Estimation of Policy Counterfactuals

Christopher Taber

University of Wisconsin

June 30, 2016

Estimation of Policy Impacts

Empirical Work can be divided into four broad categories of questions:

1. What does the world look like (descriptive)?
2. What will the world look like tommorow (forecasting)?
3. What was the impact of some policy or program (causal inference)?
4. What would happen if we implemented some new policy or program (policy counterfactual)?

I intend for for the phrases “policy” and “program” to be very broad concepts.

Jeff (I think) will focus on the third issue I want to focus on the fourth building on Chao Fu’s presentation.

Quote from Frank Knight

> *The existence of a problem in knowledge depends on the future being different from the past, while the possibility of a solution of a problem of knowledge depends on the future being like the past*

Outline

Outline

Example 1: Supply and Demand

Lets consider to the classic simultaneous equations model in a policy regime with no taxes

Supply Curve

$$Q_t = \alpha_s P_t + X_t' \beta_x + Z_{st}' \beta_s + u_t$$

Demand Curve

$$Q_t = \alpha_d P_t + X_t' \gamma_x + Z_{dt}' \gamma_d + v_t$$

We can solve for prices and quantities as

$$P_t = \frac{X_t' \left(\gamma_x - \beta_x\right) + Z_{dt}' \gamma_d - Z_{st}' \beta_s + v_t - u_t}{\alpha_s - \alpha_d}$$

$$Q_t = \frac{\alpha_s (X_t' \gamma_x + Z_{dt}' \gamma_d + v_t) - \alpha_d (X_t' \beta_x + Z_{st}' \beta_s + u_t)}{\alpha_s - \alpha_d}$$

Now suppose we want to introduce a tax on this good imposed on consumers, so now

$$Q_t = \alpha_d \left(1+\tau\right) P_t + Z_t'\gamma + v_t$$

The equilibrium effect is

$$P_t = \frac{X_t'\left(\gamma_x - \beta_x\right) + Z_{dt}'\gamma_d - Z_{st}'\beta_s + v_t - u_t}{\alpha_s - \alpha_d\left(1+\tau\right)}$$

$$Q_t = \frac{\alpha_s(X_t'\gamma_x + Z_{dt}'\gamma_d + v_t) - \alpha_d(1+\tau)(X_t'\beta_x + Z_{st}'\beta_s + u_t)}{\alpha_s - \alpha_d(1+\tau)}$$

Note that you are taking the model seriously here-all of the parameters are policy invariant

Example 2: The Roy Model

Labor Market is a Village

There are two occupations

- hunter
- fisherman

Fish and Rabbits are completely homogeneous

No uncertainty in number you catch

Let

- π_F be the price of fish
- π_R be the price of rabbits
- F number of fish caught
- R number of rabbits caught

Wages are thus

$$\begin{aligned} W_F &= \pi_F F \\ W_R &= \pi_R R \end{aligned}$$

Each individual chooses the occupation with the highest wage

Lets assume this village trades with the rest of the economy so prices fish and rabbits is taken as given.

Thats it, that is the model

Once we know the model we could think of several different policies

One is suppose we impose a minimum wage $\bar{w}$ in the fishing sector but not in the hunting sector?

What will this due to earnings inequality?

Anyone who with $W_F < \bar{w}$ will no longer be employed in the fishing sector and must now hunt where they earn lower wages.

Inequality will like rise and we can determine the magnitude from the model

Other Examples

- Effects of Affordable Care Act on labor market outcomes (Aizawa and Fang, 2015)
- Tuition Subsidies on Health (Heckman, Humphries, and Veramundi, 2015)
- Effects of extending lenth of payment for college loan programs on college enrollment (Li, 2015)
- Peer effects of school vouchers on public school students (Altonji, Huang, and Taber, 2015)
- Tax credits versus income support (Blundell, Costa Dias, Meghir, and Shaw, 2015)
- Effects of border tightening on the U.S. government budget constraints (Nakajima, 2015)
- Welfare effects of alternative designs of school choice programs (Calsamiglia, Fu, and Guell, 2014)

Outline

What does structural mean?

No obvious answer, it means different things to different people

3 Definitions:

- Parameters are policy invariant
- Estimation of preference and technology parameters in a maximizing model (perhaps combined with some specification of markets)
- The structural parameters in a simultaneous equations model

For that matter what does reduced form mean

Now for many people it essentially means anything that is not structural

What I think of as the Classic definition is that reduced form parameters are a known function of underlying structural parameters.

- fits classic Simultaneous Equation definition
- might not be invertible (say without an instrument)
- for something to be reduced form according to this definition you need to write down a structural model
- this actually has content-you can sometimes use reduced form models to simulate a policy that has never been implemented (as often reduced form parameters are structural in the sense that they are policy invariant)

Advantages and disadvantages of “structural" and “Design-Based”

Two caveats first

- To me the fact that there are advantages and disadvantages makes them complements rather than substitutes
- These are arguments that different people make, but obviously they don’t apply to all (or maybe even most) structural work or non-structural work-there are plenty of good and bad papers of any type

Advantages and disadvantages of “structural" and “design-based”

Structural	Design-Based
Better on External Validity	Better on Internal Validity
Map from parameters to implications clearer	Map from data to parameters more transparent
Formalizes conditions for external validity	Requires fewer assumptions
Forces one to think about where data comes from	Might come from somewhere else
Easier to interpret what parameters mean	Estimates more credible

(Possible) Steps for writing this type of paper

1. Identify the policy question to be answered
2. Write down a model that can simulate policy
3. Think about identification/data (with the goal being the policy counterfactual)
4. Estimate the model
5. Simulate the policy counterfactual

Other reasons to write structural models

While this is the classic use of a structural model it is not the only one.

