Hedonic Methods and Housing Markets

Chapter 5: Theoretical background and demand estimation

Edward Coulson
Department of Economics
Penn State University
fyj@psu.edu

CHAPTER 5: THEORETICAL BACKGROUND AND DEMAND ESTIMATION

5.1 The theory of equilibrium hedonic prices

This chapter is about the theoretical foundations of hedonic pricing and the modeling of attribute demand. It is far from comprehensive, and is included mostly to serve as foundation for discussing the empirical methods and results of previous chapters. It is, on that account, not rigorous and certainly not complete. Interested readers will find more rigorous and more general accounts of hedonic pricing theory and nonlinear pricing in general in Rosen(1974), Edlefsen (1978), Blomquist (1988), Epple (1987), Ekeland, Heckman and Nesheim (2004) and others.

We begin by assuming that there is one differentiated commodity, labeled housing, which is characterized by a set, or vector, of embodied attributes, x (i.e. x represents a collection of housing attributes, x_1, x_2, x_3, and so on, which might represent housing attributes such as square feet of interior space, square feet of exterior yard, and number of bathrooms, and so on). There is also assumed to be one undifferentiated numeraire commodity, z. The utility function is then

$$U=U(x, z) \qquad (2)$$

which is assumed to have the usual continuity and convexity properties[1]. We ask the question, how much is the individual willing to pay for a house with a specified vector, x*? The answer to this question of course depends on the alternatives, and for now we simply specify that there is an exogenous level of utility that is always available to the consumer in some other housing market. This level of happiness must depend on one's income, so for a specified level of income y^0, this level of utility is given as u^0. The price of z is normalized to be one, and so the individuals budget constraint is

$$z + B = y^0 \qquad (3)$$

where B is the *bid-rent,* the consumer's willingness to pay for x*. B solves

$$u^0=U(x^*, y^0-B) \qquad (4)$$

---

[1]Ignoring the problem that the attributes might not be measured continuously

which is inverted to give the bid-rent function

$$B=B(x^{*},u^{0},y^{0}). \qquad (5)$$

The amount that a household will pay for housing depends on the characteristics, the available alternatives in other housing markets (as measured by u^{0}) and the income level. This bid-rent function can then be characterized by the implicit differentiation of the utility function. Where the subscripts of B and U represent the partial derivatives of the indicated argument, we find, following Rosen (1974), that

$$B_x=U_x/U_z \qquad (6)$$

$$B_y= 1 \qquad (7)$$

$$B_u= -U_z \qquad (8)$$

Equation (6) shows that the shape of the bid-rent function is governed by the properties of the utility function. As Figure 1 demonstrates, the assumption of a strictly diminishing marginal rate of substitution means that B will increase with x, but at a decreasing rate. Each successive unit of x is of less and less value to the consumer. We can interpret the slope B_x as the price that the consumer is willing to pay for an incremental increase in the embodied quantity of x in the house, or in other words, as the hedonic price of x. Given this interpretation, equation (6) is analogous to the usual utility maximization condition which equates the price ratio of two commodities to the marginal rate of substitution between the same two commodities.

Equation (7) simply states that if utility levels are to be held constant, any increase in income must be reflected in a dollar-for-dollar increase in bids. Only then can the arguments of the utility function remain the same. Equation (8) says that a rise in the exogenous utility level results in a lowering of the bid, for only with such a reduction can the higher level of utility be achieved.

We now discuss the important step of translating these statements about *individual bid-rents* into statements about the *market hedonic* function. Additional assumptions are required about:

1. the number of consumer groups (and whether there is a continuous or discrete number of groups).
2. if there are multiple groups, the source(s) of the heterogeneity.

3. how the levels of utility are determined for each group

4. the source of supply– whether it is provided as a fixed exogenous stock or supplied endogenously.

5. If supply is endogenous, the parameters of firm behavior

As in Shepherd (1999) a few special cases of interest are examined.

Case 1: Homogeneous consumers; fixed utility

The simplest case is when consumers are identical in their resources and utility functions. Under this scenario, households bid and possibly rebid for houses with various level of x and market equilibrium is obtained when all households end up with the same level of utility. Prices for individual housing units rise and fall in the usual fashion in response to supply and demand stimuli until supply and demand are equalized for housing units at all levels of x. Then we can immediately see that the *marke*t price for housing units with different levels of x will exactly replicate the (common-to-all) *individual* bid-rent function of the type pictured in Figure 1 with the utility set according to its exogenous level. (This works because the tradeoff between price and x traced out in Figure 1 represents equal utility.) Note that the assumptions about supply are irrelevant because the restrictions on the consumers are sufficient to establish this as the equilibrium. Also note that if this assumption were fulfilled we would have very definite restrictions on the functional form of the hedonic function-- that is to say that the market hedonic relationship between housing price and housing characteristics would be such that as characteristic levels rise, the price of the housing unit rises less than proportionately, thus for example the semilogarithmic functional form would be inappropriate

Case 2: Heterogeneous consumers, exogenous supply

Case 1 obviously suffers from a severe lack of realism. So now let consumers be heterogenous with respect to income, but in the simplest possible way: with two income groups. Consider then, the bid-rent function of consumers with income y^1 . The default exogenous utility level of such consumers will be assumed to be u^1. It is now not necessary that all consumers have the same utility level. In order for both types of consumers to be able to obtain housing, the lower income group will need to outbid the higher income group for some of the housing. Therefore in equilibrium the two bid-rent curves will resemble those in Figure 2. Since houses will be allocated to whoever makes the highest bid, the level of attribute x* represents the point where the transition from one income group to another takes place. Note that $B(x^*,u^0,y^0)=B(x^*,u^1,y^1)$, so that the group with the higher income is consuming a greater amount of the numeraire good. On the other hand $B_x(x^*,u^0,y^0)<B_x(x^*,u^1,y^1)$. Thus type 1 individuals have a higher (in absolute value) marginal rate of substitution. Assuming that an increase in z, holding x constant, leads to a higher marginal rate of substitution, then it is clear that $y^1>y^o$, and correspondingly that $u^1>u^0$.[2]

