Identification

Lukáš Lafférs

Matej Bel University, Dept. of Mathematics

What can be learnt from the data?

DATA + MODEL $\rightarrow$ CONCLUSIONS

Why should we study identification?

Conceptual framework of what could we potentially learn from the data and model.

What are the crucial components of this.

What if some of the model assumptions are incorrect.

Identification

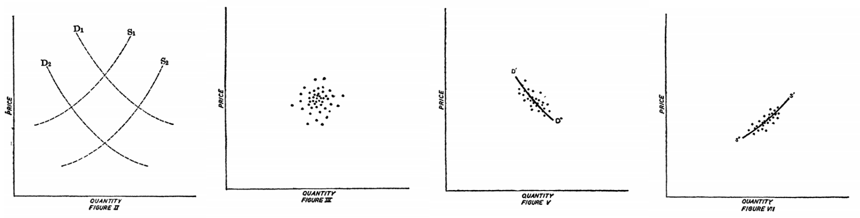
- Is a different and separate topic from the statistical inference.

Primarily based on:

Lewbel, Arthur. ”The identification zoo: Meanings of identification in econometrics.” Journal of Economic Literature 57.4 (2019): 835-903.

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)

Notation

- m - a model
- ϕ - what can be known from data
- θ - a parameter
- s - a structure

Model m

- set of <u>functions or constants</u> (regression function, utility functions, coefficient vectors)
- that satisfy given <u>restrictions</u> (linear/monotone regression function, normal errrors, parameters bounded)
- a particular model value m
- a set M of model values
- any m implies a particular DGP (data generating process)

Data ϕ

- Set of constants and/or functions about the DGP that are assumed to be known or knowable from data
- Examples: data distribution functions, conditional mean functions, linear regression coefficients, or time series autocovariances

Parameter θ

- Set of constants and/or functions that summarize relevant features of a model.
- The thing we wish to estimate.
- May include nuisance parameters that are not of direct interest, but may be necessary for identification/estimation of other objects

Structure s

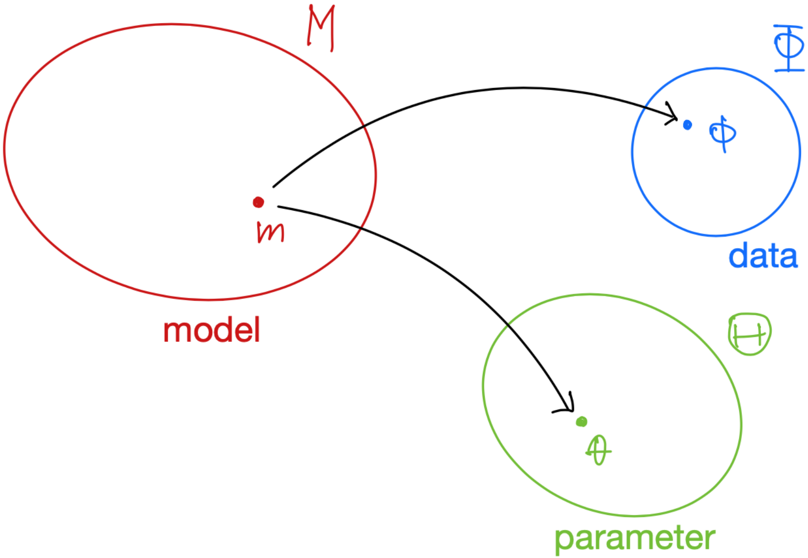
- m implies a particular value of ϕ and of θ
- BUT, there may be multiple ms that imply the same ϕ and θ
- Let Structure $s(\phi, \theta)$ be the collection of all m that imply ϕ and θ

Two parameter values θ and $\tilde{\theta}$ are said to be **observationally equivalent** if there exists ϕ such that $s(\phi, \theta)$ and $s(\phi, \tilde{\theta})$ are both not empty.

