HEDONIC THEORY AND HOUSING MARKETS

by

Bryan Ellickson

Discussion Paper No. 124

July 1978

Hedonic Theory and Housing Markets

This paper is aimed not at the specialist in urban economics but rather the outsider curious about the role of hedonic theory in the analysis of housing markets: why it is used, how it fits into the main body of economic theory and what it accomplishes. This is not intended as a comprehensive survey but instead a selective presentation of certain features of the theory and practice of hedonic analysis which I believe to be of general interest.

1.   Hedonic Theory

To most urban economists the rationale for using hedonic theory seems obvious. Housing commodities are neither homogeneous nor perfectly divisible, so some alternative to conventional microeconomic analysis is clearly needed. A natural alternative is to define consumer preferences not over housing commodities but over their characteristics: e.g., lot size, number of rooms, neighborhood quality. The consumer then seeks to

$$\max_{x,z} U(x,z) \quad \text{subject to } p_x x + h(z) = y \qquad (1)$$

where $U(x,z)$ is a utility function, x is a vector of standard private goods with prices p_x, z is a vector of housing characteristics, y is the consumer's income and $h(z)$ is the <u>hedonic price function</u> (a function giving the price of a house with characteristics z).

Since the maximization over x is a side issue, it is useful to push it into the background. Let $\phi(p_x, z, y - h(z))$ represent the result of maximizing $U(x,z)$ with respect to x subject to the constraint $p_x x + h(z) = y$, where the function ϕ is just the usual indirect utility function. Then the consumer problem characterized by (1) can be reformulated as:

$$\max_{z} \phi(p_x, z, y - h(z)); \tag{2}$$

i.e., the consumer chooses a house with characteristics which maximize

his utility. We will call (2) the <u>Muth</u> <u>formulation</u>.

For certain purposes it is useful to recast this problem in an

alternative form. By setting $\phi(p_x, z, y-V) = u$ and solving for V, we

obtain the <u>bid</u> <u>price</u> <u>function</u> (or willingness-to-pay function):

$$V = \psi(p_x, z, y, u). \tag{3}$$

The function ψ gives the price for a house with characteristics z

that will yield utility u to the consumer. For different values of u

this function defines a family of contours that plays much the same

role in this analysis as indifference contours in the conventional

theory. The solution to (2) can be described as follows in terms of

bid price functions: the consumer chooses a house with characteristics

z which places him on the bid price contour corresponding to the highest

possible level of utility subject to the constraint that $V = h(z)$.

We will call this the <u>Alonso</u> <u>formulation</u>. Figure 1 illustrates the

condition for a maximum with respect to a given characteristic z_i.

Figure 1:
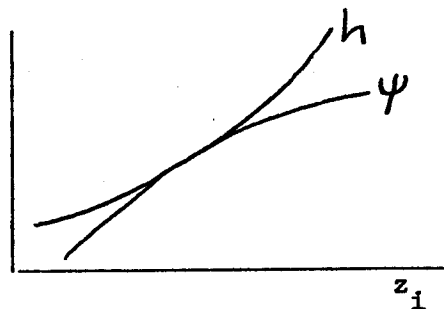
(Readers familiar with Rosen [1974] will recognize this as the funda-

mental diagram in his treatment of hedonic theory.)

By referring to these alternative (but mathematically equivalent)

characterizations as the Muth and Alonso formulations I have taken

considerable liberties with the original (Muth [1969], Alonso [1965]).

Their work predated that of Rosen by several years, and therefore their

models are not described in terms of hedonic theory. The housing character-
istics treated are limited to just two: accessibility to the CBD and lot
size (or flow of housing services). Their analysis also differs in some
details: the money cost of transportation is introduced explicitly into
the budget constraint, while I have absorbed it into the functional forms
ϕ and ψ ; indirect utility functions and hedonic price functions are not
used; and the price of a house at any given distance from the CBD is ass-
umed to be a linear function of lot size (an unnecessary restriction in the
general hedonic framework). But the point is that the basic ideas are the
same, and the general theory yields the Alonso and Muth theories as a special
case (see Ellickson [1977]). Thus, hedonic theory in its general form can be
regarded as a natural extension of the basic models of residential location.