In this simple two-group scenario, the allocation of housing consists of a stratification of the attribute x across income groups. Low income households occupy dwellings with $x<x^*$, and high income households occupy those with $x>x^*$. There is a clear generalization to multiple discrete income groups. The allocation will involve further stratification, which is governed by the fact that the relation between any two groups is as pictured in Figure 2, so that higher income groups will purchase greater levels of x. As the number of groups gets sufficiently large, it is presumably well-approximated by a continuous income density, and this is the tack taken by a number of authors, including Quigley (1982)

Similarly, if the heterogeneity arises because of differences in preferences, one could allow those preferences to be modeled by a parametric utility function where the parameters vary according to some density (and incomes are identical). Epple (1987) is a leading example of this modeling choice. In such a scheme, a greater value of the utility parameter leads to a higher marginal rate of substitution when the amount of x and z are the same. Thus, in Figure 2, replace y^j with y (i.e. the two groups have the same income), but reinterpret u^1 as being the utility received by the group with the higher parameter level. At x^*, the two groups have the same value of x and z (because they have the same income) but the slope of the bid-rent function will be higher for group 1, and so the same sort of stratification, and generalization to multiple groups, occurs.

However it has yet to be answered *which* utility levels will be obtained. One way of thinking about this question is to ask whether the housing market is in an "open" or "closed" city in the sense of Wheaton

[2]This condition (that the marginal rate of substitution increases uniformly with an increase in the commodity whose marginal utility is in the denominator of the mrs expression) is sometimes called Wheaton's Condition, having been introduced in Wheaton (1977). A sufficient condition for this is $U_{xz}\geq 0$.

(1977). In an open city, the utility level is given exogenously by the worldwide level of utility. When the utility level deviates from the world level, households enter and exit, causing the bid functions to rise and fall, and the utility level (for each income group) to adjust accordingly. This would seem to be a long-run kind of model, because it is likely that migration in response to housing price differences occurs only with a time lag. So only in the long run can such in- or out-migration have an impact on the structure of hedonic prices. This framework would seem to a certain extent incompatible with the notion of a fixed housing stock.

A more common assumption is an alternative framework, which is to assume a closed city–i.e. a fixed number of households, along with a fixed housing supply. Utility levels are then determined endogenously. The supply of housing is fixed, and characterized by the density of the attribute x.

Assume that the attribute x is continuous, and that therefore is characterized by a density function f(x). It is convenient, following the examples of Rosen (1974) and Quigley (1982) to truncate the values of x at upper and lower bounds x^u and x^L. Again following Quigley (1982) let the measure of houses be H, and the measure of people be N<H. The opportunity cost of leaving any unit vacant is assumed to be zero. Competition among landlords will ensure that among the N houses which will be rented, the vacant units are those with the lowest amount of x (Quigley, 1982). Let x_ be the level of housing such that

$$\int_{x\_}^{x^U} f(x)dx = N \qquad (9)$$

that is, that the quality of occupied units will run from x_ up to the highest level of x, and the vacant units will have levels of x from x^l up to x_. The highest bid for units with x_ will be zero, and it will come from those with the lowest income.

Then let income be distributed continuously along a bounded interval from y^L through y^h. As long as x is a normal good[3], there will then be a monotonic mapping from x to y. Let this mapping be y=M(x). Then for each type of household, characterized by y*, the mapping M determines the amount of x=x* purchased:

---

[3]Wheaton's condition above is sufficient to ensure this property.

$$\int_{x\_}^{x^*} f(x)dx = \int_{y'}^{y^*} g(y)dy \qquad (10)$$

so that solving for y* in terms of x* in (8) creates the matching function y=M(x). Using the mapping, the bid function becomes B=B(M(x*),x*, u). Recalling that the hedonic function is the upper envelope of the bid functions we know that

$$H(x^*)=B(M(x^*),x^*, u) \qquad (11)$$

and, since the hedonic function is tangent to the bid function at the consumer optimum

$$H_x(x^*)=H_x(M(x^*))=B_x(x^*)=U_x(x^*,M(x^*)-H(x^*))/U_z(x^*,y^*-H(x^*)) \qquad (12)$$

so that the hedonic function can be solved by integrating up to y*, using the initial condition that $H(x^L)=0$.