Other motivations:

- Further evaluation of an established policy: we might want to know welfare effect
- Basic Research-we want to understand the world better
    - Use data to help understand model
    - Use model to help understand data (use structural model as a lens)
- Methodological-this is a step in these directions, but we haven't gotten there yet

Outline

Data Generating Process

There are exceptions to this, but typically when we identify a structural model we first estimate the full “data generating process" and then use it to simulate our policy counterfactual

Let me define the data generating process in the following way

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

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 is infinite dimensional
  - For example if F is nonparametric we could write the model as $\theta = (\theta_1, F(\cdot))$
- To simulate a policy counterfactual your policy needs to be a known manipulation of this structural model (i.e. $\pi(\theta)$)
  - going forward I will just assume this is true as obviously it depends on the model and specific policy of interest

To relate this to our examples, for example 1 (being very loose with notation)

$$
\begin{aligned}
\Upsilon_t &= (P_t, Q_t) \\
X_t &= (X_t, Z_{dt}, Z_{st}) \\
u_t &= (u_t, v_t) \\
\theta &= (\gamma, \beta, \alpha_d, \alpha_s, G(u_t, v_t)) \\
y_0(X_i, u_i; \theta) &= \left[\begin{array}{c} \frac{X_t'(\gamma_x - \beta_x) + Z_{dt}'\gamma_d - Z_{st}'\beta_s + v_t - u_t}{\alpha_s - \alpha_d} \\ \frac{\alpha_s(X_t'\gamma_x + Z_{dt}'\gamma_d + v_t) - \alpha_d(X_t'\beta_x + Z_{st}'\beta_s + u_t)}{\alpha_s - \alpha_d} \end{array} \right]
\end{aligned}
$$

For the Roy Model we need to add some more structure to go from an economic model into an econometric model .

This means writing down the full data generation model.

First a normalization is in order.

We can redefine the units of F and R arbitrarily Lets normalize

$$\pi_F = \pi_R = 1$$

We consider the model

$$\begin{aligned} W_{fi} &= g_f(X_{fi}, X_{0i}) + \varepsilon_{fi} \\ W_{hi} &= g_h(X_{hi}, X_{0i} + \varepsilon_{hi} \end{aligned}$$

where the joint distribution of $(\varepsilon_{fi}, \varepsilon_{hi})$ is G.

Let F_i be a dummy variable indicating whether the worker is a farmer.

We can observe F_i and

$$W_i \equiv F_i W_{fi} + (1 - F_i) W_{hi}$$

Thus in this case

$$\begin{aligned}
\Upsilon_i &= (F_i, W_i) \\
X_i &= (X_{0i}, X_{fi}, X_{hi}) \\
u_i &= (\varepsilon_{fi}, \varepsilon_{hi}) \\
\theta &= (g_h, g_h, G) \\
y_0(X_i, u_i; \theta) &= \begin{bmatrix} 1\left(g_f(X_{fi}, X_{0i}) + \varepsilon_{fi} > g_h(X_{hi}, X_{0i}) + \varepsilon_{hi}\right) \\ \max\left\{g_f(X_{fi}, X_{0i}) + \varepsilon_{fi}, g_h(X_{hi}, X_{0i}) + \varepsilon_{hi}\right\} \end{bmatrix}
\end{aligned}$$

You can see the selection problem-we only observe the wage in the occupation the worker chose, we don't observe the wage in the occupation they didn't

Definition of Identification

Another term that means different things to different people

I want to think about it in an econometric way

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

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 identified

Identification of Simultaneous Equation Model

$$Q_t = \alpha_s P_t + X_t'\beta + Z_{st}'\beta_s + u_t$$
$$Q_t = \alpha_d P_t + X_t'\gamma + Z_{dt}'\gamma_d + v_t$$

Here the hard part is going to be identifying α_s and α_d

(given $(\alpha_s, \alpha_d, \beta, \beta_s, \gamma, \gamma_d)$ getting the joint distribution of the error terms is trivial)

Since this is symmetric, lets focus on identification of α_s

We will also use the assumptions

$$E(u_t \mid X_t, Z_{dt}, Z_{st}) = 0$$
$$E(\nu_t \mid X_t, Z_{dt}, Z_{st}) = 0$$

Can we just run a regression of Q_t on P_t and Z_t to estimate α_d and γ?

Think about the "reduced form equation"

$$P_t = \frac{X_t'\left(\gamma_x - \beta_x\right) + Z_{dt}'\gamma_d - Z_{st}'\beta_s + v_t - u_t}{\alpha_s - \alpha_d\left(1+\tau\right)}$$

since v_t is a direct determinant of P_t, P_t is correlated with v_t so OLS is not consistent

So is α_s identified?

