Formulating and Identifying Hedonic Models in Competitive Markets

Ivar Ekeland
University of Paris, Dauphine

James J. Heckman
University of Chicago and the American Bar Foundation

Lars Nesheim
Institute for Fiscal Studies and University College London
June 11, 2002
1 Introduction

- Study competitive hedonic markets, markets for heterogeneous goods in which characteristics of goods are priced out

- Rosen (1974) definitively established the framework for the analysis of hedonic markets in a perfectly competitive setting

- Proposed identification strategy to recover preferences and technology

- Method of identification has been severely criticized

- Widely believed that parameters of hedonic models are identified only through arbitrary functional form and exclusion assumptions, especially when they are estimated from data on a single market
Conclusions

• Hedonic model is generically nonlinear

• The economic model for which the widely used linearization methods are exact is implausible

• Commonly used linearization strategies produce identification problems

• In a wide class of additive, parametric models, model parameters are generically identified with data from a single market

• In additive nonparametric models, parameters are identified up to affine transformations
Conclusions (cont.)

• Identification analysis also applies to other non-linear pricing models
  
  – Effects of taxes on behavior when taxes are set optimally (Marrisles (1971))

  – Monopoly pricing (Mussa and Rosen (1978))

  – Taxes and labor supply (Heckman (1974); Hausman (1980))

  – Social interactions and sorting (Nesheim (2001))

• Analysis can be extended to non-additive models (Heckman, Matzkin, and Nesheim (2002))
Outline of talk – Section 2

- Section 2 presents the hedonic model and reviews an important quadratic case due to Tinbergen (1956), and used by Epple (1987)
  - Equilibrium price depends on production technology parameters, consumer preference parameters, and the distribution of heterogeneity in the population
  - Tinbergen model has closed form solution resulting in quadratic pricing function
Outline of talk – Section 3

- Section 3 focuses on quadratic model to motivate discussion of economic and empirical properties of hedonic models
  - Rosen’s empirical strategy
  - Brown and H. Rosen’s (1982) critique
    * Identification depends on arbitrary functional form restrictions
    * \( z \) is endogenous; no instruments are available
    * Multimarket data can help
  - Critique is not valid when quadratic model is perturbed
    * Non-identification is NOT generic
    * Non-identified model, is not plausible
Outline of talk – Sections 4-6

- Section 4 establishes 4 main points

  1. Hedonic model is identified within a single market for a broad class of additive parametric models

  2. Hedonic model is identified up to levels for a broad class of additive nonparametric models

  3. Analyzing all information on supply and demand jointly adds nothing to what can be identified from analyzing the two separately

  4. Extra information on levels of outcomes can aid in identifying the missing level set information

- Section 5 discusses instrumental variables estimation of the model and Section 6 concludes
2 General Hedonic Model: Supply of Workers

- Individual workers match to single worker firms

- Choose quality of job $z$ and maximize

  $$U(c, z, \theta, A)$$

- $c = P(z) + R$
  - $P(z)$ earnings of workers
  - $R$ is unearned income

- Preference parameters
  - $\theta$ varies across workers, with density $f_{\theta}(\theta)$
  - $A$ is common across all workers
Workers’ Optimization

- **FOC:**
  \[ U_c (C, z, \theta, A) P_z (z) + U_z (c, z, \theta, A) = 0 \]

- **SOC:**
  \[ U_{zz'} + U_c P_{zz'} + P_z U_{cc} (P_z)' \] is negative definite
Firms’ Technologies

- Choose quality $z$ to maximize output minus cost
  \[ \Pi(z) = F(z; \nu, B) - P(z) \]

- Technology parameters
  - $\nu$ varies across firms with density $f_\nu(\nu)$
  - $B$ is common across all firms
Firms’ Optimization

- FOC:

\[ F_z (z, \nu, B) - P_z (z) = 0 \]

- SOC:

\[ F_{zz'} - P_{zz'} \text{ is negative definite} \]
Heterogeneity in Market

Density of $\theta : f_\theta$

Density of $\nu : f_\nu$

$\dim (\theta) \geq \dim (z) ; \dim (\nu) \geq \dim (z)$
Sorting Conditions

From the FOC for the firm

$$\nu = \nu (z, P_z, B)$$

From the FOC for the consumer

$$\theta = \theta (z, P_z, P(z) + R, A)$$

Assume $R = 0$ for ease of exposition
Supply and Demand

The Demand Density is:

\[ f_\nu (\nu (z, P_z, B)) \det \left[ \frac{\partial \nu (z, P_z, B)}{\partial z} \right] d\nu \]

