The Hedonic Model of Housing: An Introduction

Holger Sieg
University of Pennsylvania

Housing as a Differentiated Product

- One popular model of housing is due to Sherwin Rosen (1974) who treated housing as a differentiated product that is valued for its characteristics.
- To start out, let us assume that houses differ by a continuous index of quality (housing services), denoted by z.
- The price of a house of quality z is given by $p(z)$.
- Note that price function does not have to be linear in z.
- To understand the basic concepts, it is useful to consider a couple of stylized examples.

The Provision of Housing: Firms

- There is a continuum of firms that differ in productivity, denoted by θ.
- The costs of building a house of quality z are given by

$$C(z) = \frac{z^{\beta}}{\theta} \quad (1)$$

- For simplicity, let's assume that $\theta \sim Uniform(0, 1)$.
- Note that we need $\beta > 1$ for the cost function to be convex.
- Each firm produces one house. Profits are thus given by:

$$\Pi(z, p(z)) = p(z) - \frac{z^{\beta}}{\theta} \quad (2)$$

- Each firms chooses z to maximize profits.

Inverse Supply

- The FOC of the profit maximization problem is given by:

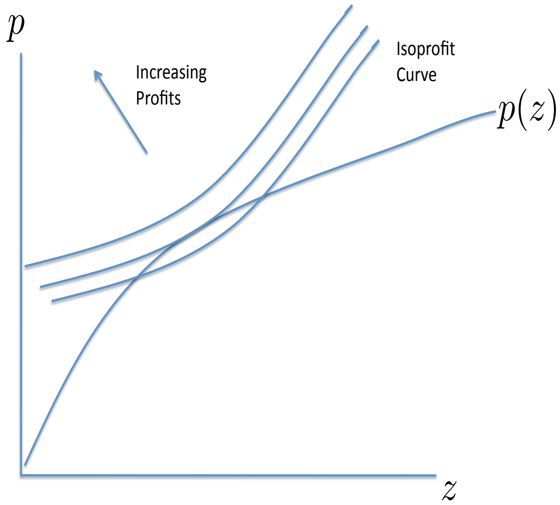
$$p'(z) = p_z = \beta \frac{z^{\beta-1}}{\theta} \tag{3}$$

- The slope of the pricing function is equal to the marginal costs of quality.
- Solving for θ gives us the inverse supply function:

$$\theta = \beta \frac{z^{\beta-1}}{p_z} \tag{4}$$

The inverse supply function tells us the productivity required for a firm to produce a house of quality z.

Hedonic Supply



Households

- Households differ in preferences for quality, $\nu \sim \mathit{Uniform}(0,1)$.
- Each household buys one house and utility is given by:

$$U(z, p(z)) \;=\; \nu\, z^{\alpha} + y \;-\; p(z) \tag{5}$$

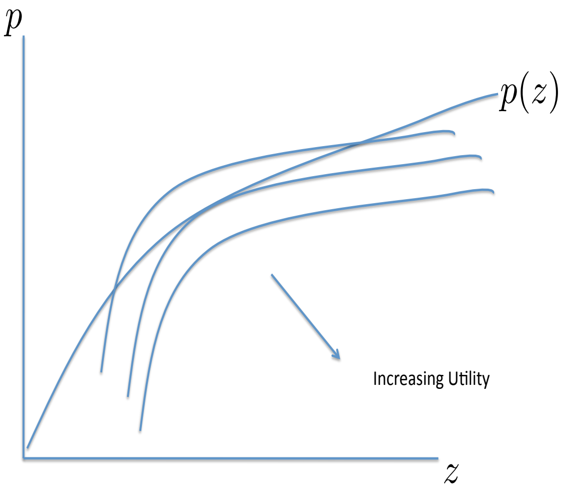
  where y denotes income.
- We need $\alpha < 1$ for the utility to be strictly concave.
- The FOC of the utility maximization problem is given by:

