has 1,980 square feet, two bathrooms, and is located in a US Census block where 10% of its population is black. About 90% of these homes are equipped with air conditioning, while only 3% have a swimming pool.

Before analyzing the descriptive statistics of the neighborhood characteristics, notice that these statistics are weighted by the number of homes sold in each Census block and do
Table 1: Descriptive statistics of completed Real Estate transactions in Albemarle and Charlottesville, VA 2000 - 2002

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Median</th>
<th>Std.Dev.</th>
<th>Min.</th>
<th>Max.</th>
<th>Obs.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Posting price ($ thousands)</td>
<td>200.6</td>
<td>181.3</td>
<td>85.5</td>
<td>52.0</td>
<td>495.0</td>
<td>2876</td>
</tr>
<tr>
<td>Transaction price ($ thousands)</td>
<td>196.4</td>
<td>179.1</td>
<td>83.1</td>
<td>50.0</td>
<td>449.3</td>
<td>2876</td>
</tr>
<tr>
<td>Days in the market</td>
<td>43.3</td>
<td>26.0</td>
<td>46.1</td>
<td>1.0</td>
<td>199.0</td>
<td>2876</td>
</tr>
<tr>
<td><strong>Home characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Finished square footage</td>
<td>1980.2</td>
<td>1910.1</td>
<td>701.9</td>
<td>261.0</td>
<td>6500.0</td>
<td>2876</td>
</tr>
<tr>
<td>Number of full bathrooms</td>
<td>2.03</td>
<td>2.00</td>
<td>0.67</td>
<td>0</td>
<td>5.00</td>
<td>2876</td>
</tr>
<tr>
<td>Number of bedrooms</td>
<td>3.35</td>
<td>3.00</td>
<td>0.80</td>
<td>0</td>
<td>9.00</td>
<td>2876</td>
</tr>
<tr>
<td>Total acres</td>
<td>0.99</td>
<td>0.26</td>
<td>2.34</td>
<td>0</td>
<td>36.20</td>
<td>2876</td>
</tr>
<tr>
<td>Age of the property</td>
<td>23.75</td>
<td>17.00</td>
<td>23.45</td>
<td>0</td>
<td>251.0</td>
<td>2876</td>
</tr>
<tr>
<td>Air conditioning</td>
<td>0.90</td>
<td>1.00</td>
<td>0.31</td>
<td>0</td>
<td>1</td>
<td>2876</td>
</tr>
<tr>
<td>Basement</td>
<td>0.55</td>
<td>1.00</td>
<td>0.50</td>
<td>0</td>
<td>1</td>
<td>2876</td>
</tr>
<tr>
<td>Central heat</td>
<td>0.97</td>
<td>1.00</td>
<td>0.18</td>
<td>0</td>
<td>1</td>
<td>2876</td>
</tr>
<tr>
<td>Sewer system</td>
<td>0.75</td>
<td>1.00</td>
<td>0.43</td>
<td>0</td>
<td>1</td>
<td>2876</td>
</tr>
<tr>
<td>Home Owner Association</td>
<td>0.51</td>
<td>1.00</td>
<td>0.50</td>
<td>0</td>
<td>1</td>
<td>2876</td>
</tr>
<tr>
<td>Pool</td>
<td>0.03</td>
<td>0.00</td>
<td>0.16</td>
<td>0</td>
<td>1</td>
<td>2876</td>
</tr>
<tr>
<td>One story only</td>
<td>0.31</td>
<td>0.00</td>
<td>0.46</td>
<td>0</td>
<td>1</td>
<td>2876</td>
</tr>
<tr>
<td>Detached</td>
<td>0.79</td>
<td>1.00</td>
<td>0.41</td>
<td>0</td>
<td>1</td>
<td>2876</td>
</tr>
<tr>
<td><strong>Neighborhood characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log of population per square mile in census block</td>
<td>7.02</td>
<td>7.50</td>
<td>1.54</td>
<td>1.42</td>
<td>10.17</td>
<td>2876</td>
</tr>
<tr>
<td>Proportion of blacks living in the census block</td>
<td>0.10</td>
<td>0.05</td>
<td>0.14</td>
<td>0.00</td>
<td>1.00</td>
<td>2876</td>
</tr>
<tr>
<td>Median age in the census block</td>
<td>38.51</td>
<td>37.70</td>
<td>7.48</td>
<td>18.50</td>
<td>80.10</td>
<td>2876</td>
</tr>
<tr>
<td>Mean household size in the census block</td>
<td>2.52</td>
<td>2.57</td>
<td>0.43</td>
<td>1.00</td>
<td>4.50</td>
<td>2876</td>
</tr>
<tr>
<td>Median household income in census block group</td>
<td>55.5</td>
<td>53.9</td>
<td>16.8</td>
<td>15.6</td>
<td>116.7</td>
<td>2876</td>
</tr>
</tbody>
</table>
not necessarily represent an accurate description of the whole population of Charlottesville City and Albemarle County. Instead, they describe only those locations where real estate transactions were made. For example, the median household income in a Census block group where a home was sold was, on average, $55,500. On the other hand, the median household income of Charlottesville City and Albemarle County was $31,007 and $50,749, respectively. With these considerations, a representative home in our sample is located in a US Census block where the median age of its inhabitants is 38.5 years and the mean household size is 2.5.

Finally, there is significant dispersion in the characteristics of the neighborhoods. For example, while there are many areas with no blacks living in them, there are several US Census blocks populated by blacks only.

4.2 The determinants of posting prices, transaction prices, and time in the market

Before estimating the structural model, it is important to understand what are the relevant factors that determine posting prices, transaction prices, and the time until a property is sold. In this section, we use simple hedonic Ordinary Least Squares (OLS) models to identify
these factors, and later we utilize these results as a benchmark to evaluate the performance of the full structural model.

We let the posting price, the transaction price, and the log of time on the market be the dependent variables of three independent linear regression models: (A), (B) and (C), respectively. The explanatory variables include the property and neighborhood’s characteristics from our matched database. Notice that we have not included time variant independent variables; that is, we have not specified year nor month dummy variables. We omit these variables to be consistent with the stationary nature of our theoretical model.

The results of the OLS regressions of models (A) and (B) are presented in Table 2. All the coefficients from (A) and (B) have the same sign, and the $R^2$ of both regressions is roughly 0.70, which is a typical level for housing price models (Mason and Quigley 1996).

Notice that we have specified linear regression models to explain posting and transaction prices. Instead, we could have used a log-linear or other type of non-parametric specification (see, for example, Bin 2004). To test the robustness of the linear specifications (A) and (B), we have estimated log-linear pricing models and found that the predictive power of the latter models was slightly higher (the Mean Square Error in A and B decreased in 2% and 2.1%, respectively). Nevertheless, we have chosen to use the linear models because the interpretation of the coefficients in a linear pricing equation is similar to the interpretation of some parameters in our structural model. In particular, the coefficients $\gamma$ (in $s = X\gamma$) in our equilibrium model represent the marginal contribution of a home characteristic to the value of the home. The coefficients of a linear hedonic pricing equation have the same interpretation. These similarities facilitate the comparison between the structural and the reduced form models.

