Heterogeneity

Krishna Pendakur

May 24, 2015

Introduction

- People are heterogeneous.
- Some heterogeneity is observed, some is not observed.
- Some heterogeneity affects cost, some affects preferences, some affects both.
- We need to account for heterogeneity when measuring inequality, poverty, the cost-of-living, etc.
- *Equivalence scales*, which measure the ratio of cost (functions) across household types, are venerable (Engel (1895), Sydenstricker and King (1921), Rothbarth (1943), Prais and Houthakker (1955), Barten (1964), Gorman (1976), Lewbel (1985).
- *Equivalent income*, equal to income scaled by the equivalence scale, gives the amount of money needed for a single individual to attain the same utility level as some reference type of individual.
- Comparing cost across people is comparing utility across people.

Equivalence Scales

$\mathbf{p}$ price vector, x expenditure, $\mathbf{z}$ characteristics vector, u utility, $\mathbf{w}$ budget-share vector.
$C(\mathbf{p}, u, \mathbf{z})$ cost to attain utility u for person with characteristics $\mathbf{z}$ when facing prices $\mathbf{p}$. $V(\mathbf{p}, x, \mathbf{z})$ indirect utility (inverse of cost over u).
Define

$$\Delta(\mathbf{p}, u, \mathbf{z}) = C(\mathbf{p}, u, \mathbf{z}) / C(\mathbf{p}, u, \bar{\mathbf{z}})$$

as the "equivalence scale" relating costs for a person with characteristics $\mathbf{z}$ to a reference person with characteristics $\bar{\mathbf{z}}$.

- Δ is not identified from behaviour: given a $\mathbf{z}$-specific monotonic transformation $\phi(u, \mathbf{z})$, $\Delta(\mathbf{p}, \phi(u, \mathbf{z}), \mathbf{z}) \neq C(\mathbf{p}, u, \mathbf{z}) / C(\mathbf{p}, u, \bar{\mathbf{z}})$
- $\phi(u, \mathbf{z})$ affects the equivalence scale, but not behaviour: thus there are an infinite number of equivalence scale functions consistent with the same behaviour.
- $\phi(u, \mathbf{z})$ structures interpersonal comparisons of utility.

What are characteristics?

- Could include characteristics of individuals, e.g., disability status.
  - this clearly affects cost, but may also affect preferences.
  - So, $\Delta(\mathbf{p}, u, \mathbf{z})$ should have both a level part (to hit cost regardless of prices) and a price response (to hit preferences).
- Could include characteristics of household, like size of household.
  - the household doesn't have utility, just the individuals inside it.
  - E.g., let $\mathbf{z} = n$, the number of household members.
  - $C(\mathbf{p}, u, n)$ then gives the amount of *household* expenditure needed to give *each* of the n members of the household a utility level of u.
  - Embedded here is the assumption that each individual in the household gets the same utility level.
  - More on this later.

IB/ESE and Identification of Equivalence Scales

Let $\Delta(\mathbf{p}, u, \mathbf{z}) = E(\mathbf{p}, \mathbf{z})$ which is independent of u. This restriction is called 'independent of base utility' (IB) or 'equivalence-scale exact' (ESE), and has been studied by Lewbel (1989), Blundell and Lewbel (1991), Blackorby and Donaldson (199), Pendakur (1999), Blundell, Duncan and Pendakur (1998) and Donaldson and Pendakur (2004, 2006).

- Given a $\mathbf{z}$-specific monotonic transformation $\phi(u, \mathbf{z})$, $\Delta(\mathbf{p}, \phi(u, \mathbf{z}), \mathbf{z}) \neq \Delta(\mathbf{p}, u, \mathbf{z})$, so there are an infinite number of equivalence scale functions consistent with behaviour.
- But, Blackorby and Donaldson (1993) show that, given IB/ESE, only one of them is independent of utility.
- The functional form restriction 'solves' the identification problem.
- Here, cost is related across z by

$$C(\mathbf{p}, u, \mathbf{z}) = E(\mathbf{p}, \mathbf{z}) C(\mathbf{p}, u, \bar{\mathbf{z}})$$

