The roles of nonparametric identification and economic theory for applied work in economics

Trial Lecture

Lukáš Lafférs

24 March 2014

Identification in econometric models maps prior assumptions and the data to information about the quantity of interest. In other words, what functions or features of functions can be recovered from the probability distribution of observable variables.

Can **economic theory** provide any useful restrictions that help to recover the quantity of interest without imposing unnecessary **parametric** structure?

Examples of restrictions from economic theory

- shape restrictions: concavity, continuity or monotonicity of functions (utility function, demand function, production function)
- implications of optimization (first order conditions)
- equilibrium conditions
- exclusion restrictions (an instrument does not appear in the equation of interest)
- long-run restrictions on covariance matrix of errors in VAR models (money-supply shock has no long-run effect on output)

Outline

- Examples
- Identification
- History of thought
- Nonparametric Identification + Applications
- Partial Identification + Applications

Example 1 (Matzkin 1994)

A firm operating in a perfectly competitive market decides whether to invest in a development of a new product.

We wish to know

- **cost function** of a typical firm
- **distribution of the revenues**

We observe input prices $(x^1, x^2, ..., x^N)$ for the N firms and whether they invested $(y^i = 1)$ or not $(y^i = 0)$.
We take revenue $(\epsilon \geq 0)$ as a random variable.

Example 1 - Model Restrictions

Properties of the production function:

- monotonous
- convex
- homogeneous of degree one in prices

Further assumptions

- revenue is independent of input prices
- the distribution of revenue ϵ, F is strictly increasing.
- the value of the cost function h is known for a particular vector of input prices. $h(x^*) = \alpha$

Question of Identification

Will the assumptions enable us to recover the cost function (h) and the distribution of revenues (F)?

It turns out that yes

$$g(x) \equiv P(y=0|x) = Pr(\epsilon \leq h(x)) = F(h(x))$$

$$F(t) \overset{norm}{=} F((t/\alpha)h(x^*)) \overset{h.o.d.1}{=} F(h((t/\alpha)x^*)) = g((t/\alpha)x^*)$$

$$h(x) \overset{mono}{=} F^{-1}g(x)$$

$\implies$ (h, F) is identified.

- parametric model: $h(x) = x'\beta,\ F \sim \ln N(\mu, \sigma^2)$
- semi-parametric model: $h(x) = x'\beta$
- nonparametric model: no parametric restrictions on both (h, F)

Application: Gandhi, Navarro and Rivers (2013)

Identification

P - true distribution of the observed data

$\mathbf{P} = \{P_\theta : \theta \in \Theta\}$ - econometric model

Correct specification: There exists θ such that $P_\theta = P$.

$P \in \mathbf{P} \implies \theta \in \Theta_0(P) = \{\theta \in \Theta : P_\theta = P\}$

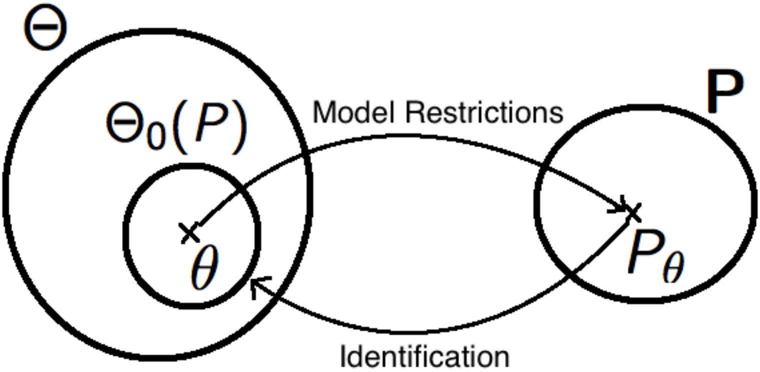
$\Theta_0(P)$ is identified set

- $\Theta_0(P) = \emptyset$ - model is refuted
- **$\Theta_0(P)$ is a singleton - point identification**
- $\Theta_0(P)$ is a set - partial identification
- $\Theta_0(P) = \Theta$ - no identification

We say θ is identified if for all $P \in \mathbf{P}$: $\Theta_0(P)$ is a singleton.