The key will be Z_{dt}

I want to think about this in three different ways.

The first two will involve the use of the “reduced form”

Lets define γ_p^*, β_p^*, and v_{pt}^* implicitly as

$$\begin{aligned} P_t =& \frac{X_t'\left(\gamma_x-\beta_x\right)+Z_{dt}'\gamma_d-Z_{st}'\beta_s+v_t-u_t}{\alpha_s-\alpha_d} \\ \equiv& X_t'\delta_{px}^*+Z_{dt}'\delta_{pd}^*+Z_{st}'\delta_{ps}^*+v_{pt}^* \end{aligned}$$

where

$$\begin{aligned} \delta_{px}^* &\equiv \frac{\gamma_x-\beta_x}{\alpha_s-\alpha_d} \\ \delta_{pd}^* &\equiv \frac{\gamma_d}{\alpha_s-\alpha_d} \\ \delta_{ps}^* &\equiv \frac{-\beta_s}{\alpha_s-\alpha_d} \\ v_{pt}^* &\equiv \frac{v_t-u_t}{\alpha_s-\alpha_d} \end{aligned}$$

Note that $E(\nu_t^* \mid X_t, Z_{dt}, Z_{st}) = 0$, so one can identify $\delta_p^* \equiv \left(\delta_{px}^*, \delta_{pd}^*, \delta_{ps}^*\right)$ and by regressing P_t on $W_t = (X_t, Z_{dt}, Z_{st})$

That is

$$
\begin{aligned}
& \left(E\left[W_t W_t'\right]\right)^{-1} E\left[W_t P_t\right] \\
= & \left(E\left[W_t W_t'\right]\right)^{-1} E\left[W_t\left(W_t' \delta_p^* + v_{pt}^*\right)\right] \\
= & \left(E\left[W_t W_t'\right]\right)^{-1} E\left[W_t W_t'\right] \delta_p^* + \left(E\left[W_t W_t'\right]\right)^{-1} E\left[W_t v_{pt}^*\right] \\
= & \delta_p^*
\end{aligned}
$$

This is called the “reduced form” equation for P_t

Note that the parameters here are not the fundamental structural parameters themselves, but they are a known function of these parameters

To me this is the classic definition of reduced form (you need to have a structural model)

How is this useful for identifying the model?

Method 1

We can also solve for the reduced for for Q_t

$$Q_t = \frac{\alpha_s(X_t'\gamma_x + Z_{dt}'\gamma_d + v_t) - \alpha_d(X_t'\beta_x + Z_{st}'\beta_s + u_t)}{\alpha_s - \alpha_d}$$
$$= X_t'\delta_{qx}^* + Z_{dt}'\delta_{qd}^* + Z_{st}'\delta_{qs}^* + v_{qt}^*$$

with

$$\delta_{qx}^* \equiv \frac{\alpha_s\gamma_x - \alpha_d\beta_x}{\alpha_s - \alpha_d}$$
$$\delta_{qd}^* \equiv \frac{\alpha_s\gamma_d}{\alpha_s - \alpha_d}$$
$$\delta_{qs}^* \equiv \frac{-\alpha_d\beta_s}{\alpha_s - \alpha_d}$$
$$v_{qt}^* \equiv \frac{\alpha_s v_t - \alpha_d u_t}{\alpha_s - \alpha_d}$$

Like the other reduced form, we can identify δ_q^* from

$$\delta_q^* = \left(E\left[W_t W_t'\right]\right)^{-1} E\left[W_t Q_t\right]$$

Notice then that

$$\frac{\gamma^*_{qd}}{\gamma^*_{pd}} = \alpha_s$$

So we can identify α_s simply by taking the ratio of the reduced form coefficients on Z_{dt}

Intuition

$$\frac{dQ^s_t(P_t)}{dZ_{dt}} = \frac{\partial Q^s_t(P_t)}{\partial P_t}\frac{\partial P_t}{\partial Z_{dt}}$$

If Z_{dt} affects Q_t in any way other than altering the price through the demand curve then this won't work

Notice that the same argument will work for α_d with the Z_{st} coefficients

Method 2

Define

$$P_t^* = X_t' \delta_{px}^* + Z_{dt}' \delta_{pd}^* + Z_{st}' \delta_{ps}^*$$

so

$$P_t = P_t^* + v_{pt}^*$$

This is identified since δ_p^* is identified

Now notice that

$$\begin{aligned} Q_t =& \alpha_s P_t + X_t' \beta_x + Z_{st}' \beta_s + u_t \\ &= \alpha_s \left[P_t^* + v_{pt}^* \right] + X_t' \beta_x + Z_{st}' \beta_s + u_t \\ &= \alpha_s P_t^* + X_t' \beta_x + Z_{st}' \beta_s + u_t + \alpha_s v_{pt}^* \end{aligned}$$

One could get a consistent estimate of α by regressing Q_t on $W_t^* = (P_t^*, X_t, Z_{st})$ then