Shape-Invariance

- Pendakur (1999) shows that, given IB/ESE, the budget-share vector-function $\mathbf{w}(\mathbf{p}, x, \mathbf{z})$ satisfies

$$\mathbf{w}(\mathbf{p}, x, \mathbf{z}) = \mathbf{w}\left(\mathbf{p}, \left(\ln x - \ln E(\mathbf{p}, \mathbf{z})\right), \bar{\mathbf{z}}\right) + \mathbf{d}(\mathbf{p}, \mathbf{z})$$

  for any integrable $\mathbf{w}(\mathbf{p}, x, \bar{\mathbf{z}})$, where the vector-function $\mathbf{d}(\mathbf{p}, \mathbf{z}) = \nabla_{\ln \mathbf{p}} \ln E(\mathbf{p}, a)$.
- Specify an indirect utility function for the reference type, e.g., QAI:

$$\mathbf{w}(\mathbf{p}, x, \bar{\mathbf{z}}) = \mathbf{a} + \mathbf{A} \ln \mathbf{p} + \mathbf{b} \ln \bar{x} + \frac{\mathbf{q} \ln \bar{x}^2}{\exp\left(\mathbf{b}' \ln \mathbf{p}\right)}.$$

  where $\ln \bar{x} = \ln x - \ln \mathbf{p}' \left(\mathbf{a} + \frac{1}{2}\mathbf{A} \ln \mathbf{p}\right)$
- Substitute the lower equation into the upper one, and estimate.

Shape-Invariant Engel Curves

- An *Engel curve* is a budget share equation evaluated at a single price regime.
- Without price variation (ie, evaluated at some $\overline{\mathbf{p}}$), and evaluating over $\ln x$, we have

$$\mathbf{w}(\ln x, \mathbf{z}) = \mathbf{w}(\ln x - \ln E(\mathbf{z}), \overline{\mathbf{z}}) + d(\mathbf{z})$$

  where $\mathbf{d}(\mathbf{z}) = \nabla_{\ln \mathbf{p}} \ln E(\overline{\mathbf{p}}, \mathbf{z})$.
- These Engel curves are 'shape-invariant', with the horizontal translation giving $\ln E$.
- You don't even need to know what shape $\mathbf{w}(\ln x, \overline{\mathbf{z}})$ has over $\ln x$.
- You can just draw pictures of $\mathbf{w}$ over $\ln x$ for different $\mathbf{z}$, and find the closest translations.
  - Blundell, Chen and Christensen (2007) do this.

Operationalising Equivalence Scales

- $\ln E$ is the log of equivalence scale.
- By definition, if you give \$$x$ to a person with characteristics $\mathbf{z}$, they have the same utility level as a reference person(with characteristics $\bar{\mathbf{z}}$) with \$$x/E$. We therefore say that \$$x/E$ is the *equivalent-expenditure* of that person.
- Equivalent expenditures structure interpersonal comparisons of utility—they make one person equivalent in a welfare sense to another person (with a different expenditure level).
- the equivalence in welfare terms derives from 2 things: 1) social welfare functions are anonymous, in that if two people have the same utility, they affect social welfare equally; and 2) the equivalence scale structures interpersonal comparisons of utility enough so that we can tell when two people have the same utility.

Using Equivalence Scales

- To measure inequality or poverty, you:
  - compute E_h for each household $h = 1, ..., H$ in the population.
  - Assign to each individual member of each household the equivalent expenditure $x_i^e = x_h / E_h$ for all i in h.
- Do your welfare analysis: poverty (count x_i^e below a threshold); inequality (what is the variance of x_i^e); welfare (what is mean of x_i^e less a variance penalty).
- Do not follow Ebert and Moyes (2006) and weight individuals by $1/E_h$. This is required to get household inequality measures to coincide with individual inequality measures. But, household measures are meaningless, since individuals have utility, not households.

Equivalence Scales and the Cost of Living

- exact equivalence scales also put structure on the cost-of-living index.
- consider the cost-of-living change for a price change given ESE:

$$C(\mathbf{p}, u, \mathbf{z}) = E(\mathbf{p}, \mathbf{z}) C(\mathbf{p}, u, \bar{\mathbf{z}}),$$

so

$$\frac{C(\mathbf{p}_1, u, \mathbf{z})}{C(\mathbf{p}_0, u, \mathbf{z})} = \frac{E(\mathbf{p}_1, \mathbf{z})}{E(\mathbf{p}_0, \mathbf{z})} \frac{C(\mathbf{p}_1, u, \bar{\mathbf{z}})}{C(\mathbf{p}_0, u, \bar{\mathbf{z}})}.$$

- So, the cost-of-living change for anybody is equal to that of the reference household multiplied by the relative size of the equivalence scale in the two price regimes.
- the cost-of-living index has an elasticity wrt utility that does not depend on characteristics. This completely characterises the behavioural restrictions of IB/ESE.

- Donaldson and Pendakur (2004, 2006) give other, more general, functional form restrictions that allow identification. These allow for equivalence scales which rise or fall with expenditure.
  - ESE:
  $$\ln x_i^e = \ln R(p) + \ln x$$
  - Generalised ESE:
  $$\ln x_i^e = \ln R(p) + c(p) \ln x$$
  - Absolute ESE:
  $$x_i^e = A(p) + x$$
  - Generalised Absolute ESE:
  $$x_i^e = A(p) + R(p)x$$

Carsten, Kouvoulatianos, and Schroeder (2005, 2007) pursue a different, more Leiden-like, strategy: they ask people how costs vary with characteristics. This circumvents the identification problem without functional form restrictions. They find some support for Generalised Absolute ESE.