It can be very difficult to solve this analytically. Epple (1987) makes the most notable contribution, following contributions to the labor economics literature by Tinbergen and Sattinger. It is worth examining his example in some detail, although his setup is a bit different than that above. As noted, in Epple's model, consumer heterogeneity is across taste, rather than income. Importantly, Epple generalizes to allow x to be multi-dimensional. The utility function is linear in the composite good and quadratic in the housing attributes:

$$U(x,z,\alpha) = \frac{1}{2}(x-\alpha)'\theta(x-\alpha)+z \qquad (13)$$

where α is a kx1 vector of taste parameters. Note that z enters linearly into the utility function; it appears to be very difficult to obtain analytical solutions unless this level of simplicity is assumed. The α are assumed to be distributed normally across the population with density $N(\tau,\Sigma)$. The matrix θ is positive definite. The supply of x is assumed, as in the description of the short run hedonic, to be fixed with distribution $N(\xi,\Omega)$. It is then possible to solve for housing price as a function of the attributes, assuming similar conditions to those described above. This solution is given by Epple[4] as

$$P(x) = \psi' x + x' \Pi x \qquad (14)$$

where

$$\psi = \theta(\tau - \Sigma^{1/2}\Omega^{1/2}\xi)$$

and

$$\Pi = -\theta(I - \Sigma^{1/2}\Omega^{-1/2})$$

so that the hedonic price of z that is obtained in the market is given by the derivative of P with respect to x:

$$P_x(x) = \psi + \Pi x \qquad (15)$$

For purposes of discussion suppose (again) that x is one-dimensional. The shape of the price function P(x) will be upward sloping, but could be concave, linear, or convex depending on the sign of Π, which in turn is opposite the sign of 1-sqrt(Σ/Ω). (Recall that θ must be positive.) Π is negative or positive depending on the relative sizes of the two numbers under the radical sign. Suppose for example that Σ is

---

[4]Ekeland, Heckman and Nesheim (2004) demonstrate that since this setup relies on quadratic utility and normally distributed characteristics, the price index can be derived in a straightforward fashion by matching the moments of the distributions.

small and Ω is large; in that case the unit price of x declines as more and more units of x are obtained. This can be interpreted as follows. Note that $\Sigma^{1/2}$ is the standard deviation of α, and α is positively associated with the marginal utility of the attribute[5], whereas $\Omega^{1/2}$ is the standard deviation of the attribute distribution. Moving from houses with (say) average amounts of x to those with incrementally larger amounts of x will engender relatively little response from consumers (because consumer preferences stray very little from the mean) but will have relatively greater supply reaction (because there is a wide distribution of the attribute). The usual mechanics of supply and demand suggest, therefore, a falling marginal price. The reverse intution is also available. Note in addition that a linear price index is in theory very rare, requiring that the two standard deviations be identical.

In any case, it can be seen the hedonic parameters are themselves functions of the parameters of the utility function, and the distributions of housing stock and tastes. Thus if it is known that the distribution of taste parameters and housing attributes are normal, and the utility function has the given linear/quadratic form, one can know the functional form of the hedonic price accordingly (i.e. that it is a quadratic function). Thus, it is possible to have the hedonic function be internally consistent with utility functions that have sensible properties, and after estimating it with market data, potentially recover the parameters of the utility function from the observations in the housing market.

However, as Ekeland, Heckman and Nesheim (2004) observe, the justification of the linear-quadratic functional form is based on rather strict assumptions about those utility (and supply) distributions, and that small changes in those underlying distributions can create large changes in the hedonic function. Their particular example is to let the utility parameters be distributed as a mixture of normals; even when the amount of "mixing" is small, the deduced hedonic function gets very far away from the linear quadratic form.

Note further that both supply and demand are more or less fixed in place, demand by the distribution of preference parameters and supply by the distribution parameters of the attributes. This supposes that the supply of housing, and therefore the supply of housing attributes, is fixed in place, exogenous. This simplifies the analysis because it reduces the problem to one of matching.

---

[5] The marginal utility for x is (a-x)

Case 3: Heterogenous consumers, endogenous, but homogenous, supply

It may be desirable to allow the supply of attributes to be endogenized. Here's why: the prices of the attributes, determined within a framework such as Epple's, may have little to do with the marginal cost of providing them to the market. Thus the equilibrium price of an attribute that comes out of equation (14) may be (say) higher than the cost of providing them. In the case of a physical attribute such as the interior square feet of the house, or a swimming pool, there are arbitrage profits to be made from supplying such attributes to the market. A model which takes this kind of thing into account needs suppliers (builders) to respond to the discrepancy between price and cost. The market may have one type of supplier, or multiple types. If one type , then these identical, competitive builders supply attributes to the market until the attribute price falls to the marginal cost of production, all profits are equal to zero, and the hedonic function will be identical to the builder cost function. Under the usual assumptions about cost, particularly the law of diminishing returns, this would imply a convex hedonic function, and a hedonic price that increases at an increasing rate (such as semilogarthmic).

Case 4: Heterogenous consumers, endogenous, heterogenous, supply

If more than one type of builder exists, the model would be constructed with heterogeneity in the builders cost function. Some of the builders would then have a comparative advantage for small amounts of x and some for large amounts of x. They will be stratified along the attribute dimension, and each builder will supply units with attribute levels at which he has a comparative advantage, in a manner similar to the way consumers stratify themselves across the amount of attribute according to their income or taste. The hedonic function would then be the envelope of both cost and bid functions, as consumers and builders are matched at particular attribute levels, as in Figure 4. As in the case with exogenous supply and heterogenous demands, the functional form of the hedonic is indeterminate, and will depend on the relative variation in firm and consumer charactersitics. Ekeland, Heckman, and Nesheim (2004) provide an example for this case (albeit in a labor market context) using underlying distributions similar to those employed by Epple (1987).

Models with endogenous supply are applicable only in the long run, as it requires a significant adjustment of the housing stock in order for such adjustments to take hold. As such, it does not seem to have found much favor in the literature. In part this may due to the difficulty of modeling such behavior, but also perhaps because in many estimates, the hedonic prices do not seem to relate to construction costs. On that account the short-run model with fixed housing stock may be preferable. (See Chapter 2 as well.)

