JOURNAL OF APPLIED ECONOMETRICS, VOL. 7, 115–129 (1992)

THE ROLE OF THE LIST PRICE IN HOUSING MARKETS: THEORY AND AN ECONOMETRIC MODEL

JOEL L. HOROWITZ

Department of Economics, University of Iowa, Iowa City, Iowa 52242, USA

SUMMARY

Houses are routinely sold at prices below, but rarely sold at prices above, their list prices. List prices appear to be price ceilings that preclude the possibility of sales at higher prices. This paper presents a theory of sellers' behaviour that explains why there are list prices in housing markets and why list prices are distinct from sellers' reservation prices. The theory forms the basis of an econometric model that has been estimated using data from the Baltimore, MD, area. The estimated model predicts sale and reservation prices conditional on list prices. The predictions of sale prices are considerably more accurate than those obtained from a standard hedonic price regression. The estimated model also explains why sellers may not be willing to reduce their list prices even after their houses have remained unsold for long periods of time.

1. INTRODUCTION

Houses are routinely sold at prices below, and are virtually never sold at prices above, their list prices. The list price of a house appears to be a price ceiling that precludes the possibility of selling at a higher price. This phenomenon is stable over time, so it is not caused by market disequilibrium.

What is the role of list prices in housing markets? Can list prices be used to predict sellers' reservation prices? Can they be used to make predictions of sale prices that are more accurate than predictions obtained from standard hedonic regressions? This paper develops an econometric model of the behaviour of sellers in housing markets that answers these questions. The model explains why there are list prices in housing markets and how sellers establish list and reservation prices. In addition, the model predicts sale and reservation prices conditional on list prices. The predictions of sale prices are considerably more accurate than those obtained from a hedonic price regression. The model also explains why sellers may not be willing to reduce their list prices even after their houses have remained unsold for long periods of time.

There have been several previous attempts to model the formation of list prices in housing markets. Guasch and Marshall (1985) and Stull (1978) developed search models in which list prices must be set, but these models do not admit the possibility that sale prices may be below list prices. Chinloy (1980) developed a model in which the seller's reservation price is assumed to be a constant fraction of the list price. But Chinloy did not explain why list and reservation prices are distinct or why they are proportional to one another. These difficulties are not present in the model developed in this paper.

The theory underlying the model is developed in Section 2. The model is presented in Section 3. Estimation results and some applications are given in Section 4. Additional applications are given in Section 5, and concluding comments are given in Section 6.

0883–7252/92/020115–15$07.50

Received July 1990
Revised May 1991

2. A THEORY OF SELLERS' BEHAVIOUR

The list price of a house, p_L, is an announced price at which the seller will trade. The reservation price, p_R, is the lowest price the seller will accept. It is not announced, and usually is known only to the seller. By definition, $p_L \geqslant p_R$, so p_L conveys information to buyers in the form of an upper bound on p_R. The goal of the theory developed here is to explain how a seller chooses p_L and p_R.

Assume that the seller's objective is to maximize the expected utility from the sale of his house. Let $V^*(p_L, p_R)$ be expected utility when the list price is p_L and the reservation price is p_R. Then the seller chooses p_L and p_R to maximize V^*. I want to derive an expression for $V^*(p_L, p_R)$ and characterize the optimal values of p_L and p_R.

A seller cannot predict with certainty whether potential buyers will make bids for his house on a given day, or what the sizes of any bids will be. Thus, from the point of view of a seller, bids are generated by a random process. I assume that:

A1. The bids are generated one at a time by independent random sampling from a probability distribution (the bid distribution). This distribution is sampled once per day, and a bid of zero, which is equivalent to no bid at all, may have positive probability. Thus, on any given day the seller receives at most one bid and may receive none. Negative bids cannot occur because a seller with a negative reservation price would abandon his house rather than place it on the market. The bid distribution depends on the list price but does not vary over time if the list price remains constant. Each bid must be accepted or rejected without knowledge of either the sizes or occurrence times of subsequent bids. Acceptance of a bid terminates the bid generation process.

A2. The bids for a house never exceed its list price.

A3. Buyers do not know sellers' reservation prices with certainty.

Assumptions A1 and A2 are idealizations of observed market operation. In real housing markets a seller sometimes receives several bids simultaneously or receives a bid slightly above his list price, and sellers may bargain with buyers. Nonetheless, A1 and A2 provide a useful starting point for achieving an improved understanding of the operation of housing markets. Assumption A3 is motivated by the observation that sellers do, in fact, reject bids.

I assume that the seller knows the bid distribution, including its dependence on the list price. Of course, real sellers are unlikely to be so well informed. The theory is unchanged if the true bid distribution is replaced by a perceived distribution that satisfies the same assumptions as the true one.

Define $Q(p \mid p_L)$ to be the probability per day that the seller of a given house receives a bid of p or more if his list price is p_L. Define $F(p \mid p_L)$ to be $1 - Q(p \mid p_L)$. Let $F_p = \partial F/\partial p$ and $F_L = \partial F/\partial p_L$. I assume that

$$F(p \mid p_L) = \begin{matrix} \alpha + (1-\alpha)\exp(z/\eta) & \text{if } z \leqslant 0 \\ 1 \quad \text{if } z > 0, & \end{matrix} \tag{1}$$

where

$$z = p + \gamma p_L + \zeta, \tag{2}$$

and α, η, γ, and ζ are constants. (In Section 3, ζ will be made a random variable and F will be conditioned on ζ.) Since F is a probability distribution function on p, it must be non-decreasing and its values must be between 0 and 1. These conditions are satisfied if $\eta > 0$ and $0 \leqslant \alpha \leqslant 1$. This specification of F is chosen mainly for its tractability in the econometric model. In addition, the specification has the reasonable properties that

1. The probability per day that the seller receives no bid can have any value between 0 and 1.
2. If $\gamma > 0$, increasing p_L decreases $Q(p_L \mid p_L)$. Thus, increasing p_L discourages buyers, other things being equal.