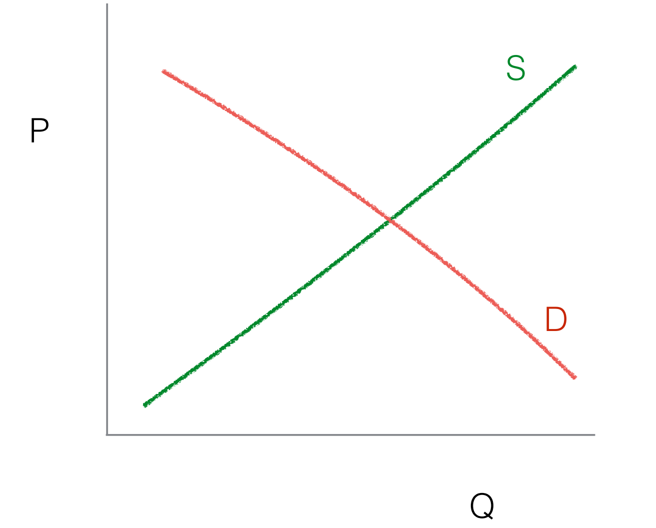
$$\left(E\left[W_t^* W_t^{*\prime}\right]\right)^{-1} E\left[W_t^* P_t\right]$$

$$= \left(E\left[W_t W_t'\right]\right)^{-1} E\left[W_t\left(W_t'\begin{bmatrix}\alpha_s\\ \beta_x\\ \beta_s\end{bmatrix} + u_t + \alpha_s v_{pt}^*\right)\right]$$

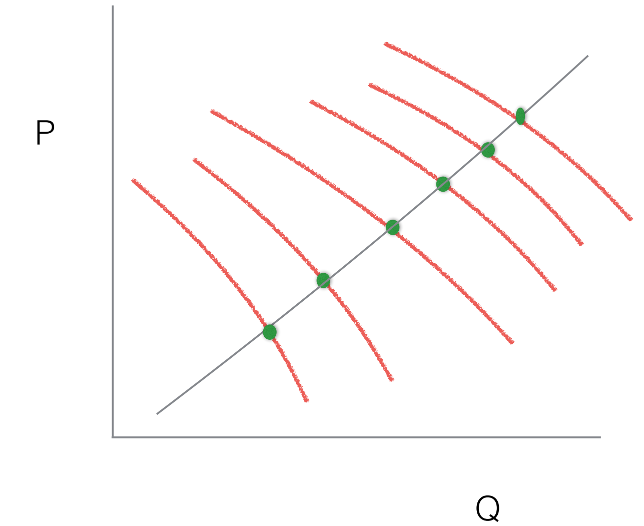
$$= \left(E\left[W_t W_t'\right]\right)^{-1} E\left[W_t W_t'\right]\begin{bmatrix}\alpha_s\\ \beta_x\\ \beta_s\end{bmatrix} + \left(E\left[W_t W_t'\right]\right)^{-1} E\left[W_t(u_t + \alpha_s v_{pt}^*)\right]$$

$$= \begin{bmatrix}\alpha_s\\ \beta_x\\ \beta_s\end{bmatrix}$$

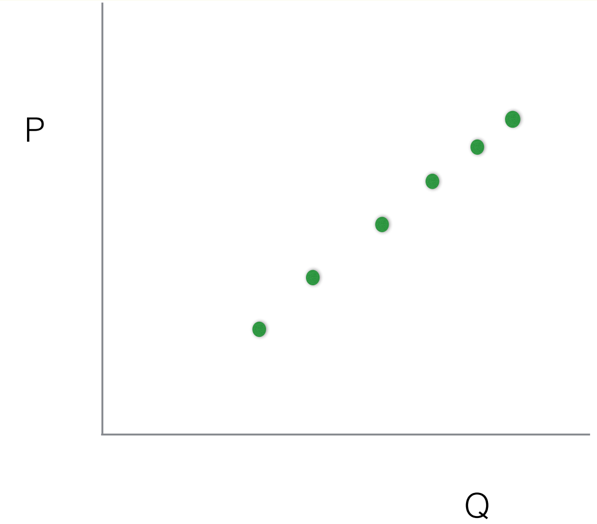
Lets see the intuition graphically

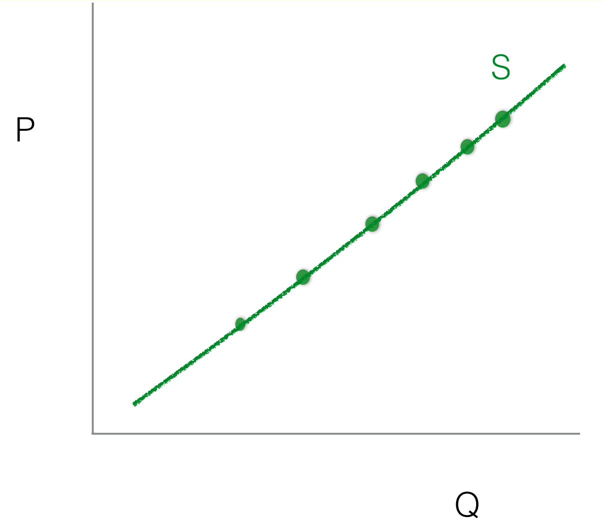


To identify the Supply curve we hold everything constant but Z_{di} which moves the price by moving demand curve