$$p_z \;=\; \nu\, \alpha\, z^{\alpha-1} \tag{6}$$

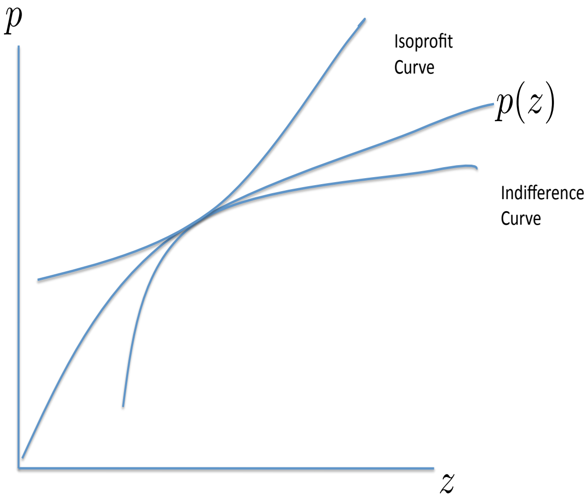
- The slope of the pricing function is equal to the marginal willingness to pay for quality.
- Solving for ν gives us the inverse demand function:

$$\nu \;=\; \frac{p_z}{\alpha\, z^{\alpha-1}} \tag{7}$$

Hedonic Demand



Matching of a Firms with a Household



Aggregate Demand and Supply

- Equilibrium requires that every firm is matched a household.
- In our model, the fraction of households that would like to purchase a house with quality less than or equal to z is given by

$$F_\nu\left(\frac{p_z}{\alpha z^{\alpha-1}}\right) = \frac{p_z}{\alpha z^{\alpha-1}} \tag{8}$$

- Similarly the fraction of firms that build houses with quality less than or equal to z is given by

$$F_\theta\left(\beta\frac{z^{\beta-1}}{p_z}\right) = \beta\frac{z^{\beta-1}}{p_z} \tag{9}$$

Equilibrium Price Function

- Equilibrium requires that demand equals supply for each z:

$$\frac{p_z}{\alpha z^{\alpha-1}} = \beta \frac{z^{\beta-1}}{p_z} \tag{10}$$

- Solving the equation above, implies that the slope of the equilibrium price function is given by:

$$p_z = (\alpha\,\beta)^{1/2}\, z^{\frac{\alpha+\beta}{2}-1} \tag{11}$$

- Integrating with respect to z gives the equilibrium price function:

$$p(z) = (\alpha\beta)^{1/2}\,\frac{2}{\alpha+\beta}\, z^{\frac{\alpha+\beta}{2}} + C \tag{12}$$

where the constant of integration is determined by the condition that the lowest productivity firm makes zero profits.

Equilibrium Housing Consumption

- Substituting the slope of the equilibrium price function into the FOC of the household problem yields:

$$z^d = \left(\frac{\alpha}{\beta}\right)^{\frac{1}{\beta-\alpha}} \nu^{\frac{2}{\beta-\alpha}} \tag{13}$$

- Similarly, firm's supply is given by:

$$z^s = \left(\frac{\alpha}{\beta}\right)^{\frac{1}{\beta-\alpha}} \theta^{\frac{2}{\beta-\alpha}} \tag{14}$$

- Note that our assumptions imply that $\beta - \alpha > 0$.
- Hence, the matching is characterized by $\nu = \theta$ along the equilibrium path.

Extending the Model

- Let us think about z as a vector of observed characteristics.
- The model can be easily extended to allow for this feature.
- Let us assume c is consumption and y income, then we have

$$\begin{aligned} \max_{c,z} \quad & U(c, z, \nu) \\ s.t. \quad & c + p(z) = y \end{aligned} \tag{15}$$

where ν is now a vector of unobserved preference parameters.

First-Order Conditions

- Substituting the budget constraint into the utility function, we obtain:

$$\max_z U(y - p(z), z, \nu) \tag{16}$$

- The FOC's are now given by:

$$U_c\,(-p_z) + U_z = 0 \tag{17}$$

or

$$p_z = \frac{U_z}{U_c} = MRS_{z,c} \tag{18}$$

- The slope of the hedonic price function is equal to the marginal willingness to pay for the attribute z.

Hedonic Equilibrium

- Similarly, we can extend the firm's problem to the multidimensional case and obtain the first-order conditions.
- We can use the FOCs to characterize the inverse demand and supply equations.
- The equilibrium conditions of the model give rise to a system of partial differential equations that typically do not have closed form solutions.
- Hence, the equilibrium price function can only be computed numerically.
- In general, the pricing function combines parameters of technology, preferences, and distributions of heterogeneity, a result that is evident in the simple example above.