Assumptions A1–A3 and equations (1) and (2) establish the role of the list price, p_L, in the model developed here: p_L is a parameter of the distribution from which the bids for a house are sampled, and this parameter is controlled by the seller. The bid distribution depends on p_L because p_L places an upper bound on the seller's reservation price and, therefore, influences buyers' estimates of whether a bid of a given size is likely to be accepted.

(1) The Expected Utility Function

The derivation of $V^*(p_L, p_R)$ is similar to the derivation of the expected utility function in a standard search model. The resulting utility maximization problem is different, however, owing to the presence of p_L.

Let w denote the alternative opportunity value of the house. That is, if the house could not be put on the market, the seller would be indifferent between trading the house for a payment of w and keeping the house. In general w and p_R are not the same since the minimum price the seller will accept when he has the option of putting his house on the market may exceed w.

Assume that w is independent of time and that the seller is risk-neutral and has an infinite time horizon.[1] Then the increase in the present value of the seller's utility if he sells his house for price p at time t is $V(p, t) = (p - w)/(1 + r)^t$, where $r > 0$ is the discount rate. Since bids occur one at a time and never exceed list prices, the expected value of $V(p, t)$ conditional on p_R, p_L and t is

$$E(V \mid p_R, p_L, t) = (1 + r)^{-t}\left[\int_{p_R}^{p_L} (p - w)F_p(p \mid p_L)\,\mathrm{d}p + (p_L - w)Q(p_L \mid p_L)\right]. \quad (3)$$

$V^*(p_L, p_R)$ is the expected discounted present value of the seller's utility:

$$V^*(p_R, p_L) = \sum_{t=1}^{\infty} E(V \mid p_R, p_L, t)[F(p_R \mid p_L)]^{t-1}. \quad (4)$$

Substitution of (3) into (4) yields

$$V^*(p_R, p_L) = [r + Q(p_R \mid p_L)]^{-1}\left[\int_{p_R}^{p_L} (p - w)F_p(p \mid p_L)\,\mathrm{d}p + (p_L - w)Q(p_L \mid p_L)\right]. \quad (5)$$

Integration by parts on the right-hand side of (5) yields

$$V^*(p_R, p_L) = [r + Q(p_R \mid p_L)]^{-1}\left[(p_L - w) - (p_R - w)F(p_R \mid p_L) - \int_{p_R}^{p_L} F(p \mid p_L)\,\mathrm{d}p\right]. \quad (6)$$

[1] Events that might reasonably place a finite bound on a seller's time horizon, such as the end of his or his house's life, are likely to be sufficiently far in the future to make the assumption of an infinite horizon a satisfactory approximation.

(b) Properties of the Optimal List and Reservation Prices

The seller chooses p_L and p_R to maximize V^* in (6). Let p_L^* and p_R^* be the resulting values of p_L and p_R. It is shown in Appendix A that if $\gamma > 0$

$$w \leqslant p_R^* < p_L^* < \infty \tag{7}$$

and

$$p_R^* = p_L^* + \eta \log \frac{\gamma + 1}{\gamma} + \eta \log\left\{1 - (1+\gamma)^{-1} \exp\left[-\frac{(1+\gamma)p_L + \zeta}{\eta}\right]\right\}. \tag{8}$$

p_L^* and p_R^* do not have closed-form analytic expressions.

(c) Why there are List Prices in Housing Markets

The list price conveys information to buyers in the form of an upper bound on the seller's reservation price. An infinite list price conveys no such information and, therefore, is equivalent to no list price at all. Since, by (7), a seller maximizes expected utility by establishing a *finite* list price, establishing a finite list price dominates establishing no list price.

The theory also explains why sellers do not establish fixed prices for houses. A fixed price is optimal only if the seller's optimal list and reservation prices are equal. However, (7) shows that the optimal list and reservation prices are unequal.

(d) Why Bids Above List Prices Sometimes Occur

Sale prices above list prices occasionally occur in housing markets. A model in which this happens can be obtained by modifying assumption A1 to permit two or more bids to occur simultaneously with positive probability. A bid equal to the list price then will not be accepted with certainty owing to the possibility that the seller receives two or more such bids together. This gives buyers with sufficiently high reservation prices an incentive to bid over the list price. The optimal list and reservation prices are obtained by maximizing V^* after replacing F with the probability distribution of the maximum bid received in a day and replacing the upper limit of integration in (6) with ∞.

In practice, sale prices above list prices usually occur infrequently and the amounts by which such sale prices exceed list prices are small. For example, in the data set used in the empirical part of this paper, only 3·8 per cent of houses have sale prices above their list prices. Among these houses, the mean difference between the sale price and the list price is 2 per cent. Therefore, little is lost (and much simplicity is gained) by making the approximation that bids and sale prices above list prices do not occur.

3. THE ECONOMETRIC MODEL

This section presents an econometric model based on the foregoing theory that can be used to predict a house's sale price and its seller's reservation price, p_R, conditional on the list price and observable attributes of the house. The main problem to be solved is estimating the parameters of F, the distribution of bids for the house. Once this is done, p_R can be predicted from (8), and the sale price can be predicted as the expected value of the first bid that exceeds p_R.

Two difficulties must be overcome in estimating F. First, the available data for a house do not include a random sample of bids drawn from F. Rather, the data contain the house's sale price. Since the sale price cannot be less than p_R, sale prices form a truncated sample. The truncation point, p_R, is not observed. Second, houses and sellers are heterogeneous, so F is not the same for all houses. Some sources of heterogeneity are observable with the available data (e.g. the number of rooms in a house), but others are not (e.g. the colour of any wall-to-wall carpeting). Thus, it is necessary to deal with the effects of both observed and unobserved heterogeneity in estimating F and making subsequent predictions.

These difficulties are similar to ones encountered in models of job search. The methods for overcoming them are different, however, because p_L provides information about F and p_R that is not available in models of job search. My methods are guided by the available data, which are described in the next section. Sections 3b and 3c develop the econometric model.

(a) Data

The data describe 1196 houses sold in Baltimore City and County, Maryland, during 1978. The houses were sampled randomly from those sold through the Multiple Listing Service in the Baltimore area during the first week of each month of 1978. Data describing the houses and their neighbourhoods were assembled from records of the Central Maryland Multiple Listing Service, Baltimore City and County tax records, the US Census Bureau, and the results of previous investigations of school quality and crime levels in the Baltimore area (Dubin and Goodman, 1982). The data include the sale price, list price, and time from listing to sale of each house as well as 46 attributes of houses and their neighbourhoods. A list of the data elements is given in Appendix B.