5. 2 The modeling of hedonic demand

In the original hedonic estimation of a developed housing market by Ridker and Henning (1967) special emphasis was laid on the estimation of the coefficient of a characteristic denoted SUL, an index of the concentration of various sulfates and sulfites in the air, variation in which they were able to measure across a number of neighborhoods in the St. Louis area.

It is no surprise that air pollution was one of the first commodities on which hedonic estimation oncentrated, because air quality is a public good. Public goods, in their purest form possess two qualities that make them different from the more ordinary private goods.

First, they are non-excludable; it is technologically infeasible (or at least quite costly) to exclude someone for consuming it, which in turn means that property rights to those commodities are not well-defined. You can't charge someone for breathing St. Louis air, clean or not. On that account, it is difficult or impossible for such good to be provided by private firms because it would be impossible to charge customers for its use. No private firm would try to provide the commodity 'air cleansing' to the residents of St. Louis because, having rid the air of sulfur, people would be able to breathe it in whether or not they paid the firm. People would be *free-riders*. While economists have developed a number of methods to overcome the free-rider problem, the quickest way around it is for the government to step in to provide such commodities because only with their power to levy compulsory taxes can a non-excludable good get paid for.

It's a large assumption but let us assume that the government wishes to provide the "right" amount of air pollution. What would this be? It would be the amount that the public would purchase if it were a commodity in a market- that is, where the intersection of the supply and demand curves occurs. The supply curve can be interpreted as the marginal cost curve for pollution reduction; we can assume this is known.

But we have more difficulty figuring out the demand curve, since there is no market there is no way of finding out how much pollution reduction people want at various prices. You might try to just ask them, but if they think that their answers have no bearing on what they would actually end up paying (in increased taxes) they would surely have a tendency to exaggerate, and if they thought their answers actually had a bearing on their tax payments they would understate their true valuation, and hope to free ride on others' tax payments.

But as long as there is some geographic variation in the levels of the public good, then the hedonic price of air pollution can be estimated., because there is no free riding in the land market. When you live in unpolluted air, you get to breathe unpolluted air, but you have to pay a price for it, a price that must be related to the benefit you receive from it. The air pollution market, such as it is, manifests itself indirectly through the housing market. Disentangling it from its larger context is just the thing that hedonic analysis is designed to do.

The second interesting characteristic of pure public goods is their non-rivalry. Non-rivalry essentially means that one person's consumption of the commodity does not impinge on another person's. Clearly this is not true of most commodities, because when you and I share a sandwich we must each consume half. But it is true of air, because you do not suffer any shortness of breath when I inhale. The importance of this quality for the interpretation of hedonic estimates can be illustrated by Ridker and Henning's analysis. As mentioned in Chapter 1, in their favorite specification of the hedonic equation the authors settle on a coefficient of -$245. That is, in comparing two neighborhoods of otherwise identical features, the one with one unit (as measured by the index) of additional concentration of sulfates in the air would have housing prices that were $245 less than the less polluted area. The direct inference is since households are willing to pay (at the margin) $245 per index unit to move from a polluted area to one that is relatively unpolluted, and so they would equivalently be willing to pay $245 to have each index unit of pollution removed from the air where they currently reside[6]. Thus the benefit to that household of cleaner

---

[6]In making these kinds of statements, the issue of moving costs that would be incurred from actually moving across neighborhoods is neglected.

air would of course be $245 per unit. The authors then go on to surmise that this benefit applies to every one of the 338,000 households in St. Louis, so that the aggregate benefit to be gained from (say) a two unit reduction in pollutant concentration is reckoned to be 2*245*338000=$165,620,000.

Note that it is only because of the non-rivalry of air that this calculation can even be suggested. With a private commodity, saying that the market price of a sandwich is $5 means that if two people are buying the sandwich, you need two sandwiohes to meet the demand (or each of them will pay $2.50, if each gets half). With air quality, we observe that the price for an individual household to get clean air is $245, but aggregate this up to the entire market we add all of the $245s up to get the aggregate benefit. This calculation can be contemplated because air's non-rivalry means that everybody consumes all of the clean air therefore everybody is willing to pay that full price.

Does this mean that Ridker and Henning's calculations are correct? Can one obtain the benefits of a hedonic good through this simple multiplication? It can't be that simple.

The endogeneity of population

Scotchmer (1985, 1986) and Kanemeto (1988) addressed the question of whether a change in property values as suggested by the hedonic coefficient actually reflects the benefit of an amenity. The basic issue raised in these papers is straightforward. When you estimate the aggregate benefit by multiplying the hedonic price by the number of people or households, the number of people or households is assumed to be exogenous in the face of variation in the amount of the amenity. This may be an invalid assumption. That is, it is possible areas with higher levels of amenities are going to have higher densities of population.

A telling example arises in the standard monocentric model. Suppose there is an improvement in the transportation system such that the per mile cost of transportation falls. In the absence of any further changes the benefits from such an infrastructure improvement could be measure by the aggregate reduction in transportation costs, which could in turn be measured by the rise in property prices at various distances from the CBD. Adding up all of these value changes would provide an estimate of the aggregate benefit of the improvement. But this would only be true if the allocation of land remained constant in the face of the transport improvement. But the theory itself says that this allocation is inefficient; the demand for land would fall near to the central city and rise in the suburbs. The "density gradient" would flatten.

More generally, the equilibrium lot size is smaller where quality of life is higher. Thus even in the simple world where individuals are homogenous, and the hedonic price index is linear, as long as people are mobile across regions and housing attributes are freely chosen, the benefits will be mis-stated (Scotchmer, 1985, 1986).

