Identification of Endogenous Social Effects: The Reflection Problem

Manski

1993

A Linear Model

Three reasons why individuals belonging to the same group tent to behave similarly.

Endogenous Effect : Behavior of the indivdiual is correlated with the behavior of the group.

Exogenous (contextual) Effect : Behavior of the individual varies with the exogenous characteristics of the group

Correlated Effects : Individuals in the same group behave similarly because they have similar individual characteristics or face similar institutional environments.

Example: High School Achievement

Endogenous Effect : Individual achievement vary with the achievement of the members of the reference group.

Exogenous (Contextual) Effect : Individual achievement is related to the socio-economic background of her reference group.

Correlated Effect : Individuals in the same reference group tend to achieve similarly because they have similar parental backgrounds or are taught by the same teacher.

Endogenous Effects generate social multiplier: Improving some students' performance increases those other students as well. Exogenous Effects or Correlated Effects do not generate social multiplier.

Model Specification

$$y = \alpha + \beta E(y|x) + E(z|x)'\gamma + z'\eta + u,\ E(u|x,z) = x'\delta \qquad (1)$$

y: scalar outcome (highschool achievement).
x: attributes of the individual's reference group.
(z, u): attributes that directly affect outcome y. For example, socioeconomic status or ability of the individual. u is unobserved
Taking expectations on both sides,

$$E(y|x,z) = \alpha + \beta E(y|x) + E(z|x)'\gamma + x'\delta + z'\eta \qquad (2)$$

$\beta E(y|x)$: Endogenous Effect
$E(z|x)'\gamma$, $E(u|x,z) = x'\delta$: Exogenous (Contextual) Effect
$x'\delta$: Correlated Effect

Solution and Identification

Social Equilibrium

$$E(y|x) = \alpha + \beta E(y|x) + E(z|x)'\gamma + x'\delta + E(z|x)'\eta$$

Derive $E(y|x)$

$$E(y|x) = \frac{\alpha}{1-\beta} + E(z|x)'\frac{\gamma+\eta}{1-\beta} + \frac{x'\delta}{1-\beta}$$

Substitute $E(y|x)$ to obtain

$$E(y|x,z) = \frac{\alpha}{1-\beta} + E(z|x)'\frac{\gamma+\beta\eta}{1-\beta} + \frac{x'\delta}{1-\beta} + z'\eta$$

Identification of the Parameters

- Composite parameters $\frac{\alpha}{1-\beta}$, $\frac{\gamma+\beta\eta}{1-\beta}$, $\frac{\delta}{1-\beta}$, η are identified if $[1, E(z|x)x, z]$ are linearly independent in the population.
- It is impossible to separately identify the endogenous effect β.

The composite social effects parameter $\frac{\gamma+\beta\eta}{1-\beta}$ is not identified if any of those conditions hold

- z is a function of x.
- $E(z|x)$ does not vary with x.
- $E(z|x)$ is a linear function of x.

A Pure Endogenous Effecs Model

Assume neither exogenous nor correlated effects are present ($\gamma = \delta = 0$)

$$E(y|x,z) = \frac{\alpha}{1-\beta} + E(z|x)'\frac{\beta\eta}{1-\beta} + z'\eta$$

- The composite parameters $\frac{\alpha}{1-\beta}$, $\frac{\beta\eta}{1-\beta}$, η are identified if the regressors $[1, E(z|x), z]$ are linearly independent in the population.
- The endogenous effects paramter β is not identified if $\eta = 0$ or $E(z|x)$ is a linear function of $[1, z]$

Reference group

Suppose $z = z(x)$, z is a function of x. That is, there is no variation in within reference group individual heterogeneity, i.e. reference group is completely defined by individual heterogeneity in the data. Then, tautologically,

$$E\left[y|x, z(x)\right] = E(y|x)$$

Therefore, the social interaction equation

$$E(y|x) = \alpha + \beta E(y|x) + E(z|x)'\gamma + x'\delta + z'\eta$$

holds with $\beta = 1$, $\alpha = 0$, $\gamma = 0$, $\delta = 0$, $\eta = 0$. A large endogenous effect, i.e. individual outcome depends on the outcome of the reference group.

On the other hand, $x = x(z)$ is also true. Then,

$$E\left[y|x, z(x)\right] = E(y|z)$$

Then the equation

$$E(y|x) = \alpha + \beta E(y|x) + E(z|x)'\gamma + x'\delta + g(z)$$

holds with $\beta = 0$, $\alpha = 0$, $\gamma = 0$, $\delta = 0$ and $g(z) = E(y|z)$. Hence, no endogenous effect.

- If the reference group equals the individuals with the same observed characteristics, then it is impossible to estimate the endogenous effect.
- The estimation results crucially depend on the referenc group of the individual. Hence, researchers need to know a priori how individuals form reference group.

Dynamic Models

$$E_r(y|x,z) = \alpha + \beta E_{t-1}(y|x) + E_{t-1}(z|x)'\gamma + x_t'\delta + z_t'\eta$$

Assumption: non-social forces act contemporaneously but social forces act with a lag.
That is, one can estimate the above model by specifying

$$z_t = c_0 z_{t-1} + c_1 x_t + u_t$$

Furthermore, specify

$$y_t = b_0 y_{t-1} + b_1 x_t + v_t$$

$$y_t = \alpha + \beta \left[b_0 y_{t-1} + b_1 x_t\right] + \left[c_0 z_{t-1} + c_1 x_t\right]' \gamma + x_t'\delta + z_t'\eta$$

But it is hard to verify from the data whether social forces act with a lag or not.