A Semi-Parametric Preference Heterogeneity Model

Hubert Wu
University of Oxford

17 November 2022

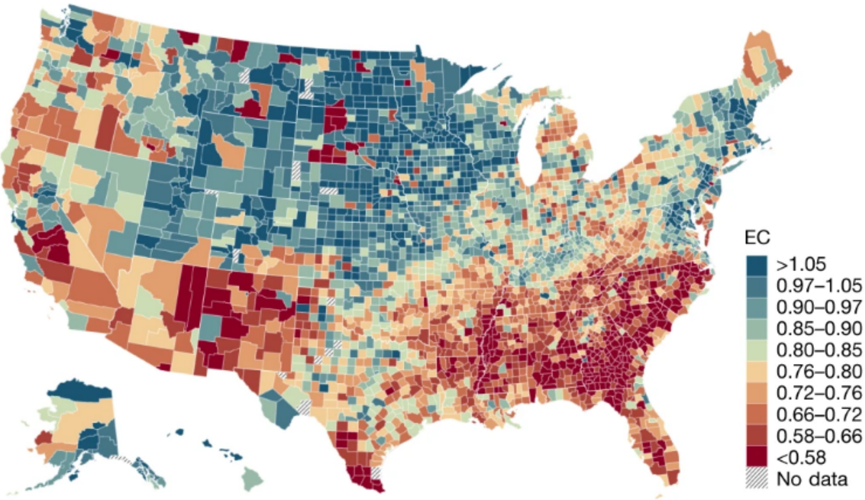
Preliminary and not for wider circulation

Heterogeneity is ubiquitous in economics...

The most important discovery [of microdata produced in the 1950s] was the evidence on the pervasiveness of heterogeneity and diversity in economic life... the long standing edifice of the representative consumer was shown to lack empirical support.

– James J. Heckman, *Microdata, Hetereogeneity, and the Evaluation of Public Policy*, Nobel Prize Lecture, 8 December 2000

Heterogeneity is ubiquitous in economics...



Chetty et al. (2022)

Heterogeneity is ubiquitous in economics...

- **Observed**: control with covariates, quantile regression, inference on ranks, etc.
- **Unobserved**: absorb via fixed effects or conduct inference based on

$$f(y|\mathbf{x},\theta) = \int f(y|\mathbf{x},\gamma)g(\gamma|\theta)d\gamma$$

with unobserved heterogeneity parameters γ_i with conditional density $g(\gamma_i|\theta)$, usually i.i.d

...but identification often necessitates trade-offs

Identifying heterogeneity often requires

- parametric restrictions
- error term rationalisation
- non-parametric 'latent classes' or 'types'

These assumptions often give up the ability of data to falsify an underlying model of interest

A semi-parametric preference heterogeneity model

- Working with Ian to develop a semi-parametric model of heterogeneity that addresses these problems
- My thesis studies this model and brings it to data

Setting

Discrete observations of prices and quantities for N agents, K goods, and T time periods:

- Data comprise $\mathbf{q}$ and $\mathbf{p}$, each an $NT \times K$ vector
- Note that this accommodates idiosyncratic prices for a given good

'NASP' preferences

Individual i has direct utility that is **non-separable, additive, and semi-parametric**:

$$v^i(\mathbf{q}) = u(\mathbf{q}) + \mathbf{a}^i.\mathbf{q} \quad (1)$$

- $\mathbf{a}$ (a $K \times 1$ vector) introduces hetereogeneity in marginal utilities
- $u(.)$ is common to everyone
- no restrictions placed on $u(.)$ other than that it is rational

'Rational?'

Definition

A utility function $u(.)$ **rationalises** the data $\{p_t, x_t\}_{t=1,2,...T}$ if for all x such that $p_t.x_t \geq p_t.x$ we have $u(x_t) \geq u(x)$

Before any kind of estimation

First consider two questions:

1. What is the empirical content of utility maximisation under (1)?
    - Put another way: can data falsify our theory? If so, what would those data look like?
2. What kinds of heterogeneity does (1) rule out?

Revealed preference preliminaries

- A **binary relation** on X is a subset of X^2 where we write xRy for the binary relation R if $(x, y) \in R$
- An **order pair** on X is a pair of binary relations $\langle R, S \rangle$ such that $S \subseteq R$
- An order pair is **acyclic** if there is no sequence $x_1, x_2, ...x_k$ such that

$$x_1 R x_2 R ... R x_k$$

Revealed preference preliminaries

Definition

A dataset's **revealed preference pair** is an order pair $\langle \succsim_{RP}, \succ_{RP} \rangle$ on X with its relations defined by:

- $x \succsim_{RP} y$ iff there exists s such that $x = x^s$ and $p^s.x^s \geq p^s.y$
- $x \succ_{RP} y$ iff there exists s such that $x = x^s$ and $p^s.x^s > p^s.y$

Revealed preference preliminaries

Definition

A dataset satisfies the **Generalized Axiom of Revealed Preference (GARP)** if its revealed preference pair is acyclic.