There is a straightforward counterargument to this-- that attributes are not, in fact freely chosen, but rather, as discussed in Chapter 3, are fixed. One of these fixed attributes is the size of the lot (and the number of such lots), so that in effect the population size to which the amenity improvement apply is fixed. Ridker and Henning themselves recognize that this assumption needs to be made in order for their calculation to make sense.

A second counterargument is that if the demands for space and amenity are independent of each other, then the simple multiplication might be valid. But this would seem to put an unreasonable restriction on the form of the utility function, although if it were true it would be possible under some circumstances to allow the estimation of long run benefits.

Demand, not price

The second problem which arises is that the observed market price (of say $245) gives a very incomplete picture of the benefits which are received from any commodity. It takes the entire demand curve to tell the story. The very standard demand curve in Figure 5 represents an individual's willingness to pay for incremental amounts of air quality. As such it presents the same information as the bid-rent function, the difference being that the bid -rentcurve represents the total willingness to pay for all of the hedonic good, while the demand curve shows the marginal willingness to pay for incremental amounts of the commodity. It is the derivative of the bid function, and, it is important to note, because the bid function represents willingness under a constant utility level, so does the demand curve in Figure 1, hence it is a Hicksian demand curve.

The point is that the total benefit to the individual of getting a level of air cleanup Q at price P must be calculated on a unit-by-unit basis. The first bit of air cleanup is valued at P*, but the price is still only P, hence the net benefit is P*-P. Because of diminishing marginal utility, the second bit of air cleanup is valued at slightly less than P*, and the price is still P, and so on, so that the total benefit to this individual can be seen as the triangle P*AP. From this it follows that if the hedonic price is simply observed to be $245, there is no way of properly gauging the total benefit, because we cannot observe what valuations are at other prices. We only observe the single point on the demand curve where the price is $245, and not the entire thing. In order to know the whole demand curve we must observe what people do when faced with different prices, but in the scenario just described there is no opportunity to do that.

In summary, knowledge of the entire demand curve is required, but in the scenario described, only one point of that demand curve is observed. Imagine running a regression that purports to estimate the (inverse) demand curve:

$$p_x = \alpha_0 + \alpha_1 x + \alpha_2 (y - p) + \alpha_3 z + v \qquad (22)$$

This demand curve (as noted) is the derivative of the winning bid function with respect to the amount of the attribute, and so the dependent variable is that derivative. The arguments are the same as those in the bid-rent function, the amount of the attribute x, the non-housing expenditure, y-p, and demographic variables z. These demographic variables are included because the default utility is unobserved but will be a function of these characteristics (such as age, race family size, education, etc.). To belabor the same point, estimating such a regression is impossible if only one value of p_x is observed. Thus a simple linear hedonic regression is not going to get the job done.

5.3 Estimating demand curves

How does the investigator obtain variation in the hedonic price? One obvious way is to have data from more than one market. If the data permits estimation from more than one housing market– and following the discussion in Chapter 4 this means data from multiple metropolitan areas or perhaps neighborhoods, then separate hedonic parameters for, say, air pollution can be estimated. This provides variation in the dependent variable in the demand curve (22), and this can be regressed on the individual values for x and the remaining independent variables. The parameters of the demand curve can be estimated.

In the estimation of demand equations, economists often worry about *simultaneous equations bias*. In simple terms, what this means is that when we run a regression of price on quantity, we are actually conflating two different kinds of behaviors: the tendency of consumers to *lower* their desired quantities in the face of higher prices, and the tendency of suppliers to *raise* their desired amount of production in the face of higher prices. In such a situation it is unclear how to interpret the coefficient α_1 in the demand regression above. However, in estimating hedonic demand equations in housing markets, it is often (though not always) acceptable to ignore this kind of difficulty, by appealing to the short-run model above, where there is no significant supply response. When the hedonic prices are obtained by mapping the distribution of an existing housing stock to the preferences of consumers then the demand curve is readily estimable by the method outlined above.

Note that because we are estimating a demand curve that is constant across markets, we are necessarily assuming that the demand parameters are the same across the markets. Therefore it can only be the supply-side that varies across markets, but as noted, it does so not because construction costs or builder characteristics vary, but because the history of construction patterns is different.

Thus, Figure 2 illustrates the estimation. In this Figure three *linear* hedonic price regressions for three markets have been estimated, yielding three hedonic prices for the attribute x: P_{x1}, P_{x2} and $P_{x3.}$ We observe this price, and the quantity purchased at each price, Q_1, Q_2, and Q_3 by consumers with demographic variables z^* and the income measure $(y\text{-}p)^*$, and this allows us to trace out the points of the demand curve D(z, y-p) using the standard regression tools. We can also measure the response of demand to changes in income and other variables. The intercept of the demand curve is (for any particular individual)

$$\alpha_0 + \alpha_2(y - p) + \alpha_3 z + v$$

and so it shifts vertically in response to changes in income, z or the error term, ν, and thus shifts quantity to the left or right along the appropriate horizontal price line. In this way all of the parameters can be estimated– in econometrics jargon, the parameters are identified.

As Chapter 2 discussed at some length, there is no reason to presume that the hedonic regression is linear, indeed there are good reasons for assuming nonlinearity, and this nonlinearity creates both opportunities and pitfalls for the estimation of hedonic demand curves. One of the things that happens is that as the quantity changes, so does the (marginal) price price paid, so that even within a single market, we can observe price variation. Figure three demonstrates. The upward sloping line P_x is the graphical representation of a concave hedonic price estimation, while the three downward sloping lines are bid functions for three consumers with varying intercepts (that is, varying demographics). Note that the three households face three different prices at three different quantities[7]. Therefore variation in prices, demographics, etc. are all observed and it appears that the demand curves can be traced out accordingly. But, as noted, with the opportunity to observe price variation comes serious pitfalls.