The coefficients’ estimates of models (A) and (B) suggest that one additional square foot increases the posting and transaction price by $68 and $65 respectively. A surprising finding is that, after conditioning on square footage, one fewer bedroom adds over $5,100 to the transaction price of a property. This fact suggests that agents prefer homes with larger bedrooms. Because we expect that the age of the property would affect its value in a nonlinear way, we include a quadratic and a cubic term in our specifications.$^{25}$ The estimates

---

$^{25}$Although the coefficient on the linear term is not statistically significant, we reject at the 1% significance
Table 2: Hedonic price OLS regression models

<table>
<thead>
<tr>
<th>Dependent variable ($ thousands)</th>
<th>Model (A)</th>
<th>Model (B)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Posting price</td>
<td>Transaction price</td>
</tr>
<tr>
<td>Constant</td>
<td>-15.12 (15.74)</td>
<td>-17.01 (15.03)</td>
</tr>
<tr>
<td>Finished square footage</td>
<td>0.068* (0.00)</td>
<td>0.065* (0.00)</td>
</tr>
<tr>
<td>Number of full bathrooms</td>
<td>5.20* (2.40)</td>
<td>4.36 (2.32)</td>
</tr>
<tr>
<td>Number of bedrooms</td>
<td>-5.74* (2.08)</td>
<td>-5.16* (2.05)</td>
</tr>
<tr>
<td>Log of acreage (zero if none)</td>
<td>7.25* (1.01)</td>
<td>6.71* (0.97)</td>
</tr>
<tr>
<td>Log of age of the property (zero if new)</td>
<td>-1.61 (6.33)</td>
<td>-5.49 (6.14)</td>
</tr>
<tr>
<td>Log of age of the property ^2</td>
<td>-10.89* (3.70)</td>
<td>-9.25* (3.57)</td>
</tr>
<tr>
<td>Log of age of the property ^3</td>
<td>2.19* (0.58)</td>
<td>1.96* (0.56)</td>
</tr>
<tr>
<td>Air conditioning</td>
<td>11.96* (4.10)</td>
<td>12.93* (4.00)</td>
</tr>
<tr>
<td>Basement</td>
<td>5.40* (2.21)</td>
<td>5.38* (2.16)</td>
</tr>
<tr>
<td>Central heat</td>
<td>12.63* (6.38)</td>
<td>13.50* (5.88)</td>
</tr>
<tr>
<td>Sewer system</td>
<td>-3.99 (3.72)</td>
<td>-4.13 (3.63)</td>
</tr>
<tr>
<td>Home Owner Association</td>
<td>14.54* (2.62)</td>
<td>15.02* (2.53)</td>
</tr>
<tr>
<td>Pool</td>
<td>15.64* (7.44)</td>
<td>15.05* (7.30)</td>
</tr>
<tr>
<td>One story only</td>
<td>-9.63* (2.41)</td>
<td>-9.64* (2.31)</td>
</tr>
<tr>
<td>Detached</td>
<td>42.18* (2.75)</td>
<td>41.88* (2.66)</td>
</tr>
<tr>
<td>Log of population per square mile in census block</td>
<td>3.06* (1.02)</td>
<td>3.45* (0.99)</td>
</tr>
<tr>
<td>Proportion of blacks living in the census block</td>
<td>-57.89* (7.47)</td>
<td>-57.76* (7.27)</td>
</tr>
<tr>
<td>Median age in the census block</td>
<td>1.00* (0.19)</td>
<td>0.97* (0.18)</td>
</tr>
<tr>
<td>Mean household size in the census block</td>
<td>-5.99* (2.90)</td>
<td>-4.92 (2.78)</td>
</tr>
<tr>
<td>Median household income in census block group</td>
<td>0.48* (0.08)</td>
<td>0.47* (0.08)</td>
</tr>
<tr>
<td>R²</td>
<td>0.705</td>
<td>0.707</td>
</tr>
<tr>
<td>Error term’s standard deviation</td>
<td>46.6</td>
<td>45.2</td>
</tr>
<tr>
<td>Number of observations</td>
<td>2876</td>
<td>2876</td>
</tr>
</tbody>
</table>

Standard errors are in parentheses. Asterisks indicate those parameters significant at the 5 percent significance level. The covariance matrix was calculated using the White-Heteroscedasticity-Consistent Method.
suggest that the transaction price of a new property declines rapidly at a decreasing rate for the first 40 years and slowly appreciates after that. For example, model (B) predicts that a ten year old property sells for $37,100 less than a new property, a twenty year old property sells for $9,000 less than the a ten year old home, and a 190 year old home has the same value as a new property. Buyers are willing to pay more for homes that are located in high-populated or high-income areas. As the mean age of the neighborhood increases, so does the price of the homes; the opposite is true for household size. It is difficult, however, to give a meaningful interpretation about the coefficients of these two variables because, most likely, they are capturing unobserved variation in income at the block level. Finally, there is a significant “premium” that individuals pay for living in a non-black neighborhood which averages more than $57,000.

To explain the determinants of the time that a house stays on the market (Time on the Market, TOM), we present in Table 3 the OLS estimates of a log-linear duration model (Model C). Model (C)’s $R^2$ is significantly lower than (A)’s or (B)’s, but it is consistent with other findings in the literature (see for example, Horowitz 1992). We also find that the size and age of the house, the population density of the block, and the share of blacks living in the neighborhood, are statistically significant predictors of the time that a house stays on the market. Bigger properties stay longer on the market, that is, a 10% increase in square footage implies a 6.1% increase in the expected TOM. New and old properties sell quicker than middle age ones. Finally, our results suggest that properties located in black neighborhoods stay 58% longer on the market than properties in non-black neighborhoods. The rest of the coefficients are not statistically significant.

A common approach in the literature when estimating TOM hedonic equations -as in Jansen and Jabson (1980), Kang and Gardner (1989) and Yavas and Yang (1995)- consists of including the percent difference between the actual posting price and the predicted posting price from an OLS model, as an explanatory variable in (C). These models conjecture that, after controlling for the property’s characteristics, higher than average posting prices should lead to longer TOM. However, if there are any unobserved (for the econometrician) housing characteristics that affect both the posting price and the TOM, there may be important level the joint null hypothesis that these three coefficients are equal to zero.
Table 3: Linear duration regression models

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>Model (C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log days in the market</td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-0.33</td>
</tr>
<tr>
<td>Log of finished square footage</td>
<td>0.61*</td>
</tr>
<tr>
<td>Number of full bathrooms</td>
<td>0.02</td>
</tr>
<tr>
<td>Number of bedrooms</td>
<td>-0.004</td>
</tr>
<tr>
<td>Log of acreage (zero if none)</td>
<td>-0.001</td>
</tr>
<tr>
<td>Log of age of the property (zero if new)</td>
<td>-1.38*</td>
</tr>
<tr>
<td>Log of age of the property $^2$</td>
<td>0.51*</td>
</tr>
<tr>
<td>Log of age of the property $^3$</td>
<td>-0.06*</td>
</tr>
<tr>
<td>Air conditioning</td>
<td>-0.17</td>
</tr>
<tr>
<td>Basement</td>
<td>-0.20</td>
</tr>
<tr>
<td>Central heat</td>
<td>-0.02</td>
</tr>
<tr>
<td>Sewer system</td>
<td>-0.15</td>
</tr>
<tr>
<td>Home Owner Association</td>
<td>-0.07</td>
</tr>
<tr>
<td>Pool</td>
<td>0.08</td>
</tr>
<tr>
<td>One story only</td>
<td>-0.003</td>
</tr>
<tr>
<td>Detached</td>
<td>0.05</td>
</tr>
<tr>
<td>Log of population per square mile in census block</td>
<td>-0.059*</td>
</tr>
<tr>
<td>Proportion of blacks living in the census block</td>
<td>0.58*</td>
</tr>
<tr>
<td>Median age in the census block</td>
<td>0.006</td>
</tr>
<tr>
<td>Mean household size in the census block</td>
<td>0.084</td>
</tr>
<tr>
<td>Median household income in census block group</td>
<td>0.000</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.133</td>
</tr>
<tr>
<td>Number of observations</td>
<td>2876</td>
</tr>
</tbody>
</table>

Standard errors are in parentheses. Asterisks indicate those parameters significant at the 5 percent significance level. The covariance matrix was calculated using the White-Heteroscedasticity-Consistent Method.
endogeneity biases with this approach. In fact, we expect that unobserved home features will be negatively correlated with TOM and positively correlated with the posting price, causing this coefficient to be biased downwards. Thus, we choose not to include this variable in our specification. The structural model we propose solves this problem by explicitly modeling and controlling for unobserved housing heterogeneity.