The Tinbergen Model

- An exception is the linear-quadratic-normal model (Tinbergen, 1956) which is based on the following three assumptions:
  1. preferences and technology are quadratic;
  2. the heterogeneity parameters enter preferences and technology linearly,
  3. unobserved preference and technology parameters are normally distributed.
- Let's consider the Tinbergen model with $\dim(z) = 1$.
- The general model is, for example, presented in Epple (1987).

Utility with dim(z)=1

- Preferences are given by

$$U(z, p(z), \nu) \;=\; \nu \; z \;-\; \frac{1}{2} \; B \; z^2 - p(z)$$

  where $\nu = \beta y + \eta$, where y is an observed demand shifter (income) and η is an unobserved shock
- We assume that both are normally distributed such that $\nu \sim N(\mu_\nu, \sigma_\nu^2)$.
- Hence the FOC is given by

$$\nu \;-\; B \; z \;-\; p_z \;\;=\;\; 0$$

Profits with dim(z)=1

- Profits are given by

$$\Pi(z, p(z), \theta) \ = \ p(z) \ + \ \theta \ z \ - \ \frac{1}{2} \ A \ z^2$$

  where $\theta = \alpha \ x + \epsilon$, where x is an observed cost shifter and ϵ is an unobserved shock.
- Both are normally distributed such that $\theta \sim N(\mu_\theta, \sigma_\theta^2)$.
- Hence, the FOC is given by

$$p_z \ + \ \theta \ - \ A \ z \ \ = \ \ 0$$

Pricing

- Let's conjecture that pricing function is quadratic:

$$p(z) \;=\; \pi_0 + \pi_1 z + \frac{1}{2} \pi_2 z^2$$

- The FOCs for firms and households can be written as:

$$\begin{aligned} \pi_1 + \pi_2 z + \theta - A z &= 0 \\ \nu - B z - \pi_1 - \pi_2 z &= 0 \end{aligned}$$

Demand and Supply

- Hence demand is given by

$$z^d = \frac{\nu - \pi_1}{B + \pi_2} = \frac{\beta y + \eta - \pi_1}{B + \pi_2}$$

  which linear in y and normally distributed.

- And supply is given by

$$z^s = \frac{\theta + \pi_1}{A - \pi_2} = \frac{\alpha x + \epsilon + \pi_1}{A - \pi_2}$$

  which is linear in x and normally distributed.

Equilibrium

- Equilibrium requires that mean and the variances of demand and supply are the same given that both distributions are normal.
- Solving these two equations for π_1 and π_2, we obtain.

$$\pi_1 = \frac{-\mu_\theta \sigma_\nu + \mu_\nu \sigma_\theta}{\sigma_\nu + \sigma_\theta}$$

$$\pi_2 = \frac{A\sigma_\nu - B\sigma_\theta}{\sigma_\nu + \sigma_\theta}$$

- π_0 is nailed down by the initial conditions that profits have to be nonnegative.

Equilibrium Matching

- Also note that equilibrium matching function is given by:

$$\frac{\theta + \pi_1}{A - \pi_2} = \frac{\nu - \pi_1}{B + \pi_2}$$

- Plugging in our linear model, we obtain:

$$\frac{\alpha x + \epsilon + \pi_1}{A - \pi_2} = \frac{\beta y + \eta - \pi_1}{B + \pi_2}$$

- Note that the equilibrium matching imposes some dependence between η and x as well as ϵ and y conditional on z.
- As a consequence, we cannot simply use demand shifters as instruments for the supply estimation and supply shifters as instruments for demand estimation as we will discuss in more detail below.

Identification and Estimation of the Pricing Function I

- Most of the proposed methods treat z and p as observed.
- Consider the following additively separable pricing function:

$$p = p(z, \psi) + u \tag{19}$$

  where ψ is a parameter vector to be estimated.
- We can interpret u as an unobserved product characteristic u.
- Alternatively u can be interpreted as measurement error.
- Assuming $E[z|u] = 0$ the pricing function can be estimated using standard parametric and non-parametric techniques.