This allows us to trace out the supply curve




S
P
Q

Method 3

For a generic random variable Y_t define

$$\widetilde{Y}_t = Y_t - E\left(P_t \mid X_t, Z_{st}\right)$$

since

$$\begin{aligned} Q_t &= \alpha_s P_t + X_t' \beta_x + Z_{st}' \beta_s + u_t \\ E\left(Q_t \mid X_t, Z_{st}\right) &= \alpha_s E\left(P_t \mid X_t, Z_{st}\right) + X_t' \beta_x + Z_{st}' \beta_s \end{aligned}$$

we know that

$$\widetilde{Q}_t = \alpha_s \widetilde{P}_t + u_t$$

Now notice

$$cov\left(\widetilde{Z}_{dt},\widetilde{Q}_t\right) = \alpha_s cov\left(\widetilde{Z}_{dt},\widetilde{P}_t\right) + cov\left(\widetilde{Z}_{dt},u_t\right)$$

but since $cov\left(\widetilde{Z}_{dt},u_t\right) = 0,$

$$\alpha_s = \frac{cov\left(\widetilde{Z}_{dt},\widetilde{Q}_t\right)}{cov\left(\widetilde{Z}_{dt},\widetilde{P}_t\right)}$$

Lets think about the denominator $cov\left(\widetilde{Z}_{dt},\widetilde{P}_t\right)$ We also know that

$$
\begin{aligned}
P_t =& X_t'\delta_{px}^* + Z_{dt}'\delta_{pd}^* + Z_{st}'\delta_{ps}^* + v_{pt}^* \\
E\left(P_t \mid X_t, Z_{st}\right) =& X_t'\delta_{px}^* + E\left(Z_{dt}' \mid X_t, Z_{st}\right)\delta_{pd}^* + Z_{st}'\delta_{ps}^* \\
\widetilde{P}_t =& \widetilde{Z}_{dt}'\delta_{pd}^* + v_{pt}^*
\end{aligned}
$$

Thus

$$
cov\left(\widetilde{Z}_{dt},\widetilde{P}_t\right) = \delta_{pd}^* var\left(\widetilde{Z}_{dt}\right)
$$

so the denominator will be non-zero as long as $\delta_{pd}^* \neq 0$ and $var\left(\widetilde{Z}_{dt}\right) \neq 0$

One can see the importance of the assumption that $cov\left(\widetilde{Z}_{dt}, u_t\right) = 0$

$$\alpha_{sIV} \equiv \frac{cov\left(\widetilde{Z}_{dt}, \widetilde{Q}_t\right)}{cov\left(\widetilde{Z}_{dt}, \widetilde{P}_t\right)}$$
$$= \alpha_s + \frac{cov\left(\widetilde{Z}_{dt}, u_t\right)}{cov\left(\widetilde{Z}_{dt}, \widetilde{P}_t\right)}$$

This formula is helpful. In order for the model to be consistent you need

- $cov(\widetilde{Z}_{dt}, \widetilde{u}_t) = 0$
- $cov\left(\widetilde{Z}_{dt}, \widetilde{P}_t\right) \neq 0$

But more generally for the asymptotic bias to be small you want

- $cov(\widetilde{Z}_{dt}, \widetilde{u}_t)$ to be small
- $|cov\left(\widetilde{Z}_{dt}, \widetilde{P}_t\right)|$ to be large

This means that in practice there is some tradeoff between them.

If your instrument is not very powerful, a little bit of correlation in $cov(\widetilde{Z}_{dt}, \widetilde{u}_t)$ could lead to a large asymptotic bias.

Identification of the Roy Model

Lets think about identifying this model

This is discussed in Heckman and Honore (EMA, 1990)

We will follow the discussion in French and Taber, Handbook of Labor Economics, 2011

While the model is about the simplest in the world, identification is difficult

Why is thinking about nonparametric identification useful?

- Speaking for myself, I think it is. I always begin a research project by thinking about nonparametric identification.
    - Literature on nonparametric identification not particularly highly cited
    - At the same time this literature has had a huge impact on empirical work in practice. A Heckman two step model without an exclusion restriction is often viewed as highly problematic these days- because of nonparametric identification
    - It is also useful for telling you what questions the data can possibly answer. If what you are interested is not nonparametrically identified, it is not obvious you should proceed with what you are doing

Next we consider nonparametric identification of the Roy model

We consider the model above

$$\begin{aligned} W_{fi} &= g_f(X_{fi}, X_{0i}) + \varepsilon_{fi} \\ W_{hi} &= g_h(X_{hi}, X_{0i}) + \varepsilon_{hi}, \end{aligned}$$

where the joint distribution of $(\varepsilon_{fi}, \varepsilon_{hi})$ is G.