Unobserved Preference Heterogeneity

- Up to now, heterogeneity has been observable. What if it is not observed by the researcher?
- It screws up pretty much everything.
- E.g., if there is unobserved heterogeneity in the level of cost, then $\ln E$ has an unobserved component, like $\ln E = \ln \overline{E} + \nu$.
- So, our nonlinear Engel curve equations

$$\mathbf{w}(\ln x, \mathbf{z}) = \mathbf{w}(\ln x - \ln \overline{E}(\mathbf{z}) - \nu, \bar{\mathbf{z}}) + d(\mathbf{z})$$

  have a regressor $(\ln x)$ with measurement error (ν).
- If there is unobserved heterogeneity in preferences, then could also be error terms in $d(\mathbf{z})$.
- But, since $d(\mathbf{z})$ is the price elasticity of $E(\mathbf{p}, \mathbf{z})$, this has implications on price responses of equivalence scales.

More Unobserved Preference Heterogeneity

- Hoderlein (2007, 2008) shows that unobserved preference heterogeneity also makes it hard to test rationality restrictions with cross-sectional data.
- Although tests of homogeneity are robust to the presence of such heterogeneity, tests are symmetry and concavity are totally ruined.
- Matzkin (so many papers) and Beckert (a couple of papers) show that one can estimate demand models that allow for arbitrary unobserved preference heterogeneity by using quantile estimators.
- But, Slutsky symmetry and concavity tend to be pretty hard to impose in these contexts.

Hicks Demands

- Let ϵ be a vector of random utility parameters. (Forget observed heterogeneity for the moment).
- Log total expenditure $\ln x$, Log cost function $\ln x = \ln C(\mathbf{p}, u, \epsilon)$
- Shephard's lemma: $\mathbf{w} = \mathbf{\Omega}(\mathbf{p}, u) = \nabla_{\mathbf{p}} C(\mathbf{p}, u, \epsilon)$.
- These are Hicks demands, easy to have flexible $\mathbf{p}$ and u effects, linear in parameters, and additive errors. e.g., C a poly in $\mathbf{p}$, u plus $\mathbf{p}'\epsilon$:

$$\ln C = \ln \mathbf{p}' \left(\mathbf{a} + \frac{1}{2}\mathbf{A} \ln \mathbf{p} + \mathbf{b}u + \frac{1}{2}\mathbf{B} \ln \mathbf{p}u + \mathbf{c}u^2 + ... + \epsilon \right)$$

$$\mathbf{w} = \mathbf{\Omega}(\mathbf{p}, u, \epsilon) = \mathbf{a} + \mathbf{A}\mathbf{p} + \mathbf{b}u + \mathbf{B}\mathbf{p}u + \mathbf{c}u^2 + ... + \epsilon$$

- If u were observable, this would be totally easy to do. Alas, it isn't observable.
- Pendakur and Sperlich (2009, 2010) provide a difficult semiparametric way to estimate u and then $\mathbf{w}$. Allows for arbitrary $\mathbf{A}(u)$.

Implicit Marshallian Demands

- Problem with Hicks demands: u not observed.
- Solution: Implicit Marshallian Demands.
- Lewbel and Pendakur 2009, "Tricks with Hicks: The EASI Demand System".
- The idea: construct $C(\mathbf{p}, u, \epsilon)$ so that $u = g\left[\boldsymbol{\Omega}(\mathbf{p}, u, \epsilon), \mathbf{p}, x\right]$ for a simple g.
- Then let $y = g(\mathbf{w}, \mathbf{p}, x)$ and estimate Implicit Marshallian demand functions:

$$\mathbf{w} = \boldsymbol{\Omega}(\mathbf{p}, y, \epsilon)$$

- In our applications y is linear in x, and $y \approx$ a log money metric utility measure so will call y log real expenditures.

Implicit Marshallian Demands: Trivial Example

- let $\ln x = \ln C(\mathbf{p}, u, \epsilon) = u + \ln \mathbf{p}' [\mathbf{m}(u) + \epsilon]$
- By Shephard's lemma, $\mathbf{w} = \mathbf{\Omega}(\mathbf{p}, u, \epsilon) = \mathbf{m}(u) + \epsilon$
- so $u = \ln x - \ln \mathbf{p}' [\mathbf{m}(u) + \epsilon] = \ln x - \ln \mathbf{p}' \mathbf{\Omega}(\mathbf{p}, u, \epsilon)$,
- so let $y = g(\mathbf{w}, \mathbf{p}, x) = \ln x - \ln \mathbf{p}'\mathbf{w}$.
- $\ln \mathbf{p}'\mathbf{w}$ is the log of the geometric mean of prices. The geometric mean of prices was proposed by Stone (1954) as an intuitive price index, and is named after him.
- y is the log of Stone index deflated x.
- Implicit-Marshallian budget shares are

$$\begin{aligned} \mathbf{w} &= \mathbf{m}(\ln x - \ln \mathbf{p}'\mathbf{w}) + \epsilon \\ \mathbf{w} &= \mathbf{m}(y) + \epsilon \end{aligned}$$

- Endogenous y, but instruments $\ln \mathbf{p}, \ln x$ are available.
- Here, y is Exact Stone Index (ESI) deflated expenditure.