To avoid the computational difficulties involved in working with a large number of attribute variables, the econometric analysis was carried out using a vector of explanatory variables consisting of 20 principal components of the 46-element attribute set. The 20 components explain 91 per cent of the variance of the attribute data. No attempt was made to assign behavioural meanings or interpretations to the principal components.

(b) The Likelihood Function

The parameters of F can be estimated by maximum likelihood from observations of houses' list prices, sale prices, times from listing to sale and observed attributes. The likelihood function is complicated, however, owing to the presence of truncation (a house is in the data set only if its seller received a bid over his reservation price), censoring (the sale price cannot exceed the list price), a latent dependent variable (the seller's reservation price) and unobserved heterogeneity. This section derives the likelihood function.

I assume that the attributes of houses affect the distribution of bids F through the parameter ζ in (2). Specifically,

$$\zeta = x\beta + \varepsilon, \tag{9}$$

where x is a row vector of observed attributes, β is a conformable column vector of constant parameters, and ε is a random variable that represents the effects of unobserved variables that influence the distribution of bids. Let $\theta = (\alpha, \gamma, \eta, \beta')'$, and let $F(p \mid p_L, x, \varepsilon, \theta)$ denote the distribution of bids conditional on p_L, x and ε that is obtained by combining (1), (2) and (9).

Let $l_{pt}(p, t \mid p_L, p_R, x, \varepsilon, \theta)$ denote the likelihood that a house is sold for price p after t days on the market, conditional p_L, p_R, x, ε and θ. The sale price of the house is the first bid that exceeds p_R. Therefore, if $0 \leqslant p_R \leqslant p < p_L$, $l_{pt}(p, t \mid p_L, p_R, x, \varepsilon, \theta)$ is the probability that no bid equal to or greater than p_R occurs before day t times the probability density of a bid equal to p. Thus, if $0 \leqslant p_R \leqslant p < p_L$,

$$l_{pt}(p, t \mid p_L, p_R, x, \varepsilon, \theta) = \{F[p_R \mid p_L, x, \varepsilon, \theta]\}^{t-1} F_p(p \mid p_L, x, \varepsilon, \theta), \tag{10}$$

where $F_p = \partial F(p \mid p_L, x, \varepsilon, \theta)/\partial p$. If $0 \leqslant p_R \leqslant p = p_L$, $l_{pt}(p, t \mid p_L, p_R, x, \varepsilon, \theta)$ is the probability that no bid equal to or greater than p_R occurs before day t times the probability of a bid equal to p_L. Thus, if $0 \leqslant p_R \leqslant p = p_L$,

$$l_{pt}(p, t \mid p_L, p_R, x, \varepsilon, \theta) = \{F[p_R \mid p_L, x, \varepsilon, \theta]\}^{t-1}[1 - F(p_L \mid p_L, x, \varepsilon, \theta)]. \tag{11}$$

$l_{pt}(p, t \mid p_L, p_R, x, \varepsilon, \theta) = 0$ if $p_R > p$ or $p_R < 0$. Sales of houses with $p_R < 0$ cannot occur because a seller with a negative reservation price will abandon his house rather than hold it for sale.

To obtain an estimable model, the dependence of l_{pt} on the unobserved variables p_R and ε must be removed. The dependence on p_R can be removed by using (8). Let $p_R^*(p_L, x, \varepsilon, \theta)$ denote the right-hand side of (8) when (9) holds. Then substitution of $p_R^*(p_L, x, \varepsilon, \theta)$ for p_R in (10) and (11) eliminates the dependence of l_{pt} on p_R.

The dependence of l_{pt} on ε can be removed by taking the expected value of l_{pt} over ε conditional on the observed value of p_L and subject to the restriction $\varepsilon \in S \equiv \{\varepsilon: 0 \leqslant p_R^*(p_L, x, \varepsilon, \theta) \leqslant p_L\}$. Conditioning on p_L is needed because ε and p_L are likely to depend on common unobserved attributes of houses and, therefore, to be correlated. The restriction $\varepsilon \in S$ is needed because, under the theory of Section 2, very large or small ε values may be incompatible with optimality of the observed list price. For example, a value of ε that makes high bids likely may be incompatible with a low observed list price. In other words, the observed value of p_L restricts the range of possible values of ε. I assume that any dependence of ε on x (due, for example, to the possibility that houses with 'better' observed attributes also have better unobserved ones) can be subsumed in the dependence of p_L on x. Therefore, the distribution of ε is not explicitly conditioned on x.

To obtain the conditional expectation of l_{pt} over ε it is necessary to specify the probability density of ε in a way that restricts ε to S. To do this, let ε^* be an auxiliary random variable whose probability density conditional on p_L and a vector of unknown parameters ψ is $g(\varepsilon^* \mid p_L, \psi)$. Define $P(S) \equiv Pr(\varepsilon^* \in S)$, and let ε be ε^* truncated to S. Then, the probability density of ε conditional on p_L is $g(\varepsilon \mid p_L, \psi)/P(S)$ if $\varepsilon \in S$ and 0 otherwise. Therefore, the expectation of l_{pt} over ε, conditional on p_L, x, θ, and ψ is

$$l(p, t \mid p_L, x, \theta, \psi) = [P(S)]^{-1} \int_{S_p} l_{pt}[p, t \mid p_L, p_R^*(p_L, x, \varepsilon, \theta), x, \varepsilon, \theta]\, g(\varepsilon \mid p_L, \psi)\, d\varepsilon, \tag{12}$$

where S_p is the set of ε values such that $0 \leqslant p_R^*(p_L, x, \varepsilon, \theta) \leqslant p$. The integral in (12) is over S_p rather than S because $l_{pt}(p, t \mid p_L, p_R^*, x, \varepsilon, \theta) = 0$ if $p < p_R^*$.