The Supply Density is:

\[ f_{\theta} (\theta (z, P_z, P (z), A)) \det \left[ \frac{\partial \theta (z, P_z, P (z), A)}{\partial z} \right] d\theta \]
Equilibrium in Hedonic Market

\[ f_{\nu} (\nu (z, P_z, B)) \det \left[ \frac{\partial \nu (z, P_z, B)}{\partial z} \right] = \]

\[ f_{\theta} (\theta (z, P_z, P (z), A)) \det \left[ \frac{\partial \theta (z, P_z, P (z), A)}{\partial z} \right] \]

Equivalent Formulation in terms of CDF's

\[ F_{\nu} (\nu (z, P_z, B)) = F_{\theta} (\theta (z, P_z, P (z), A)) \]
3  Linear-Quadratic Example

Consumer Side

- Preferences quadratic in $z$ and linear in $P(z)$
  \[ U(c, z, \theta, A) = P(z) + \theta'z - \frac{1}{2}z'Az \]

- FOC:
  \[ \theta - Az + P_z = 0 \quad (2) \]

- SOC:
  \[ (P_{zz'} - A) \text{ is negative definite} \]

- Worker heterogeneity
  \[ \theta \sim N(\mu_{\theta}, \Sigma_{\theta}) \]
Linear Quadratic Example

Firm Side

• Production quadratic in $z$

$$\Pi(z, \nu, B, P(z)) = \nu'z - \frac{1}{2}z'bz - P(z)$$  \hspace{1cm} (3)

• FOC:

$$\nu - Bz - P_z = 0$$  \hspace{1cm} (4)

• SOC:

$$-(B + P_{zz'})$$ is negative definite

• Firm heterogeneity

$$\nu \sim N(\mu_{\nu}, \Sigma_{\nu})$$
Equilibrium

- Equilibrium must satisfy (1)

- Given special structure, one can guess (correctly) that

\[ P(z) = \pi_0 + \pi'_1 z + \frac{1}{2} z' \pi_2 z \]

- Then check that the guess is correct

- Firm FOC:

\[ \nu - Bz - \pi_1 - \pi_2 z = 0 \]

- Consumer FOC:

\[ \theta - Az + \pi_1 + \pi_2 z = 0 \]
Sorting conditions

\[ z_D = (B + \pi_2)^{-1}(\nu - \pi_1) \]

\[ z_S = (A - \pi_2)^{-1}(\theta + \pi_1) \]

Equate average demand to average supply

\[ E^D(z) = (B + \pi_2)^{-1} E(\nu - \pi_1) \]

\[ E^S(z) = (A - \pi_2)^{-1} E(\theta + \pi_1) \]

- One vector equation in unknown coefficients

\[ (B + \pi_2)^{-1}(\mu_\nu - \pi_1) = (A - \pi_2)^{-1}(\mu_\theta + \pi_1) \]
Equate variance of supply and demand

\[ V^D(z) = (B + \pi_2)^{-1} \Sigma \nu (B + \pi_2)^{-1'} \]
\[ V^S(z) = (A - \pi_2)^{-1} \Sigma \theta (A - \pi_2)^{-1'} \]

- Second matrix equation

\[ (A - \pi_2)^{-1} \Sigma \theta (A - \pi_2)^{-1'} = (B + \pi_2)^{-1} \Sigma \nu (B + \pi_2)^{-1'} \]

- Initial conditions

\[ U(z) \geq U_0 \]
\[ \Pi(z) \geq 0 \]

- Implies

\[ \pi_0 = 0 \]
• Solution depends on
  
  – Production and preference parameters $A$, $B$
  
  – Heterogeneity $\mu_\nu$, $\mu_\theta$, $\Sigma_\nu$, and $\Sigma_\theta$

• Except in polar cases, price function does not directly reveal any individual structural parameters

• Note equilibrium matching implies

\[
(B + \pi_2)^{-1}(\nu - \pi_1) = (A - \pi_2)^{-1}(\theta + \pi_1)
\]

• Functional and statistical dependence between $\nu$ and $\theta$
3.1 Identifying and Estimating the Model

Two Step Estimation Procedure

Estimate $P(z)$ from market data

Use first-order conditions (2) and (4) in conjunction with the marginal prices obtained from step 1 to recover preferences and technology respectively
• \( \hat{\pi}_1 \) and \( \hat{\pi}_2 \) fitted price coefficients

• \( x \) and \( y \) are observable worker and firm regressors

\[
\hat{\pi}_1 + \hat{\pi}_2 z = \mu_\nu(y) - Bz + \varepsilon_\nu \tag{6}
\]

\[
\hat{\pi}_1 + \hat{\pi}_2 z = -\mu_\theta(x) + Az - \varepsilon_\theta \tag{7}
\]