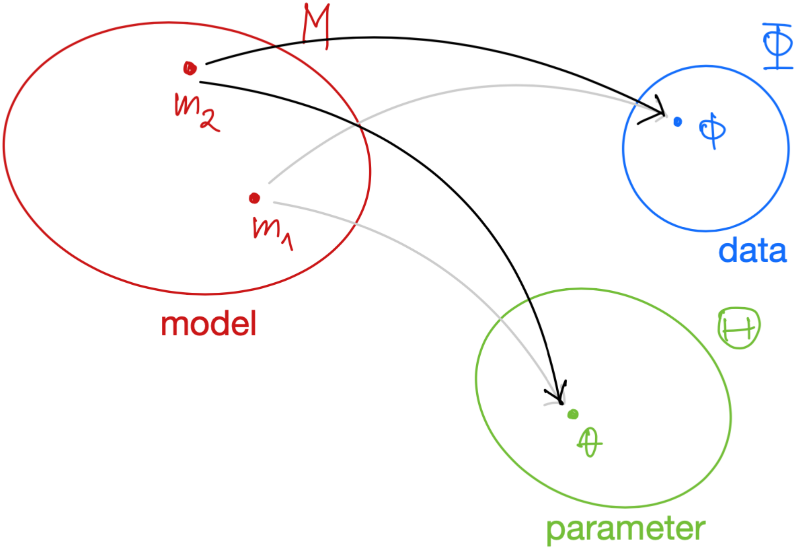
(in other words: both θ and $\tilde{\theta}$ could be true, based on observed ϕ)

Notation

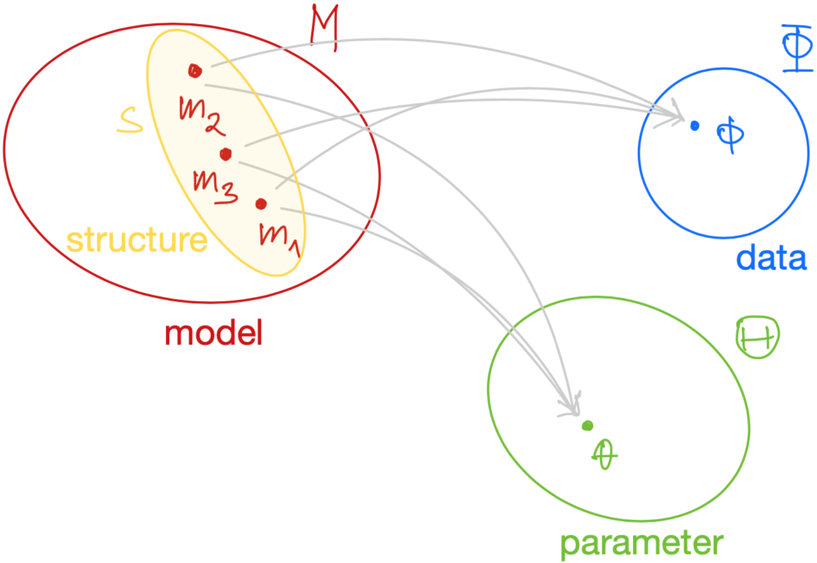
- m - a model
- ϕ - what can be known from data
- θ - parameter
- s - structure



