Identifying multi-attribute hedonic equilibrium models

Victor Chernozhukov, MIT

Alfred Galichon, Sciences-Po

Marc Henry, Penn State

Hedonic approach

Motivations for the introduction of hedonic models

- Construction of price indexes that track quality improvements
- Explain price differentiation in art, luxury and other highly differentiated goods
- Infer the value of life, environmental features and other non marketable goods

Hedonic regression vs. hedonic equilibrium

Two related but distinct strands of literature:

- Hedonic regressions
  - Aim at recovering consumer WTP for good attributes/qualities from price regressions
  - Endogeneity issue: consumers with greater taste for attribute/quality z consume more of it
- Hedonic equilibrium models
  - General equilibrium model of consumption and production of attributes/qualities

Hedonic equilibrium models

- Perfectly competitive market

- Distribution of consumers with heterogeneous tastes for attributes/qualities determined by consumer characteristics $\tilde{x}$

- Distribution of producers with heterogeneous production capabilities, determined by producer characteristics $\tilde{y}$

- Distribution of traded attributes/qualities z and their prices $p(z)$ arise endogenously at equilibrium

Examples

- Wine market:
  - Wine qualities z include expert ratings, acidity, sugar content...
  - Wine maker's characteristics $\tilde{y}$ include grapes used, amount of sunshine, harvesting technology...
  - Consumer characteristics $\tilde{x}$ include age, gender, income...
- Labor market
  - Job characteristics z include riskiness, degree of interaction with co-workers...
  - Worker's (i.e., producer's) characteristics $\tilde{y}$ include education, age, experience...
  - Firm's (i.e., consumer's) characteristics $\tilde{x}$ include industry, size...

Existence of equilibrium

- Producers and Consumers are price takers
  - Consumers maximize $U(\tilde{x}, z) - p(z)$
  - Producers maximize $p(z) - C(\tilde{y}, z)$
- Equilibrium: collection $(P_{\tilde{x}z}, P_{\tilde{y}z}, p(z))$ such that markets clear
  - $P_{\tilde{x}z}$: distribution of who consumes what
  - $P_{\tilde{y}z}$: distribution of who produces what
  - $\sum_z P_{\tilde{x}z} = P_{\tilde{x}}$, $\sum_z P_{\tilde{y}z} = P_{\tilde{y}}$ and $\sum_{\tilde{x}} P_{\tilde{x}z} = \sum_{\tilde{y}} P_{\tilde{y}z}$
- Equilibrium exists: Ekeland (2010), Chiappori et al. (2010)

Identification problem

- Recovering preferences and technology from market data

$$\left\{ \begin{array}{c} P_{\tilde{x}z} \\ P_{\tilde{y}z} \\ p \end{array} \right\} \Longrightarrow \left\{ \begin{array}{c} U(\tilde{x}, z) \\ C(\tilde{y}, z) \end{array} \right\}$$

- Producers/consumers symmetric: concentrate on $U(\tilde{x}, z)$

- Introduce unobserved heterogeneity: $U(\tilde{x}, z) = \overline{U}(x, z) + \zeta(x, \varepsilon, z)$
  - $\tilde{x} = (x, \varepsilon)$, x observed, ε unobserved, $\varepsilon \perp\!\!\!\perp x$
  - ε is the vector of unobserved tastes for each quality dimension in z
  - We make assumptions on ζ and P_ε and recover $\overline{U}(x, z)$

State of the art

Case of a single quality dimension $z \in \mathbb{R}$

- Ekeland, Heckman and Nesheim (JPE 2004)
  - $\overline{U}(x,z)$ identified with additively separable marginal utility:
    1. $\overline{U}(x,z) = z\alpha(x) - \beta(z)$
    2. $\zeta(x,\varepsilon,z) = z\varepsilon$
    3. $E\varepsilon = 0$ and $E\varepsilon^2 = 1$
- Heckman, Matzkin and Nesheim (Econometrica 2010)
  - $\overline{U}(x,z)$ identified under a single crossing condition:
    1. $\zeta(x,\varepsilon,z)$ is known
    2. Single crossing: $\zeta_{\varepsilon z}(x,\varepsilon,z) > 0$
    3. Normalized distribution P_ε of unobserved heterogeneity