5 Estimation

5.1 The Likelihood Function

Before we estimate the equilibrium model specified in the previous section, we need to solve it. That is, given all the exogenous variables and parameters of the model, we shall find the function $\phi^*$ (or equivalently $q^*, \gamma_s^* \text{ and } \gamma_b^*$) such that both buyers and sellers have no incentives to deviate from their optimal endogenous strategies. First, we need to specify a functional form to the exogenous distributions $K_s(\xi_s)$, $K_s(\xi_b)$, $G_o(\sigma_o)$, and $G_u(\sigma_u)$, where $\xi_s$, $\xi_b$, $\sigma_o$, and $\sigma_u$ are parameter vectors that characterize these distributions. Then, for a given set of parameter values $\Theta = \{\gamma, \theta, c_b, \delta, \xi_s, \xi_b, \sigma_o, \sigma_u\}$ and exogenous housing characteristics $X = \{X_i\}_{i=1}^N$, we use numerical algorithms to solve our model according to the steps presented in the appendix.

The set of parameters of our structural model $\Theta$ can be estimated using Maximum Likelihood.\textsuperscript{26} As we described in the previous section, for each transaction $i$, we observe the posted price $p_{si}$, the transaction price $p_{mi}$, the number of days the property stays in the market $t_i$, and a set of home characteristics $X_i^o$. In this section, we use the information in our dataset to specify the relevant likelihood function.

It is important to recognize that certain features of the property, which are displayed in pictures and/or detailed comments in a Multiple Listing Service (MLS) ad, can be observed only by the agents in the market and not by the econometrician. Thus, when estimating the model, we need to control for unobserved housing characteristics. We model this unobserved

\textsuperscript{26}In what follows, we condition on a set of parameter values $\Theta$. But, for expository purposes, we omit it from our notation.
heterogeneity by letting \( s_i = X_i^\circ \gamma + u_i \) and assuming that \( u_i \) is an i.i.d. mean zero error with density \( f_u \). In addition, we assume that \( u \) is independent of the seller’s discount factor \( \beta_s \).

First, let \( d_i \) equal one if the posting price equals the transaction price, and zero otherwise, and let us derive the likelihood contribution of an observation where the transaction price was below the posting price. Notice that, given a posting and a transaction price, we can use the structure of the model (equations 4 and 3 and the optimal \( \phi^* \)) to calculate \( u_i \) and \( \beta_{si} \). That is, with information on \( \{p_{si}, p_{mi}, X_i^\circ\} \) we compute \( u_i \) and \( \beta_{si} \) as the unique values that simultaneously solve

\[
\begin{align*}
p_{mi} &= p_{si} \frac{1 - \phi^*(p_{si}|X_i^\circ \gamma + u_i)}{\phi^*(p_{si}|X_i^\circ \gamma + u_i)} \\
\beta_{si} &= \frac{p_{mi} - X_i^\circ \gamma - u_i}{p_{mi} - X_i^\circ \gamma - u_i + (p_{si} - p_{mi}) \theta [1 - \phi^*(p_{si}|X_i^\circ \gamma + u_i)]}.
\end{align*}
\]

For any given \( u_i \), we can determine the corresponding \( s_i = X_i^\circ \gamma + u_i \); and, using \( q_s^*, \gamma_s^*, \) and \( \gamma_b^* \), the well defined functions determined by the solution of the equilibrium model, we find the probability that trade occurs at a transaction price below the posting price

\[
l_m(p_{si}, p_{mi}, X_i^\circ, u_i) = (1 - \theta)q_s^*(p_{si}|s_i)\gamma_b^*(p_{mi}|s_i).
\]

In a similar way, notice that

\[
l_p(p_{si}, X_i^\circ, u_i) = \theta q_s^*(p_{si}|s_i)\gamma_b^*(p_{si}|s_i)
\]

is the probability that in a period a buyer visits a seller and purchases the property for the posting price. Thus, the unconditional probability that a property does not sell in one period is defined by

\[
l_o(p_{si}, p_{mi}, X_i^\circ, u_i) = 1 - l_p(p_{si}, X_i^\circ, u_i) - l_m(p_{si}, p_{mi}, X_i^\circ, u_i),
\]

and the probability of observing a property staying \( t_i \) periods in the market is

\[
l_{om}(t_i|p_{si}, p_{mi}, X_i^\circ, u_i) = l_o(p_{si}, p_{mi}, X_i^\circ, u_i)^{t_i - 1} l_m(p_{si}, p_{mi}, X_i^\circ, u_i).
\]

Using the previous definitions, we construct the likelihood contribution of an observation \( \{t_i, p_{si}, p_{mi}, X_i^\circ, d_i = 0\} \) as

\[
l_{tpm}(t_i, p_{si}, p_{mi}, X_i^\circ) = l_{om}(t_i|p_{si}, p_{mi}, X_i^\circ, u_i) k_s(\beta_{si}) f_u(u_i) |F_i|,
\]

26
where $k_s$ is the density of the seller’s $\beta_s$, and $|J_i|$ is the absolute value of the determinant of the Jacobian of the transformation implied by equations (12) and (13).

Now, let us focus on evaluating the likelihood contribution of an observation where the home sold for the posting price. Notice that, because we do not observe the seller’s reservation value, we cannot recover the value of the unobserved home characteristics $u_i$ as we previously did. However, if we condition on a particular value of $u$, we can use equation (12) to estimate the seller’s reservation value $p_m(u) = p_m(p_{si}^o, X_i^o, u)$, and compute the probability of observing a property staying $t_i$ periods in the market

$$l_{op}(t_i|p_{si}, p_{mi}, X_i^o, u) = l_o(p_{si}, p_m(u), X_i^o, u)^{t_i - 1}l_p(p_{si}, X_i^o, u).$$

Thus, the likelihood contribution of observation $\{t_i, p_{si}, p_{mi}, X_i^o, d_i = 1\}$ is

$$l_{tp}(t_i, p_{si}, p_{mi}, X_i^o) = \int l_{op}(u)k_s(\beta_s(u)) |J_i(u)| f_u(u|p_m(u) > X_i^o \gamma + u)du.$$

Note that we integrate the likelihood contributions for all values of $u$ that are consistent with the model. That is, only those values of $u$ that satisfy the condition that $p_m(u) > s$ are considered.

