Nonparametric Estimation of Nonadditive Hedonic Models

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

Rosa Matzkin
Northwestern University

Lars Nesheim
Institute for Fiscal Studies and University College London
1 Introduction

• Common threads in the models being presented are:
  – sorting (across locations or products)
  – pricing of locations (characteristics, products)
  – consumer and product heterogeneity
  – general equilibrium
  – empirical applications
• Locational equilibrium models
  – Sorting across geographic locations
  – Pricing of locations
  – Social interactions and pricing of social interactions (price discrimination)

• Hedonic models
  – characteristics of location (good)
  – reduces parameters
  – enables predictions for new goods
  – structure on substitution patterns
  – adjust welfare indices for quality change
• Applications

  – public finance

  – residential location

  – environment

  – markets with differentiated products

  – labor economics

• This afternoon hear about asset pricing models and search models
• Hedonic labor market models


  – Explain wage as a function of risk, skill, education

  – Seek estimation of preferences and technologies to make predictions and to do counterfactual policy analysis
• General hedonic model: theory and empirics

• 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
  – Assumed smooth equilibrium exists

• 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

- Much of this criticism is misplaced, fails to adequately consider properties of hedonic equilibrium

- Much is based on linearizations of general hedonic model: 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 general nonadditive hedonic models, identification in a single market does not appear feasible

  – Need more structure on the problem

  – 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

  – In nonadditive models with some structure, preference and technology functions are identified up to normalizations
Outline of talk

- General hedonic model

- Equilibrium

- Some analysis of equilibrium: curvature of pricing function and bunching

- Identification and estimation of the model
  - linear additive
  - nonlinear additive
  - nonadditive
2 General Hedonic Model: Supply of Workers

- Individual workers match to single worker firms
- Choose quality of job $z$ and maximize
  \[ P(Z) - U(z, x, \varepsilon) \]
- $P(z)$ earnings of workers
- $x$ is vector of observable characteristics of workers with density $f_x(x)$
- $\varepsilon$ is vector of unobservable worker characteristics with density $f_\varepsilon(\varepsilon)$, independent of $x$
- Assume $U_{z\varepsilon} < 0$
Workers’ Optimization

- FOC:

\[ P_z(z) - U_z(z, x, \varepsilon) = 0 \]

- SOC:

\[ P_{zz} - U_{zz} < 0 \]
Firms’ Technologies

- Choose quality $z$ to maximize output minus cost
  \[ \Pi(z) = \Gamma(z, y, \eta) - P(z) \]

- $y$ is vector of observable firm attributes with density $f_y(y)$

- $\eta$ is a vector of unobservable firm attributes with density $f_\eta(\eta)$, independent of $y$

- Assume $\Gamma_{z\eta} > 0$
Firms’ Optimization

• FOC:

\[ \Gamma_z(z, y, \eta) - P_z(z) = 0 \]

• SOC:

\[ \Gamma_{zz} - P_{zz} < 0 \]
Goals of hedonic analysis

- Analyze first order conditions
  
  Workers \[ P_z (z) = U_z (z, x, \varepsilon) \]
  
  Firms \[ P_z (z) = \Gamma_z (z, y, \eta) \]

- Solve theoretical problem
  
  - Characterize equilibrium pricing function

- Solve econometric problem
  
  - Given data on \((P (z), z, x)\) (or equivalently for firm side) and recover estimates of \(U_z\) and the distribution of \(\varepsilon\)

- Everything we do focuses on these two equations
Workers’ Sorting Conditions

- For each worker \((x, \varepsilon)\), FOC implicitly defines quality supplied (or location chosen)

\[
z = s(x, \varepsilon)
\]

- i.e., the mapping from worker \((x, \varepsilon)\) to location \(z\)

- Define the inverse mapping

\[
\varepsilon = \tilde{s}(z, x)
\]

- Note \(\tilde{s}(z, x)\) depends on the function \(P\) and

\[
\frac{\partial \tilde{s}}{\partial z} = \frac{P_{zz} - U_{zz}}{U_{z\varepsilon}} > 0 \quad (1)
\]
Firms’ Sorting Conditions

• For every firm \((y, \eta)\), FOC implicitly defines quality demanded

\[ z = d(y, \eta) \]

• a mapping from firm \((y, \eta)\) to location \(z\)

• Define the inverse mapping

\[ \eta = \tilde{d}(z, y) \]

• \(\tilde{d}(z, y)\) also depends on \(P\) and

\[
\frac{\partial \tilde{d}}{\partial z} = \frac{P_{zz} - \Gamma_{zz}}{\Gamma_{z\eta}} > 0
\] (2)
Supply and Demand

The Supply Density is:

$$\int_{\tilde{X}} f_\varepsilon (\tilde{s} (z, x)) \cdot \frac{\partial \tilde{s} (z, x)}{\partial z} \cdot f_x (x) \, dx$$