To evaluate $P(S)$, let $G(\cdot \mid p_L, \psi)$ be the cumulative distribution function corresponding to $g(\cdot \mid p_L, \psi)$, and let $\varepsilon_{\min}$ and $\varepsilon_{\max}$, respectively, be the smallest and largest values of ε consistent with $\varepsilon \in S$. Then

$$P(S) = G(\varepsilon_{\max} \mid p_L, \psi) - G(\varepsilon_{\min} \mid p_L, \psi). \tag{13}$$

The values of $\varepsilon_{\min}$ and $\varepsilon_{\max}$ are determined by solving $p_R^*(p_L, x, \varepsilon_{\min}, \theta) = 0$ and $p_R^*(p_L, x, \varepsilon_{\max}, \theta) = p_L$ for $\varepsilon_{\min}$ and $\varepsilon_{\max}$, where $p_R^*(p_L, x, \varepsilon, \theta)$ is given by (8) and (9). The results are

$$\varepsilon_{\max} = -(1 + \gamma)p_L - x\beta \tag{14}$$

$$\varepsilon_{\min} = -\eta \log[1 + \gamma - \gamma \exp(-p_L/\eta)] - (1 + \gamma)p_L - x\beta. \tag{15}$$

To complete the derivation of the log-likelihood function, let $l_{\mathrm{L}}(p_{\mathrm{L}} \mid x, \delta)$ denote the probability density function of p_{L} conditional on x and a vector of parameters δ. It is necessary to specify l_{L} and $g(\varepsilon^* \mid p_{\mathrm{L}}, \psi)$. I assume that $g(\varepsilon^* \mid p_{\mathrm{L}}, \psi) l_{\mathrm{L}}(p_{\mathrm{L}} \mid x, \delta)$ is a bivariate normal density with marginal means 0 for ε^* and $E(p_{\mathrm{L}} \mid x, \delta)$ for p_{L}, marginal standard deviations σ_{ε^*} for ε^* and σ_ν for p_{L}, and correlation coefficient ρ. The conditional and marginal distributions associated with the bivariate normal are normal, so g is a normal density function, G is a cumulative normal distribution function, and p_{L} is univariate normally distributed. Since p_{L} has mean $E(p_{\mathrm{L}} \mid x, \delta)$ and standard deviation σ_ν, p_{L} satisfies

$$p_{\mathrm{L}} = E(p_{\mathrm{L}} \mid x, \delta) + \nu, \tag{16}$$

where ν is a normally distributed random variable with mean 0 and standard deviation σ_ν. It follows from the properties of the bivariate normal distribution that the mean and standard deviation of the conditional distribution $g(\varepsilon^* \mid p_{\mathrm{L}}, \psi)$ are

$$E(\varepsilon^*) = (\rho\sigma_{\varepsilon^*}/\sigma_\nu)[p_{\mathrm{L}} - E(p_{\mathrm{L}} \mid x, \delta)] \tag{17}$$

$$\mathrm{Var}(\varepsilon^* \mid p_{\mathrm{L}}, x) = (1 - \rho^2)\sigma_{\varepsilon^*}^2. \tag{18}$$

I have specified $E(p_{\mathrm{L}} \mid x, \delta)$ as

$$E(p_{\mathrm{L}} \mid x, \delta) = x\delta. \tag{19}$$

Of course, (19) is a reduced form for $E(p_{\mathrm{L}} \mid x, \delta)$. A structural model for p_{L} that is based on the theory of Section 2 produces a highly nonlinear dependence of p_{L} on the unobservable variables w and ε and is intractable computationally. Equations (16) and (19) imply that

$$l_{\mathrm{L}}(p_{\mathrm{L}} \mid x, \delta) = (1/\sigma_\nu)\phi[(p_{\mathrm{L}} - x\delta)/\sigma_\nu], \tag{20}$$

where ϕ is the standard normal density function.

The joint distribution of (p, p_{L}, t) conditional on x is $l(p, t \mid p_{\mathrm{L}}, x, \theta, \psi) l_{\mathrm{L}}(p_{\mathrm{L}} \mid x, \delta)$. Therefore, the log-likelihood of a random sample of N observations of (p, p_{L}, t) conditional on x is

$$L_{\mathrm{N}}(\theta, \psi, \delta) = \sum_{n=1}^{N} [\log l(p_n, t_n \mid p_{\mathrm{L}n}, x_n, \theta, \psi) + \log l_{\mathrm{L}}(p_{\mathrm{L}n} \mid x_n, \delta)], \tag{21}$$

where the subscript n denotes values pertaining to the nth sampled house, $l(p_n, t_n \mid p_{\mathrm{L}n}, x_n, \theta, \psi)$ is given by (12), and $l_{\mathrm{L}}(p_{\mathrm{L}} \mid x, \delta)$ is given by (20).

(c) Estimating the Model

The unknown parameters α, β, δ, γ, ρ, σ_{ε^*}, σ_ν, and η can be estimated by maximum likelihood using the log-likelihood function (21). This is burdensome computationally, however, since it entails maximizing a complicated function over a large number of parameters. An alternative approach, which yields asymptotically equivalent estimates at considerably less computational cost, consists of taking one Newton step from an initial set of consistent estimates towards the maximum of the log-likelihood function. Initial consistent estimates of δ and σ_ν can be obtained by least-squares estimation of (16). Initial consistent estimates of the remaining parameters can be obtained by maximizing the log-likelihood function with respect to these parameters while holding the values of δ and σ_ν at their least-squares estimates. This reduces the number of parameters over which the log-likelihood function must be maximized by nearly 50 per cent. The reduction is much larger when the simplification described below in Section 4(d) is made. Accordingly, the estimation results presented in this paper are based on the one-step method.

(d) Simplifying the Model

The vector β contains 20 slope coefficients and an intercept. Therefore, the log-likelihood function (21) depends on 26 unknown parameters when δ and σ_ν are held at their least-squares estimates. Maximizing a complicated log-likelihood function over such a large number of parameters is difficult computationally, so it is useful to seek modifications of the specification that reduce the number of unknown parameters. The terms $x\beta$ and $x\delta$ in (9) and (19) are a promising source of such modification. These terms are linear functions of the same variable x and, as has been discussed by Dawes and Corrigan (1974), such linear functions tend to be highly correlated. Since correlation is a measure of linear association, this suggests that $x\beta$ may be well approximated by a linear function of $x\delta$.