The log-likelihood contribution $L_i$ of observing $\{t_i, p_{si}, p_{mi}, X_i, d_i\}$ is

$$L_i(t_i, p_{si}, p_{mi}, X_i^o, d_i; \Theta) = (1 - d_i) \log(l_{tpm}) + d_i \log(l_{tp}),$$

and the MLE parameter estimates are the ones that maximize the log-likelihood of observing the sample

$$\Theta^{MLE} = \arg \max_{\Theta} \left\{ \sum_{i=1}^{n} L_i(p_{si}, p_{mi}, t_i, X_i^o, d_i; \Theta) \right\}. \quad (14)$$

### 5.2 Functional form assumptions and identification

Before we solve the model, we need to specify functional forms for all the relevant exogenous distributions. We let both $\beta_b$ and $\beta_s$ be a transformation of a normally distributed random variable. That is,

$$\log \left( \frac{\beta_j}{1 - \beta_j} \right) \sim N(u_{\beta_j}, \sigma_{\beta_j}^2); \quad j = b, s.$$
In addition, we let $b^o$, $b^u$, and $u$ be mean zero normal random variables with standard deviations $\sigma_{b^o}$, $\sigma_{b^u}$ and $\sigma_u$ respectively. Given these functional form assumptions and a set of parameters values, we use the numerical methods described in the previous section and in the appendix to solve the model.

To make estimation feasible, we make the additional identifying assumptions that the distributions of $\beta_b$ and $\beta_s$ are the same, that is $K_s = K_b = K$. We need to make this assumption because, in our sample, we do not observe buyer’s behavior, and thus, cannot infer anything about the buyer’s motivation to trade. This assumption is reasonable to the extent that both buyers and sellers in the real estate market are equally influenced, on average, by the same kind of factors that could affect their motivation to trade (for example, changes in life style, income, and employment). Thus, when making comparative statics analysis, we need to be aware of this normalization to provide a correct interpretation of our results.

With the previous assumption, we use MLE to estimate the model’s parameter vector

$$\Theta = \{\mu_{\beta}, \sigma_{\beta}, \sigma_{b^o}, \sigma_{b^u}, \sigma_u, \gamma, \delta, c_b, \theta\}.$$ 

It is important to provide some intuition about how we identify $\Theta$ in our likelihood function. Variation in prices and home characteristics identify the vector $\gamma$, in a similar manner as OLS does in a linear regression. Differences between posting prices and transaction prices and the structure of the model identify $\mu_{\beta}, \sigma_{\beta},$ and $\sigma_u$. The joint distribution of posting prices and time in the market and the observed agent’s behavior identify $\sigma_{b^o}, \sigma_{b^u}$ and $c_b$. Finally, the share of transactions where the transaction price was the posting price identifies $\theta$.

6 Results

6.1 Parameter estimates

In this section, we present and discuss the parameter estimates that result from the estimation procedure described in the previous one.
Table 4: Parameter estimates of the structural model: baseline parameters

<table>
<thead>
<tr>
<th>Variable</th>
<th>Baseline model’s parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta$</td>
<td>0.47* (0.009)</td>
</tr>
<tr>
<td>$\mu_\beta$</td>
<td>0.99967* (0.004)</td>
</tr>
<tr>
<td>$\sigma_\beta$</td>
<td>0.0053* (0.00012)</td>
</tr>
<tr>
<td>$c_b$</td>
<td>0.506* (0.0226)</td>
</tr>
<tr>
<td>$\delta$</td>
<td>1.08* (0.0035)</td>
</tr>
<tr>
<td>$\sigma_u$</td>
<td>45.694* (0.750)</td>
</tr>
<tr>
<td>$\sigma_{\tilde{b}}$</td>
<td>80.50* (0.029)</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>0.031* (0.00063)</td>
</tr>
</tbody>
</table>

Number of observations 2876

Standard errors are in parentheses. Asterisks indicate those parameters significant at the 5 percent significance level.

In Tables 4 and 5, we present the Maximum Likelihood estimates of the structural model parameters, and, in the rest of the section, we discuss the implications of these results. The estimate of $\theta$ implies that there is a 0.47 probability that a home sells for the posting price, and, because $\theta$ is close to one half, buyers and sellers have a similar level of market power.

Since we are measuring periods in calendar days, we expect the mean discount factor $\mu_\beta$ to be close to one. As we anticipated, the estimated daily mean discount factor is 0.9997, which implies a monthly mean discount factor of 0.989 and a yearly one of 0.88. These results suggest that, on average, agents value an opportunity to sell their home one year from today 12% less than having the same opportunity now.

The estimate of $\sigma_\beta$ is small. However, small variation in the agents’ discount-factors implies important differences in both buyer’s and seller’s optimal behavior. To illustrate this statement, we present in Figure 3 the seller’s optimal posting prices and reservation values for different values of $\beta_s$ (and given an estimated mean home quality $\bar{s}$ of $160,929$). The smallest (largest) discount factor we picked is 0.99963 (0.9999), which corresponds to the 0.2 th (99.8th) percentile of the distribution $K_s$. Within this interval, there exists significant variation in both posting prices and reservation values, implying that the transaction price of such a property could be anywhere between $170,000$ and $220,000$. As in Glower et al.(1998), we find that more motivated agents post cheaper prices and sell their homes at
lower prices.\textsuperscript{27}

Our results suggest that buyers spend $506 each time they visit and consider purchasing a real estate asset. This result is consistent with Shimizu et al.\textsuperscript{(2003)} who find that, in the Tokyo housing market, buyers spend about $1,700 (207,900 Japanese Yen) each time they visit and consider buying a home. Thus, this estimate provides further evidence that search costs in the housing market are significantly large and may be an important source of market inefficiency.

The visiting cost includes time, transportation, and other nonpecuniary (emotional) costs that are spent during the process of visiting, touring, and deciding to buy a home. The per-search time cost is the value of the time spent in one visit, which is the opportunity cost per hour times the time interval required for one search. We assume that the average time that a buyer’s household uses in one search adds up to 9 hours.\textsuperscript{28} Furthermore, as

\textsuperscript{27}Our results, however, are not fully comparable with those found by Glower et al (1998). While these authors assess the effect of five (observed) variables -related to the seller’s motivation to trade- in posting and transaction prices, we use a structural model to evaluate how equilibrium prices change as the (unobserved) discount rate varies.

\textsuperscript{28}Based on our interview with the Charlottesville and Albemarle Real Estate Association’s Director and on Shimizu et. al (2003), we postulate the following decomposition of the visiting time:
a) Travelling time: one hour.
b) Inspection time: two hours. (On average, buyers visit homes-for-sale in pairs (couples) for approximately one hour per visit).
in Shimizu et al. (2003), we let the buyer’s wage rate be her per-hour opportunity cost.  

And, assuming that the median visitor has an hourly income of $19.2 per hour, the related buyer’s time-opportunity costs add up to approximately $173. To estimate the buyer’s transportation costs, we have assumed that the average (one way) transportation time in the area is 20 minutes and the corresponding taxi-cab rate is approximately $20. Thus, the per-visit transportation cost is $40. Time and transportation costs add to $213 which is far lower than our visiting cost estimate of $506. This means that other non pecuniary (emotional) costs account for the majority of it (58%).

The estimate of \( \delta \) implies that buyers value homes 8% more than sellers. Considering that the average real estate agent’s commission fee in the area is 6%, we may argue that the agents capture the majority of the buyers’ and sellers’ gains from trade.

\( \sigma_u \) represents the dispersion of all unobserved factors (to the econometrician) that affect the value of a home. The errors’ standard deviation of the OLS hedonic models in Table 2 have a quite similar interpretation and, thus, we expect \( \sigma_u \) to be close to them. In fact, the estimate of \( \sigma_u \) is $45.7, and the differences with the OLS errors’ standard deviation are less than 2%.

The estimate of \( \sigma_{\tilde{b}} \), the standard deviation of the idiosyncratic buyer-home match value \( \tilde{b} \), is relatively large ($80.5). This means that there is significant heterogeneity in buyer’s tastes in the housing market.