The Demand Density is:

$$\int_{\tilde{Y}} f_\eta (\tilde{d} (z, y)) \cdot \frac{\partial \tilde{d} (z, y)}{\partial z} \cdot f_y (y) \, dy$$
Equilibrium in Hedonic Market

1. Price function equates supply and demand at all $z$

$$
\int_{\tilde{X}} f_{\varepsilon}(\tilde{s}(z, x)) \frac{\partial \tilde{s}(z, x)}{\partial z} f_x(x) \, dx = \quad (3)
$$

$$
\int_{\tilde{Y}} f_{\eta}(\tilde{d}(z, y)) \frac{\partial \tilde{d}(z, y)}{\partial z} f_y(y) \, dy
$$

2. SOC for all firms and workers are satisfied (local condition)

3. $z = s(x, \varepsilon)$ and $z = d(y, \varepsilon)$ are globally optimal
Importance of (3)

- Compute equilibria of sample economies

- Understand how primitives of model influence sorting and pricing
  - Relations between curvature of pricing function and curvatures of preferences and technology

- Understand when bunching occurs

- Analyze identification of model
Curvature of pricing function

General Nonadditive case

- Substitute (1) and (2) into (3)

\[
P_{zz} = \frac{\int_{\tilde{Y}} \left( \frac{f_{zz}}{f_{z\eta}} \right) f_{\eta} f_y dy - \int_{\tilde{X}} \left( \frac{U_{zz}}{U_{z\varepsilon}} \right) f_{\varepsilon} f_x dx}{\left( \int_{\tilde{Y}} \frac{f_{\eta} f_y}{f_{z\eta}} dy - \int_{\tilde{X}} \frac{f_{\varepsilon} f_x}{U_{z\varepsilon}} dx \right)}
\]  

(4)

- \( P_{zz} \) is a weighted average of curvature of technology and preferences

- Curvature of technology and preferences only matters at points actually chosen, i.e. \( d(y, \eta) \) and \( s(x, \varepsilon) \)
Special Case 1

Linear FOC

Workers \( P_z (z) = U_{zz} z + U_{zx} x - \varepsilon \)

Firms \( P_z (z) = \Gamma_{zz} z + \Gamma_{zy} y + \eta \)

- \( \Gamma_{zz} \) and \( U_{zz} \) are constants

- \( \Gamma_{z\eta} = 1 \) and \( U_{z\varepsilon} = -1 \)

- \( \Gamma_{zy} \) and \( U_{zx} \) are constants

\[
P_{zz} = \frac{\Gamma_{zz} \int f_{\eta} f_y dy + U_{zz} \int f_{\varepsilon} f_x dx}{\int f_{\eta} f_y dy + \int f_{\varepsilon} f_x dx}
\]

\( \tilde{Y} \)

\( \tilde{X} \)

(5)
• $P_{zz}$ is a simple weighted average of $\Gamma_{zz}$ and $U_{zz}$

• When all consumer and firm characteristics are distributed normally this simplifies even further

$$P_{zz} = \frac{\Gamma_{zz}\sigma_f + U_{zz}\sigma_w}{\sigma_f + \sigma_w}$$


– In this very special (nongeneric) case the price function is quadratic
Special Case 2

Additive FOC

Workers  \[ P_z(z) = m_w(z) + n_w(x) - \varepsilon \]

Firms  \[ P_z(z) = m_f(z) + n_f(y) + \eta \]

- \( \Gamma_{zz}, \Gamma_{zy}, U_{zz}, \) and \( U_{zx} \) are NOT-constants

- \( \Gamma_{z\eta} = 1 \) and \( U_{z\varepsilon} = -1 \)

- \( \Gamma_{zz\eta} = 0 \) and \( U_{zzx} = 0 \)

\[
P_{zz} = \frac{\int \Gamma_{zz} f_\eta f_y dy + \int U_{zz} f_\varepsilon f_x dx}{\int f_\eta f_y dy + \int f_\varepsilon f_x dx}
\]

\[(6)\]
Bunching

- Classic hedonic model assumes equilibrium sorting of agents is smooth

- No location has positive mass of people

- Equilibrium pricing function is $C^2$

- Are there conditions on primitives that rule out bunching?
SOC of Consumer

- For every consumer who chooses $z$, it must be the case that $U_{zz}(z, x, \varepsilon) > P_{zz}(z)$

$$\frac{\int_{\tilde{Y}} \left( \frac{\Gamma_{zz}}{\Gamma_{z\eta}} \right) f_{\eta} f_{Y} dY - \int_{\tilde{X}} \left( \frac{U_{zz}}{U_{z\varepsilon}} \right) f_{\varepsilon} f_{X} dX}{\left( \int_{\tilde{Y}} \frac{f_{\eta} f_{Y}}{\Gamma_{z\eta}} dY - \int_{\tilde{X}} \frac{f_{\varepsilon} f_{X}}{U_{z\varepsilon}} dX \right)} < U_{zz}(z, x, \tilde{s}(z, x))$$

- For every firm who optimally chooses $z$, it must be the case that $\Gamma_{zz}(z, y, \eta) < P_{zz}(z)$