The NASP form provides a restriction for panel data

Proposition 1 (Crawford, 2022)

The following statements are equivalent:

1. The data $\{\mathbf{p}, \mathbf{q}\}$ for which each individual i satisfies GARP are jointly rationalisable by a utility function of the form in (1)
2. There exists a rationalisation of the form in (1) that is continuous, concave, and strictly monotonic
3. There exist real numbers and vectors $\{U_t^i, \lambda_t^i > 0, \mathbf{a}^i\}_{t=1,\dots T}^{i=1,\dots,N}$ such that

$$U_s^i \leq U_t^j + \lambda_t^j \mathbf{p}_t^j . \left(\mathbf{q}_s^i - \mathbf{q}_t^j\right) - \mathbf{a}^j \left(\mathbf{q}_s^i - \mathbf{q}_t^j\right) \tag{2}$$

$$\lambda_t^j \mathbf{p}_t^j - \mathbf{a}^j \geq \mathbf{0}_K \tag{3}$$

Proposition 1 is a version of Afriat's Theorem (Afriat, 1967) for (1).

The restriction is a rationalisation

- The Afriat inequalities in (2)-(3) define a linear program that can be checked for a solution with well-known algorithms
- Without these we would have to check the data against every function satisfying (1)... of which there are an infinite number :'(

Economic intuition for the Afriat inequalities (2)-(3)

Take Part 2 of Proposition 1 to be true: $\{\mathbf{p}, \mathbf{q}\}$ has a 'nice' rationalisation in the form of (1).

Concavity of $\upsilon^i(\mathbf{q})$ implies that

$$\upsilon^i(\mathbf{q}_s^j) \leq \upsilon^i(\mathbf{q}_t^i) + \nabla u^i(\mathbf{q}).(\mathbf{q}_s^j - \mathbf{q}_t^i), i, j \in \{1, ..., N\}; s, t \in \{1, ..., T\}$$

Optimising behaviour requires the FOC

$$\nabla u^i(\mathbf{q}) = \lambda_t \mathbf{p}_t^i - \mathbf{a}^i$$

(2)-(3) request utility levels, shadow prices, and heterogeneity parameters to satisfy these conditions.

However, we're interested in heterogeneity

- Our primary object of interest are the vectors $\mathbf{a}^i$
- A solution to (2)-(3) means we can proceed as if (1) holds, and restrict the possible values these vectors can take
- Can we go further?

Finding bounds on heterogeneity

A natural next question is "what is the 'minimum amount' of heterogeneity consistent with the data?"

- Why not the maximum?

One way to formulate this problem is:

$$\min_{\{\mathbf{a}^1,\mathbf{a}^2,\dots,\mathbf{a}^N\}} \sum_{i=1}^{N} ||\mathbf{a}^i||_2 \tag{4}$$
$$\text{subject to } A\mathbf{x} \leq \mathbf{b}$$

where $||.||$ is the ℓ^2-norm and $A\mathbf{x} \leq \mathbf{b}$ are the (re-written) Afriat inequalities for (1).

Some comments about (4)

Pub trivia

- (4) is a special case of the 'general Fermat problem' of minimising a sum of heterogeneous norms
- the problem is "the earliest example of duality in the mathematical programming literature" (see Anderson et al., 2000; Kuhn, 1974 for more)

Some comments about (4)

Computational issues

- The number of constraints in the problem $((NT)^2 + NTK)$ rises faster than the square of N and T
- Cf. no. choice variables: $N(3T + K)$
- Potentially of more concern: (4) is no longer an LP

How hard is it to solve (4)?

The answer may not be that bad because:

Fact

(4) is a convex optimisation problem.

Why? (1) Any norm $||.||$ is a convex function; and (2) the weighted sum of m convex functions $f_1, ..., f_m$, $f := \sum_{i=1}^{m} \alpha_i f_i$ is convex if the weights $\alpha_1, ..., \alpha_m$ are non-negative.

The point of all this: although we "cannot yet claim solving general convex optimization problems is a mature technology" (Boyd and Vandenberghe, 2004), we are not without options.