Recall from the theoretical model that \( \tilde{b} \) has two independent components: a) \( b^o \), the match value learned when the buyer observes an online ad, and b) \( b^u \), the match value learned when the buyer visits and tours the home. To evaluate how much relevant information online ads provide to real estate customers, we estimate \( \lambda = \frac{\sigma^2_{b^o}}{\sigma^2_{\tilde{b}}} \), the ratio between the variance of \( b^o \) and the variance of \( \tilde{b} \). The estimate of \( \lambda \) is 0.03 suggesting that, on average, only 3% of the relevant information that buyers collect before making a purchase decision is obtained through online ads. To a certain extent, this is a surprising result. However, notice that

---

29Here, we implicitly assume that buyers choose between work and search. This is a standard practice but has been strongly criticized (Shaw 1992).

30To construct this estimate, we assume that the median household supplies an average of 60 hours per week. Then, we divide the area’s yearly median household income ($55,500) by the total number of hours per year worked by this median household.

---

31Post-search consideration period: six hours.
many of the current online advertising features, such as virtual tours for example, were not available during the first years of our sample.

As we stated in the introduction, the amount of information displayed in online ads has notably increased in the last five years, meaning that the value of $\lambda$ should have increased accordingly. To test this hypothesis, we divide our sample in four groups that correspond to the year when the transaction was made. Then, for each year, we estimate $\lambda$ while keeping constant the rest of the previously estimated parameters. The results of this exercise are displayed in Table 6 and provide evidence that the importance of online information raises with time.

The estimates of $\gamma$ (Table 7) have the same sign and are close in magnitude to the coefficients of the OLS price hedonic regressions. These results suggest that one additional square foot increases the value of a home by $69, while an extra bathroom raises it by $2,176. As was the case with the hedonic regressions, an additional bedroom reduces the value of a property, suggesting that sellers dislike small bedrooms. Homes that have a yard are more valuable, and, on average, a property that has one additional acre than the median is worth an extra $13,400. The coefficients on age suggest that the value of a new property declines rapidly at a decreasing rate for the first 31 years and slowly appreciates after that. In particular, the model predicts that a new property is $48,000, $58,000, and $0 more valuable than a ten, twenty and 220-year-old property, respectively. Having air conditioning and central heat adds approximately $8,600 and $19,000 to the value of a home, respectively.

All coefficients on the neighborhood variables have the expected signs. Agents value more homes that are located in high-populated or high-income areas. As the mean age of the people in the neighborhood increases, so does the intrinsic value of the home; the

<table>
<thead>
<tr>
<th>Year</th>
<th>$\lambda$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>0.026 (0.019)</td>
</tr>
<tr>
<td>2001</td>
<td>0.024* (0.004)</td>
</tr>
<tr>
<td>2002</td>
<td>0.053* (0.001)</td>
</tr>
</tbody>
</table>

Standard errors are in parentheses. Asterisks indicate those parameters significant at the 5 percent significance level.
Table 6: Parameter estimates of the structural model: seller’s home value

<table>
<thead>
<tr>
<th>Variable</th>
<th>Seller’s home value $s = X\gamma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-53.802* (12.4)</td>
</tr>
<tr>
<td>Finished square footage</td>
<td>0.069* (0.0013)</td>
</tr>
<tr>
<td>Number of full bathrooms</td>
<td>2.176 (1.63)</td>
</tr>
<tr>
<td>Number of bedrooms</td>
<td>-9.119* (1.18)</td>
</tr>
<tr>
<td>Log of acreage (zero if none)</td>
<td>8.492* (1.06)</td>
</tr>
<tr>
<td>Log of age of the property (zero if new)</td>
<td>-11.30* (4.34)</td>
</tr>
<tr>
<td>Log of age of the property $^2$</td>
<td>-8.735* (2.16)</td>
</tr>
<tr>
<td>Log of age of the property $^3$</td>
<td>2.006* (0.30)</td>
</tr>
<tr>
<td>Air conditioning</td>
<td>8.529* (3.05)</td>
</tr>
<tr>
<td>Basement</td>
<td>1.311 (1.99)</td>
</tr>
<tr>
<td>Central heat</td>
<td>18.969* (4.40)</td>
</tr>
<tr>
<td>Sewer system</td>
<td>-5.154 (3.49)</td>
</tr>
<tr>
<td>Home Owner Association</td>
<td>11.348* (2.70)</td>
</tr>
<tr>
<td>Pool</td>
<td>12.555* (4.95)</td>
</tr>
<tr>
<td>One story only</td>
<td>-10.371* (2.17)</td>
</tr>
<tr>
<td>Detached</td>
<td>36.876* (3.26)</td>
</tr>
<tr>
<td>Log of population per square mile in census block</td>
<td>3.318* (0.82)</td>
</tr>
<tr>
<td>Proportion of blacks living in the census block</td>
<td>-60.442* (7.46)</td>
</tr>
<tr>
<td>Median age in the census block</td>
<td>0.906* (0.12)</td>
</tr>
<tr>
<td>Mean household size in the census block</td>
<td>-13.707* (2.68)</td>
</tr>
<tr>
<td>Median household income in census block group</td>
<td>0.436* (0.06)</td>
</tr>
</tbody>
</table>

Number of observations 2876

Standard errors are in parentheses. Asterisks indicate those parameters significant at the 5 percent significance level.
opposite is true for household size. As was the case with the OLS models, it is difficult to
give a meaningful interpretation about the coefficients of these two variables because, most
likely, they are capturing unobserved variation in income at the block level. Finally, living
in a non-black neighborhood is worth $44,000 more than living in a black neighborhood.

6.2 Goodness of fit

6.2.1 Posting and transaction prices

Within our sample, we use the structural model and simulation methods to predict posting
and transaction prices. That is, given our parameter estimates, we solve the model using
the numerical algorithms detailed in the previous section. Then, for each observation \( i \) and
for many realizations of the random variables \( u \) and \( \beta \), we use equations 4 and 3 to predict
posting and transaction prices.

Overall the model fits the pricing data reasonably well. The mean posting price of our
final dataset was $200,600, and our model’s predictions are only 3.5%, or $7,200, higher.
Furthermore, the model predicts that the mean posting price is approximately $203,000,
which is only 3.8% higher than the mean transaction price of our sample data.

To have a measure of how well our model fits the observed data, we construct pseudo-
\( \hat{R}^2 \) indexes. This indexes are computed in a similar way as the \( R^2 \) of an OLS regression.
For example, let \( E[p_{si}^{*}|X_i] \) be the expected posting price predicted by our model of a property
with observed characteristics \( X_i \), and let \( p_{si} \) be the observed posting price. Then, the
corresponding pseudo-\( \hat{R}^2 \) index is defined as follows

\[
\hat{R}^2 = 1 - \frac{\sum_{i=1}^{n} (p_{si} - E[p_{si}^{*}|X_i])^2}{\sum_{i=1}^{n} (p_{si} - \bar{p}_{si})^2},
\]

where \( \bar{p}_{si} \) is the posting price’s sample mean, and \( n \) is the sample size.

For both the posting price and the transaction price, the pseudo-\( \hat{R}^2 \)s were about 0.69,
which is slightly below the \( R^2 \)s derived from the OLS models. Despite the variety of as-
sumptions and the complexity of our structural model, posting and transaction prices are
predicted almost with the same accuracy of a linear regression model.
In Figures 4 and 5 we present the densities of the actual versus the predicted posting and transaction prices, respectively. In both cases, the model displays a reasonably good representation of our sample data. This is not a feature of most empirical equilibrium search models in the literature (see for example Eckstein and Wolpin 1990).