To see whether $x\beta$ is well approximated by a linear function of $x\delta$ in the Baltimore data, I estimated β and δ by maximizing the log-likelihood function (21) using 10 principal components instead of the full set of 20. The ratios of the estimated values of the non-intercept components of β and δ were approximately equal, as would be expected if $x\beta$ is approximately a linear function of $x\delta$. Using the estimated values of β and δ, I computed the values of $x\beta$ and $x\delta$ for each house in the data set and regressed $x\beta$ on $x\delta$. The value of R^2 for this regression is 0·99998, which confirms the expectation of high correlation between $x\beta$ and $x\delta$ and implies that $x\beta$ is well approximated by a linear function of $x\delta$. It is convenient to write this linear function in the form

$$x\beta = \vartheta_0 + [\vartheta_1 - (1+\gamma)]\, x\delta, \tag{22}$$

where ϑ_0 and ϑ_1 are scalar constants. Subsequent estimation was carried out under the hypothesis that (22) holds for all 20 principal components. Using this modified specification, estimation of the 21-component vector β is replaced by estimation of the scalar constants ϑ_0 and ϑ_1, thereby reducing by 19 the number of parameters that must be estimated.

4. ESTIMATION RESULTS

The estimates of α, η, γ, ϑ_0, ϑ_1, σ_{ε^*}, σ_ν and ρ are shown in Table I. The estimates of the coefficients δ in $E(p_L \mid x, \delta)$ are not shown since the principal components x and, therefore, δ have no economic interpretations. The estimated values of η, α, and γ are consistent with the conditions $\eta > 0$, $0 \leqslant \alpha \leqslant 1$, and $\gamma > 0$ that were discussed in connection with (1) and (2).

Table I. Estimated parameters of the bid generation function[a]

Parameter	Estimate	Standard error
ϑ_0	−0·1368	0·0117
ϑ_1	−0·0149	0·0021
α	0·9507	0·0035
η	0·6979	0·0713
γ	0·6280	0·0389
σ_{ε^*}	1·8752	0·0454
σ_ν	1·138	0·0530
ρ	−0·9997	$7{\cdot}4663 \times 10^{-5}$

[a] The parameters are as defined in equations (16), (23), and (26). Prices are in tens of thousands of dollars.

The estimate of ρ is close to -1, which implies that, in the Baltimore data, the bid distribution depends on attributes of houses only through p_L. To see why this is so, observe that since $\rho \approx -1$ and correlation is a measure of linear association, $\varepsilon^* \approx a_0 + a_1\nu$ for some a_0 and a_1 with $a_1 < 0$. In addition, $E(\varepsilon^*) \approx a_0 + a_1E(\nu)$, and $\sigma_{\varepsilon^*}/\sigma_\nu \approx |a_1|$. Since $E(\varepsilon^*) = E(\nu) = 0$ by assumption, $a_0 \approx 0$. It can be seen from Table I that the estimated values of $|\vartheta_1 - (1+\gamma)|$ and $\sigma_{\varepsilon^*}/\sigma_\nu$ are approximately equal (1·6429 and 1·6475, respectively), and the estimated value of $[\vartheta_1 - (1+\gamma)]$ is negative. Therefore, $a_1 \approx [\vartheta_1 - (1+\gamma)]$ and, since $\varepsilon = \varepsilon^*$ in S, $\varepsilon \approx [\vartheta_1 - (1+\gamma)]\nu$. This result and (22) imply that $x\beta + \varepsilon \approx \vartheta_0 + [\vartheta_1 - (1+\gamma)]p_L$, so that in the Baltimore data the bid distribution depends on attributes of houses only through p_L.

Figures 1–3 illustrate the ability of the estimated model to fit the data. The figures compare the observed and predicted list prices, sale prices and times from listing to sale of a sample of 50 houses in the estimation data set. The predicted list price is the expected value conditional on x. The predicted sale price is the expected value conditional on x and the observed list price. The predicted time is the median conditional on x and the observed list price. The median, rather than the mean, is used for predicting time because, according to the specified model, the mean time conditional on x and p_L does not exist.[2]

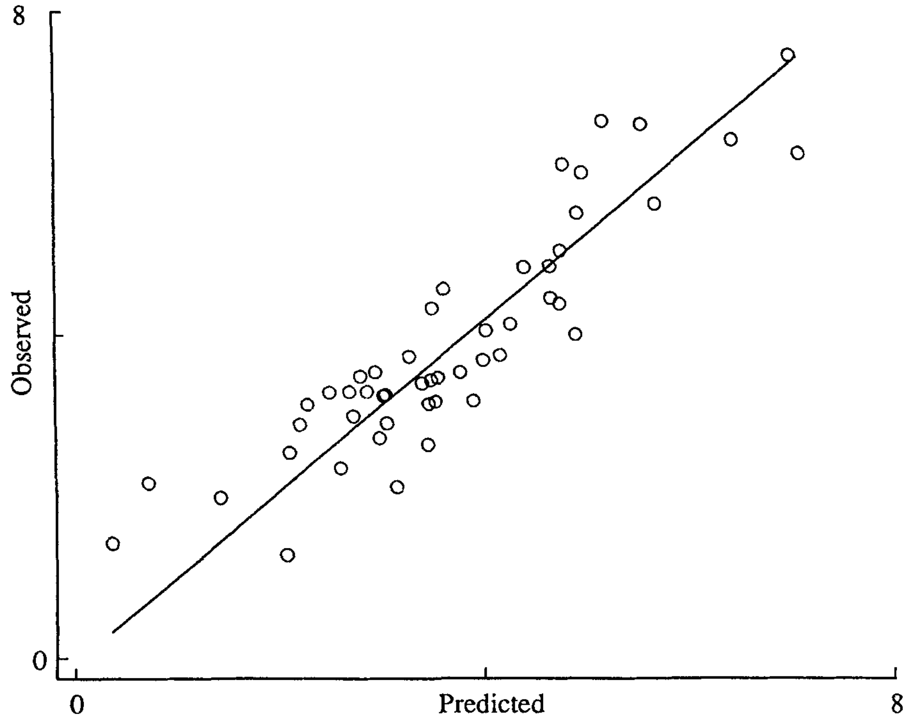

Figure 1. Predicted and observed list prices. Prices are in units of \$10,000

[2] The mean time to sale is $E(t) = E_\varepsilon[1/Q(p_R^* \mid p_L, x, \varepsilon, \theta) \mid 0 < p_R^* \leqslant p_L]$, where p_R^* is given by (8) and (9). As $\varepsilon \to \varepsilon_{\max}$, $1/Q(p_R^* \mid p_L, x, \varepsilon, \theta) \to \infty$, but the probability density of ε does not approach 0 if ε is normally distributed. As a result, the expected value integral is infinite.