While the hedonic theory of housing markets may seem the obvious
way to proceed for most urban economists, the outsider may well ask
why it is necessary to introduce characteristics. After all, char-
acteristics are not needed in most general equilibrium analysis (the
reason for the qualification "most" will be apparent in a moment).
And hedonic analysis as usually presented leaves open the question of
whether what is being described is really a competitive equilibrium.
The reason why these issues seem cloudy is that hedonic theory intro-
duces two complications into the analysis at the same time: indivis-
ible commodities (since a consumer either lives in a house or not) and
a continuum of commodities (since the set of available characteristics
is generally not assumed finite). To gain some insight into what is
going on, therefore, it is useful to introduce one complication at a
time.

Suppose that the set of possible housing types is finite: i.e., $z \in K$, where K is a finite set. A considerable amount is now known about competitive equilibrium with indivisible commodities: if the set of consumers is finite, then it is possible to prove existence of an "approximate" competitive equilibrium (see Broome [1972]), with the approximation improving as the number of consumers increases. In the limit with a continuum of agents, an exact equilibrium exists (see Mas-Colell [1977]). What happens to hedonic theory in this context? Characteristics are superfluous, simply a label for the type of house. The maximization described by (1) is just a way to describe preference maximization in an economy with indivisible commodities, as the reader can verify by rewriting it with preferences defined over the housing commodities rather than characteristics. However, while characteristics are not needed to establish existence of equilibrium, they may serve as a useful vehicle for parameterizing utility functions and, therefore, for describing the demand for various types of houses in a form better suited to empirical estimation. An example will be given shortly.

If the set K of possible housing types is infinite, then formulating a model of competitive equilibrium is not so straightforward. As Mas-Colell [1975] has observed, in the general context of an economy with indivisible commodities, the usual justification for the competitive hypothesis may fail: the core may be larger than the set of competitive allocations. Heuristically, what is required for core equivalence is that the markets for commodities be "thick," but if there is an infinite number of commodities this requirement will generally fail. However, by introducing the notion of characteristics, just as in

hedonic theory, he is able to demonstrate the existence of a competitive equilibrium with core equivalence. This work has yet to be applied to the context of housing markets, but at least it suggests that hedonic theory is the appropriate way to proceed and that the use of characteristics is not superfluous. It also provides a new perspective on what is being-accomplished-by assuming that the hedonic function is a <u>continuous</u> function of characteristics: a uniformity on the collection of consumer preferences is imposed, an assumption that houses with similar characteris are regarded as close subsitutues by all consumers in the economy.

2. The Empirical Application of Hedonic Theory.

We now have a sense of why hedonic theory is used in the analysis of housing markets and how it fits into the main body of economic theory. We turn to the question of whether it has anything useful to contribute to empirical studies of the housing market. Recall that we presented two versions of the theory, the formulations of Muth and Alonso. Since the two formulations are mathematically equivalent, presenting both seems redundant. However, the two approaches have generated largely disjoint literatures and, as we shall see, their empirical implementation leads in quite distinct directions. The Muth formulation is the cornerstone of the "new urban economics," the development of rigorous models of residential location. The Alonso formulation is central to the "soft" side of urban economics, the articulation of less rigorous models of residential segregation, jurisdictional fragmentation and filtering.

a. The Muth Formulation

Turning first to the Muth formulation, perhaps the most natural procedure is to use the constrained maximization problem (1) to derive the consumer's "demand" for characteristics. This is essentially the procedure adopted by Straszheim [1974]. The problem with this is that the hedonic function $h(z)$ need not be linear; therefore, nothing plays the role of prices in the demand equations, and -- as a result -- the estimates seem of limited value. If we are willing to aggregate houses into a finite number of types, on the other hand, it is possible to derive equations closer in spirit to the demand functions of conventional microeconomic theory. In this case, the problem becomes a natural candidate for the logit analysis of McFadden [1974], an approach adopted within the context of housing market analysis by Friedman [1975], Lerman [1977], and Quigley [1976].