It seems, however, that our model is unable to replicate the skewness of the actual pricing data. We may be predicting a distribution of prices that is too symmetric around the mean for one particular reason. We have assumed that all the exogenous distributions in our model are directly related to the normal distribution, and this assumption is imposing a certain degree of symmetry in the distribution of prices. In particular, changes in the distribution of $\beta_s$ and $\beta_b$ may cause significant variations in the shape of the pricing distributions. These normality assumptions were taken only for computational convenience and could be relaxed by assuming a more flexible functional form of the errors’ densities.

6.2.2 Time on the market

As in Horowitz (1992), our model does not perform well when predicting the duration data. Figure 6 compares the density of the duration that a house stays on the market predicted by
Figure 5: Nonparametric density of actual and predicted transaction prices.

Both densities have been evaluated at 100 equally-spaced points in the range of the data using a normal kernel function.

The model with the density of the observed data. To a certain extent, our model replicates the mode of our sample. However, it overpredicts both the median and the mean.31

There may be several sources of misspecification that may lead the model to inflate the average and median time on the market. First, our assumptions about the meeting technology imply that the number of days that a property stays on the market follows a geometric distribution. However, this current specification fails to describe the actual duration data. In particular, it inflates the right tail of the simulated distribution by overpredicting the number of properties that stay in the market for more than 200 days.

Second, notice that our model is stationary. That is, we assume that the seller’s posting and reservation prices do not change with time. This may not be a good assumption in the housing market because there may be, in fact, time dependent variables that affect posting and reservation prices. For example, a seller may lower her posting price as her moving date approaches or as the listing-contract period is about to expire (Glower et al. 1998). If posting and transaction prices decline with time, it may be less likely to observe properties staying for more than 200 days.

31The median and the mean of the observed duration data is 26 and 43 days, respectively. Our model overpredicts both of these variables. According to the model’s predictions, the median and the mean time that a property stays on the market is 79 days and 178 days, respectively. Furthermore, our estimate of the pseudo-$R^2$ is negative (-0.28), meaning that, on average, our model does even worst than the sample mean.
Figure 6: Nonparametric density of actual and predicted time on the market.

Another possible source of misspecification may be in the way we specify the meeting technology. Recall that our meeting technology is random and exogenous and does not include other important elements of buyer’s search behavior. For example, the buyer’s searching intensity should depend upon her motivation to trade, and our model assumes that the rate at which buyers observe ads online is the same for every buyer and that it is exogenous. Furthermore, we assume that buyers search for ads sequentially; that is, they observe one new ad each period (calendar day). This may be a weak assumption since it is more plausible that buyers observe more than one new ad every day.

One could easily extend the model and allow buyers to receive multiple ads in a period using a Poisson-type ad viewing rate. We did not follow this approach for identification issues, since we do not observe data on the buyer’s search intensity. In another more complicated and more computationally expensive approach, one could also model the buyer’s decision of how many ads to pick endogenously (using an extension of Stern’s 1989 model). However, the model becomes intractable and highly expensive to compute.

Finally, notice that the assumptions about the stationary environment and the nature of the meeting technology in our model imply that there is no duration dependence.\textsuperscript{32} However,
Table 7: Effects of an increase in the amount of online information

<table>
<thead>
<tr>
<th>Variable</th>
<th>Scenario 1 (λ = 0.03)</th>
<th>Scenario 2 (λ = 0.50)</th>
<th>Scenario 3 (λ = 1.00)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean posting price ($ thousands)</td>
<td>207.8</td>
<td>204.5</td>
<td>201.8</td>
</tr>
<tr>
<td>Mean transaction price ($ thousands)</td>
<td>204.0</td>
<td>199.1</td>
<td>197.6</td>
</tr>
<tr>
<td>Mean time on the market (days)</td>
<td>67.2</td>
<td>69.2</td>
<td>71.3</td>
</tr>
</tbody>
</table>

this may not be a suitable assumption for the housing market.

There may be several ways to add duration dependence into our model. For example, one may assume that the individual’s discount rate varies with time. In this case, posting and reservation prices change each period and the probability of trade becomes time dependent. We did not use this approach for identification issues, because we do not observe changes in posting prices in our sample. Another more complicated way to generate duration dependence is to assume that the motivation to trade of new buyers and sellers into the market are realizations from a stationary distribution. This extension is not trivial and will be explored in the future.

6.3 Counterfactual experiment

One of the primary benefits of estimating a structural model is that it may be used to conduct counterfactual experiments. In this section, we use the estimated model to evaluate how equilibrium market outcomes change when the amount of information displayed in online ads increases.

Our experiment raises the amount of the relevant information that buyers obtain when looking at an online ad, and it can be described in terms of one parameter change. That is, we let $\lambda$ increase from 0.03 to 0.50 and to 1.00, and evaluate the effect of these changes on posting prices, transaction prices, and time on the market. The results of this exercise are displayed in Table 8.

Our results suggest that additional online information decreases market prices. For instance, if buyers could obtain from an online ad all the relevant information they need to make a purchase decision ($\lambda = 1$), posting and transaction prices would decrease by 2.9% of trading in a period is constant across time. Second, it is easy to see that the hazard rate of a random variable with a geometric distribution is time independent.
and 3.1%, respectively. These results provide evidence that the internet increases market competition and are consistent with other findings in the literature.33

We also find evidence that, as more information is displayed in real estate online ads, the time that a property stays on the market increases. In particular, when \( \lambda \) equals one, the average time on the market increases in four days compared to our benchmark case (\( \lambda = 0.03 \)). As online information increases, buyers have more options to choose from and take more time to make a final purchase decision. This result is consistent with that found by D’urso (2002).

Notice that we have assumed that the other parameter estimates are invariant to changes in \( \lambda \). This may be a weak assumption, since changes in the information technology take place in a relative large amount of time, and it is unlikely that the buyers’ and sellers’ search behavior remains constant during this large period. Thus, one should be aware of these limitations and interpret the results of the counterfactual experiment with caution.

7 Conclusions

In this research we have conducted an empirical examination of the role of online information in the real estate market. This work contributes to the existing literature in several ways. We specified and estimated the first two-sided equilibrium search model with posting prices that explicitly incorporates the availability of an online advertising technology. Unlike previous studies, the model specified in this dissertation explains posting prices, transaction prices, and time on the market in an integrated context and using an equilibrium approach. Furthermore, the estimation techniques add to the growing literature on estimation of equilibrium search models.

By modeling the decision rules of buyers and sellers, and assuming that the parameter estimates are time-invariant, we have used our model to assess the effects of improvements in the amount of information displayed in online ads on market outcomes. Thus, our findings may be of great interest to the literature that seeks to explain how the use of the internet has influenced online and offline markets.

33See, for example, Brown and Goolsbee (2002).
Parameter estimates have a reasonable interpretation, and the model is able to replicate the pricing data remarkably well. In particular, the parameter estimates suggest that, on average, only 3% of the relevant information that home-buyers collect before making a purchase decision is obtained through on-line ads. In addition, through the counterfactual experiment, we may conclude that additional online information decreases posting and transaction prices; on the other hand, the average time that a home stays on the market would increase. Despite the variety of assumptions and the complexity of our structural model, posting and transaction prices are predicted almost with the same accuracy of a linear regression model.

