ECON6300/7320/8300 Advanced Microeconometrics Cross-Sectional Dependence

Christiern Rose [1]

[1]University of Queensland

Identification of Endogenous Social Effects: The Reflection Problem

CHARLES F. MANSKI
University of Wisconsin-Madison

First version received December 1991; final version accepted December 1992 (Eds.)

This paper examines the reflection problem that arises when a researcher observing the distribution of behaviour in a population tries to infer whether the average behaviour in some group influences the behaviour of the individuals that comprise the group. It is found that inference is not possible unless the researcher has prior information specifying the composition of reference groups. If this information is available, the prospects for inference depend critically on the population relationship between the variables defining reference groups and those directly affecting outcomes. Inference is difficult to impossible if these variables are functionally dependent or are statistically independent. The prospects are better if the variables defining reference groups and those directly affecting outcomes are moderately related in the population.

Introduction

- The paper studies identification of endogenous social effects.
- The propensity of an individual to behave in some way varies with the prevalence of that behaviour in some reference group containing the individual.
- Such phenomena have been often called "social norms", "peer influences", "neighbourhood effects", "conformity", "imitation", "contagion", "epidemics", etc.
- Endogenous social effects in Economics:
  - The output chosen by each firm is a function of aggregate industry output
  - When decision making is costly, people may want to imitate the behaviour of other persons who are better informed.
- Manski (1993) studies identification of endogenous social effects

Three Hypotheses

- Three hypotheses on the common observation that individuals belonging to the same group tend to behave similarly.
  - Endogenous effects: the propensity of an individual to behave in some way varies with the behaviour of the group.
  - Exogenous effects: the propensity of an individual to behave in some way varies with the exogenous characteristics of the group.
  - Correlated effects: individuals in the same group tend to behave similarly because they have similar individual characteristics or face similar institutional environments.

Three Hypotheses: High school achievements

- Endogenous effects: individual achievement varies with the average of average achievements in the youth's school, ethnic group, or other reference group $\Rightarrow$ social multiplier, reflection

- Exogenous effects: individual achievement depends on the average of individual characteristics in the reference group
- Correlated effects: similar family background, same teachers, etc.

Setup

- Each member i in the reference group g is characterised by a value for $(y_{ig}, x_g, z_{ig}, u_{ig})$.
  - $y_{ig} \in \mathbb{R}$: scalar individual outcome (exam score, etc)
  - $x_g \in \mathbb{R}^J$: characteristics of the reference group (school, race, etc.)
  - $(z_{ig}, u_{ig}) \in \mathbb{R}^K \times \mathbb{R}$: individual characteristics
- Econometrician observes (y_{ig}, x_g, z_{ig}), but not u_{ig}.
- Assumption on DGP:

$$y_{ig} = \alpha + \beta E[y_{ig}|x_g] + E[z_{ig}|x_g]'\gamma + z_{ig}'\eta + u_{ig}$$

with $E[u_{ig}|x_g, z_{ig}] = x_g'\delta$.

Setup

- The regression model is given as

$$E[y_{ig}|x_g, z_{ig}] = \alpha + \beta E[y_{ig}|x_g] + E[z_{ig}|x_g]'\gamma + x_g'\delta + z_{ig}'\eta$$

- β measures the endogenous effects: $E[y_{ig}|x_g]$ is the average exam score in the reference group
- γ measures the exogenous effects: $E[z_{ig}|x_g]$ is the average socio-economic characteristics in the reference group, i.e., average family income, average parents' education, etc.
- δ measures the correlated effects
- η measures direct effects
- We are interested in the identification of $(\alpha, \beta, \gamma, \delta, \eta)$.

Identification

- Rewrite the structural model:

$$E[y_{ig}|x_g, z_{ig}] = \alpha + \beta E[y_{ig}|x_g] + E[z_{ig}|x_g]'\gamma + x_g'\delta + z_{ig}'\eta$$

- We assume that the conditional expectation functions above $E[y_{ig}|x_g, z_{ig}]$, $E[y_{ig}|x_g]$, and $E[z_{ig}|x_g]$ are all identified, and focus on identification of $(\alpha, \beta, \gamma, \delta, \eta)$.