In this case $\theta = (g_f, g_h, G)$

Assumptions

- $(\varepsilon_{fi}, \varepsilon_{hi})$ is independent of $X_i = (X_{0i}, X_{fi}, X_{hi})$
- Normalize $E(\varepsilon_{fi})$=0
  To see why this is a normalization we can always subtract $E(\varepsilon_{fi})$ from ε_{fi} and add it to $g_f(X_{fi}, X_{0i})$ making no difference in the model itself
- Normalize the median of $\varepsilon_{fi} - \varepsilon_{hi}$ to zero.
  A bit non-standard but we can always add the median of $\varepsilon_{fi} - \varepsilon_{hi}$ to ε_{hi} and subtract it from $g_h(X_{hi}, X_{0i})$
- $supp(g_f(X_{fi}, x_0), g_h(X_{hi}, x_0)) = \mathbb{R}^2$ for all $x_0 \in supp(X_{0i})$

Step 1: Identification of Reduced Form Choice Model

This part is well known in a number of papers (Manski and Matzkin being the main contributors) We can write the model as

$$\begin{aligned} Pr(F_i = 1 \mid X_i = x) &= Pr(\varepsilon_{ih} - \varepsilon_{if} \leq g_f(x_f, x_0) - g_h(x_h, x_0)) \\ &= G_{h-f}(g^*(x)), \end{aligned}$$

where G_{h-f} is the distribution function for $\varepsilon_{ih} - \varepsilon_{if}$ and

$$g^*(x) \equiv g_f(x_f, x_0) - g_h(x_h, x_0).$$

We can not separate $g*$ from G_{h-f}, but we can identify this as a function of x

We also know that for any two values x_1 and x_2, if

$$Pr(F_i = 1 \mid X_i = x_1) = Pr(F_i = 1 \mid X_i = x_2)$$

then

$$g^*(x_1) = g^*(x_2)$$

Step 2: Identification of the Wage Equation g_f

Next consider identification of $g_{f.}$ This is basically the standard selection problem.

Notice that we can identify the distribution of W_{fi} conditional on $(X_i = x, F_i = 1.)$

In particular we can identify

$$E(W_i \mid X_i = x, F_i = 1) = g_f(x_f, x_0) + E(\varepsilon_{if} \mid \varepsilon_{ih} - \varepsilon_{if} < g^*(x)).$$

An exclusion restriction is key, we need a variable x_h that allows us move (x_f, x_0) holding $g^*(x)$ and thus $E(\varepsilon_{ih} - \varepsilon_{if} \mid X_i = x, \varepsilon_{ih} - \varepsilon_{if} < g^*(x)))$ fixed.

By holding $g^*(x)$ fixed and varying (x_f, x_0) we can identify g_f up to location

Identification at Infinity

What about the location?

Notice that

$$\begin{aligned}
\lim_{Pr(F_i=1|X_i=x)\to 1} & E(W_i \mid X_i = x, F_i = 1) \\
&= g_f(x_f, x_0) + \lim_{g^*(x)\to\infty} E(\varepsilon_{fi} \mid \varepsilon_{ih} - \varepsilon_{if} < g^*(x))) \\
&= g_f(x_f, x_0) + E(\varepsilon_{fi}) \\
&= g_f(x_f, x_0)
\end{aligned}$$

Thus we are done

Another important point is that the model is not identified without identification at infinity.

To see why suppose that $g^*(x)$ is bounded from above at g^u then if $\varepsilon_{ih} - \varepsilon_{if} > g^u$, $F_i = 0$. Thus the data is completely uninformative about

$$E(\varepsilon_{fi} \mid \varepsilon_{ih} - \varepsilon_{if} > g^u)$$

so the model is not identified.

Parametric assumptions on the distribution of the error term is an alternative.

Who cares about Location?

Actually we do, a lot

- Without our intercept we know something about wage variation within fishing
- However we can not compare the level of fishing to the level of hunting
- If our policy involves moving people from one to the other we need the intercepts

Step 3: Identification of g_h

What will be crucial is the other exclusion restriction (i.e. X_{fi}).

For any (x_r, x_0) we want to find an x_f so that

$$Pr(F_i = 1 \mid X_i = (x_f, x_r, x_0)) = 0.5.$$

This means that

$$0.5 = \Pr\left(\varepsilon_{hi} - \varepsilon_{fi} \leq g_f(x_f, x_0) - g_h(x_h, x_0)\right).$$

But the fact that $\varepsilon_{hi} - \varepsilon_{fi}$ has median zero implies that

$$g_h(x_h, x_0) = g_f(x_f, x_0).$$

Since

g_f is identified, clearly g_h is identified from this expression.

Step 4: Identification of G

Relatively straight forward given everything else

To identify the joint distribution of $(\varepsilon_{fi}, \varepsilon_{hi})$ note that from the data one can observe

$$\begin{aligned} & \Pr(F_i = 1, W_i < s \mid X_i = x) \\ = \; & \Pr(g_h(x_h, x_0) + \varepsilon_{hi} \leq g_h(x_h, x_0) + \varepsilon_{hi}, g_f(x_f, x_0) + \varepsilon_{fi} \leq s) \\ = \; & \Pr(\varepsilon_{hi} - \varepsilon_{fi} \leq g_f(x_f, x_0) - g_h(x_h, x_0), \varepsilon_{fi} \leq s - g_f(x_f, x_0)) \end{aligned}$$

which is the cumulative distribution function of $(\varepsilon_{hi} - \varepsilon_{fi}, \varepsilon_{fi})$ evaluated at the point $(g_f(x_f, x_0) - g_r(x_r, x_0), s - g_f(x_f, x_0))$