The first concerns functional form. In the example in Epple (1987) a quadratic hedonic price function is shown to be congruent with a quadratic utility function, and therefore linear bid/demand curves. Simplifying this to a single attribute, the hedonic price regression is

$$P = \beta_0 + \beta_1 X + \beta_2 X^2$$

whereupon the hedonic price is

$$P_x = \beta_1 + 2\beta_2 X$$

On the one hand, the nonlinearity of the hedonic creates variation in the marginal hedonic price of the commodity. That is an opportunity, since it would appear to be no longer necessary to obtain data from multiple markets. But, since the hedonic price is a linear function of the characteristic (and only the characteristic) the demand regression (22) would immediately break down. Both the right and left hand sides of (22) are linear functions of x. Thus the α parameters of the demand equation would simply replicate the parameters of the price calculation ($\alpha_0= \beta_1$,$\alpha_1=2\beta_2$ $\alpha_3=\alpha_4=0$) and the errors would all be zero. We are back to square one: variation from the single market is not sufficient to identify the demand function. But

[7]It should be remembered that the line P_x is not the supply curve but merely the set of prices faced by consumers in the housing market (nor, obviously would it be the demand curve if it were derived from a convex hedonic function). This was the cause of some confusion in the early literature on this subject.

as Ekeland, Heckman, and Nesheim (2004) have emphasized, the import of this result says as much about the unsuitability of the linear-quadratic-normal assumptions maintained by Epple (1987) as anything else.

We return to this issue in due course, but it is appropriate to discuss a second issue that has received attention in the literature. Epple (1987) and particularly Bartik (1987) emphasize the fact that when the hedonic price regression is nonlinear, and non-constant prices are available to the households, households choose price and quantity simultaneously. Thus, as their demographics shift their bid functions back and forth, they (in the case of the concave pricing given in Figure 7) automatically lower the marginal price as they lower their quantity. This would not be a problem if all of the variables that govern those shifts were observable. But suppose not. Then the error term would contain (say) an unobservable taste component. Now imagine that the shifts in the bid that are given in Figure 7 were caused by this taste component. Since the price and quantity are jointly chosen, the quantity variable on the right hand side is correlated with the unobserved taste components, and on that account the coefficient of quantity, x , is biased. The problem is similar to the "omitted attributes problem" discussed in Chapter 2.

We need a replacement for quantity, an "instrumental variable" to take the place of quantity in the demand regression. Such an instrument must be uncorrelated with taste (because if it weren't it would give a biased coefficient as well) but must be significantly correlated with quantity (because if it weren't it would not be a very capable substitute). Bartik's (1987) solution to this is interesting; he notes that the quantity that is chosen depends not only on the tastes of the household but the budget constraint as well, so variables that move the budget constraint would be suitable instruments. One obvious candidate is income, y. This works because income is not part of the demand equation, (y-p) is[8]. But also, if data is indeed available from more than one market, and separate hedonic regressions are estimated for each market, then indicator variables for each of the markets will be appropriate instruments, as will indicator variables multiplied by other variables (such as income).

Epple (1987) expands this analysis in several directions, including estimating more complete systems

---

[8]And in fact because p is part of that expression, and is jointly chosen, y-p also has a biased coefficient, and needs its own instrument.

of demand equation (i.e. estimating demand equations for several characteristics simultaneously) , and cases where the attributes are missing or imperfectly measured. The general idea is that you need at least as many instruments as you have "bad" variables in the demand equation, and that this will be more likely to occur when you have lots of hedonic markets with distinct parameters.

However, all of analysis in Epple (1984) was conducted using quadratic utility functions and normal distributions of taste parameters and characheristics. Such cases are solvable in closed form, and are on that account quite enlightening. In some recent work Ekeland, Heckman and Nesheim (2004) have asked what can be gained when those assumptions on functional form are loosened. The short answer is that our need of parameter variation from multiple markets is vastly reduced when the hedonic index is "more nonlinear". We can obtain sufficient price variation from a single hedonic market to identify the demand curve, using nonlinear functions of the instruments (such as demographic characteristics) as well. In fact, they argue that because linearity is "nongeneric" (meaning, intuitively, that if you picked the functional form "at random" it would almost surely not be linear) it is a very poor assumption to make when there are no theoretical reasons to do so.

Applications

There have been a few applications of the above Epple-Bartik modeling strategy applied to the problem of estimating demand for environmental and other "neighborhood" characteristics, including Coulson and Bond (1990), Cheshire and Sheppard (2000), Coulson, Hwang and Imai (2003), and Kiel and Zabel (2000), the last of which is a helpful case study.

These authors wish to estimate the demand curve for air quality, and follow the two-stage process discussed above. They use American Housing Survey data to estimate hedonic price indexes for four US cities (Chicago, Denver, Philadelphia and Washington, DC) thus obtaining cross-sectional variation in the relevant hedonic parameters. Notably they also have data over 14 years of the AHS sample waves (1974 through 1991) so that they can allow the parameters to vary over time as well. The hedonic regression has structural characteristics such as age, room composition, and structural amenities such as air conditioning. The authors also take advantage of their special access to AHS files to observe each individual housing unit's census tract, and from this create several tract-based neighborhood variables, such as average income. Finally, they include four measures of air pollution. They estimate a linear, log-linear and semilog version of the hedonic price index for each market[9]. Since the regressions are stratified by both time and place, the linear model induces a formidable amount of price variation across the market, even though the linear model allows no within-market variation in the price.