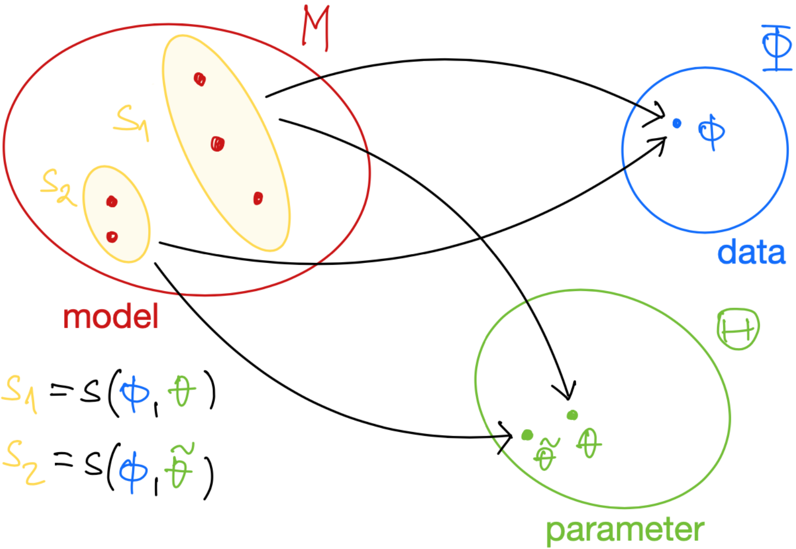
M
m
model
Φ
ϕ
data
Θ
θ
parameter



M
m_2
m_1
model
Φ
ϕ
data
Θ
θ
parameter



M
S
m_2
m_3
m_1
structure
model
Φ
ϕ
data
Θ
θ
parameter



$$S_1 = s(\phi, \theta)$$

$$S_2 = s(\phi, \tilde{\theta})$$

Types of identification

- Point identification of θ
- Global identification
- Point identification of m
- Local identification
- Partial identification
- Parametric/Semi-/Non- identification

Notation

- m - a model
- ϕ - what can be known from data
- θ - parameter
- s - structure

Point identification (of a parameter θ)

There do not exist any pairs θ and $\tilde{\theta}$ that are different and observationally equivalent.

Notation

- m - a model
- ϕ - what can be known from data
- θ - parameter
- s - structure

Global identification (of a parameter θ)

- Let $\theta \in \Theta$ and let θ_0 be the true value
- θ_0 is point identified if there isn't any $\theta \in \Theta_0$ that is observationally equivalent to θ_0
- But we don't know what θ_0 is.
- So, if we require that no two elements of Θ are obs. equivalent.
- Then θ_0 is identified no matter what it happens to be. (hence the word <u>global</u>)

Notation

- m - a model
- ϕ - what can be known from data
- θ - parameter
- s - structure

Point identification (of a model m)

There do not exist any pairs m and $\tilde{m}$ that are different and observationally equivalent (now treating the whole models m and $\tilde{m}$ as parameters).

Stronger than a point identification of θ.

Notation

- m - a model
- ϕ - what can be known from data
- θ - parameter
- s - structure

Local identification (of a parameter θ)

There exists a neigborhood of θ_0 so that for all values of $\theta \neq \theta_0$ in this neighborhood, θ is not observationally equivalent to θ_0

Notation

- m - a model
- ϕ - what can be known from data
- θ - parameter
- s - structure

Partial (set) identification (of a parameter θ)

There exist some parameter values θ that are not observationally equivalent to θ_0 (so that not all $\theta \in \Theta$ are obs. equivalent).

The collection of all θ that are obs. equivalent to θ_0 is called an identified set.

Notation

- m - a model
- ϕ - what can be known from data
- θ - parameter
- s - structure

Partial Identification

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)

Semi-/Non- parametric identification (of a parameter θ)

- Parametric - ϕ and θ are finite
- Non-parametric - θ includes functions or infinite sets
- Semi-parametric - θ includes both vector of constants and functions
- (Not always easy to distinguish between them)

Notation

- m - a model
- ϕ - what can be known from data
- θ - parameter
- s - structure

Nonparametric Models

Pros

- more credible assumptions
- more flexible
- economic restrictions

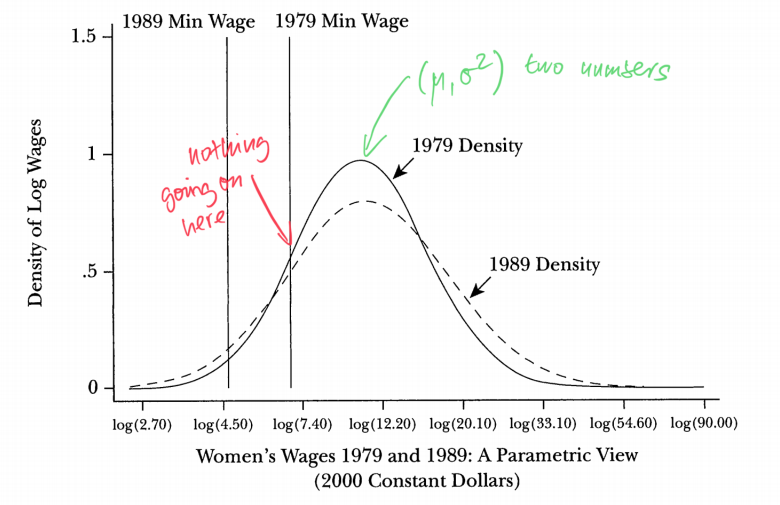
Cons

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

Why Non-parametric?