In order to describe this approach, we write the indirect utility function for a consumer of type t in the form $\tilde{\phi}_t(z,h(z))$ where we have suppressed the price vector p_x (assumed invariant throughout the metropolitan area) and the income parameter y_t (since households of the same type have the same income). To translate the model into a form suitable for econometric estimation, this deterministic indirect utility function is replaced by a stochastic indirect utility function:

$$\tilde{\phi}_t[z,h(z)] + \varepsilon_z \tag{4}$$

where ε_z is a random variable associated with a house of type z. The probability that a consumer of type t will choose a house of type z is then given by

$$p(z|t) = \text{prob}\{\tilde{\phi}_t[z,h(z)] + \varepsilon_z > \tilde{\phi}_t[z',h(z')] + \varepsilon_{z'}; z \neq z'; z, z' \in K\} \tag{5}$$

If the random variables ε_z, $z \in K$, are independently and identically distributed Weibull, McFadden [1974] has shown that equation (5) will take the form:

$$p(z|t) = \frac{\exp\{\phi_t(z, h(z))\}}{\sum\limits_{z' \in K} \exp\{\phi_t(z', h(z'))\}} \tag{6}$$

If the indirect utility functions are linear in the parameters, we obtain:

$$p(z|t) = \frac{\exp\{\beta_t z + \gamma_t h(z)\}}{\sum\limits_{z' \in K} \exp\{\beta_t z' + \gamma_t h(z')\}} \tag{7}$$

This is essentially the model estimated by Friedman, Lerman and Quigley.[1] For an extensive discussion of this work, see McFadden [1977].

    b.   **The Alonso Formulation**

     Logit provides, therefore, an attractive means for translating the Muth formulation of hedonic theory into a form suitable for empirical estimation. However, as noted earlier a substantial portion of the urban economics literature is cast in terms of the Alonso formulation, and the various hypotheses that have been advanced do not fit easily into the Muth mold. The most natural way to interpret such models is in terms of a prediction of what sort of consumer is most likely to occupy a house with a specified set of characteristics. The house will be occupied by the consumer offering the highest bid price. Thus, the traditional accessibility model predicts that houses located far away from the central business district will be occupied by households with low marginal commuting costs and relatively high demand for

---

[1] Lerman also incorporates choice of transport mode into his model.

housing space. The filtering model predicts that newer housing will be occupied by wealthier households. And the Tiebout models predict that houses in communities offering higher quality public services will be occupied by households with high income or a strong preference for public services.

To translate the Alonso formulation into an empirical model, we will first write the bid price function for a consumer of type t in the form $\tilde{\psi}_t(z)$ where we have suppressed the price vector p_x (assumed invariant throughout the market), the income parameter y_t (since consumers of the same type have the same income) and the equilibrium level of utility u_t (since consumers of the same type have identical incomes and preferences and hence attain the same level of utility). In the empirical version of the Muth model the indirect utility function is replaced by a stochastic indirect utility function. In parallel fashion here we replace the bid price function by a <u>stochastic bid price function</u>:

$$V_t = \tilde{\psi}_t(z) + \varepsilon_t \qquad (8)$$

where ε_t is a random disturbance term. Then the deterministic proposition that a house with characteristics z will be occupied with probability one by a particular household type is replaced by the probabilistic statement that

$$p(t|z) = \text{prob}\{\tilde{\psi}_t(z) + \varepsilon_t > \tilde{\psi}_{t'}(z) + \varepsilon_{t'}; t' \neq t; t, t' \in T\} \qquad (9)$$

where T is an index set for household types. If, following McFadden [1974], we assume that the disturbance terms are independently and identically distributed Weibull, equation (9) takes the form

$$p(t|z) = \frac{\exp[\tilde{\psi}_t(z)]}{\sum_{t' \in T} \exp[\tilde{\psi}_{t'}(z)]}. \qquad (10)$$

Assuming that the bid price functions are linear in the parameters we obtain

$$p(t|z) = \frac{\exp(\alpha_t z)}{\sum\limits_{t' \varepsilon T} \exp(\alpha_{t'} z)} \qquad (11)$$

as the model to be estimated.