The demand equations (one for each pollutant) use the marginal price derived from the hedonic equation as the dependent variable; in the case of their double-log regressions this is (as noted in Chapter 2) the estimated coefficient times the total house price, divided by the attribute quantity. This is regressed on the appropriate quantity of pollutant and several individual or household characteristics including race indicators, marital and gender indicators as well as income. As noted the quantity is endogenous so instrumental variables estimation is required. The authors perform a two-stage least squares procedure in which the quantities are regressed on city and time dummies, which works because the hedonic index varies with both time and place. The fitted values from these preliminary regressions are used in place of the quantities themselves.

For the most part the demand regressions give back coefficient values which are sensible, in particular the quantity variables have negative signs, as would be expected. Having estimated the demand curves, Kiel and Zabel are in a position to estimate the benefits from reduced pollution. For any given metropolitan area, they calculate the reduction in the given pollutant needed to bring the air quality into compliance with national air quality standard guidelines. They then integrate the willingness to pay curve (i.e. the demand curve from Figure 3) over the region from the current level of pollutant to the standard level, for each census tract. This figure is then multiplied by the number of households in the tract. Calculating this for each census tract, and aggregating across census tracts provides a benefit calculation for

---

[9]Actually they estimate a model of changes in housing values, arguing that the first-differencing eliminates unobservable housing characteristics (thus ameliorating the missing data problem discussed earlier).

the reduction in that pollutant for each MSA. The authors find that this value depends of course on the functional form and specification of the demand function and the original hedonic index. For Chicago, the most highly parameterized forms of the model yield a benefit calculation of approximately $300 million to $500 million.

5.4 Further Developments

The past several years have seen several new approaches that are not always in the strictest sense hedonic, do approach the problem of uncovering information about household preferences for hedonic commodities, particularly public and neighborhood characteristics.

Perhaps closest to the methods outlined above is Bajari and Kahn (2005)[10]. Unlike the demand estimation discussed previously, these authors specify the functional form of the utility function (as log-linear to be precise) but allow the values of the parameters to vary across individuals. With a logarithmic estimate of the hedonic function in hand as well, they can solve for (not estimate) the value of the utility parameter *for each household* using a version of equation (6) above. They transform this to a willingness-to-pay measure (again relying on functional form assumptions) by multiplying the utility parameter by the relevant change in attribute level. They then can regress this willingness-to-pay measure on a vector of demographic characteristics to obtain the joint distribution of tastes and household attributes. This last regression is thus like the demand estimation above, but it avoids the problem of instrumenting the hedonic attribute level since that level is absent from the specification. They apply their method to the question of how black and white households end up in different locations. They ask, for example, whether whites are willing to pay more for larger housing units, and find that this is the case, and suggest that this is a reason, (along with their greater willingness to pay for educated neighbors, and black households' higher demand for residential locations nearer central city workplaces) for the concentration of blacks and whites in different parts of metropolitan areas.

Epple and Sieg (1999) take a rather different approach to modeling the allocation of households to locations. The focus on the idea that hedonic equilibria make use of the "single-crossing" properties of

[10]See also Bajari and Benkard (2005).

(some) utility functions, as in Figures 2,3 and 4. When the attribute in question is a local public good, the stratification is observable as self-selection into particular communites. With knowledge of the crossing points, the model can predict the distribution of income within each of the communities, again using assumptions on the functional forms of the distributions of income and tastes, as well as the utility function. The parameters of these functions are then estimated by matching the theoretical distribution of income in each jurisdiction with the actual distribution of income. With this, one can also predict each communities level of public good provision, although this is not derived from hedonic demand methods.

While Epple and Sieg (1999) is not a hedonic method, per se, Sieg, Smith, Banzhaf and Walsh (2002) discuss the implications of location self-stratification for the construction of hedonic price indexes. They estimate hedonic functions using only structural characteristics, and then include dummy variables for each community in the metropolitan are. Using a variant of the Epple-Sieg model, they propose that the ordering of the communities, according to the value of this fixed effect, ought to be reflected in the ordering of the public good provision. Furthermore they state that one's choice of functional form for the hedonic should be based on its adherence to this property. In the event, it does not seem to matter very much whether the hedonic function is log-linear, semi-log or whatever, since the ordering of the fixed effects is roughly invariant to this choice.