(DiNardo and Tobias 2001)

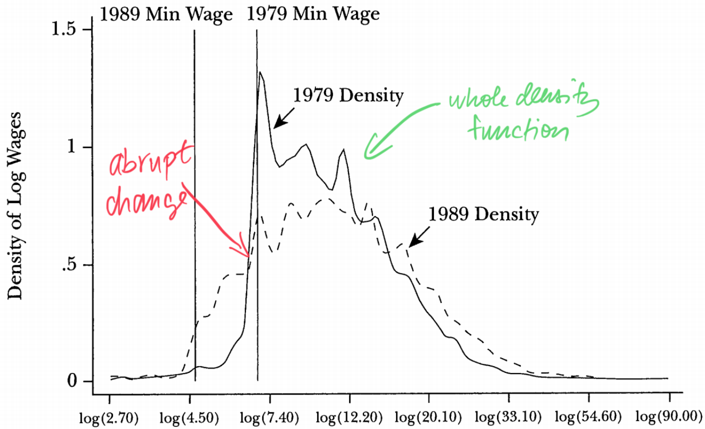
Parametric model:



Why Non-parametric?

(DiNardo and Tobias 2001)

Non-parametric model:



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

Example 1 - Median

- M - set of all possible distributions of rv W with strictly increasing distribution function.
- ϕ is the distribution function of W, $F(w)$
- θ is the median of W

- Structure $s(\phi, \theta)$ contains a single element if $F(\theta) = 1/2$ where $\phi = F$ or is empty if $F(\theta) \neq 1/2$.
- $F(\theta) = 1/2$ and $F(\tilde{\theta}) = 1/2$ implies $\theta = \tilde{\theta}$ and hence θ is point identified

Notation

- m - a model
- ϕ - what can be known from data
- θ - parameter
- s - structure

Example 2a - Linear regression

- M - Set of joint distributions of (ε, X) that satisfy
  - $y = X\theta + \varepsilon$
  - $E(X^T\varepsilon) = 0$
  - $E(X^TX)$ is non singular
  - both ε and X have finite first and second moments
- ϕ is the set of first and second moments of X and y
- θ is the vector of parameters

- $s(\phi, \theta)$ is non-empty when $E[X^T(y - X\theta)] = 0$ is satisfied.
- θ is uniquely determined by $\theta = E(X^TX)^{-1}E(X^Ty)$ and hence it is point identified

parametric or semi-parametric?

Example 2b - Linear regression

- M - Set of joint distributions of (ε, X) that satisfy
  - $y = X\theta + \varepsilon$
  - $E(X^T\varepsilon) = 0$
  - $E(X^TX)$ is non singular
  - both ε and X have finite first and second moments
- ϕ is the **joint distribution of** (X, y)
- θ is the vector of parameters **and the distribution function of** ε

- $s(\phi, \theta)$ is non-empty when $E[X^T(y - X\theta)] = 0$ is satisfied.
- θ is uniquely determined by $\theta = E(X^TX)^{-1}E(X^Ty)$ and hence it is point identified.

parametric or semi-parametric?

Example 3a- Treatment effects

- M - all possible joint distributions of $(Y(1), Y(0), T)$
  - $(Y(1), Y(0)) \perp T$
  - if $T = t$ then $Y = Y(t)$
- ϕ is the joint distribution of (Y, T) (alternatively $E[Y|T=1]$ and $E[Y|T=0]$)
- θ is the average treatment effect $\theta = E[Y(1) - Y(0)]$

- $s(\phi, \theta)$ is non-empty whenever $\theta = E[Y|T=1] - E[Y|T=0]$ is satisfied.
- Under $(Y(1), Y(0)) \perp T$ we have that $\theta = E[Y|T=1] - E[Y|T=0]$ and hence it is point identified.

Does there exist an unique value of θ for every possible ϕ?

