Identification by Nonlinearity & Social Interactions

Abi Adams

HT 2017

Outline

Aim: extend the linear social interactions model to a non-linear version to explore ‘identification by functional form’ and the potential identifying power of multiplicity

1. Discrete Choice
2. Nonlinear models & the Reflection Problem
3. Identification by Nonlinearity
4. Examples

Recap

- In Lecture 4, we covered the reflection problem in the context of identification results for linear simultaneous equation models
- In Lecture 5, we discussed how more complicated simultaneous models might result in additional identification challenges arising from incompleteness
- Today, we look at how multiple equilibria can provide identifying power in some contexts
- This will provide insights into 'identification by nonlinearity' or 'identification by functional form'

Social Interactions

- Linear models of social interaction typically have a unique equilibrium and are complete
- However, most micro-founded models suggest more complicated non-linear functional forms
- Consider identification for Brock and Durlaf's (2001) model of social interactions
  - 'Large' game of incomplete information

Discrete Choice

- Many of the contexts in which peer effects are studied have discrete outcomes:
  - Smoke / Don't Smoke
  - Get Pregnant / Don't Get Pregnant
  - Play Truant / Don't Play Truant
  - &c
- (Many studies on the role of peer effects and youth choice)

Discrete Choice

- Brock & Durlauf (2001, 2007): leading framework for studying social interactions in discrete choice environments
- General framework for capturing many types of social interactions as well as characterising equilibrium (and multiple equilibria) and identification
- Simplify the framework here to show how the discrete choice environment allows one to separately identify endogenous and endogenous peer effects

Peer Effects: Recap

- In Lecture 4, we looked at a very simple example of a peer effects model

$$\begin{aligned} s_{1g} &= \theta_0 + \theta_1 b_{1g} + \theta_2 s_{2g} + \theta_3 b_{2g} + u_{1g} \\ s_{2g} &= \theta_0 + \theta_1 b_{2g} + \theta_2 s_{1g} + \theta_3 b_{1g} + u_{2g} \end{aligned} \quad (1)$$

- θ_2: endogenous effects
- θ_3: exogenous effects
- As we saw these effects are not generally separately identified in the linear framework

Example

Example

- Imagine we are interested in modelling the decision of university students of whether to go and protest in various towns/cities
- Let $y_i = \{0, 1\}$ be an indicator of whether engage in a protest at least once in the last 2 months
- Suppose the propensity for demonstrating is a function of:
  - the student's characteristics,
  - the (exogenous) characteristics of the student's classmates
  - their parent's demographic characteristics

Example

- Endogenous peer effect: the influence on peers decision to protest on an individual's own propensity to protest
- Exogenous peer characteristics: the family background of the other students in the peer group
- b_{ig}: the political leaning of the parents of student i in university g
- $b_{\neg i,g}$: the mean political leaning of the parents of students in i's peer group

Example

- Let the peer group be of size N_g and the mean action of i's peers is:

$$m_{ig} = \frac{1}{N_g - 1} \sum_{j \neq i} y_j \tag{2}$$

Utility

- The utility for action y_i is given as:

$$V_i(y_i) = u_i(y_i) + S(y_i, E(m_{ig})) + \epsilon_i(y_i) \tag{3}$$

where:

  - $u_i(y_i)$ is the private component of utility
  - $S(y_i, E(m_{ig}))$ is the impact on utility of the mean action taken by others in the peer group
  - $\epsilon_i(y_i)$ is an idiosyncratic component

Utility

$$V_i(y_i) = u_i(y_i) + S(y_i, E(m_{ig})) + \epsilon_i(y_i) \quad (4)$$

- Note that the peer interaction term is expressed in term's of *i*'s expectations of the actions of her peers
- BD discuss the dynamics of expectation formation and study equilibrium properties for the case of self-consistent, rational expectations

Utility

- Let the private component of utility be dependent on an individual's own characteristics and on the exogenous characteristics of the peer group

$$u_i(y_i) = y_i \left(\theta_0 + \theta_1 b_{ig} + \theta_3 b_{\neg i,g}\right) \tag{5}$$

- Let the peer interaction component of utility be:

$$\begin{aligned} S(y_i, E(m_{ig})) &= y_i \theta_2 E(m_{ig}) \\ &= y_i \theta_2 s_{\neg i,g} \end{aligned} \tag{6}$$

Utility

- The utility from choosing y_i is then:

$$V_i(y_i) = y_i \left(\theta_0 + \theta_1 b_{ig} + \theta_2 s_{\neg i,g} + \theta_3 b_{\neg i,g}\right) + \epsilon_i(y_i) \qquad (7)$$

- Choose to protest if:

$$\theta_0 + \theta_1 b_{ig} + \theta_2 s_{\neg i,g} + \theta_3 b_{\neg i,g} + \epsilon_{i1} > \epsilon_{i2} \qquad (8)$$