Quantile transform identification

Heckman, Matzkin and Nesheim (2010) strategy

- Consumer F.O.C.: $\zeta_z(x, \varepsilon, z) = p'(z) - \overline{U}_z(x, z)$
  - Defines implicitely an inverse demand function $z \mapsto \varepsilon(x, z)$
- From further differentiation and the single crossing condition,

$$\varepsilon_z(x, z) = -\frac{\overline{U}_{zz}(x, z) + \zeta_{zz}(x, \varepsilon, z) - p''(z)}{\zeta_{\varepsilon z}(x, \varepsilon, z)} > 0$$

- As an increasing function of z, the inverse demand $\varepsilon(x, z)$ is uniquely defined by the quantile transform

$$\varepsilon(x, z) = F_\varepsilon^{-1}\left(F_{z|x}(z|x)\right)$$

- Recover $\overline{U}(x, z) = p(z) - \int_0^z \zeta_z(x, \varepsilon(x, z'), z')dz'$ up to a constant

Dimension free reformulation

- Assortative matching: increasing inverse demand $z \mapsto \varepsilon(x,z)$ means consumers with higher taste for quality receive more
  - With increasing $z \mapsto \varepsilon(x,z)$, total surplus $\int_z \zeta(x,\varepsilon(x,z),z)\, dP_{z|x}$ maximizes $\int_{(z,\varepsilon)} \zeta(x,\varepsilon,z)\, dP_{z\varepsilon|x}$ over all possible joint distributions

  $\Rightarrow$ Efficiency of assortative matching

- With multiple quality dimensions, this efficiency formulation still holds and carries our identification results
  - Inverse demand $z \mapsto \varepsilon(x,z)$ solves the surplus maximization problem with marginal distributions P_ε and $P_{z|x}$ fixed

$$\max_{P_{z\varepsilon|x}} \int_{(z,\varepsilon)} \zeta(x,\varepsilon,z)\, dP_{z\varepsilon|x}$$

  - Uniqueness of the solution to this optimization problem yields identification

Separable marginal utility

Consider first the special case $\zeta(x, \varepsilon, z) = z'\varepsilon$

- Consumers solve $V^*(x, \varepsilon) = \sup_z \left\{ z'\varepsilon - \left[p(z) - \overline{U}(x, z) \right] \right\}$

- $V^*(x, \varepsilon)$ is the convex conjugate of $V(x, z) := p(z) - \overline{U}(x, z)$
  - Shape restriction: V convex

- F.O.C.: $\varepsilon(x, z) = \nabla p(z) - \nabla_z \overline{U}(x, z) = \nabla_z V(x, z)$
  - Inverse demand is the gradient of a convex function

  $\rightarrow$ Increasing inverse demand in the scalar case

As we shall see, this shape restriction guarantees uniqueness, hence identification of $\overline{U}(x, z)$

Optimal transport

- Consider a planner maximizing total surplus from unobserved tastes

$$\max_{P_{z\varepsilon|x}} \int_{(z,\varepsilon)} z'\varepsilon \, dP_{z\varepsilon|x}$$

- The dual of this problem is

$$\min_{V} \int_{z} V(x,z)\, dP_{z|x} + \int_{\varepsilon} V^{*}(x,\varepsilon)\, dP_{\varepsilon}$$

- There is a unique gradient of convex function $\nabla_z V(x,z)$ such that V solves the dual and $\varepsilon(x,z) = \nabla_z V(x,z)$ (F.O.C. of consumer problem) solves the primal

$\Rightarrow$ $\nabla_z V(x,z)$, hence $\nabla \overline{U}(x,z)$, is identified (and can be efficiently computed with convex programming)

Nonlinear simultaneous equations

As a corollary, the simultaneous equations system

$$z = f(x, \varepsilon),\ x \in \mathbb{R}^{d_x},\ \varepsilon \text{ and } z \in \mathbb{R}^{d_z}$$

is nonparametrically identified under the following:

