Identification

Christopher Taber

University of Wisconsin

February 11, 2018

Definition of Identification

Another term that means different things to different people

I will base my discussion on Matzkin’s (2007) formal definition of identification but use my own notation and be a bit less formal

This will all be about the **Population** in thinking about identification we will completely ignore sampling issues

We first need to define a data generation process

Data Generating Process

Let me define the data generating process in the following way

$$\begin{aligned} X_i &\sim H_0(X_i) \\ u_i &\sim F_0(u_i;\theta) \\ \Upsilon_i &= y_0(X_i, u_i;\theta) \end{aligned}$$

The data is (Υ_i, X_i) with u_i unobserved.

We know this model up to θ

To think of this as non-parametric we can think of θ as infinite dimensional

For example if F_0 is nonparametric we could write the model as $\theta = (\theta_1, F_0(\cdot))$

Point Identification of the Model

The model is identified if there is a unique θ that could have generated the population distribution of the observable data (X_i, Υ_i)

A bit more formally, let Θ be the parameter space of θ and let θ_0 be the true value

- If there is some other $\theta_1 \in \Theta$ with $\theta_1 \neq \theta_0$ for which the joint distribution of (X_i, Υ_i) when generated by θ_1 is identical to the joint distribution of (X_i, Υ_i) when generated by θ_0 then θ is not identified
- If there is no such $\theta_1 \in \Theta$ then θ is (point) identified

Set Identification of the Model

Define Θ_I as the identified set.

I still want to think of there as being one true θ_0

Θ_I is the set of $\theta_1 \in \Theta$ for which the joint distribution of (X_i, Υ_i) when generated by θ_1 is identical to the joint distribution of (X_i, Υ_i) when generated by θ_0.

So another way to think about point identification is the case in which

$$\Theta_I = \{\theta_0\}$$

Identification of a feature of a model

Suppose we are interested not in the full model but only a feature of the model: $\psi(\theta)$

The feature is identified if there is a unique value of it consistent with the observed data

More formally

$$\Psi_I \equiv \{\psi(\theta) : \theta \in \Theta_I\}$$

Most interesting cases occur when Θ_I is a large set but Ψ_I is a singleton

In practice $\psi(\theta)$ could be something complicated like a policy counterfactual in which we typically need to first get θ and then simulate $\psi(\theta)$

However, often it is much simpler and we can just write it as a known function of the data.

I can think of the Imbens/Angrist framework as a data generation process.

The data is (Z_i, T_i, Y_i)

Here Z_i is my exogenous X_i so $H_0(X_i)$ in that case is

$$
\begin{aligned}
Pr(Z_i = 1) =& \rho \\
Pr(Z_i = 0) =& 1 - \rho
\end{aligned}
$$

They are not explicit about u_i but I can add it as

$$u_i = \begin{bmatrix} S_i \\ Y_{0i} \\ Y_{1i} \end{bmatrix}$$

and we can write this joint distribution as

$$Pr(S_i = s)$$

and $F_0(Y_{0i} \mid S_i = s)$ and $F_1(Y_{1i} \mid S_i = s)$ as the distributions of Y_{0i} and Y_{1i} respectively conditional on S_i.

Then finally $\Upsilon_i = (T_i, Y_i)$.

Once we have done this the y_0 is trivial

$$\begin{aligned}\Upsilon_i =& y_0(Z_i, u_i) \\ =& \begin{cases} (1, Y_{1i}) & S_i = \clubsuit \\ (0, Y_{0i}) & S_i = \heartsuit \\ (0, Y_{0i}) & S_i = \diamondsuit, Z_i = 0 \\ (1, Y_{1i}) & S_i = \diamondsuit, Z_i = 1 \end{cases}\end{aligned}$$

What can we hope to identify?

What we get from data is essentially:

- Dist of $Z_i : \rho$
- Dist of T_i conditional on Z_i. This is $\mu_\clubsuit$ when $Z_i = 0$ and $\mu_\clubsuit + \mu_\diamondsuit$ when $Z_i = 1$, so the $\mu's$ are identified
- Then
  1. The distribution of Y_i conditional of $Z_i = 0, T_i = 0$, which is the distribution of Y_{0i} conditional on $S_i \in \{\heartsuit, \diamondsuit\}$
  2. The distribution of Y_i conditional of $Z_i = 0, T_i = 1$, which is the distribution of Y_{1i} conditional on $S_i \in \{\clubsuit\}$
  3. The distribution of Y_i conditional of $Z_i = 1, T_i = 0$, which is the distribution of Y_{0i} conditional on $S_i \in \{\heartsuit\}$
  4. The distribution of Y_i conditional of $Z_i = 1, T_i = 1$, which is the distribution of Y_{1i} conditional on $S_i \in \{\clubsuit, \diamondsuit\}$

Notice that clearly the full model and the average treatment effect is not identified. The data is silent about

- the distrubution of Y_{0i} for $S_i = \clubsuit$
- the distrubution of Y_{1i} for $S_i = \heartsuit$

From that sense if we want to get a conditional treatment effect our only hope is the $\diamondsuit s$.