- Integrate z_{ig} out:

$$E[y_{ig}|x_g] = \alpha + \beta E[y_{ig}|x_g] + E[z_{ig}|x_g]'(\gamma + \eta) + x_g'\delta$$

- If $\beta \neq 1$, we can isolate $E[y_{ig}|x_g]$ on the LHS:

$$E[y_{ig}|x_g] = \left(\frac{\alpha}{1-\beta}\right) + E[z_{ig}|x_g]'\left(\frac{\gamma + \eta}{1-\beta}\right) + x_g'\left(\frac{\delta}{1-\beta}\right)$$

- Plug the last equation into the one in the first bullet point.

Identification

- Then, we have the reduced form regression

$$E[y_{ig}|x_g, z_{ig}] = \left(\frac{\alpha}{1-\beta}\right) + E[z_{ig}|x_g]'\left(\frac{\gamma+\beta\eta}{1-\beta}\right) + x_g'\left(\frac{\delta}{1-\beta}\right) + z_{ig}'\eta$$

- Even if the four parameters are all identified, we cannot identify all the structural parameters $(\alpha, \beta, \gamma, \delta, \eta)$ without further assumptions, as we have 5 structural parameters but only 4 reduced form parameters.
- But, we can still learn about presence of social effects (i.e. $\beta = \gamma = 0$ or not).

Identification of composite social effects

- **Result 1:** Provided that $\beta \neq 1$, the four composite parameters above, $\left(\frac{\alpha}{1-\beta}\right), \left(\frac{\gamma+\beta\eta}{1-\beta}\right), \left(\frac{\delta}{1-\beta}\right), \eta$ are identified if the regressors $(1, E[z_{ig}|x_g], x_g, z_{ig})$, are linearly independent.
- Result 1 implies that
  - we identify the presence of social effects because if $\left(\frac{\gamma+\beta\eta}{1-\beta}\right) \neq 0$, it must be that either $\gamma \neq 0$ or $\beta \neq 0$.
  - if $E[z_{ig}|x_g]$ is linear in $(1, x_g, z_{ig})$, we cannot even identify $\left(\frac{\gamma+\beta\eta}{1-\beta}\right)$.

Identification of composite social effects

- Specifically, the social effect parameter $\left(\frac{\gamma+\beta\eta}{1-\beta}\right)$ is not identified if either
  - z_{ig} is a function of x_g
  - $E[z_{ig}|x_g]$ does not depend on x_g, or
  - $E[z_{ig}|x_g]$ is linear in x_g
- Therefore, the social effect parameter is identified only when $E[z_{ig}|x_g]$ is nonlinear in x_g

Identification of composite social effects

- Let's investigate each case.
- First, if $z_{ig} = h(x_g)$, the reduced form regression is written as

$$E[y_{ig}|x_g, z_{ig}] = \left(\frac{\alpha}{1-\beta}\right) + E[h(x_g)|x_g]' \left(\frac{\gamma + \beta\eta}{1-\beta}\right) + x_g' \left(\frac{\delta}{1-\beta}\right) + h(x_g)'\eta$$

$$= \left(\frac{\alpha}{1-\beta}\right) + h(x_g)' \left(\frac{\gamma + \eta}{1-\beta}\right) + x_g' \left(\frac{\delta}{1-\beta}\right)$$

- Now we have only 3 reduced form parameters, and 5 structural parameters
- It can be $\left(\frac{\gamma+\eta}{1-\beta}\right) \neq 0$, when $\eta \neq 0$ but $\gamma = \beta = 0$ (no social effects).

Identification of composite social effects

- Second, if $E[z_{ig}|x_g] = E[z_{ig}]$, we have

$$
\begin{aligned}
E[y_{ig}|x_g, z_{ig}] &= \left(\frac{\alpha}{1-\beta}\right) + E[z_{ig}]' \left(\frac{\gamma + \beta\eta}{1-\beta}\right) + x_g' \left(\frac{\delta}{1-\beta}\right) \\
&\quad + z_{ig}'\eta \\
&= \left(\frac{\alpha + E[z_{ig}]'[\gamma + \beta\eta]}{1-\beta}\right) + x_g' \left(\frac{\delta}{1-\beta}\right) + z_{ig}'\eta
\end{aligned}
$$

  The intercept can be nonzero if $\alpha \neq 0$ and $\gamma = \beta = 0$.