Example 3b- Treatment effects

- M - all possible joint distributions of $(Y(1), Y(0), T)$
  - $(Y(1), Y(0)) \perp T$
  - if $T = t$ then $Y = Y(t)$
- ϕ <u>**consists of** $E[Y|T = 1]$ **and** $E[Y|T = 0]$</u>
- θ is the average treatment effect $\theta = E[Y(1) - Y(0)]$

- $s(\phi, \theta)$ is non-empty whenever $\theta = E[Y|T = 1] - E[Y|T = 0]$ is satisfied.
- Under $(Y(1), Y(0)) \perp T$ we have that $\theta = E[Y|T = 1] - E[Y|T = 0]$ and hence it is point identified.

Example 3c- Treatment effects

- M - all possible joint distributions of $(Y(1), Y(0), T)$
  - $E[Y(t)|T] = E[Y(t)]$ **(mean unconfoundedness)**
  - if $T = t$ then $Y = Y(t)$
- ϕ **consists of** $E[Y|T=1]$ **and** $E[Y|T=0]$
- θ is the average treatment effect $\theta = E[Y(1) - Y(0)]$

- $s(\phi, \theta)$ is non-empty whenever $\theta = E[Y|T=1] - E[Y|T=0]$ is satisfied.
- Under $(Y(1), Y(0)) \perp T$ we have that $\theta = E[Y|T=1] - E[Y|T=0]$ and hence it is point identified.

Example 3d- Treatment effects

- $\textcolor{orange}{M}$ - all possible joint distributions of $(Y(1), Y(0), T)$
  - $y_{min} \leq Y(t) \leq y_{max}$ $\cancel{(Y(1), Y(0)) \perp T}$ **no randomization here!**
  - if $T = t$ then $Y = Y(t)$
- ϕ consists of $E[Y|T = 1]$, $E[Y|T = 0]$ **and** $Pr(T = 1)$
- θ is the average treatment effect $\theta = E[Y(1) - Y(0)]$

- $E[Y(1)] = E[Y(1)|T = 1]Pr(T = 1) + \underbrace{E[Y(1)|T = 0]}_{unobserved}Pr(T = 0)$
- $y_{min} \leq E[Y(1)|T = 0] \leq y_{max}$
- $\theta \in [\theta_L, \theta_H]$ and θ is partially identified

Bounds on Average Treatment Effect

$$E[Y(t)] = \underbrace{E[Y|T=t] \cdot P(T=t)}_{\textit{Observed}} + \underbrace{E[Y(t)|T \neq t]}_{\textit{Unobserved}} \cdot \underbrace{P(Z \neq t)}_{\textit{Observed}}$$

Observed quantities
Unobserved quantities

Assumption of Bounded support

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

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

$E[Y(t)]$ (and hence also ATE) is **partially** identified and the interval $\left(LB_{E[Y(t)]}, UB_{E[Y(t)]}\right)$ is called an **identified set**.

Example 4 - Supply and Demand

Demand: $Q = b \cdot P + c \cdot Z + U$,
Supply: $Q = a \cdot P + \varepsilon$

- M - all possible joint distributions of (I, U, ε) and coeffs (a, b, c)
  - $E(U) = E(\varepsilon) = 0$ and $(U, \varepsilon) \perp Z$
- ϕ is coeffs (ϕ_1, ϕ_2) from $Q = \phi_1 Z + V_1$ and $P = \phi_1 Z + V_2$, where $E(V_1) = E(V_2)$ and $(V_1, V_2) \perp Z$
- $\theta = a$ - coeff of price in supply eqn.

- For any m in $s(\phi, \theta)$, we need to have $\theta = a$, $\phi_1 = \frac{ac}{a-b}$, $\phi_2 = \frac{c}{a-b}$
- if $c \neq 0$ we get $a = \frac{\phi_1}{\phi_2}$ and $s(\phi, \theta)$ contains many elements
- if $c = 0$ then any θ and $\tilde{\theta}$ will be obs. equivalent with $\phi = (0, 0)$

In other words: we need the instrument Z to appear in the demand eqn.

Point identification?

So far ”by construction”:

- Ex 1: $\theta = F^{-1}(0.5)$
- Ex 2: $\theta = E(X^T X)^{-1} E(X^T y)$
- Ex 3: $\theta = E[Y|T=1] - E[Y|T=0]$
- Ex 4: $\theta = \phi_1/\phi_2$

Other strategies?

True θ_0 is an <u>unique</u> maximizer of a optimization problem defined by the model. (e.g. Likelihood function is globally concave)

Identification logically precedes estimation.