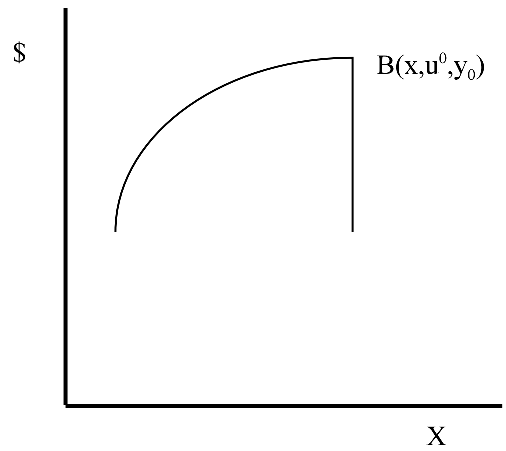

Figure 1: The Bid-Rent Function

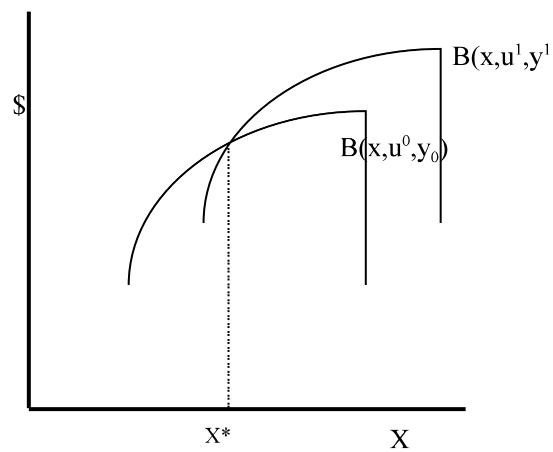

Figure 2: Heterogenous Consumers

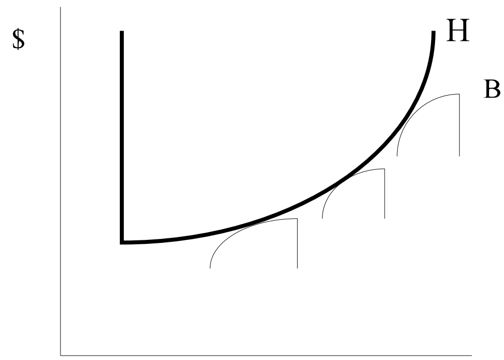

Figure 3: Hedonic Price Index with Multiple Consumer Types

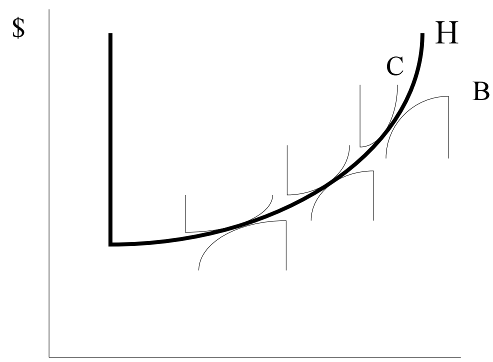

Figure 4: Hedonic Price Determination with Multiple Consumer and Firm Types

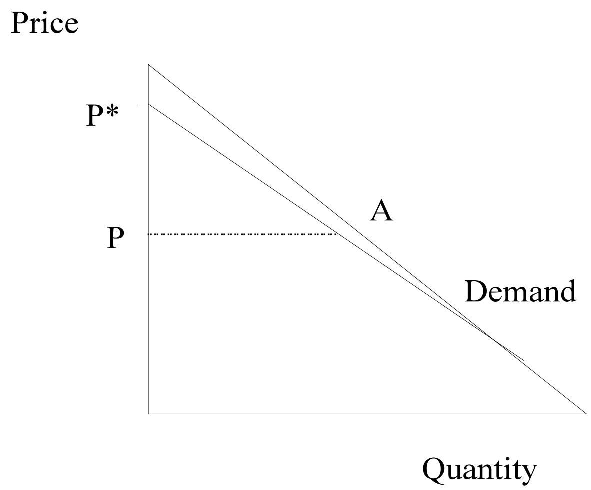

Figure 5: Demand curve

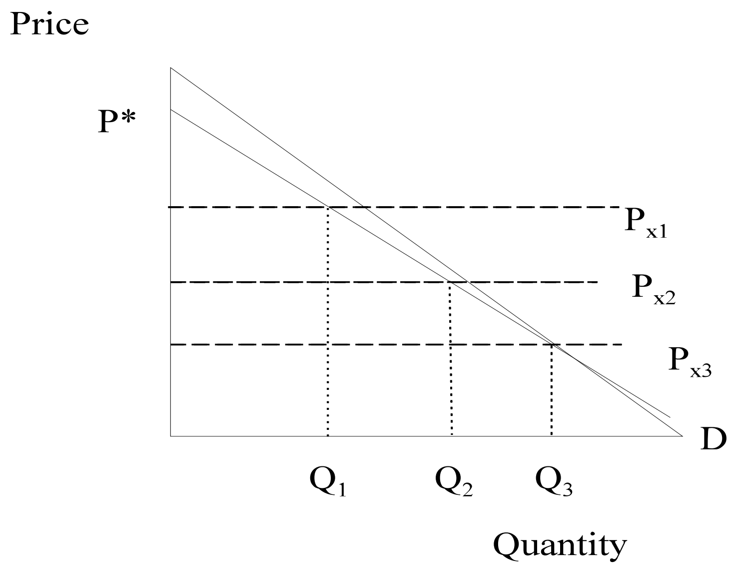

Figure 6

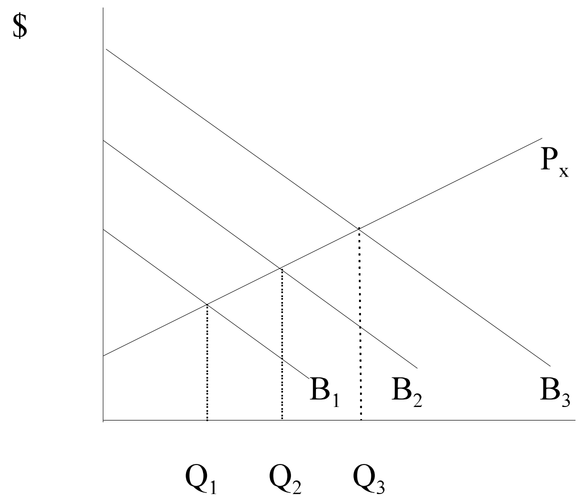

Figure 7