We acknowledge that our theoretical model has several limitations and suggest several avenues for future research in order to overcome them.
A Theorem’s proofs

Proof of Theorem 1: Substituting (4) in (3) and rearranging, we obtain the following expression:

\[
h(p^*_s | s)(p^*_s - s) = \frac{1 - \beta_s [1 - \theta(1 - \phi(p^*_s | s))]}{1 - \beta_s}.
\]  

(16)

Furthermore, assume that the hazard function is non-decreasing in \( p^*_s \). Note that the left hand side of equation (16) evaluated at \( p^*_s = s \) is zero, and that it is increasing in \( p^*_s \); the right hand side of equation (16) evaluated at \( p^*_s = s \) is a positive value greater or equal than one, and it is decreasing in \( p^*_s \). Thus, there exists a unique solution for the posting price. Replacing \( p^*_s \) in (4) we solve for the optimal reservation value \( R^*_s \), and, since the hazard function is nonnegative, it must be the case that \( p^*_s \geq R^*_s \) (for any value of \( \beta_s \) and \( s \)). Using equation (16) and the implicit function theorem, we take the derivative of \( p^*_s \) w.r.t. \( \beta_s \) and obtain

\[
\frac{\partial p^*_s}{\partial \beta_s} = \frac{\frac{\partial h}{\partial p^*_s} + \frac{\beta_s}{1 - \beta_s} \theta \frac{\partial \phi}{\partial p^*_s}}{(p^*_s - s) \frac{\partial h}{\partial p^*_s} + h + \frac{\beta_s}{1 - \beta_s} \theta \frac{\partial \phi}{\partial p^*_s}} > 0.
\]

It is also easy to see from equation (4) that \( \frac{\partial R^*_s}{\partial \beta_s} = 1 + \frac{\partial h}{\partial p^*_s} > 0 \). Thus, \( \frac{\partial R^*_s}{\partial \beta_s} = \frac{\partial R^*_s}{\partial p^*_s} \frac{\partial p^*_s}{\partial \beta_s} > 0 \).

We let \( \beta_s = 0 \) in equation (3) and find that \( R^*_s = s \). Because \( 0 \leq \beta_s \leq 1 \) and \( \frac{\partial R^*_s}{\partial \beta_s} \geq 0 \), it must be the case that \( p^*_s \geq R^*_s \geq s \).

To prove Theorem 1, we still need to show that \( p^*_s \) and \( R^*_s \) are increasing in \( s \). To show this, we use equation (16) and the implicit function theorem to find

\[
\frac{\partial p^*_s}{\partial s} = \frac{h - \left( \frac{\partial h}{\partial p^*_s} (p - s) + \frac{\partial \phi}{\partial p^*_s} \right)}{(p^*_s - s) \frac{\partial h}{\partial p^*_s} + h + \frac{\beta_s}{1 - \beta_s} \theta \frac{\partial \phi}{\partial p^*_s}} > 0.
\]

Notice that the expression in parenthesis in the numerator is always negative, because both \( \frac{\partial \phi}{\partial s} \) and \( \frac{\partial h}{\partial s} \) are negative. Furthermore, \( \frac{\partial R^*_s}{\partial s} = \frac{\partial R^*_s}{\partial p^*_s} \frac{\partial p^*_s}{\partial s} > 0 \) QED.

Proof of Theorem 2

To prove Theorem 2, we need to show the details of integration of equation (6), and the monotonicity of \( W^*_b \) and \( p^*_b \).

1) Details of the integration of equation (6)
Let \( v_p = b^o + \delta s - p_o^s \) and integrate by parts the first term of (6)

\[
E[\max\{b - p_o^s, W_b\}|p_o^s, s, b^o] = \int \max\{b^u + b^o + \delta s - p_o^s, W_b\} dG_u(b^u)
\]

\[
= \int_{W_b-v_p} (b^u + v_p) dG_u(b^u) + W_bG_u(W_b - v_p)
\]

\[
= - [(b^u + v_p)(1 - G_u(b^u))]_{W_b-v_p}^{\infty}
\]

\[
+ \int_{W_b-v_p} [1 - G_u(b^u)] db^u + W_bG_u(W_b - v_p)
\]

\[
= W_b[1 - G_u(W_b - v_p)]
\]

\[
+ \int_{W_b-v_p} [1 - G_u(b^u)] db^u + W_bG_u(W_b - v_p)
\]

\[
= W_b + \int_{W_b+p_o^s-\delta s-b^o} [1 - G_u(b^u)] db^u.
\]

Using the same procedure, the second term of \( U^e(p_o^s, s, b^o, W_b) \) becomes

\[
E[\max\{b + \delta s - R_o^s(p_o^s, s), W_b\}|p_o^s, s] = W_b + \int_{W_b+R_o^s(p_o^s, s)-\delta s-b^o} [1 - G_u(b^u)] db^u.
\]

2) Monotonicity of \( W_b^* \) and \( p_o^s W_b \)

Define \( g(W_b^*) = \int \int \max\{j(p_o^s, s, b^o, W_b^*), 0\} d\Gamma(p_o^s, s) dG_o(b^o) \) and rewrite equation (8) as \( W_b^* = \frac{\beta}{1-\beta} g(W_b^*) \). Clearly, \( g(W_b^*) \) is a non-negative decreasing function. Then, we use the implicit function theorem to show that

\[
\frac{\partial W_b^*}{\partial \beta_b} = \frac{g(W_b^*)}{1-\beta_b} [g(W_b^*)] + 1 > 0.
\]

Furthermore, we use equation (9) and Leibnitz rule to show that

\[
\frac{\partial p_o^{s*}}{\partial W_b^*} = - \frac{\theta(1 - G_u(V_p)) + (1 - \theta)(1 - G_u(V_r))}{\theta(1 - G_u(V_p)) + (1 - \theta) \frac{\partial R_o^s}{\partial p_o^{s*}}(1 - G_u(V_r))} < 0,
\]

where \( V_p = W_b + p_o^{s*} - \delta s - b^o \), and \( V_r = W_b + R_o^s(p_o^{s*} | s) - \delta s - b^o \). Thus, \( \frac{\partial p_o^{s*}}{\partial \beta_b} = \frac{\partial p_o^{s*}}{\partial W_b^*} \frac{\partial W_b^*}{\partial \beta_b} < 0 \).

Steps one and two complete the proof. QED.

Proof of Theorem 2a:

42
First, note that because \( s \) is non random and \( b^o = 0 \), equation (5) becomes

\[
W_b = \beta_b \int \max \{U^c(p^*_s), W_b \} \, d\Gamma(p^*_s)
\]

\[
\frac{W_b}{\beta_b} = \int_{p^*_s: U^c(p^*_s) > W_b} U^c(p^*_s) d\Gamma(p^*_s) + W_b \int_{p^*_s: U^c(p^*_s) < W_b} d\Gamma(p^*_s)
\]

\[
= \int_{p^*_s: U^c(p^*_s) < U^{c-1}(W_b)} U^c(p^*_s) d\Gamma(p^*_s) + W_b \int_{p^*_s: U^c(p^*_s) > U^{c-1}(W_b)} d\Gamma(p^*_s)
\]

where \( U^{c-1}(\cdot) \) is the inverse image of \( U^c(\cdot) \), which is well defined since we have already shown that \( U^c \) is monotone. Let us integrate by parts and rearrange to obtain

\[
\frac{W_b}{\beta_b} = \int_{U^{c-1}(W_b)} U^c(p^*_s) d\Gamma(p^*_s) + W_b \int_{U^{c-1}(W_b)} d\Gamma(p^*_s)
\]

\[
= \int_{U^{c-1}(W_b)} [\Gamma(p^*_s) \{U^c(p^*_s)\}]_{\infty}^{U^{c-1}(W_b)} - \int_{U^{c-1}(W_b)} \Gamma(p^*_s) \frac{\partial U^c(p^*_s)}{\partial p_s} dp_s + W_b [1 - \Gamma(U^{c-1}(W_b))]
\]

\[
= W_b \Gamma(U^{c-1}(W_b)) + \int_{U^{c-1}(W_b)} \Gamma(p^*_s) \left(\frac{\partial U^c(p^*_s)}{\partial p_s}\right) dp_s + W_b [1 - \Gamma(U^{c-1}(W_b))]
\]

\[
= W_b + \int_{U^{c-1}(W_b)} \Gamma(p^*_s) \left(\frac{\partial U^c(p^*_s)}{\partial p_s}\right) dp_s.
\]