• \( \varepsilon_\nu \) and \( \varepsilon_\theta \) are unobservable

\[
\varepsilon_\nu = \nu - \mu_\nu(y)
\]

\[
\varepsilon_\theta = \theta - \mu_\theta(x)
\]
Point One: Identification Can Only be Obtained Through Arbitrary Functional Form Assumptions

- If fitted $P(z)$ is quadratic, linear functions of $z$ on left and right sides

- If fitted $P(z)$ is not quadratic, nonlinearity can help with identification

- This nonlinearity is arbitrary

- However, small perturbations of above model lead to non-quadratic $P(z)$

- Economics of the problem suggests that for many applications this quadratic equilibrium model not very good any way we want to move away

- Analysis of equilibrium equation shows that in fact, nonquadratic $P(z)$ is generic
Example 1

- Perturb scalar version of quadratic model

- Suppose heterogeneity distributed as mixture of normals with weights (0.999, 0.001) or (0.99, 0.01)

- Marginal price function is clearly nonlinear

- Unattractive features of quadratic model
  - Negative and positive $z$
  - Negative marginal products
Example 2

- Restrict marginal product to be positive
- Restrict marginal utility of work to be negative
- The non-identified case is uninteresting as well as unlikely
Point Two: Endogeneity Problem

- $z$ is endogenous in (6) and (7)

- Rewrite (3)

\[ \varepsilon_\nu = \varepsilon_\theta + (A - B)z + \mu_\theta(x) - \mu_\nu(y) \]

- Conditional on $z$, there is a functional and statistical dependence between $x, y, \varepsilon_\nu,$ and $\varepsilon_\theta$
Point Three: Use of Multimarket Data

- Why do prices vary across markets?

- Must specify what elements vary across markets, and what elements do not

- Our results below show there is no need for multimarket data
4 Generic identification of general additive scalar model

- Establish points made in previous section more formally and more generally
  - Generic identification
  - Recover all structural parameters up to location

- Results apply to any model in which FOC reduce to additive first-order conditions

- Generalization of quadratic model

- Can constrain to have economically meaningful interpretations
Firm side

• Production technology

\[ \Pi(z, y, \varepsilon_1) = \Phi^1(z) + z(\eta_1(y) + \varepsilon_1) - P(z) \]

• \((y, \varepsilon_1) = \nu\) with distribution

\[ q_1(y) g_1(\varepsilon_1) \]

• \(\varepsilon_1\) is independent of \(y\)

• FOC:

\[ \varphi_1(z) + \eta_1(y) + \varepsilon_1 - P'(z) = 0 \quad (8) \]

• SOC:

\[ \varphi'_1(z) - P''(z) < 0 \]
Consumer side

• Preferences

\[ U(z, x, \varepsilon_2) = P(z) - \Phi^2(z) + z(\eta_2(x) + \varepsilon_2) \]

• \((x, \varepsilon_2) = \theta\) with distribution

\[ q_2(x) g_2(\varepsilon_2) \]

• \(\varepsilon_2\) is independent of \(x\)

• FOC:

\[ P'(z) - \varphi_2(z) + \eta_2(x) + \varepsilon_2 = 0 \]

• SOC:

\[ P''(z) - \varphi'_2(z) < 0 \]
Equilibrium Sorting

- The first-order conditions define one-to-one mappings \((y, \varepsilon_1) \longleftrightarrow (y, z)\) and \((x, \varepsilon_2) \longleftrightarrow (x, z)\)

- Equilibrium defines mapping from \((x, \varepsilon_1, y, \varepsilon_2) \longrightarrow (x, y, z)\)

- Firms:
  \[
  \varepsilon_1 = P'(z) - \varphi_1(z) - \eta_1(y) \\
  y = y
  \]

- Workers:
  \[
  \varepsilon_2 = \varphi_2(z) - P'(z) - \eta_2(x) \\
  x = x
  \]
Equilibrium

\[
\left( P''(z) - \varphi_1'(z) \right) \cdot \\
\int_X g_1 \left( P'(z) - \varphi_1(z) - \eta_1(\tilde{x}) \right) \cdot q_1(\tilde{x}) \, d\tilde{x}
\]

\[
= \left( \varphi_2'(z) - P''(z) \right) \cdot \\
\int_Y g_2 \left( \varphi_2(z) - P'(z) - \eta_2(\tilde{y}) \right) q_2(\tilde{y}) \, d\tilde{y}
\]

\[
F \left( z, P', P''; \theta \right) = 0 \text{ for all } z
\]