- $P_{z|x}$ is absolutely continuous
- $\varepsilon \perp\!\!\!\perp x$ and P_ε is normalized
- f is the gradient of a convex function

$\rightarrow$ In the scalar case, this is Matzkin (Econometrica 2003)

General case

Consider now the general case $U(x, \varepsilon, z) = \overline{U}(x, z) + \zeta(x, \varepsilon, z)$

- Twist condition: $\zeta_z(x, \varepsilon, z)$ is an injective function of ε
  - Multivariate version of the single crossing condition

- Shape restriction: $V(x, z)$ is ζ-convex
  - Defined from a generalized notion of convex conjugacy:

$$V(x, z) = \sup_{\varepsilon}\{\zeta(x, \varepsilon, z) - V^{\zeta}(x, \varepsilon)\}$$

$$\text{with} \qquad V^{\zeta}(x, \varepsilon) = \sup_{z}\{\zeta(x, \varepsilon, z) - V(x, z)\}$$

Then, $\overline{U}(x, z)$ is nonparametrically identified

Optimal transport (continued)

- Consider a planner maximizing total surplus from unobserved tastes

$$\max_{P_{z\varepsilon|x}} \int_{(z,\varepsilon)} \zeta(x,\varepsilon,z)\ dP_{z\varepsilon|x}$$

- The dual of this problem is

$$\min_{V} \int_{z} V(x,z)\ dP_{z|x} + \int_{\varepsilon} V^{\zeta}(x,\varepsilon)\ dP_{\varepsilon}$$

- There is a unique gradient of ζ-convex function $\nabla_z V(x,z)$ such that V solves the dual and $\nabla_z V(x,z)$ solves the F.O.C.

$$\nabla_z \zeta(x,\varepsilon(x,z),z) = \nabla_z V(x,z),$$

  where the inverse demand $\varepsilon(x,z)$ solves the primal

$\Rightarrow \nabla_z V(x,z)$, hence $\nabla \overline{U}(x,z)$, is identified

Identification implications of Monge-Ampère

So far, preferences are just identified under the assumption that the distribution of unobserved heterogeneity is known. We now investigate what can be recovered if we relax this assumption.

From Brenier's Theorem, the inverse demand

$$\varepsilon(x,z) = \nabla_z \left[p(z) - \overline{U}(x,z) \right]$$

satisfies:

$$\int \zeta(\varepsilon) f_\varepsilon(\varepsilon) d\varepsilon = \int \zeta \left(\nabla_z p(z) - \nabla_z \overline{U}(x,z) \right) f_{z|x}(z|x) dz.$$

for all bounded continuous functions ζ.

Taking ζ equal to the identity and assuming only that P_ε has mean zero, instead of fixing the whole distribution, yields identification of averaged partial effects

$$\mathbb{E} \left[\nabla_z \overline{U}(x,Z) | x \right].$$

Additively separable marginal utilities

- In the scalar case, Ekeland, Heckman and Nesheim (2004) identify preferences based on the equality:

$$F_{z|x}(z|x) = F_\varepsilon(\nabla_z[p(z) - \overline{U}(x,z)]). \qquad (1)$$

  Differentiation with respect to x and to z respectively allows to eliminate the unknown F_ε and identify α and β.

- In the multivariate case, (2) no longer holds and has to be replaced with the Monge-Ampère equation:

$$f_{z|x}(z|x) \;=\; f_\varepsilon(\nabla_z[p(z) - \overline{U}(x,z)]) \text{ det } D^2[p(z) - \overline{U}(x,z)].$$

- Individual characteristics do not appear in the Jacobian for the change of variables from ε to z.

- Hence, differencing or differentiating the Monge-Ampère equation with respect to x eliminates the Jacobian term, so that further differentiation eliminates the unknown unobserved taste distribution f_ε.

This yields identification of J_b from knowledge of J_α.

Multiple market identification

Another source of identification to avoid normalization of P_ε is data from multiple markets.

- Distributions of producers and consumers vary with the market,
- Hence price functions $P(z)$ vary across markets
- Preferences, costs and the distribution of unobservables are supposed fixed.

Identified set: $\{P_\varepsilon : \overline{U}(x, z)$ is constant across markets$\}$.