- Finally, if $E[z_{ig}|x_g] = x_g'\kappa$, we have

$$
E[y_{ig}|x_g, z_{ig}] = \left(\frac{\alpha}{1-\beta}\right) + x_g' \left(\frac{\kappa[\gamma + \beta\eta] + \delta}{1-\beta}\right) + z_{ig}'\eta
$$

  where $\left(\frac{\kappa[\gamma+\beta\eta]+\delta}{1-\beta}\right)$ can be nonzero if $\delta \neq 0$ and $\gamma = \beta = 0$.

Identification of pure endogenous effects

- Empirical studies of endogenous effects typically assume that $\gamma = \delta = 0$; no exogenous and no correlated effects.
- Then, the reduced form regression reduces to

$$E[y_{ig}|x_g, z_{ig}] = \left(\frac{\alpha}{1-\beta}\right) + E[z_{ig}|x_g]' \left(\frac{\beta\eta}{1-\beta}\right) + z_{ig}'\eta$$

- **Result 2:** Provided $\beta \neq 1$ and $\gamma = \delta = 0$, the composite parameters $\left(\frac{\alpha}{1-\beta}\right)$, $\left(\frac{\beta\eta}{1-\beta}\right)$, and η are identified if $(1, E[z_{ig}|x_g], z_{ig})$ are linearly independent. Moreover, the endogenous effect β is identified if $\eta \neq 0$.
- As before, β is not identified, if $\eta = 0$ or $E[z_{ig}|x_g]$ is linear in $(1, z_{ig})$, i.e.,
  - if z_{ig} is a function of x_g,
  - if $E[z_{ig}|x_g]$ does not depend on x_g, or
  - if $E[z_{ig}|x_g]$ is linear in x_g and x_g is linear in z_{ig}.

Tautological models

- We have seen that even when the parameters are not identified, we could have some testable restrictions. Recall the reduced model

$$E[y_{ig}|x_g, z_{ig}] = \left(\frac{\alpha}{1-\beta}\right) + E[z_{ig}|x_g]'\left(\frac{\gamma+\beta\eta}{1-\beta}\right) + x_g'\left(\frac{\delta}{1-\beta}\right) + z_{ig}'\eta$$

where $\left(\frac{\gamma+\beta\eta}{1-\beta}\right) \neq 0$ implies some social effects, either $\gamma \neq 0$ or $\beta \neq 0$.

- But, some specifications of z_{ig} and x_g may lead to a tautological model that is consistent with <u>any</u> observed behaviour.

Tautological models

- For example, if $z_{ig} = h(x_g)$, then the structural model

$$E[y_{ig}|x_g, z_{ig}] = \alpha + \beta E[y_{ig}|x_g] + E[z_{ig}|x_g]'\gamma + x_g'\delta + z_{ig}'\eta$$
$$\implies E[y_{ig}|x_g] = \alpha + \beta E[y_{ig}|x_g] + E[z_{ig}|x_g]'\gamma + x_g'\delta + z_{ig}'\eta$$

  which always holds with $\beta = 1$ & $\alpha = \gamma = \delta = \eta = 0$. So, $E[y_{ig}|x_g] = E[y_{ig}|x_g] \ldots$ (We regress the outcome on itself!)

- Similarly, if $x_g = h(z_{ig})$, then

$$E[y_{ig}|x_g, z_{ig}] = \alpha + \beta E[y_{ig}|x_g] + E[z_{ig}|x_g]'\gamma + x_g'\delta + z_{ig}'\eta$$
$$\implies E[y_{ig}|z_{ig}] = \alpha + \beta E[y_{ig}|x_g] + E[z_{ig}|x_g]'\gamma + x_g'\delta + z_{ig}'\eta$$

  which always holds with $\alpha = \beta = \gamma = \delta = 0$. So, $E[y_{ig}|z_{ig}] = z_{ig}'\eta$ and therefore only testable restriction is the linearity assumption.

Prior knowledge on reference group

- The econometrician must know, a priori, how individuals form reference groups $i \in g$ with x_g for studying social effects.