I have estimated this model (Ellickson [1977]) using data drawn from a sample survey of 28,000 households in the San Francisco Bay Area conducted in 1965. Table 2 presents the results for white owners classified into six groups, three income categories (Y1 = under $7000, Y2 = $7,000 - $9,999, Y3 = $10,000 or more) each subdivided according to whether children were present (C) or not present (NC) in the household. The ten characteristics used are described in Table 1.

The results provide strong confirmation of several hypotheses that have appeared in the housing market literature. To interpret these results, we begin with a comparison across income classes, family composition held constant. With minor exceptions, the coefficients of the first six characteristics and Z10 (an index of housing quality) exhibit the pattern one expects. The coefficients of Z1 (commuting time to San Francisco) tend to become increasingly negative as income increases, precisely the result one expects if higher income households attach a higher value to commuting time. Higher income households also prefer newer housing (characteristic Z2), larger lots (Z3), more rooms (Z4), a better neighborhood (as represented by Z5, median tract income in 1960), and those aspects of housing quality captured by the hedonic residual (Z10). As household income increases, owners with children attach more value to housing within the attendance area of elementary

Table 1

HOUSING CHARACTERISTICS USED IN THE EMPIRICAL ANALYSIS

$Z1$ Log (travel time to San Francisco in minutes)

$Z2$ Log (age of dwelling unit in years)

$Z3$ Log (lot size in square feet)

$Z4$ Log (number of rooms)

$Z5$ Log (median tract income in 1960)

$Z6$ Log (elementary median income)

$Z7$ Percent of students in elementary school who are black

$Z8$ Percent of students in junior high school who are black

$Z9$ Percent of households in census tract in 1960 who are black

$Z10$ Hedonic residual

Table 2

LOGIT ESTIMATES: WHITE OWNERS

Parameter Estimates

Group	Constant	Z1	Z2	Z3	Z4	Z5	Z6	Z7	Z8	Z9	Z10	Number of Observations
Y1,C	0	0	0	0	0	0	0	0	0	0	0	236
Y2,C	-11.1	-.16	-.08	.19	.25	.14	.08	.09	-.28	-.10	.36	498
Y3,C	-35.5	-.49	-.37	.35	.71	.41	.34	-.03	-.14	-.05	.73	698
Y1,NC	5.0	-.46	.57	.22	-.26	-.43	.43	.23	-.11	-.18	.47	180
Y2,NC	-9.4	-.38	.25	.28	-.32	-.08	.38	.09	-.12	-.10	.50	236
Y3,NC	-33.5	-.84	-.16	.43	.14	.40	.48	.05	-.07	0	.68	466

Asymptotic t Statistics (Group i Relative to Group j)[a]

Group i	Constant	Z1	Z2	Z3	Z4	Z5	Z6	Z7	Z8	Z9	Z10	Group j
Y3,C	-6.40**	-4.06**	-3.75**	3.50**	7.93**	2.73**	2.13*	-.22	-1.24	-.43	7.54**	Y1,C
Y3,NC	-6.23**	-3.19**	-4.73**	2.03*	4.37**	4.55**	.27	-1.38	.42	1.71	2.02*	Y1,NC
Y1,C	-.68	3.23**	-3.45**	-1.82	2.52*	2.11*	-2.12*	-1.59	.88	1.70	-3.90**	Y1,NC
Y2,C	-.31	1.84	-2.90**	-.92	6.52**	1.34	-1.81	0	-1.32	-.89	-1.51	Y2,NC
Y3,C	-.59	4.40**	-2.83**	-1.23	8.58**	.11	-1.37	-.75	-.87	-.48	.98	Y3,NC

Sample Standard Deviations		.39545	.78676	.72274	.21851	.18434	.19259	.18783	.24782	.08017	.77803	

[a] A single asterisk indicates significance (two-tailed) at the .05 level; a double asterisk, at the .01 level.