Utility

- The probability individual i goes out to protest is then:

$$
\begin{aligned}
Pr(y_i = 1) = Pr\left(V_i(y_i = 1) > V_i(y_i = 0)\right) \\
s_{1g} = Pr\left(\epsilon_{i2} < \theta_0 + \theta_1 b_{ig} + \theta_2 s_{\neg i,g} + \theta_3 b_{\neg i,g} + \epsilon_{i1}\right)
\end{aligned} \tag{9}
$$

Utility

- Brock & Durlauf allow for arbitrary error distribution but to keep things simple, we'll assume that errors are distributed iid Type 1 extreme value

$$\begin{aligned} f(\epsilon_{ij}) &= e^{-\epsilon_{ij}} e^{e^{-\epsilon_{ij}}} \\ F(\epsilon_{ij}) &= e^{e^{-\epsilon_{ij}}} \end{aligned} \tag{10}$$

- This leads to the beautiful Logit form for the probabilities of choosing to protest:

$$Pr(y_i = 1) = \frac{\exp\left(\theta_0 + \theta_1 b_{ig} + \theta_2 s_{\neg i,g} + \theta_3 b_{\neg i,g}\right)}{1 + \exp\left(\theta_0 + \theta_1 b_{ig} + \theta_2 s_{\neg i,g} + \theta_3 b_{\neg i,g}\right)} \tag{11}$$

Utility

- In our simple, two person world, we would then have:

$$
\begin{aligned}
s_{1g} &= \frac{\exp\left(\theta_0 + \theta_1 b_{1g} + \theta_2 s^e_{2g} + \theta_3 b_{2g}\right)}{1 + \exp\left(\theta_0 + \theta_1 b_{1g} + \theta_2 s^e_{2g} + \theta_3 b_{2g}\right)} \\
s_{2g} &= \frac{\exp\left(\theta_0 + \theta_1 b_{2g} + \theta_2 s^e_{1g} + \theta_3 b_{1g}\right)}{1 + \exp\left(\theta_0 + \theta_1 b_{2g} + \theta_2 s^e_{1g} + \theta_3 b_{1g}\right)}
\end{aligned}
\tag{12}
$$

Key Point: Nonlinearity

- The discrete choice framework doesn't allow one to express mean endogenous peer effects as a linear function of exogenous characteristics
- This is different from the linear framework in which outcomes can be solved as a linear function of exogenous characteristics
- This nonlinearity is what permits exogenous and endogenous peer effects to be separately identified

Key Point: Nonlinearity

- Intuitively suppose one moves an individual from one peer group to another and observes the difference in her behaviour
- If the characteristics and behaviour of peers always move in proportion as you move across peer groups, then not surprising that you can't determine the respective roles of characteristics as opposed to group behaviour in determining individual outcomes
- However, with binary choice, the expected group behaviour is bounded between 0 and 1 and thus choice probabilities are inherently nonlinear in controls

Key Point: Nonlinearity

- Proposition 12 in Brock & Durlauf (2001) proves this formally
- Intuitively require the following:
  - Data must have sufficient intra-neighbourhood variation in individual exogenous characteristics for identification of θ_1
  - Require sufficient inter-neighbourhood variation in peer group composition for the separate identification of θ_2 and θ_3

Proposition 12

- We first generalise our framework slightly to also allow for individual characteristics (e.g. age, option choice &c), X, as well as (exogenous) peer effects, b, and also consider a general specification for peer group behaviour, s_g (with the expectation given model parameters, m_g)

$$s_{ig} = \frac{\exp\left(\theta_0 + \theta_1 X_{ig} + \theta_2 s_g + \theta_3 b_g\right)}{1 + \exp\left(\theta_0 + \theta_1 X_{ig} + \theta_2 s_g + \theta_3 b_g\right)} \tag{13}$$

Proposition 12

- Identification: the model is identified if for all parameter pairs $(\theta_0, \theta_1, \theta_2, \theta_3)$ and $\left(\theta_0', \theta_1', \theta_2', \theta_3'\right)$, if

$$\theta_0 + \theta_1 X_{ig} + \theta_2 s_g + \theta_3 b_g = \theta_0' + \theta_1' X_{ig} + \theta_2' s_g + \theta_3' b_g \quad (14)$$

and

$$\begin{aligned} s_g &= m_g \\ &= \int \int y_i dF(y_i|\theta) dF_{X|b_g} \\ &= \int \int y_i dF(y_i|\theta') dF_{X|b_g} \end{aligned} \quad (15)$$

then $(\theta_0, \theta_1, \theta_2, \theta_3) = \left(\theta_0', \theta_1', \theta_2', \theta_3'\right)$

Identification Conditions

Proposition 12. (*Sufficient conditions for identification to hold in the binary choice model with interactions and self-consistent beliefs*). *Assume*;

(i) supp $(\mathbf{X}_i, \mathbf{Y}_{n(i)})$ *is not contained in a proper linear subspace of* R^{r+s}.[9]
(ii) supp $(\mathbf{Y}_{n(i)})$ *is not contained in a proper linear subspace of* R^s.
(iii) *No element of* $\mathbf{X}_i$ *or* $\mathbf{Y}_{n(i)}$ *is constant.*
(iv) *There exists at least one neighbourhood* n_0 *such that conditional on* $\mathbf{Y}_{n_0}$, $\mathbf{X}_i$ *is not contained in a proper linear subspace of* R^r.
(v) *None of the regressors in* $\mathbf{Y}_{n(i)}$ *possesses bounded support.*
(vi) $m_{n(i)}$ *is not constant across all neighbourhoods* n.