- We will show this implies a property about \( P \)
Definition of genericity:

Property $P(\theta)$, depending on a parameter $\theta \in \Theta$, is called generic if the set $\Omega \subset \Theta$ of values of the parameter for which it holds true contains a countable intersection of open dense subsets.
Theorem 1

Theorem 1: Generically, with respect to any of the parameter pairs, \((g_1, g_2), (\eta_1, \eta_2),\) and \((\varphi_1, \varphi_2),\) the equilibrium equations have no solution of the form \(P'(z) = a_1 + b_1 \varphi_1(z),\) nor any solution of the form \(P'(z) = a_2 + b_2 \varphi_2(z).\)

- As a consequence, Brown-Rosen Point One is not generically correct.
• Proof: Technical details are in the appendix, but
  
  – Equilibrium ODE is infinite dimensional

  – Suppose the theorem is not true

  – Then, the solution to the ODE has only two unknown parameters

    \[ F(z, a_1, b_1; \theta) = 0 \text{ for all } z \]

  – Infinite set of equations in two unknowns
• In addition, we go on to show that the unknown functions \((g_1, g_2, \eta_1, \eta_2, \varphi_1, \varphi_2)\) are identified using data on \((P(z), z, x, y)\) in a single market.

• Rewrite FOC (8)

\[ P'(z) - \varphi_1(z) = \eta_1(y) + \varepsilon_1 \]

• Focus on firm side, analogous argument applies to worker side.
Transformation models

- Define
  \[ T_1 (z) = P'(z) - \varphi_1 (z) \]

- Substitute into FOC (8):
  \[ T_1 (z) = \eta_1 (y) + \varepsilon_1 \]

- Transformation model (See Horowitz (1998), Powell, Stock and Stoker (1989))
Statistical implications

• Then

\[ F^1(z \mid y) = G_1(T_1(z) - \eta_1(y)) \]

• \( F^1(z \mid y) \) conditional CDF of observed data

• \( G_1, T_1, \) and \( \eta_1 \) are unknown functions

• Note special structure implied by separability
Ratio of derivatives

\[ F^1_z(z \mid y) = g_1(T_1(z) - \eta_1(y)) \cdot T'_1(z) \]  \hspace{1cm} (9)

\[ F^1_{y_i}(z \mid y) = -g_1(T_1(z) - \eta_1(y)) \cdot \frac{\partial \eta_1}{\partial y_i} \]  \hspace{1cm} (10)

\[
\left( \frac{-F^1_z(z \mid y)}{F^1_{y_i}(z \mid y)} \right) = \frac{T'_1(z)}{\frac{\partial \eta_1(y)}{\partial y_i}}
\]

- Sign \( F^1_{y_i} \) = \(-\) sign \( \left( \frac{\partial \eta_i}{\partial y_i} \right) \). Assume, without loss of generality, that \( \frac{\partial \eta_i}{\partial y_i} > 0 \).
• Then,

\[ \frac{\partial}{\partial z} \log \left[ \frac{-F_{z}^1(z \mid y)}{F_{y_i}^1(z \mid y)} \right] = \frac{T_1''(z)}{T_1'(z)}. \]  \hspace{1cm} (11)

• Define

\[ h(z, y) = \log \left[ \frac{-F_{z}^1(z \mid y)}{F_{y_i}^1(z \mid y)} \right] \]

• Testable

\[ \frac{\partial^2 h(z, y)}{\partial z \partial y'} = 0 \text{ for all } z, y \]
\( h(z, y) \) is known

- \( h(z, y) \) is function of observable data

- Moreover, due to separability (see (11)),

\[
h(z, y) = h_0 + h_1(z) + h_2(y)
\]

where \( h_1(0) = 0, h_2(0) = 0, \) and \( h_0 \) is a constant. \( h_0, h_1(z) \) and \( h_2(y) \) are known empirically.
Identification of $T(z)$

- Solving (11)

$$T'_1(z) = K_1 \exp(h_1(z)) \quad (12)$$

where $K_1$ is a constant of integration.

- Further,

$$T_1(z) = C_1 + K_1 \int_0^z \exp(h_1(s))ds$$

where $C_1$ is another constant of integration.
Identification of $\eta_1(y)$

- Substituting and integrating

$$\eta_1(y) = R_1 + K_1 \int_0^y \exp(-h_0 - h_2(s))ds$$

where $R_1$ is a constant of integration and the multiple integral is taken over all the dimensions of $y$.