11

schools drawing from a higher income population (characteristic Z6), while households without children do not. Income differences appear to have no effect on the reaction of white households to racial composition of the schools (Z7 and Z8) or the census tract (Z9).

The difference in parameter estimates for households with and without children, income held constant, is also reasonable with one major exception. Owners with children put much more weight on number of rooms and newer houses at the expense of accessibility to the center of San Francisco. The only anomolous result involves characteristic Z6: Owners without children attach a higher value to areas served by elementary schools with higher median income than do their counterparts with school age children, though the difference is only significant for the lowest income group.

The model performs as well when applied to white renters. Perhaps the only surprise in either case is the absence of significant differences in the coefficients for the racial variables (Z7, Z8 and Z9). But when the analysis is applied to a classification of black and white households, highly significant differences appear in the coefficients for Z7 and Z9 between the racial groups, income held constant, and the pattern is in the expected direction.

The usual χ^2 tests indicate that the equations all are significant at better than .001 level and, when used to predict the type of household occupying each house in the sample, the models classify correctly for 34.5% of the cases for white owners, 38.1% for white renters, 59.7% for black and white owners and 59.2% for black and white renters. A χ^2 test indicates that these classifications perform significantly better than chance at the .001 level.

3. Conclusions

In this paper I have argued that hedonic theory provides a useful model of housing markets, serving both as a generalization of the earlier work of Alonso and Muth and an application of recent advances in the theory of competitive equilibrium with indivisible commodities. The theory can also be given empirical content, leading to demand functions of the form given by (7) under the Muth formulation and to functions giving the conditional probability that a house with character- istics z will be occupied by a household of type t (equation (11)) under the Alonso formulation. I see no reason to prefer one empirical model to the exclusion of the other. The former approach yields demand functions while the latter does not. The latter provides sharper and more easily interpreted tests of a number of hypotheses in the urban economics literature than does the former. As I have indicated both emerge quite naturally from the general hedonic theory of housing markets, and the empirical evidence gives strong support for the theory.

REFERENCES

Alonso, William, 1965, _Location and Land Use_, Harvard University Press, Cambridge.

Broome, John, 1972, "Approximate Equilibrium in Economies with Indivisible Commodities," _Journal of Economic Theory_, 5, 224-249.

Ellickson, Bryan (with Barry Fishman and Peter A. Morrison), 1977, _Economic Analysis of Urban Housing Markets: A New Approach_, The Rand Corporation, R-2024-NSF.

Friedman, Joseph, 1975, _Housing Location and the Supply of Local Public Services_, The Rand Corporation, P-5421.

Lerman, Steven R., 1977, "Location, Housing, Automobile Ownership and the Mode to Work: A Joint Choice Model," _Transportation Research Board Record_, No. 610, 6-11.

Mas-Colell, Andreu, 1975, "A Model of Equilibrium with Differentiated Commodities," _Journal of Mathematical Economics_ 2, 263-295.

Mas-Colell, Andreu, 1977, "Indivisible Commodities and General Equilibrium Theory," _Journal of Economic Theory_ 16, 443-456.

McFadden, Daniel, 1974, "Conditional Logit Analysis of Qualitative Choice Behavior," in _Frontiers in Econometrics_, edited by Paul Zarembka, Academic Press, New York.

McFadden, Daniel, 1977, "Modeling the Choice of Residential Location," Cowles Foundation Discussion Paper No. 477.

Muth, Richard, 1969, _Cities and Housing_, University of Chicago Press, Chicago.

Quigley, John M., 1976, "Housing Demand in the Short Run: An Analysis of Polytomous Choice," _Explorations in Economic Research_ 3, 76-102.

Rosen, Sherwin, 1974, "Hedonic Prices and Implicit Markets: Product

Differentiation in Pure Competition," Journal of Political Economy

82, 34-55.

Straszheim, Mahlon R., 1974, An Econometric Analysis of the Urban Housing

Market, National Bureau of Economic Research, New York.