One option: quadratic programming

With a non-negativity assumption for $\mathbf{a}^i$ (4) can be reformulated as the constrained least-squares problem,

$$\min_{\mathbf{x}} \mathbf{x}' Q \mathbf{x} \tag{5}$$
$$\text{subject to } A\mathbf{x} \leq \mathbf{b}$$

where

- $\mathbf{x}$ is the $\mathbb{R}^{N(3T+K)}$ vector of decision variables
- Q is a $N(3T+K) \times N(3T+K)$ square 'selection matrix' that takes the value of 0 except for in the diagonal entries corresponding to where the as are in $\mathbf{x}$; the latter entries equal 1.

Issues with solving (5) via CLS

Aside from the additional assumption, (5) can be involved to implement because its dual,

$$\max_{\lambda} \frac{1}{2}\lambda' A Q^{-1} A' \lambda - \lambda' b \qquad (6)$$
$$\text{subject to } \lambda \geq \mathbf{0}$$

asks for an inverse of Q which is positive semidefinite in our context.

There are ways around this (e.g. pseudo-inverses, as Yang is working with) but these can be ill-behaved from high correlation between the rows of Q.

A better option: conic programming

The path we have chosen involves **cones**. Recall from convex analysis that

Definition

$C \in \mathbb{R}^n$ is a **convex cone** if

- for all $x, y \in C$ we have $x + y \in C$
- for all $x \in C$ and $\alpha \geq 0$ we have $\alpha x \in K$

Cones are abundant in economics. For example, under constant returns to scale the production set is a cone.

Cones are also useful in optimization. For example, the dual of a closed convex cone is easy to derive and is itself a convex cone.

Conic formulation of (4)

The conic programming version of (4) is:

$$\min_{\mathbf{x}} \sum_{i=1}^{N} t_i \tag{7}$$
$$\text{subject to } (t_i, F_i\mathbf{x}) \in \mathcal{Q}_i^{K+1}$$
$$\text{and } A\mathbf{x} \geq \mathbf{b}$$

where $\mathcal{Q}_i^{K+1}$ is the $K+1$-dimensional quadratic cone defined as

$$\mathcal{Q}^{K+1} = \{x \in \mathbb{R}^{K+1} | x_1 \geq \sqrt{x_2^2 + ... x_{K+1}^2}\} \tag{8}$$

Intuitively, individuals' 'aggregated heterogeneity contributions' $||\mathbf{a}^i||_2$ live within a cone with dimension defined by the goods space K.

Applications: empirical work

1. Minimizing heterogeneity in micro panel data
   - CentERpanel: representative, weekly, stratified survey of c. 2,000 households and 5,000 individuals
2. (Speculative) Finding feasible Pareto weights in Dworczak, Kominers, and Akbarpour (2021)

Applications: dealing with other forms of heterogeneity

When $\mathbf{a}^i$ is time-invariant, there are some forms of heterogeneity that (1) cannot represent.

Consider for example the case where $i = \{1, 2\}$, $\mathbf{q} = (x, y, z)'$, $u(\mathbf{q}) = x$, and where the true preferences for the agents are respectively represented by

$$\upsilon^1(\mathbf{q}) = x + \sqrt{yz} \tag{9}$$

$$\upsilon^2(\mathbf{q}) = x + y + z \tag{10}$$

Now suppose each agent's choice set comprises the bundles $\mathbf{q}_1 = (1, 0, 3)'$ and $\mathbf{q}_2 = (1, 1, 1)'$. Adding a third bundle $\mathbf{q}_3 = (1, 1, 9)'$ to the choice set cannot be rationalised by (1).

Applications: dealing with other forms of heterogeneity

Here are two helpful results about such cases:

1. If the data $\{\mathbf{p}_t^i, \mathbf{q}_t^i\}$ can be rationalised by a utility function, then they can also be rationalised by a utility function of the form in (1).
2. In any cross section of N individuals, the dataset $\{\mathbf{p}_t^i, \mathbf{q}_t^i\}$ is rationalisable by a utility function of the form in (1)

Corollary

Forms of preference heterogeneity that cannot be rationalised by (1) can be rationalised in the form of (1) when $\mathbf{a}^i$ is time-varying.

Applications: welfare analysis

The idiosyncratic part of utility $\mathbf{a}^i.\mathbf{q}$ can be used as the basis of interpersonal welfare comparisons since the difference in utility for individuals i and j is,

$$\Delta v^{ij}(\mathbf{q}) = v(\mathbf{q})^i - v(\mathbf{q})^i = (\mathbf{a}^i - \mathbf{a}^j)\mathbf{q} \tag{11}$$

Note that since $\Delta v^{ij}(\mathbf{q}) = 0$ has an economic meaning, (11) implicitly restricts the set of acceptable utility profile transformations for this purpose.