Thus we know the joint distribution of $(\varepsilon_{hi} - \varepsilon_{fi}, \varepsilon_{fi})$

Outline

Estimation

So how do we estimate the model and do policy analysis? There are really 3 different approaches

1. Estimate full structural model (and thus data generating process) and simulate policy effect
2. Estimate reduced form of data generating process and simulate policy effect
3. Try to estimate policy directly without estimating full DGP

By far the most common is the first so I will focus on that

There are really two basic ways of estimating the data generation process

1. Maximum Likelihood
2. Simulation Methods (SMM,Indirect Inference)

Background maximum likelihood

For some random variable Y, let $f(Y;\theta)$ be the density of Y if it is generated by a model with parameter θ

The likelihood function just writes the function the other way:

$$\ell(\theta;Y) = f(Y;\theta).$$

Let θ_0 represent the true parameter

The key result is that

$$E\left(\frac{\ell(\theta;Y_i)}{\ell(\theta_0;Y_i)}\right) = \int \frac{\ell(\theta;Y_i)}{\ell(\theta_0;Y_i)} f(Y_i;\theta_0)dY_i$$
$$= \int \frac{f(Y_i;\theta)}{f(Y_i;\theta_0)} f(Y_i;\theta_0)dY_i$$
$$= \int f(Y_i;\theta)dY_i$$
$$=1$$

because $f(Y_i;\theta)$ is a density.

We use Jensen's inequality which implies that for any random variable X_i, the fact that log is concave implies that:

$$E(log(X_i)) \leq log(E(X_i))$$

We apply this with

$$X_i = \frac{\ell(\theta; Y_i)}{\ell(\theta_0; Y_i)}$$

Thus

$$
\begin{aligned}
E\left(log\left(\frac{\ell(\theta;Y_i)}{\ell(\theta_0;Y_i)}\right)\right) =& E\left(log\left(\ell(\theta;Y_i)\right)\right) - E\left(log\left(\ell(\theta_0;Y_i)\right)\right) \\
\leq& log\left(E\left(\frac{\ell(\theta;Y_i)}{\ell(\theta_0;Y_i)}\right)\right) \\
=& log(1)
\end{aligned}
$$

or

$$
E\left(log\left(\ell(\theta;Y_i)\right)\right) \leq E\left(log\left(\ell(\theta_0;Y_i)\right)\right)
$$

thus we know that the true value of θ maximizes $E\left(log\left(\ell(\theta;Y_i)\right)\right)$

Maximum likelihood is just the sample analogue of this

Choose $\widehat{\theta}$ as the argument that maximizes

$$\frac{1}{N}\sum_{i=1}^{N} log(\ell(\theta; Y_i))$$

The most important result for MLE is that it is efficient

In particular no alternative estimator can have a lower asymptotic variance

Likelihood for Roy

Lets assume that

$$\begin{bmatrix} \varepsilon_{fi} \\ \varepsilon_{hi} \end{bmatrix} \sim N\left(0, \begin{bmatrix} \sigma_{ff} & \sigma_{fh} \\ \sigma_{fh} & \sigma_{hh} \end{bmatrix}\right)$$

For a fisherman we observe whether you fish

$$\varepsilon_{ih} - \varepsilon_{if} \leq g_f(X_{fi}, X_{0i}) - g_h(X_{hi}, X_{0i})$$

and their wage

$$W_i = g_f(X_{fi}, X_{0i}) + \varepsilon_{if}$$

so

$$\varepsilon_{ih} \leq W_i - g_h\left(X_{hi}, X_{0i}\right)$$

The likelihood is

$$\int_{-\infty}^{W_i - g_h(X_{hi}, X_{0i})} \phi(W_i - g_f(X_{fi}, X_{0i}), \varepsilon_{ih}); \Sigma) d\varepsilon_{ih}$$

We get an analogous expression for hunters which gives the log likelihood function

$$\frac{1}{N}\sum_{i=1}^{N}\left[F_i \log\left(\int_{-\infty}^{W_i-g_h(X_{hi},X_{0i})}\phi(W_i-g_f(X_{fi},X_{0i}),\varepsilon_{ih});\Sigma)d\varepsilon_{ih}\right)\right.$$

$$\left.+(1-F_i)\log\left(\int_{-\infty}^{W_i-g_f(X_{fi},X_{0i})}\phi(\varepsilon_{if},W_i-g_h\left(X_{hi},X_{0i}\right);\Sigma)d\varepsilon_{if}\right)\right]$$

Generalized Method of Moments

Another way to estimate such a model is by GMM, simulated method of moments, or indirect inference

I am not sure these terms mean the same thing to everyone, so I will say what I mean by them but recognize it might mean different things to different people.

Lets continue to assume that the econometrician observes (Υ_i, X_i) which are i.i.d. and both X_i and Υ_i are potentially large dimensional.