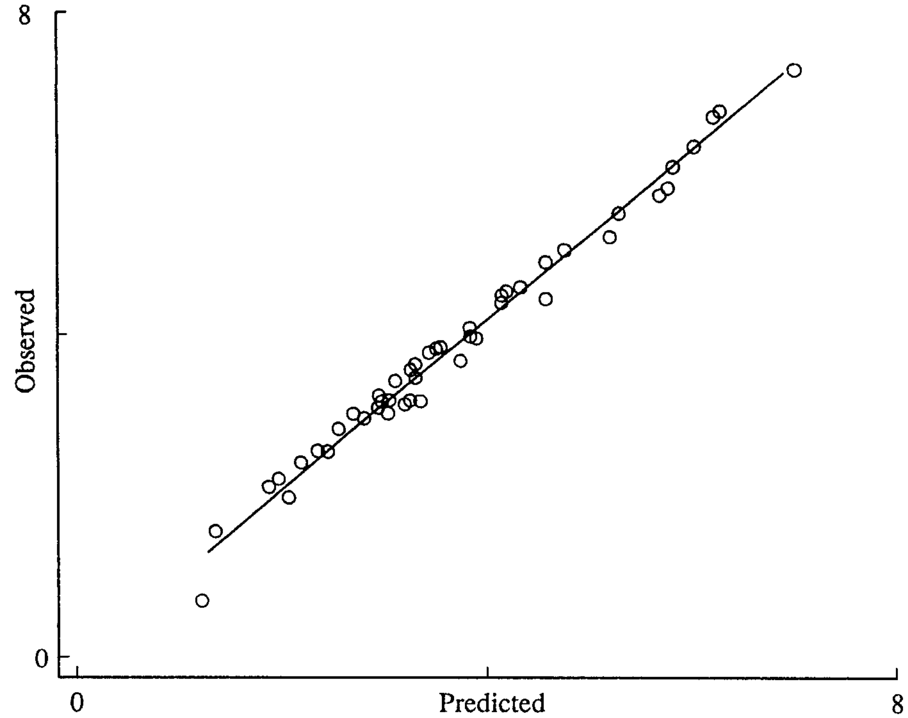

Figure 2. Predicted and observed sale prices. Prices in units of $10,000

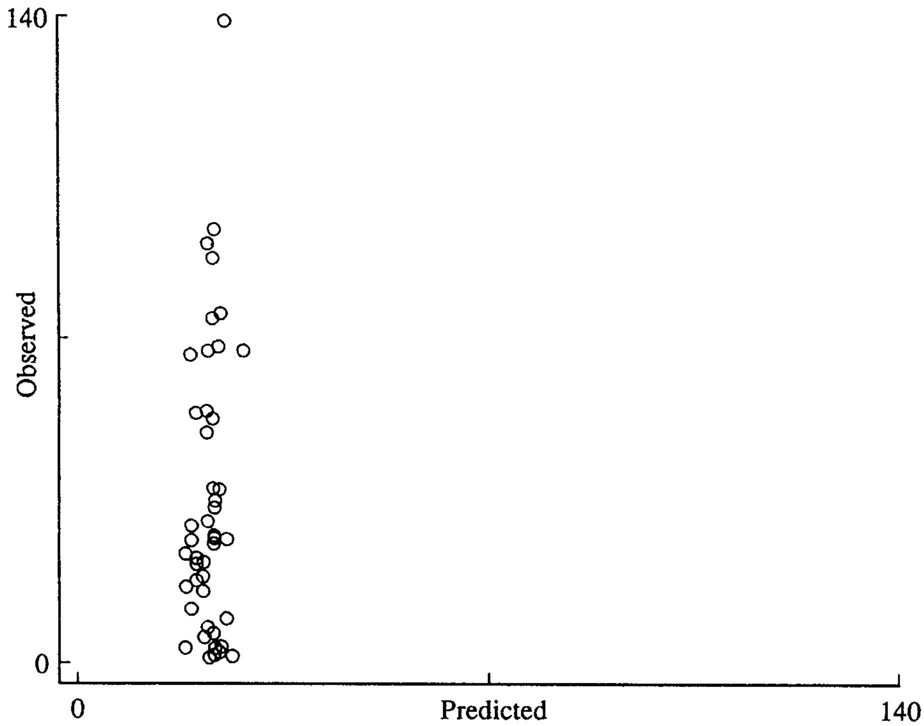

Figure 3. Predicted and observed times to sale

The figures indicate that the model fits the price data well. There is no evidence of systematic lack of fit or heteroscedasticity. In the full sample of 1196 houses the root-mean-square (RMS) error in the prediction of the list price is $11,283, and the RMS error in the prediction of the sale price conditional on the list price is $2960. The model of sale prices conditional on list prices and observed attributes explains 98 per cent of the variation in observed sale prices.

By contrast, hedonic price regressions of the sale price and its logarithm on x have RMS prediction errors of $10,361 and $10,380, respectively. Thus, the econometric model developed here gives predictions of sale prices that are considerably more accurate than those of a hedonic price regression.