We may be interested in a certain feature of θ, $\psi(\theta)$.

Identification



What if the identification fails?

If we treat unidentified model as if it was identified:

- Parameters, Tests and Confidence sets have no clear interpretation
- Consistent estimation is not possible
- Statistical inference methods are not valid
- Numerical problems (inverting singular matrices)

Back to the production function

$\theta = (h, F) \in \Theta$

$\mathbf{P} = \{P_\theta$: h does not depend on ϵ, $h(ax) = ah(x)$, $h(x^*) = \alpha$, F is strictly increasing, $Pr(y = 0|x) = F(h(x))\}$

Example 2 - Linear Regression Model

$$Y = X'\beta + \epsilon$$

$\theta = (P_X, \beta, P_{\epsilon|X})$
$\mathbf{P} = \{P_\theta : Y = X'\beta + \epsilon\}$

Under these identifying assumptions:

- $E_{P_\theta}[\epsilon|X] = 0$,
- $P_\theta(X \text{ has full rank}) = 1$,

θ is indentified.

Example 3 - Generalized Regression Model

$$Y = g(X) + \epsilon$$

$\theta = (P_X, g, P_{\epsilon|X})$
$\mathbf{P} = \{P_\theta : Y = g(X) + \epsilon\}$

Under these identifying assumptions:

- $E_{P_\theta}[\epsilon|X] = 0$,
- $P_\theta(X \text{ has full rank}) = 1$,

θ is indentified.

Example 3 - Generalized Regression Model - 2

$$Y = g(X) + \epsilon$$

$X = (p, I)$ (price and income), Y - quantity demanded

If $g(X)$ is a demand function we might also be interested in the **utility function** that generates such demand.

Mas-Colell (1977) - Recoverability of U

- Let $\mathcal{W}$ denotes a set of monotone increasing, continuous, concave and strictly quasi-concave functions, such that no two functions are strictly increasing transformation of each other.
- Let $U \in \mathcal{W}$ and get $g_U(X)$ denotes the demand function generated by U.

Then U is identified $(\forall U, U' \in \mathcal{W} : g_U(X) = g_{U'}(X) \implies U = U')$

Parametric, Semi-parametric, Nonparametric Models

- Parametric model: θ is finite dimensional
- Semi-parametric model: θ consists of finite dimensional parameter and functions (Example 2: Linear Regression Model)
- Nonparametric model: θ consists of functions

Nonparametric Models

Pros

- more credible assumptions
- more flexible
- economic restrictions

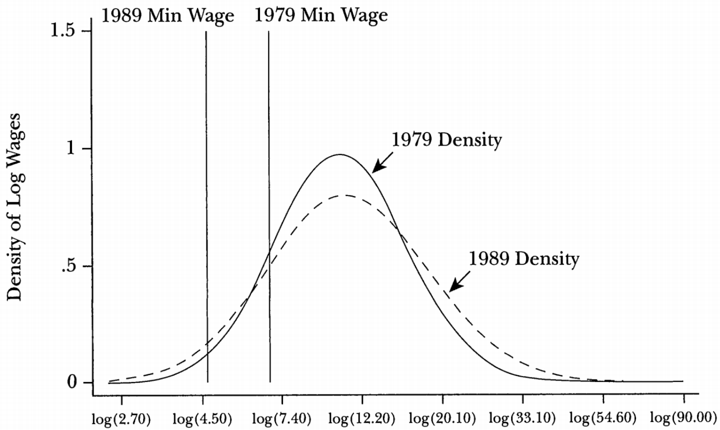
Cons

- curse of dimensionality
- more difficult to implement
- sometimes harder to interpret

Why Non-parametric?