Then, (k, c, d, J) *is identified relative to any distinct alternative* $(\bar{k}, \bar{c}, \bar{d}, \bar{J})$.

Proposition 12

- Proof by contradiction, imagine that there does exist an alternative such that on supp(X, Y, s_g), we have

$$(\theta_0 - \theta_0') + (\theta_1 - \theta_1')X_{ig} + (\theta_2 - \theta_2')s_g + (\theta_3 - \theta_3')b_g = 0 \tag{16}$$

with

$$s_g = m_g \tag{17}$$

Proposition 12

- However, it is the case that we can identify θ_1 by condition (iv)
  - Within at least one neighbourhood g, X_i is of full rank
  - There is sufficient variation in X within a peer group to identify θ_1
- Given identification of θ_1, if $\theta_2 \neq \theta_2'$, that s_g is a linear function of b_g unless
  - s_g is always equal to zero: ruled out by assumption (vi)
  - $\theta_3 = \theta_3'$

Proposition 12

- However, linear dependence is ruled out by the differences in the support of s_g and b_g
  - By definition, $s_g \in [0, 1]$ but from assumption (v), the support of each element of b_g is unbounded
  - Thus, b_g can assume values with positive probability that violate the bounds on s_g
  - Thus θ_2 is identified

Proposition 12

- Finally, if θ_1 and θ_2 are identified, it must be the case that:

$$(\theta_3 - \theta_3')b_g = -(\theta_0 - \theta_0') \tag{18}$$

- Given that a constant is excluded from b_g, this would again violate the rank condition on b_g imposed by Assumption 2
- Thus, $\theta_1 = \theta_1'$ and identification of θ_0 follows

Identification by Nonlinearity

- This is a specific example of the wider concept of 'identification by nonlinearity' or, more generally, 'identification by functional form'
- You only get identification by assuming some functions in the model have specific parametric or semi-parametric forms.
- While relatively natural in the discrete choice case we just considered, typically depends on strong modelling assumptions that, if only just identified, are not testable — thus generally thought to be undesirable

Nonlinearity and Local Identification

- That nonlinearities help identification is by no means a general statement!
- The definition of identification laid out in the first lecture is sometimes referred to as **global** identification

$$F_{\theta}^{S_0} = F_{\theta}^{S} \quad \rightarrow \quad S = S_0 \tag{19}$$

for any $S \in \mathcal{M}_{\Gamma}$

- i.e. we can identify the unique S_0 over the entire range of possible values

Nonlinearity and Local Identification

- Local identification is a weaker condirion that this
  - S_0 identified if we restrict attention only to structures in the neighbourhood of the true value
$$F_\theta^{S_0} = F_\theta^{S} \quad \rightarrow \quad S = S_0 \tag{20}$$
  for any S in some abritrarily small open neighbourhood of S_0
- Note that this is not ‘local’ in the sense of LATE

Nonlinearity and Local Identification

- Suppose that $m(X)$ is a known continuous function, where we know that $m(\theta) = 0$
  - 1. Assume that $m(X)$ is strictly monotonic.
    - Then θ (if it exists) is globally identified
    - Strict monotonicity ensures that only one value of θ can satisfy the equation $m(\theta) = 0$.

Nonlinearity and Local Identification

- Suppose that $m(X)$ is a known continuous function, where we know that $m(\theta) = 0$
  - 2. Assume that $m(X)$ is a J^{th} order polynomial
    - Then θ (if it exists) is not typically globally identified
    - Up to J values of θ that can satisfy the equation $m(\theta) = 0$
    - However, we do have local identification — there will exist a neighbourhood of Θ close to θ_0 small enough to exclude all other roots

Nonlinearity and Local Identification

- Suppose that $m(X)$ is a known continuous function, where we know that $m(\theta) = 0$
  - 2. Assume that $m(X)$ is continuous
    - Then θ might not even be locally identified
    - $m(x)$ could equal zero for all values of x in some interval

Nonlinearity and Local Identification

- Suppose that $m(X)$ is a known continuous function, where we know that $m(\theta) = 0$
  - 2. Assume that $m(X)$ is continuous
    - Then θ might not even be locally identified
    - $m(x)$ could equal zero for all values of x in some interval

Conclusion

- Nonlinearities can sometimes aid identification
- However, with nonlinear models might have to weaken identification concept to that of local identification
- To finish, I will run through the technical material that you need to have mastered before the exam