- Thus, we identify

$$\tilde{\eta}_1(y) = \frac{\eta_1(y) - R_1}{K_1}$$

$$\tilde{T}_1(z) = \frac{T_1(z) - C_1}{K_1}.$$ 

- Since we know $P'(z)$, we can identify

$$\tilde{\varphi}_1(z) = P'(z) - K_1 \tilde{T}(z) - C_1.$$
Identification of distribution of error term

• In this notation,

\[ \varepsilon_1 = T_1(z) - \eta_1(y) \]
\[ = (C_1 - R_1) + K_1(\tilde{T}(z) - \tilde{\eta}_1(y)) \]

• By assuming \( E(\varepsilon_1) = 0 \) or median \( (\varepsilon_1) = 0 \) (or fixing some quantile of \( \varepsilon_1 \)) we get \( C_1 - R_1 \), leaving \( K_1 \) undetermined. This leaves \( K_1 \) undetermined.

• Define,

\[ \tilde{\varepsilon}_1 = (\varepsilon_1/K_1) \]
\[ g_1(\varepsilon_1)d\varepsilon_1 = K_1g_1(\tilde{\varepsilon}_1K_1)d\tilde{\varepsilon}_1 = \tilde{g}_1(\tilde{\varepsilon}_1) \]

• \( \tilde{g}_1(\tilde{\varepsilon}_1) \) is known from (9)
Identification of scale $K_1$

- Lack of identification of the scale of the utility function is a classical result.

- If we observe output or utility, we can determine the missing parameters by using (8) as a replacement function in the sense of Heckman and Robb (1985) or as a control function in the sense of Blundell and Powell (2001)
• Substitute (8) into production to obtain

\[ F(z, x, \varepsilon_1) = \Phi^1(z) + z(\eta_1(y) + \varepsilon_1) = \Phi^1(z) + zP'(z) - z\varphi_1(z) \]

• Define

\[ \psi(z) = F(z, y, \varepsilon_1) - zP'(z) \]

• Then, \( \psi(z) \) is observed and

\[ \psi(z) = \Phi^1(z) - z\varphi_1(z) \]

\[ = \int_0^z \varphi_1(t)\,dt - z\varphi_1(z) \]

• Hence we may estimate

\[ \frac{\partial \psi(z)}{\partial z} = -z\varphi_1'(z). \]

and integrate to get

\[ C_0 + \int \left[ -\frac{1}{z} \frac{\partial \psi(z)}{\partial z} \right] \,dz = \varphi_1(z) \]
Alternative identification of $K_1$

- Suppose $\varphi_1, \varphi_2 \in E$, a *known* finite dimensional vector space

- i.e. $\varphi_1 = \sum a_k \varphi_k, \varphi_2 = \sum b_k \varphi_k$

- Then we have

Theorem 4.1 *Generically with respect to any of the parameter pairs in Theorem 1, no solution $P(z)$ of the equilibrium equation belongs to $E$, and $\varphi_1, \varphi_2$ are identified up to additive constants*
Estimation strategy

1. Estimate joint distribution of \((z, y) : F^1(z \mid y)\)

2. Recover \(\tilde{T}_1(z), \tilde{\eta}_1(y)\), and the distribution of \(\tilde{\varepsilon}_1\)

3. Estimate \(P_z(z)\)

4. Recover \(\tilde{\varphi}_1(z)\)
• Note same can be done for worker side of market \((z, x)\)

• Might think that the joint density \((z, x, y)\) contains more information

• It does not

• Theorem 3: The joint density of \((z, x, y)\) provides no more information than the marginal densities \(f^1(z, y)\) and \(f^2(z, x)\)
5 Instrumental variables

- Alternative to the approach outlined is IV

- Have FOC (8)

\[ P'(z) = \varphi_1(z) + \eta_1(y) + \varepsilon_1 \]

- We prove that generically \( E(\varphi_1(z) | Y) \) is linearly independent of \( \eta_1(y) \)

- Hence, can use parametric non-linear IV to estimate

- Construct \( E(\varphi_1(z) | Y) \) to instrument for \( \varphi_1(z) \)

- We conjecture nonparametric IV (Darolles et. al, 2001, Florens, Heckman, Meghir and Vytlacil, 2000, or Newey and Powell, 2000). is also valid

- Exploration of this conjecture is left for future
6 Summary and conclusions

1. Linearization is artificial and is not robust

2. Nonlinearity is intrinsic and is implied by the economics of the problem

3. Parameters are identified in a single market

4. Valid instruments exist

5. Analysis extends to peer effects model and non-linear pricing models

6. Additive assumption, while strong, is testable

7. Alternative to additive assumption, need some knowledge about technology. See Heckman, Matzkin, Nesheim (2002) for non-additive model results