(DiNardo and Tobias 2001)

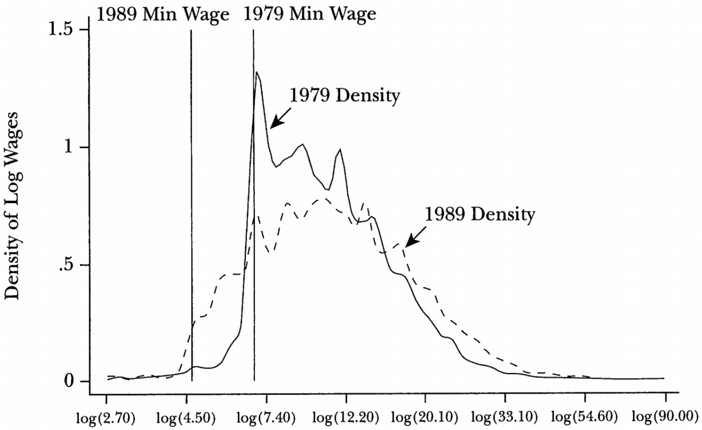
Parametric model:



Women's Wages 1979 and 1989: A Parametric View
(2000 Constant Dollars)

Why Non-parametric?

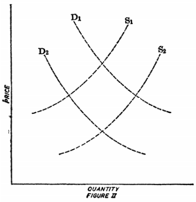
(DiNardo and Tobias 2001)
Non-parametric model:

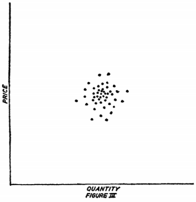


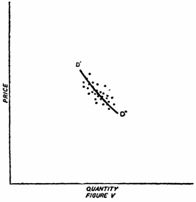
Women's Wages 1979 and 1989: A Nonparametric View
(2000 Constant Dollars)

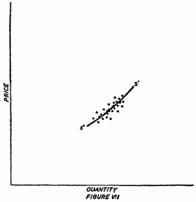
History of Identification

- Working (1925, 1927): *"By intelligently applying proper refinements, and making corrections to eliminate separately those factors which cause demand curves to shift and those factors which cause supply curves to shift, it may be possible even to obtain both a demand curve and a supply curve for the same product and from the same original data."*









History of Identification 2

- Frisch (1934, 1938) - confluency in linear regression
- Hurwicz (1950) - introduced the term "structure"
- Koopmans and Reiersol (1950): *"Scientific honesty demands that the specification of a model be based on prior knowledge of the phenomenon studied and possibly on criteria of simplicity, but not on the desire for identifiability of characteristics that the researcher happens to be interested in"*
- Phillips (1989): *"it seems important that we should understand the implications of identification failure for statistical inference. Yet, this is a subject that seems to be virtually untouched in the literature"*

Reviews: Dufour and Hsiao (2008), Tamer (2010)

Nonparametric Identification

Nonparametric econometrics

- **Identification**
- Estimation
- Testing

Reviews: Matzkin (1994, 2007, 2012), Horowitz (1998, 2009), Pagan and Ullah (1999), Li and Racine (2007), Racine (2008).

Application: Demand Function under Slutsky Condition

Blundell, Horowitz and Parey (2013)

- Heterogenous demand function for gasoline in the U.S.
- Additive separability only under very restrictive assumptions about preferences
- Nonparametric estimate is noisy (DWL < 0)