Identification and Estimation of the Pricing Function II

- Of course, there is no reason to believe that the unobserved product characteristic enters into the utility function in an additively separable way.
- Hence, the hedonic pricing regression is given by:

$$p = p(z, u) \tag{20}$$

- Bajari and Benkard (2005, JPE) discuss how to identify and estimate these types of pricing functions using techniques developed by Matzkin (2002, ECA).
- These techniques are taught in an advanced micro-econometrics class.

Estimating Hedonic Price Functions in Practice

- A commonly used approach is to regress the log of the price on the observed housing characteristics:

$$\ln(p) \;=\; \psi' z \,+\, u \tag{21}$$

- We can estimate ψ using OLS if we have a random sample of housing transactions in a local housing market.
- Often, we have data on multiple markets.

The Price of Land in the New York Metropolitan Area

- Conventional wisdom holds that vacant land is rare in urban areas, particularly in the New York area.
- Andrew Haughwout, James Orr, and David Bedoll (2008) studied the price of land in the NY metro area based a sample of 6,186 land sales between 1999 and mid-2006.
- 623 transactions, or roughly 10 percent, were in Manhattan.
- 1,639, or about 25 percent, took place in the other parts of New York City.
- The remaining sales took place in northern and central New Jersey.

Land Prices in NYC

Parameter	Estimate	Std Error
Location:		
log of distance from ESB	-0.95	(0.05)
log of distance from ESB * residential land	-0.32	(0.04)
Characteristics of transaction:		
Lot sold as part of expansion plans by buyer	0.17	(0.07)
Foreclosure transaction	-0.38	(0.17)
Eminent domain transaction	0.38	(0.18)
Lot has significant environmental problems	-0.81	(0.14)
Lot was not sold on the open market	0.04	(0.06)
Intended use:		
Buyer intends to hold lot for investment	-0.21	(0.07)
Lot is intended for public use	-0.48	(0.08)
Lot will be held as open space	-1.24	(0.08)
Intended use unknown	-0.19	(0.07)

Land Prices in NYC

Parameter	Estimate	Std Error
Type of Property:		
Residential land	0.09	(0.25)
Industrial land	-0.75	(0.23)
Condition of Property:		
Lot is graded	0.45	(0.06)
Lot is paved	0.45	(0.09)
Lot is "finished"	0.45	(0.05)
Lot is "fully improved"	0.38	(0.07)
Lot was "previously developed"	0.55	(0.06)
Lot is currently "partially developed"	0.55	(0.31)
Lot is platted and engineered	0.23	(0.37)
Lot has a structure present	-0.11	(0.19)
Structure present	0.03	(0.07)
Improvements not available	0.23	(0.05)

Housing Hedonics

- When we apply the hedonic model to housing we can either use rents or home prices as dependent variable.
- We include structural characteristics of the house such as lot size, size of the housing unit, number of bedrooms, etc.
- Some researchers also include neighborhood characteristics such as measures of school quality, access to parks and areas of recreation, distance to work, etc.
- Once we have estimated the hedonic price function, we can compute the slope of the pricing function.
- The FOC of the hedonic model imply that the estimated slopes of the hedonic price function are equal to the marginal willingness to pay for a marginal change in each characteristic.

Estimation of Preferences: Cobb-Douglas

- Let's assume that preferences of individual i for product j are Cobb-Douglas:

$$u_{ij} = \sum_{k=1}^{K} \beta_{ik} \ln(z_{jk}) + c_i \quad (22)$$

- Then the FOC of the consumer choice problem implies:

$$\beta_{ik} = \frac{\partial p}{\partial z_{jk}} z_{jk} \quad (23)$$

- If the price function is known, household i's preference parameters, can be recovered even if only a single choice of the household i is observed.
- By aggregating the decisions of all the households in a single market, we can estimate the distribution of taste coefficients in market. (Bajari and Benkard, 2005).