An EASI model

- Make a model where y is exactly affine in stone-index (EASI) deflated expenditure:
- Consider the cost function

$$\begin{aligned}\ln C(\mathbf{p}, u, \mathbf{z}, \epsilon) &= u + \ln \mathbf{p}' \left[\sum_{r=0}^{M} \mathbf{b}_r u^r + \mathbf{C}\mathbf{z} + \mathbf{D}\mathbf{z}u \right] \\ &\quad + \frac{1}{2} \sum_{l=0}^{L} z_l \ln \mathbf{p}' \mathbf{A}_l \ln \mathbf{p} + \frac{1}{2} \ln \mathbf{p}' \mathbf{B} \ln \mathbf{p} u + \ln \mathbf{p}' \epsilon.\end{aligned}$$

Has Implicit Marshallian demands:

$$\mathbf{w} = \sum_{r=0}^{M} \mathbf{b}_r y^r + \mathbf{C}\mathbf{z} + \mathbf{D}\mathbf{z}y + \sum_{l=0}^{L} z_l \mathbf{A}_l \ln \mathbf{p} + \mathbf{B} \ln \mathbf{p} y + \epsilon,$$

where

$$y = g(\mathbf{w}, \mathbf{p}, x, \mathbf{z}) = \frac{x - \ln \mathbf{p}' \mathbf{w} + \sum_{l=0}^{L} z_l \ln \mathbf{p}' \mathbf{A}_l \ln \mathbf{p}/2}{1 - \ln \mathbf{p}' \mathbf{B} \ln \mathbf{p}/2}.$$

Nice and EASI

- Engel curves are arbitrary functions in $y, \mathbf{z}$. The rank of demand is M. There are no Gorman type rank restrictions.
- Additive errors are (coherent, invertible) random preference heterogeneity.
- Like AI demands, linear in parameters up to y. Approximate model

  takes $y = x - \ln \mathbf{p}'\mathbf{w}$. Could get via linear estimation.
- Closed form expressions for consumer surplus, cost of living indices, etc.,.

EASI Estimation

- The empirical model in Lewbel and Pendakur 2009 is

$$\mathbf{w} = \sum_{r=0}^{5} \mathbf{b}_r y^r + \mathbf{Cz} + \mathbf{Dz}y + \sum_{l=0}^{L} z_l \mathbf{A}_l \ln \mathbf{p} + \mathbf{B} \ln \mathbf{p} y + \epsilon,$$

where

$$y = g(\mathbf{w}, \mathbf{p}, x, \mathbf{z}) = \frac{x - \ln \mathbf{p}'\mathbf{w} + \sum_{l=0}^{L} z_l \ln \mathbf{p}'\mathbf{A}_l \ln \mathbf{p}/2}{1 - \ln \mathbf{p}'\mathbf{B} \ln \mathbf{p}/2}.$$

- So, nonlinear system GMM would make instruments r as close as possible to orthogonal to a residual vector

$$\mathbf{e} = \mathbf{w} - \sum_{r=0}^{5} \mathbf{b}_r y^r + \mathbf{Cz} + \mathbf{Dz}y + \sum_{l=0}^{L} z_l \mathbf{A}_l \ln \mathbf{p} + \mathbf{B} \ln \mathbf{p} y$$

- y is the endogenous variable.
- Instruments are natural: $y = g(\mathbf{w}, \mathbf{p}, x, \mathbf{z})$, so $\mathbf{p}, x, \mathbf{z}$ are the instruments.

What's the Point?

- Heterogeneity is everywhere.
- Measurement of the cost-of-living, cost-of-characteristics, inequality, poverty, etc, have to take it seriously.
- We've made progress over the last decade.
- Observed heterogeneity can be accomodated in demand.
  - We can estimate its effect on cost and on demand
- Unobserved heterogeneity can be accomodated in demand, too.

Where to Go?

- Heterogeneity of unknown form—what can we learn, eg, about averages? (Hoderlein 2009)
- What form does unobserved heterogeneity take?
  - panels (Christensen 2008);
  - Crawford and Pendakur (2009)
- Is it 'almost ignorable', like in approximate representative agent macro models?
- Collective Household models