We wrote the data generating process

$$\begin{aligned} \Upsilon_i \equiv & y_0(X_i, u_i; \theta) \\ u_i \sim & F(u_i; \theta) \\ X_i \sim & H(X_i) \end{aligned}$$

The standard GMM model would come up with a set of moments

$$m(X_i, \Upsilon_i, \theta)$$

for which

$$E[m(X_i, \Upsilon_i, \theta_0)] = 0$$

the sample analogue comes from recognizing that

$$\frac{1}{N}\sum_{i=1}^{N} m(X_i, \Upsilon_i, \theta_0) \approx 0$$

But more generally we are overidentified so we choose $\widehat{\theta}$ to minimize

$$\left[\frac{1}{N}\sum_{i=1}^{N} m(X_i, \Upsilon_i, \theta)\right]' W \left[\frac{1}{N}\sum_{i=1}^{N} m(X_i, \Upsilon_i, \theta)\right]$$

Relationship between GMM and MLE

Actually in one way you can think of MLE as a special case of GMM

We showed above that

$$\theta_0 = argmax \left[E \left(log \left(\ell(\theta; Y_i) \right) \right) \right]$$

but as long as everything is well behaved this means that

$$E \left(\frac{\partial log \left(\ell(\theta; Y_i) \right)}{\partial \theta} \right) = 0$$

We can use this as a moment condition

The one very important caveat is that this is only true if the log likelihood function is strictly concave

Otherwise there might be multiple solutions to the first order conditions, but only one actual maximum to the likelihood function

Simulated Method of Moments

The classic reference is "A Method of Simulated Moments of Estimation of Discrete Response Models Without Numerical Integration," McFadden, EMA, 1989

However, I will present it in a different way

Take any function of the data that you like say $g(\Upsilon_i, X_i)$ (where the dimension of g is often large)

Then notice that since y_0 and F represent the data generating process

$$E[g(\Upsilon_i, X_i)] = \int\int (g(y_0(X, u; \theta_0), X_i) dF(u; \theta_0) dH(X)$$

So this means that we can do GMM with

$$m(\Upsilon_i, X_i, \theta) = g(\Upsilon_i, X_i) - \int\int (g(y_0(X, u; \theta), X_i) dF(u; \theta) dH(X)$$

So what?

Here is where things get pretty cool

Notice that at the true value the estimator is

$$\begin{aligned} & \frac{1}{N}\sum_{i=1}^{N} g(\Upsilon_i, X_i) - \frac{1}{S}\sum_{s=1}^{S}(g(y_0(x_s, u_s; \theta_0)) \\ \approx & E[g(\Upsilon_i, X_i)] - \int\int (g(y_0(X, u; \theta_0))dF(u; \theta_0)dH(X) \\ = & 0 \end{aligned}$$

The nice thing about this is that we didn't need S to be large for every N, we only needed S to be large for the one integral.

For MLE we had to approximate the integral well for every single observation

This makes this much easier computationally

Indirect inference

The classic reference here is "Indirect Inference" Gourieroux, Monrort, and Renault, Journal of Applied Econometrics, 1993

Again I will think about this in a different way then them

Think about the intuition for the SMM estimator

$$\frac{1}{N}\sum_{i=1}^{N} g(X_i, Y_i) \approx \frac{1}{S}\sum_{s=1}^{S}(g(y_0(x_s, u_s; \theta_0))$$

If I have the right data generating model taking the mean of the simulated data should give me the same answer as taking the mean of the actual data

But we can generalize that idea

If I have the right data generating model, if I use the true parameter value, the simulated data should look the same as the actual data

That means whatever the heck I do to the real data-if I do exactly the same thing to the simulated data I should get the same answer

Indirect Inference Procedure

- Estimate auxiliary parameter $\widehat{\beta}$ using some estimation scheme in real data
- for any particular value of θ
    - Simulate data using data generation process: $y_0(x,u;\theta), H(X), G(u;\theta)$
    - Estimate $\widehat{B}(\theta)$ using exactly the same estimation scheme on simulated data
- Choose θ to minimize

$$\left(\widehat{B}(\theta)-\widehat{\beta}\right)'\Omega\left(\widehat{B}(\theta)-\widehat{\beta}\right)$$

This is consistent because

$$\widehat{B}(\theta_0)-\widehat{\beta}\xrightarrow{p}0$$

The most important thing: this can be misspecified, it doesn't have to estimate a true causal parameter

Creates a nice connection with reduced form stuff, we can use 2SLS or Diff in Diff as auxiliary parameters and it is clear where identification comes from

Can think of the analogue to the forecasting out of the sample-we use Indirect Inference to extend the convincing identification scheme into a structural framework

Examples of $\widehat{\beta}$:

- Moments
    - Regression models
    - Misspecified MLE
    - Misspecified GMM
    - IV
    - Difference in Differences
    - Regression Discontinuity
    - Even Randomized Control Files

Maximum Likelihood versus Indirect Inference

- MLE is efficient
- Indirect inference you pick auxiliary model

Which is better is not obvious.

Picking auxiliary model is somewhat arbitrary, but you can pick what you want the data to fit.

MLE essentially picks the moments that are most efficient-a statistical criterion

- Indirect inference is often computationally easier because of the simulation approximation of integrals
  - With confidential data, Indirect Inference often is easier because only need to use the actual data to get $\widehat{\beta}$
  - A drawback of simulation estimators is that they often lead to nonsmooth objective functions
  - Indirect inference preserves some of the advantages of design based estimation
    - Map from data to parameters is more transparent
    - Becomes like the forecasting experiment where we are forecasting out of the range of the data