Estimation of Preferences: The General Case I

- In general, identifying the preferences (and technology parameters) is much harder than the Cobb-Douglas example suggests.
- To illustrate the problem consider the Tinbergen model and assume that $dim(z) = 1$.
- Given that the pricing function is quadratic the FOC can be written as:

$$\pi_1 + \pi_2 z \; = \; \nu - Bz \; = \; \beta' y - Bz \; + \; \eta$$

- Note that the left hand side can be estimated as discussed above using prices and characteristics.
- We are also willing to assume that $E[\eta | y = 0]$.
- Does that mean that we can identify B and β from the regression above?
- Note that regressing $\hat{\pi}_1 + \hat{\pi}_2 z$ on y and z only identifies π_1 and π_2.

Estimation of Preferences: The General Case II

- Moreover, there are no obvious exclusion restrictions.
- In particular, we cannot use the supply shifter x as an instrument for z in the demand equation .
- To show this negative result, recall that the matching equation derived above implies that:

$$\frac{\alpha x + \epsilon + \pi_1}{A - \pi_2} = \frac{\beta y + \eta - \pi_1}{B + \pi_2}$$

- Hence, the error in the demand equation η is correlated with x conditional on z.
- Hence, we cannot use x as an instrument for z.

Estimation of Preferences: The General Case III

- It turns out, the situation is much worse in the Tinbergen model. In particular, Epple (1987) shows that the Tinbergen model is not identified at all.
- Ekeland, Heckman, and Nesheim (2004) argue that the Tinbergen model is a knife-edge case.
- Generically speaking, the preferences, technology, and the pricing function will not have the same curvature, i.e. these nonlinearities are generic features of equilibrium in hedonic models.

Estimation of Preferences: The General Case IV

- Let's assume that there is nonlinearity in $p_z(z)$. So suppose, we have:

$$\pi_1 + \pi_2 z + \pi_3 z^2 \; = \; \beta' y - Bz \; + \; \eta$$

  where by assumption $E[\eta|y] = 0$.
- Maybe the cost function of the firms is not quadratic in z or the productivity shocks are non-normal.
- Note that z and the η are correlated due to the matching in equilibrium. We need an instrument for z.
- Similarly, we cannot use x as an instrument since x and η are correlated due to matching in equilibrium as well.
- The obvious choice for an instrument is then $E[z|y]$ which works as long as this is a nonlinear function in y.

Estimation of Preferences: The General Case V

- Heckman, Matzkin and Nesheim (2010, ECA) discuss identification and estimation of non-additively separable models.
- These papers use econometric techniques that are outside the scope of the class, but are taught in an advanced micro-econometric class.
- Talk to Wayne Gao if you are interested in learning more about these methods.

Estimation of Preferences: The General Case VI

- Progress can also be made if one observes repeated purchases as pointed out by Bishop and Murphy (2018).
- Suppose we observe two equilibria in period 1 and 2. Let $p_1(z)$ and $p_2(z)$ denote the two price functions.
- Let z_1 and z_2 denote the corresponding two purchase of the same individual.
- If we approximate the marginal willingness to pay function of that individual by a linear function then we have the following two equations:

$$\left.\frac{\partial p_1(z)}{\partial z}\right|_{z=z_1} = \delta_0 + \delta_1 z_1$$
$$\left.\frac{\partial p_2(z)}{\partial z}\right|_{z=z_2} = \delta_0 + \delta_1 z_2$$

  which can be solved for δ_0 and δ_1.
- This procedure rests on the assumption that preferences do not chance over time.

Summary

- Hedonic models are used to characterize the demand and supply for heterogeneous or differentiated products such as housing.
- We can estimate hedonic price functions using data on housing prices and observed characteristics.
- The estimated slopes of the hedonic price functions are informative about household preferences, especially in the Cobb-Douglas case.
- Identifying and estimating hedonic models without imposing strong functional form assumptions can also be done, but requires more advanced econometric methods.

Advanced Topics

- We can use hedonics to construct a time series of a price index in a single market using repeated sales (Case-Shiller).
- Sieg, Smith, Banzhaf, and Walsh (2002) show that we can also use hedonic regression to construct cross-sectional housing price indices, which is necessary to estimate spatial models.
- Epple, Quintero and Sieg (2021, JPE) show how to estimate a hedonic model of rental markets when there is a single unobserved latent characteristic.
- Langvoigt, Piazzesi and Schneider (2015 AER) discuss show how to extend the hedonic model to a dynamic, non-stationary framework and study homeownership decisions during the "bubble" period.