It can be seen from Figure 3 that there is little variation in the predicted (median) times from listing to sale. The predictions all are between 19 and 29 days, whereas the observed times vary from 1 to 138 days. Thus, the model explains little of the observed variation in time from listing to sale. This result is consistent with the results of an exploratory data analysis in which a median regression of time from listing to sale on list price and the principal components was found to explain only 1.6 per cent of the variation in time.[3]

There is a simple explanation for the large unexplained variation in observed times from listing to sale. Conditional on the list price and attributes of a house, the time from listing to sale has the geometric distribution with probability per day of 'success' equal to the probability per day that the seller receives a bid above his reservation price. Median times from listing to sale in the range 19–29 days correspond to success probabilities in the range 0·024–0·038 per day.[4] When the success probability of the geometric distribution is in this range, the 5th percentile of the distribution is 2–3 days and the 95th percentile is 78–122 days. The large unexplained variation in times from listing to sale is, therefore, a reflection of the fact that any geometrically distributed random variable whose median is in the range 19–29 days will have a range of variation similar to that observed in the Baltimore sample. It does not imply that the estimated model is misspecified and, in fact, any model of the relation between t, p_L, and x would give a comparable unexplained variation in t.

5. FURTHER APPLICATIONS

(a) Predicting the Seller's Reservation Price

The expected value of the seller's reservation price can be obtained by taking the expected value of $p_R^*(p_L, x, \varepsilon, \theta)$, conditional on $0 \leqslant p_R^* \leqslant p_L$. Figure 4 shows expected values of reservation prices according to the estimated model as functions of list prices for the 50 sampled houses. The expected reservation prices are approximately $5500 below the list prices.

[3] A median regression was used instead of least-squares to maintain consistency with the specified model. According to the model the time from listing to sale has no moments, thereby violating the regularity conditions of least-squares estimation. The proportion of the variation explained is defined as $1 - d_R/d_M$, where d_R is the mean absolute residual of the estimated model and d_M is the mean absolute difference between the observed values of the dependent variable and the sample median.

[4] The cumulative geometric distribution function with success probability q is $Pr(t < T) = 1 - (1-q)^{T-1}$ $(T = 1, 2, \ldots)$. The λ quantile of this distribution is the largest integer T such that $Pr(t < T) \leqslant \lambda$. The median is obtained by setting $\lambda = 0{\cdot}5$. It is easily verified that the median is 19 if $q = 0{\cdot}038$ and 29 if $q = 0{\cdot}024$. The other quantiles given in the text can be verified using the cumulative distribution function and the definition of a quantile.

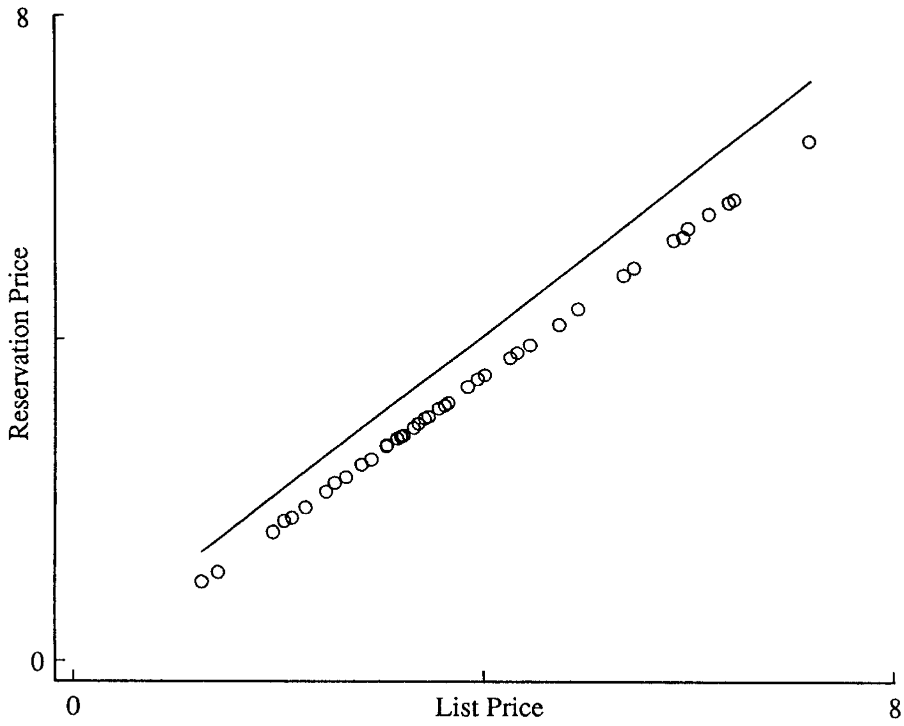

Figure 4. Predicted reservation prices plotted against observed list prices. Prices in units of $10,000

(b) Why a Seller whose House has Remained Unsold for a Long Time May Not Want to Reduce the List Price

It is well established in real estate folklore that sellers whose houses have remained unsold for long periods of time often are unwilling to reduce their list prices, even if doing so would be likely to lead to quick sales. The results presented in Section 5 provide an explanation for this phenomenon. As has been discussed, the time from listing to sale is a geometrically distributed random variable with a low success probability (that is, a low probability per day that an offer above the reservation price occurs). Because the geometric distribution with low success probability is highly skewed to the right, long periods of time from listing to sale can occur even if a house is optimally priced. In such cases a seller cannot increase his expected utility by reducing the list price.

6. CONCLUSIONS

The theory developed in this paper explains why sellers of houses establish list prices that are price ceilings. The econometric model makes possible the prediction of a seller's reservation price conditional on the list price and observed attributes of the house. It gives predictions of sale prices that are considerably more accurate than those of a standard hedonic price regression. In addition, it explains why sellers may not be willing to reduce their list prices even after their houses have remained unsold for long periods of time.