What is ”knowable” ϕ?

- Distribution based on IID data: Glivenko-Cantelli theorem
- Expected values: Law of Large Numbers

In many cases, it is assumed that the parameter is identified (GMM).

Example:
Preferences θ may be identified from demand functions ϕ.
But how do we identify these demand functions?

ϕ is the starting point. We assume this is knowable from the data.

Reasons for not (point) identification

- model is incomplete
- perfect collinearity
- non-linearity
- simultaneity
- endogeneity
- unobservability

Some remarks

- We keep asking this: ”Does there exist an unique value of θ for every possible ϕ ?”
- There are different ways how to achieve identification.
- Stronger assumptions are more difficult to defend but easier to work with.
- Weaker assumptions may not be sufficient to guarantee identification.
- Some assumptions are difficult to interpret (mean unconfoundedness is sensitive to transformation of Y)

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)

”Harmful econometrics” vs. ”Cuteonomics”

Structural

- model of economic behaviour is built up based on economic theory
- focus on <u>deep</u> parameters
- allows to answer rich set of questions

Reduced form

- as few assumptions as possible
- focus on reduced form parameters (e.g. ATE, ATT, MTE, LATE, QTE)
- attempts to do or mimic RCT
- prefers simplicity and transparency

Lewbel’s JEL Zoo paper (section 5.1) suggests to use both and gives many examples.

Example: $Y = a + bT + e$

Structural model

- variables U_1, U_0, V_1, V_0
- individual effect U_1
- $y = U_0 + U_1 T$ and $T = V_0 + V_1 Z$
- $E(V_1) \neq 0$
- $(U_1, U_0, V_1, V_0) \perp Z$
- $cov(e, Z) = 0$ (this implies $cov(U_1, V_1) = 0$)

$\Longrightarrow$

$$E[Y(1) - Y(0)] = E(U_1) = b$$

Reduced form model

- variables $Y(1), Y(0), T(1), T(0)$
- individual effect $Y(1) - Y(0)$
- $Y(t, z)$ satisfies $Y(t, 0) = Y(t, 1)$
- $E(T(1) - T(0)) \neq 0$
- $(Y(1), Y(0), T(1), T(0)) \perp Z$
- $P(T(1) = 0, T(0) = 1) = 0$

$\Longrightarrow$

$$E[Y(1) - Y(0) | T(1) = 1, T(0) = 0] = \frac{cov(Z, Y)}{cov(Z, T)}$$

Structural model

- identifies ATE
- $cov(U_1, V_1) = 0$ is a restriction on the heterogeneity of the treatment effect U_1
- stronger assumptions about **the outcome** Y

- can we justify $cov(e, Z) = 0$?

Reduced form model

- identifies LATE
- No defiers condition is a restriction on the heterogeneity of types, **not** about outcomes
- stronger assumptions about **the treatment** T

- who are the compliers?
- what do we know about the rest?
- how about non-binary treatments?

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)

Example 5 - Cost function and Revenue distribution

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 ($\varepsilon \geq 0$) as a random variable.

Example 5 - 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. $h(x^*) = \alpha$

- M - Set of joint distributions of (x, y, ε), cost function h and production function that they jointly satisfy all the assumptions.
- ϕ is the probability of not investing given prices x: $P(y = 0|x)$
- θ is the cost function h and the distribution of revenues F

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

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

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

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

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

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

Example 6: Demand Function under Slutsky Condition

Blundell, Horowitz and Parey (2012)

- 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: $\frac{\partial g(P,Y,\alpha)}{\partial P} + g(P, Y, \alpha)\frac{g(P,Y,\alpha)}{\partial Y} \leq 0$

Results:

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

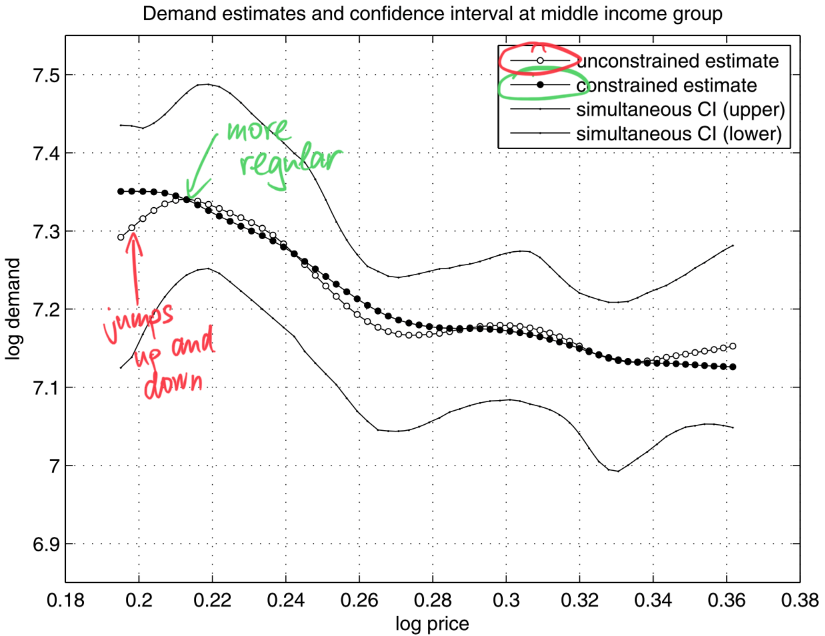
Slutsky restriction

$$\underbrace{\underbrace{\frac{\partial g(P,Y,\alpha)}{\partial P}}_{total\ effect} + \underbrace{g(P,Y,\alpha)\frac{g(P,Y,\alpha)}{\partial Y}}_{-income\ effect}}_{substitution\ effect} \leq 0$$

Slutsky matrix is negative semi-definite.

Simpler: In one dimension: own price elasticity is negative.

Even simpler: cost minimizing consumer will buy less of a certain good if it gets more expensive.


Demand estimates and confidence interval at middle income group
unconstrained estimate
constrained estimate
simultaneous CI (upper)
simultaneous CI (lower)
log demand
log price
7.5
7.4
7.3
7.2
7.1
7
6.9
0.18
0.2
0.22
0.24
0.26
0.28
0.3
0.32
0.34
0.36
0.38
more regular
jumps up and down

Thank you for your attention!

References

- Lewbel, Arthur. ”The identification zoo: Meanings of identification in econometrics.” Journal of Economic Literature 57.4 (2019): 835-903.
- Working, Holbrook. ”The statistical determination of demand curves.” The Quarterly Journal of Economics 39.4 (1925): 503-543.
- Working, Elmer J. ”What do statistical “demand curves” show?.” The Quarterly Journal of Economics 41.2 (1927): 212-235.
- Frisch, Ragnar. ”Circulation planning: proposal for a national organization of a commodity and service exchange.” Econometrica, Journal of the Econometric Society (1934): 258-336.
- Frisch, Ragnar. ”The double-expenditure method.” Econometrica: Journal of the Econometric Society (1938): 85-90.
- Koopmans, Tjalling C., and Olav Reiersol. ”The identification of structural characteristics.” The Annals of Mathematical Statistics 21.2 (1950): 165-181.
- Phillips, Peter CB. ”Partially identified econometric models.” Econometric Theory 5.2 (1989): 181-240.
- Hsiao, Cheng and Jean-Marie Dufour. ”The New Palgrave Dictionary of Economics Online.” (2008).
- Tamer, Elie. ”Partial identification in econometrics.” Annu. Rev. Econ. 2.1 (2010): 167-195.
- Manski, Charles F. Partial identification of probability distributions. Springer Science & Business Media, 2003.
- DiNardo, John, and Justin L. Tobias. ”Nonparametric density and regression estimation.” Journal of Economic Perspectives 15.4 (2001): 11-28.
- Matzkin, Rosa L. ”Restrictions of economic theory in nonparametric methods.” Handbook of econometrics 4 (1994): 2523-2558.
- Gandhi, Amit, Salvador Navarro, and David A. Rivers. ”On the identification of gross output production functions.” Journal of Political Economy 128.8 (2020): 2973-3016.
- Blundell, Richard, Joel L. Horowitz, and Matthias Parey. ”Measuring the price responsiveness of gasoline demand: Economic shape restrictions and nonparametric demand estimation.” Quantitative Economics 3.1 (2012): 29-51.