To see this is identified note that

- From 2 $E\left(Y_{1i} \mid S_i = \clubsuit\right)$ is identified
- From 4 we can identify

  $$E\left(Y_i \mid Z_i = 1, T_i = 1\right) = \frac{\mu_{\clubsuit} E\left(Y_{1i} \mid S_i = \clubsuit\right) + \mu_{\diamondsuit} E\left(Y_{1i} \mid S_i = \diamondsuit\right)}{\mu_{\clubsuit} + \mu_{\diamondsuit}}$$

  So

  $$E\left(Y_{1i} \mid S_i = \diamondsuit\right) = \frac{\left(\mu_{\clubsuit} + \mu_{\diamondsuit}\right) E\left(Y_i \mid Z_i = 1, T_i = 1\right) - \mu_{\clubsuit} E\left(Y_{1i} \mid S_i =\right.}{\mu_{\diamondsuit}}$$

  is identified
- This is symmetric so

  $$E\left(Y_{0i} \mid S_i = \diamondsuit\right) = \frac{\left(\mu_{\heartsuit} + \mu_{\diamondsuit}\right) E\left(Y_i \mid Z_i = 0, T_i = 0\right) - \mu_{\heartsuit} E\left(Y_{0i} \mid S_i =\right.}{\mu_{\diamondsuit}}$$

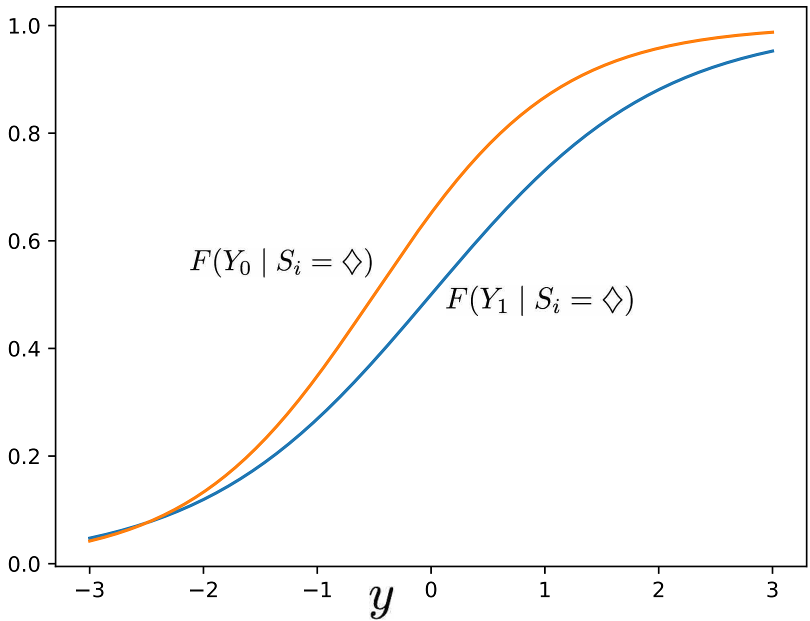
- And thus

  $$E\left(\alpha_i \mid S_i = \diamondsuit\right) = E\left(Y_{1i} \mid S_i = \diamondsuit\right) - E\left(Y_{0i} \mid S_i = \diamondsuit\right)$$

  is identified

Observations

- pretty nice that this is what IV actually converges to
- Not this simple if covariates or either Z_i or T_i is not binary
- however if Z_i take on lots of values you can get more, if it varies enough the average treatment effect is identified
- This doesn't work with other features of the distribution like the median. (difference of the medians is not the median of the difference)
- However mean is pretty general, we can define Y_i however we want so we can identify the mean of any function of Y_i.
  - One such function is $1(Y_i \leq y)$.
  - Do that at all y and we can identify the conditonal cdf of Y_{1i} for the $\diamondsuit$ and the conditional cdf of Y_{0i} for the $\diamondsuit$



1.0
0.8
0.6
0.4
0.2
0.0
−3
−2
−1
0
1
2
3
y
$F(Y_0 \mid S_i = \Diamond)$
$F(Y_1 \mid S_i = \Diamond)$