APPENDIX A: PROOFS OF (7) AND (8)

Proof that $p_L^* < \infty$: Given any p_L, let p_R^* maximize $V^*(p_R, p_L)$. The total derivative of $V^*(p_R^*, p_L)$ with respect to p_L is

$$dV^*(p_R^*, p_L)/dp_L$$
$$= [r + Q(p_R^* \mid p_L)]^{-1}\left\{F_L(p_R^* \mid p_L)[V^*(p_R^*, p_L) - (p_R^* - w)] + Q(p_L \mid p_L) - \int_{p_R^*}^{p_L} F_L(p \mid p_L)\,dp\right\}. \tag{A1}$$

Let $\tilde{p_L}$ be so large that $Q(\tilde{p_L} \mid \tilde{p_L}) = 0$, and suppose that $p_L \geqslant \tilde{p_L}$. Then

$$dV^*(p_R^*, p_L)/dp_L$$
$$= [r + Q(p_R^* \mid p_L)]^{-1}\left\{F_L(p_R^* \mid p_L)]^{-1}[V^*(p_R^*, p_L) - (p_R^* - w)] - \int_{p_R^*}^{p_L} F_L(p \mid p_L)\,dp\right\}. \tag{A2}$$

Since p_R^* is the optimal reservation price corresponding to p_L, the expected utility of selling at price p_R^* must be at least as large as the expected utility of leaving the house on the market. Therefore, $V^*(p_R^*, p_L) - (p_R^* - w)$ is non-positive. In addition, $F(p_R^* \mid p_L) < 1$ (otherwise, bids equal to or greater than p_R^* would have probability 0 and p_R^* could not be optimal), so it follows from (1) and (2) that $F_L(p_R^* \mid p_L) > 0$ and that $F_L(p \mid p_L) \geqslant 0$ if $p_R^* \leqslant p \leqslant p_L$. Therefore, continuity of $F_L(p \mid p_L)$ as a function of p implies that

$$\int_{p_R^*}^{p_L} F_L(p \mid p_L)\,dp > 0,$$

so the right-hand side of (A2) is negative. It follows that p_L^* cannot exceed $\tilde{p}_L$. Because p_L^* cannot be less than w, the search for the optimal value of p_L can be restricted to the set $w \leqslant p_L \leqslant \tilde{p}_L$. Since $V^*(p_R^*, p_L)$ is a continuous function of p_L, an optimal p_L exists. Q.E.D.

Proof that $w \leqslant p_R^* < p_L^*$: It is clear from the definitions of w, p_R, and p_L that the search for optimal values of p_R can be confined to the interval $w \leqslant p_R \leqslant p_L^*$. Since V^* is a continuous function of p_R when $w \leqslant p_R \leqslant p_L^*$, there exists an optimal reservation price p_R^* satisfying $p_R^* \leqslant p_L^*$. Left-differentiation of the right-hand side of (6) with respect to p_R at $p_R = p_L = p_L^*$ yields

$$(\partial V^*/\partial p_R)_{p_R = p_L^{*-}} = -rF_p(p_L^{*-} \mid p_L^*)(p_L^* - w)/[r + Q(p_L^* \mid p_L^*)]^2, \tag{A3}$$

where p_L^{*-} signifies left differentiation. The quantity on the right-hand side of (A3) is strictly negative unless $F_p(p_L^{*-} \mid p_L^*) = 0$. Therefore, $p_R^* < p_L^*$ unless $F_p(p_L^{*-} \mid p_L^*) = 0$, in which case $p_R^* = p_L^*$ may be possible. By (1) and (2), $F_p(p_L^{*-} \mid p_L^*) = 0$ implies that $F(p_L^* \mid p_L^*) = 1$, in which case bids equal to or greater than p_L^* have probability 0, so $p_R^* = p_L^*$ cannot be optimal. Q.E.D.

Proof of (8): The first-order conditions for maximizing V^* imply after some algebra that

$$1 - F(p_L^* \mid p_L^*) = \int_{p_R^*}^{p_L^*} F_L(p \mid p_L^*)\,dp. \tag{A4}$$

Equation (8) results from substituting (1) and (2) into (A4). Q.E.D.

APPENDIX B: CONTENTS OF THE ESTIMATION DATA SET

Attributes of the house
Sale price
List price
Time from listing to sale
Location (city or county)
Type of dwelling (attached or detached)
Number of rooms
Number of bathrooms
Basement (yes or no)
Swimming pool (yes or no)
Patio (yes or no)
Fireplace (yes or no)
Garage (yes or no)
Air conditioning (yes or no)
Age
Plot size
Floor space
Harmonic mean distance to major shopping locations
Harmonic mean distance to motorway junctions.

Attributes of census tract or neighbourhood in which house is located.
Mean household income (1970 and 1980)
Elementary and high school students per household (1970 and 1980)
Percentage of population that is black (1970 and 1980)
Percentage of population over age 25 who have completed high school (1970 and 1980)
Median age of population (1970 and 1980)
Percentage of employed residents with blue-collar occupations (1970 and 1980)
Percentage of housing units that are owner-occupied (1970 and 1980)
Six indices of quality of the school district in which house is located
Three indices of crime levels in crime reporting district in which house is located
(Plot size) × (Elementary and high school students per household in 1980)
(Floor space) × (persons per dwelling unit in 1980)
(Education indices) × (Elementary and high school students per household in 1980)

ACKNOWLEDGEMENTS

This research was supported in part under contract no. CR-811043-02-0 from the US Environmental Protection Agency to the University of Maryland and subcontract no. 56361-A from the University of Maryland to the University of Iowa. I thank Andrew Daughety, Benjamin Eden, Gary Fethke, Donald McCloskey, Forrest Nelson, Thomas Pogue, Gene Savin, and seminar participants at Australian National University and the University of Sydney for their comments and suggestions concerning this work, Allen Goodman for providing the data used in the empirical analysis, and Jean-Claude Thill and John Fortney for research assistance.

REFERENCES

Chinloy, P. T. (1980). 'An empirical model of the market for resale homes', *Journal of Urban Economics*, **7**, 279–292.

Dawes, R. M., and B. Corrigan (1974). 'Linear models in decision making', *Psychological Bulletin*, **81**, 95–106.

Dubin, R. A., and A. C. Goodman (1982). 'Valuation of crime and education neighborhood characteristics in a hedonic housing price model', *Population and Environment*, **5**, 66–81.

Guasch, J. L., and R. C. Marshall (1985). 'An analysis of vacancy patterns in the rental housing market', *Journal of Urban Economics*, **17**, 208–229.

Stull, W. J. (1978). 'The landlord's dilemma: asking rent strategies in a heterogeneous housing market', *Journal of Urban Economics*, **5**, 101–115.