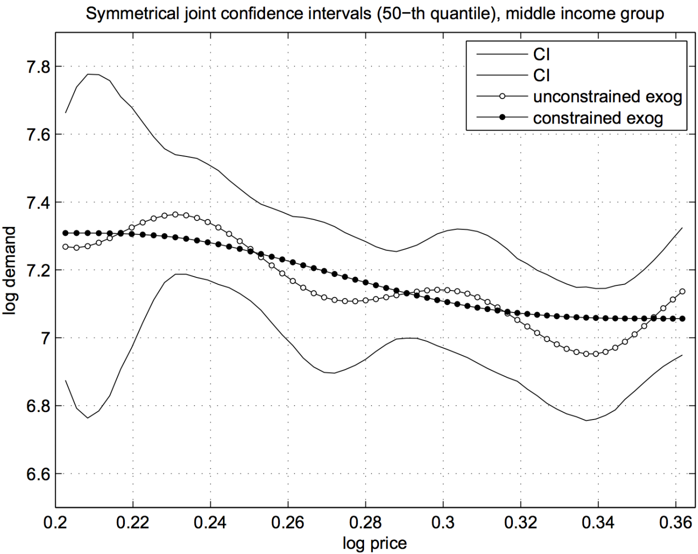
Identification:

- Q - Demand, P - Price, Y - Income, U - Unob. Heterogeneity
- $Q = g(P, Y, U)$ increasing in U
- U is independent of (P, Y)
- Slutsky restriction: $\frac{\partial G_\alpha(P,Y)}{\partial P} + G_\alpha(P, Y)\frac{\partial G_\alpha(P,Y)}{\partial Y} \leq 0$

Result:

- Middle income group shows
    - strongest price responsiveness
    - highest DWL

Application: Demand Function under Slutsky Condition



Partial Identification

Assumptions do not (point) identify the quantity of interest but only restrict it to lie in a set.

$\Theta_0(P)$ is a set.

Manski (2003): *"...it has been commonplace to think of identification as a binary event – a parameter is either identified or not – and to view point identification as a pre-condition for inference. Yet there is enormous scope for fruitful inference using data and assumptions that partially identify population parameters."*

Reviews: Manski (2003), Tamer(2010)

Example : Bounds on Average Treatment Effect

Say we are interested in $E[y(t)] - E[y(s)]$

$$E[y(t)] = E[y|z = t] \cdot P(z = t) + E[y(t)|z \neq t] \cdot P(z \neq t)$$

Observed quantities
Unobserved quantities

Assumption of Bounded support

Suppose that $y_{min} \leq y_i(t) \leq y_{max}$

$$
\begin{aligned}
LB_{E[y(t)]} &= E[y|z=t] \cdot P(z=t) + y_{min} \cdot P(z \neq t) \\
&\leq \\
E[y(t)] &= E[y|z=t] \cdot P(z=t) + E[y(t)|z \neq t] \cdot P(z \neq t) \\
&\leq \\
UB_{E[y(t)]} &= E[y|z=t] \cdot P(z=t) + y_{max} \cdot P(z \neq t)
\end{aligned}
$$

Under this assumption, $E[y(t)]$ is **partially** identified and the interval $(LB_{E[y(t)]}, UB_{E[y(t)]})$ is called an **identified set**.

Different assumptions

- Monotone Treatment Response (MTR) assumption
  $\forall i, t_2 \geq t_1 : y_i(t_2) \geq y_i(t_1)$
- Monotone Treatment Selection (MTS) assumption
  $\forall z_2 \geq z_1 : E[y(t)|z = z_2] \geq E[y(t)|z = z_1]$
- Monotone Instrumental Variable (MIV) assumption
  $\forall v_2 \geq v_1 : E[y(t)|v = v_2] \geq E[y(t)|v = v_1]$

Analytical bounds on $E[y(t)]$ under MTR, MTS, MIV, MTR+MTS are available. These then translate to bounds on $E[y(t)] - E[y(s)]$.