SOC of Firm

$$\Gamma_{zz}(z, y, \tilde{d}(z, y)) < \frac{\int_{\tilde{Y}} \left( \frac{\Gamma_{zz}}{\Gamma_{z\eta}} \right) f_{\eta} f_{Y} dY - \int_{\tilde{X}} \left( \frac{U_{zz}}{U_{z\varepsilon}} \right) f_{\varepsilon} f_{X} dX}{\left( \int_{\tilde{Y}} \frac{f_{\eta} f_{Y}}{\Gamma_{z\eta}} dY - \int_{\tilde{X}} \frac{f_{\varepsilon} f_{X}}{U_{z\varepsilon}} dX \right)}$$
\[ \Gamma_{zz} (z, y, \tilde{d}(z, y)) < U_{zz} (z, x, \tilde{s}(z, x)) \]

- Conditions depend in complicated way on curvatures of preferences and technology and on distribution of preferences and technology

- Special case: Additive FOC

\[ \Gamma_{zz} (z) < U_{zz} (z) \]
2.1 Identifying and Estimating the Model

- Focus on firm side

- \[ P_z(z) = \Gamma_z(z, y, \eta) \]

- Given data on \((z, y, P)\)

- \(F_{Z|Y=y}(\cdot)\) and \(P(z)\) are known

- Estimate \(\Gamma_z(z, y, \eta)\) and the distribution \(F_\eta\)

- In a single market (single cross section) you need more structure

- For each consumer \((y, \eta)\), there is no variance in prices
• You can estimate \( z = d(y, \eta) \) without further structure using Matzkin (1999,2002) but this is not structural, invariant to changes in economic environment

• Two sets of results
  
  – Additive structure identifies up to some constants

  – Separability structure identifies up to normalization
Linear additive structure

\[ P_z(z) = \Gamma_{zz}z + \Gamma_{zy}y + \eta \]


- If \( P_z(z) = P_zz \), no identification except in polar cases

- \( z \) is endogenous and there are no instruments

- EHN (2002) results
  - \( P_z(z) \) is generically non-linear
  - One estimation method is IV: use \( E(z|y) \) as instrument for \( z \)
    * \( E(z|y) \neq a + by \)
– More generally, when nonlinear additive structure,

\[ P_z(z) = m_f(z) + n_f(y) + \eta \]

can identify up to additive constant and scale

– Additionally, if \( m_f(z) \) is a polynomial (or element of known finite dimensional function space), can identify scale (generically)
Nonadditive structure

• In single market must add structure to get identification

• In the paper, we show several different specifications that will work

• Here, I develop one example
• Suppose

\[ \Gamma_z(z, y, t) = m(q(z, y), t) \]

- \( q \) is a known function
- \( m \) is an unknown function with \( \frac{\partial m}{\partial \eta} > 0 \)

• Make a normalization so that

\[ \Gamma_z(t, \bar{y}, t) = P_z(t) \]

• Then \( \Gamma_z(z, y, t) \) and \( F_\eta \) are identified
• Stage one: estimate \( z = d(y, \eta) \)

• \( F_{Z|Y=y} \) is known.

• Since

\[
\Gamma_z(t, \bar{y}, t) = P_z(t) \\
d(\bar{y}, t) = t
\]

•

\[
F_{\eta}(e) = \Pr(\eta \leq e) \\
= \Pr(\tilde{d}(z, \bar{y}) \leq e) \\
= \Pr(z \leq d(\bar{y}, e))
\]

• \( d(\bar{y}, e) = e \) by construction

•

\[
\Pr(z \leq d(\bar{y}, e) | \bar{y}) = F_{Z|Y=\bar{y}}(e)
\]

\[
\Pr(z \leq e | \bar{y}) = F_{Z|Y=\bar{y}}(e)
\]
- Result

\[ F_\eta(e) = F_{Z|Y=y}(e) \]

- Use this to get \( d(y, e) \)

\[
F_{Z|Y=y}(d(y, e)) = \Pr(z \leq d(y, e) | y) \\
= \Pr(\eta \leq e) \\
= F_\eta(e)
\]

- Hence

\[ d(y, e) = F_{Z|Y=y}^{-1}(F_\eta(e)) \]

- Not structural, still want \( m \), use the fact that \( q(z, y) \) is known
• Stage 2: estimate $\Gamma_z$

• $d(y, e)$ is known from stage 1.

• Let $t_1$ and $t_2$ be two numbers in supports of $z$ and $\eta$.

• Let $y^*$ solve

$$q(d(y^*, t_2), y^*) = t_1$$

•

$$m(t_1, t_2) = m(q(d(y^*, t_2), y^*), t_2)$$

• But when $z = d(y^*, t_2)$ then

$$m(q(d(y^*, t_2), y^*), t_2) = P_z(d(y^*, t_2))$$

so

$$m(t_1, t_2) = P_z(d(y^*, t_2))$$

• $y^*$ is a known function of $t_1$ and $t_2$
3 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 nonlinear 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