- To see this, suppose the econometrician tries to infer the reference group from the observed behaviour, i.e., the econometrician forms x_g using the observed characteristics z_{ig}.

- Then, x_g is determined by z_{ig}. So, the model becomes tautological.

Estimation Strategies

- Typically, researchers assume that there is no correlated effect ($\delta = 0$) and there is either only exogenous effects ($\beta = 0, \gamma \neq 0$) or only endogenous effects ($\beta \neq 0, \gamma = 0$)
- Studies of exogenous effects use two stage method to estimate (γ, η) restricting $(\beta = \delta = 0)$.
- Under the parameter restrictions, the reduced form model becomes

$$E[y_{ig}|x_g, z_{ig}] = \alpha + E[z_{ig}|x_g]'\gamma + z_{ig}'\eta$$

- **Stage 1:** estimate $E[z_{ig}|x_g]$ nonparametrically.
- **Stage 2:** regress y_{ig} on 1, $\widehat{E}[z_{ig}|x_g]$, and z_{ig}.
- Note here that often x_g is discrete, and $\widehat{E}[z_{ig}|x_g]$ is simply the cell average of z_{ig}.

Estimation Strategies

- Studies of endogenous effects also use two stage method to estimate (β, η), restricting $(\gamma = \delta = 0)$.
- The structural model reduces to

$$E[y_{ig}|x_g, z_{ig}] = \alpha + \beta E[y_{ig}|x_g] + z_{ig}'\eta$$

- **Stage 1:** estimate $E[y|x]$ nonparametrically
- **Stage 2:** regress y_{ig} on 1, $\widehat{E}[y|x_g]$, and z_{ig}.

- Many nonparametric estimates $\widehat{E}[y|x_g]$ are in the form of weight average (LOWESS), i.e., $\widehat{E}[y|x] := \sum_{ig} \omega_{ig}(x) y_{ig}$
- Then, the representation above has the form of the spatial correlation model

$$y_{ig} = \alpha + \beta \left\{ \sum_{ig} \omega_i(x_g) y_{ig} \right\} + z_{ig}'\eta + u_{ig}.$$

Nonparametric endogenous effects model

- The regressions $E[y_{ig}|x_g, z_{ig}]$ does not have to be linear.
- For some unknown function $f : \mathbb{R} \times \mathbb{R}^K \to \mathbb{R}$, we have

$$E[y_{ig}|x_g, z_{ig}] = f(E[y_{ig}|x_g], z_{ig})$$

- The endogenous effects can be measured by the difference

$$f(E[y_{ig}|x_g], z_{ig}) - f(E[y_{ig}|\tilde{x}_g], z_{ig})$$

at two different points x_g and $\tilde{x}_g$, holding z_{ig} at a certain point.

- Manski (1993) does not provide identification conditions, but discusses conditions under which the endogenous effects cannot be nonparametrically identified; see the reference.

Demand analysis

- The endogenous social effects model can be used for demand analysis:
- individual demand y_{ig} for a product varies with price $p(x_g)$, which is partly determined by aggregate demand in the relevant market x_g.
- So, the individual demand model can be written as

$$E[y_{ig}|x_g, z_{ig}] = D[p(x_g), z_{ig}]$$

where z_{ig} is individual characteristics and D is the mean demand.
- Equilibrium price $p(x)$ is determined by the aggregate demand and supply condition of market g,

$$p(x_g) = \pi\{E[y_{ig}|x_g]m(x_g), s(x_g)\}$$

where $m(x_g)$ is the size of market g and $s(x_g)$ is the supply condition.

Demand analysis

- Then, the individual demand model is

$$E[y_{ig}|x_g, z_{ig}] = D[\pi\{E[y_{ig}|x_g]m(x_g), s(x_g)\}, z_{ig}]$$

which is different from the endogenous effects model we have studied,

$$E[y_{ig}|x_g, z_g] = f(E[y_{ig}|x_g], z_{ig})$$

- But, if we assume that $m(x_g)$ and $s(x_g)$ do not depend on x_g, all markets have the same size and homogenous supply conditions, then the demand model can be written as

$$E[y_{ig}|x_g, z_{ig}] = D[E[y_{ig}|x_g], z_{ig}]$$

and analysed in the framework of endogenous social effects model.