It is optimal for buyers to set a reservation strategy such that \( U^c(p^*_b) = W_b^* \). Then, we replace this optimality condition in the former equation and obtain

\[
\frac{\beta_b}{1 - \delta_b} \int_{p^*_b = p^*_s - \delta s} p^*_s \Gamma(p^*_s) \left(\frac{\partial U^c(p^*_s)}{\partial p_s}\right) dp_s = W_b^*.
\]

Finally, we let \( U^c(p^*_b) = W_b^* \) in equation (7) to find

\[
\theta \int_{W_b^* + p^*_s - \delta s} [1 - G_a(b^n)] \, db^n + (1 - \theta) \int_{W_b^* + \tilde{R}_s(p^*_s) - \delta s} [1 - G_u(b^n)] \, db^n = c_b.
\]

Thus, the solution to the search problem is equivalent to finding the pair \( \{W_b, p^*_b\} \) that solves the previous equation system. QED.

**Proof that** \( \phi^*(p_s|s) \) **is well defined and decreasing.**
The function \( q(p_s|s, b^o) \) is well defined by the exogenous distribution \( K_b \) and the optimal buyers reservation price function \( p_s^* \).

\[
q(p_s|s, b^o) = \Pr\{p_b^*(\beta_b, s, b^o) > p_s|s, b^o\} = K_b(p_b^{*s-1}(p_s|s, b^o))
\]

where \( p_b^{*s-1}(p_s) \) is the function that determines \( \beta_b \) for any value \( p_s \) (which exists because we have shown that \( p_b^*(\beta_b) \) is monotone). Then

\[
\gamma_s(p_s|s, b^o) = \frac{\Pr\{b^o + b^n + \delta s - p_s > W_b^s(\beta_b)|p_b^{*s}(\beta_b, s, b^o) > p_s\}}{K_b(p_b^{*s-1}(p_s|s, b^o))}
\]

\[
\gamma_b(p_s|s, b^o) = \frac{\int (1 - G_u[W_b^s(\beta_b) + p_s - \delta s - b^o]) dK_b(\beta_b)}{K_b(p_b^{*s-1}(p_s|s, b^o))},
\]

and

\[
\phi^*(p_s) = 1 - \int q(p_s|s, b^o) \ast \gamma_s(p_s|s, b^o) dG_o(b^o)
\]

\[
= 1 - \int \int (1 - G_u[W_b^s(\beta_b) + p_s - \delta s]) dK_b(\beta_b)dG_o(b^o) \quad (17)
\]

which is clearly non-increasing in \( p_s \). QED.

**Heuristic proof about the existence of an equilibrium**

What follows should not be considered a formal proof of existence. However, it sketches the steps needed to construct a more rigorous and technical proof.

**Step 1:** Let us construct a compact convex Banach space spanned by our functions \( q, \gamma_s, \gamma_b \).

Define \( W_1, W_2, W_3 \), as the following space of distributions:

\[
W_1=\{q(\beta_s|s) : q \text{ is continuous }, 0 \leq q \leq 1, 0 \leq \beta_s \leq 1, -\infty < s < \infty \},
\]

\[
W_2=\{\gamma_s(\beta_s|s) : \gamma_s \text{ is continuous }, 0 \leq \gamma_s \leq 1, 0 \leq \beta_s \leq 1, -\infty < s < \infty \},
\]

\[
W_3=\{\gamma_b(\beta_s|s) : \gamma_b \text{ is continuous }, 0 \leq \gamma_b \leq 1, 0 \leq \beta_b \leq 1, -\infty < s < \infty \}.
\]

Under an appropriate norm, it can be shown that \( W_1, W_2, \) and \( W_3 \) are a Banach space. Furthermore, we conjecture (without proof) that \( W_1, W_2, \) and \( W_3 \) are convex and compact.
If this is the case, the space $W = W_1 \times W_2 \times W_3$ spanned by the set of functions $\Phi = \{q, \gamma_s, \gamma_b\}$ is a compact convex Banach space.

**Step 2:** Let us define a continuous mapping of $W$ into itself.

Define $\Omega_{\Phi P} : \Phi \rightarrow P$, as the function that maps the set of distributions $\Phi$ into the set of optimal seller’s pricing strategies $P = \{p^*_s, R^*_s\}$; $P$ is well defined and continuous as shown in Theorem 1. Define $\Omega_{PB} : P \rightarrow B$, as the function that maps $P$ into the set of optimal buyer’s strategies $B = \{W^*_s, p^*_b\}$, which are well defined and continuous as shown in Theorem 2. Define $\Omega_{B\Phi} : B \rightarrow \Phi$, as the function that maps $B$ into $\Phi$. This relationship is well defined and continuous as shown in this appendix.

Define: $\Omega : W \rightarrow W$ as $\Omega(\Phi) = \Omega_{PB}(\Omega_{PB}(\Phi))$.

Since $\Omega_{B\Phi}, \Omega_{PB}, \Omega_{\Phi P}$ are continuous, so is $\Omega$. The function $\Omega$ maps into itself. The existence of an equilibrium follows from Shauder’s fixed-point theorem.
B Sketch of the numerical methods used to solve the empirical model

<table>
<thead>
<tr>
<th>Step</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1)</td>
<td>Using $X$ and $\gamma$, determine $\Psi$ (the distribution of $s$).</td>
</tr>
<tr>
<td>2)</td>
<td>Create a grid of size $n_2$ for values of $s$ in equally spaced points that depends on the distribution $\Psi$.</td>
</tr>
<tr>
<td>3)</td>
<td>Create grids of size $n_1$ for $\beta_s$ and $\beta_b$ of not necessarily equally spaced points.</td>
</tr>
<tr>
<td>4)</td>
<td>Specify an initial guess for the function $\phi(p_s</td>
</tr>
<tr>
<td>5)</td>
<td>For each combination of $\beta_s$ and $s$, use equations (3), (4) and the array $\phi(p_s</td>
</tr>
<tr>
<td>6)</td>
<td>Use equation (8) to solve for $W_b^*(\beta_b)$, and store it in a $(n_1 \times 1)$ array.</td>
</tr>
<tr>
<td>7)</td>
<td>Use equation (9) and $W_b^<em>(\beta_b)$ to find $p_b^</em>(\beta_b, s</td>
</tr>
<tr>
<td>8)</td>
<td>Given $p_b^*(\beta_b, s</td>
</tr>
<tr>
<td>9)</td>
<td>Using a discrete approximation, evaluate $\phi^*(p_s</td>
</tr>
<tr>
<td>10)</td>
<td>Stop if $\phi^*(p_s</td>
</tr>
</tbody>
</table>
References


50