Applications : Bounds on Average Treatment Effect

- Gonzales (2005) - Returns to language skills
- Kreider and Pepper (2007) - Effects of food stamps on children's health outcomes
- Gundersen and Kreider (2009) - Effects of food insecurity on children's health outcomes
- Bhattacharya, Shaikh and Vytlacil (2011) - Effect of Swan-Ganz catheterization on mortality
- deHaan (2011) - Effects of parent's schooling on child's schooling
- Siddique (2013) - Effects of mandatory arrest policy on domestic violence

Example: English Auctions

Oral ascending bid

Button auction model - Milgrom and Weber (1982)

- the price rises continuously and exogenously
- depress a button to indicate their willingness to buy at the current price
- bidders exit by releasing their buttons
- auction ends when only one bidder remains

Identification

- Economic theory:
  - Behavioural economics: Bayesian Nash equlibirum
  - Restrictions on preferences: Risk neutrality
- Functional form assumptions: convenient for identification and estimation

Reviews: Athey and Haile (2002, 2006)

Application: English Auctions

Haile and Tamer (2003) - studied symmetric independent private values model of English auction and relies on two assumptions

- Bidders do not bid more than they are willing to pay.
- Bidders do not allow an opponent to win at a price they are willing to beat.

It allows for

- a jump bidding
- a bidder's highest bid lies below his valuation
- some bidders do not bid at all
- ranking of the bidders to be different from their valuations

There is **no**

- unique distribution of bids given a distribution of valuations,
- unique distribution of valuations given a distribution of bids.

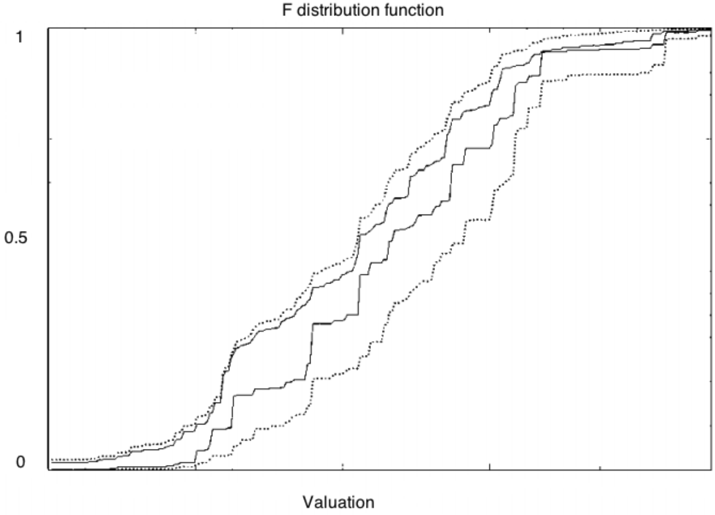
Application: English Auctions - results

Auctions on Timber-harvesting contracts held by the U.S. Forest Service

- contract is the right to harvested timber in the specified area
- consider short term contracts
- reserve price set to ensure "fair market value" but also other objectives (meet U.S. wood demands, improving forest health, preventing fires)
- Report of the U.S. Forest Service: "studies indicate it is nearly impossible to use sale records to determine if marginal sales made in the past would have been purchased under a different [reserve price] structure."
- $\rightarrow$ Haile and Tamer (2003) addressed precisely this issue
- the reserve price can be **doubled** and still at least 85% of timber would actually be sold

Application: English Auctions - results 2

Bounds on the distribution of valuation



Application: Changes in the Wage Distribution of Male and Female Wages

Blundell, Gosling, Ichimura and Meghir (2007)

Setup

- We are interested in the wage distribution.
- Non-random selection into employment
- No obvious point-identifying strategies

Economic Restrictions:

- the probability of work is higher for those with higher wages
- wage distribution is independent of out-of-work benefit income
- or alternatively, higher values of such income are positively associated with wages

Application: Changes in the Wage Distribution of Male and Female Wages

Results:

- 1978 - 2000 UK Family Expenditure Survey
- inequality in wage distribution has increased (overall and within educational groups)
- educational differentials increased for both 25 and 45yrs old
- gender wage differentials
    - decreased for 25yrs old unskilled group
    - may have decreased for 40yrs old unskilled group
    - have not decreased for college group

Conclusion

Economic theory **can** provide useful restrictions for nonparametric identification (point, partial) in econometric models.

These models are relevant for a wide range of economic applications.

Thank you for your attention!