Jan 09, 2019

1 Model of peer effects with an endogenous network

1.1 Algebric

$$y_i = \beta_1^0 \sum_{j=1, j\neq i}^{N} g_{ij,N} y_j + \mathbf{x'}_{1i}\beta_2^0 + \left(\sum_{j=1, j\neq i}^{N} g_{ij,N}\mathbf{x}_{1i} \right) \beta_3^0 + \nu_i \tag{1}$$

Suppose $d_{ij,N}$ are the observed links among individuals $i \in 1, \ldots, N$, $d_{ij,N} = 1$ if i and j are directly connected and 0 otherwise. Where $\mathbf{x'}_{1i}$ are observed individual characteristics that affect the outcome y_i, ν_i are unobserved individual characteristics, and $g_{ij,N} = 0$ if $i = j$, $g_{ij,N} = \frac{d_{ij,N}}{\sum\limits_{j\neq i} d_{ij,N}}$ otherwise which is the weight of the peer effect.

β_1^0 captures the endogenous social effect, and β_3^0 measures the exogenous social effect.

1.2 Matrix

$\mathbf{D}_N$ be the $(N \times N)$ adjacency matrix of the network whose $(i, j)^{th}$ element is $d_{ij,N}$. $\mathbf{G}_N$ be the $(N \times N)$ adjacency matrix of the network whose $(i, j)^{th}$ element is $g_{ij,N}$. $\mathbf{X}_{1N} = (\mathbf{x'}_{11}, \ldots, \mathbf{x'}_{1N})'$, $\mathbf{y}_N = (y_1, \ldots, y_N)'$, $\boldsymbol{\nu}_N = (\nu_1, \ldots, \nu_N)'$

$$\mathbf{y}_N = \beta_1 \mathbf{G}_N \mathbf{y}_N + \mathbf{X}_{1N}\beta_2 + \mathbf{G}_N \mathbf{X}_{1N}\beta_3 + \boldsymbol{\nu}_N \tag{2}$$

$$\mathbf{y}_N = (\mathbf{I}_N - \beta_1 \mathbf{G}_N)^{-1}(\mathbf{X}_{1N}\beta_2 + \mathbf{G}_N \mathbf{X}_{1N}\beta_3 + \boldsymbol{\nu}_N) \tag{3}$$

assuming that the peer group (or the network) is exogenous $E[\nu_i|\mathbf{X}_{1N}, \mathbf{G}_N] = 0$. the fact that the regressor $\sum\limits_{j=1}^{N} g_{ij,N} y_j$ is correlated with the error term ν_i. For example, if $\nu_i \sim i.i.d.(0, \sigma^2)$, it is true that

$$E[(\mathbf{G}_N \mathbf{y}_N)' \boldsymbol{\nu}_N] = E[(\mathbf{G}_N (\mathbf{I}_N - \beta_1 \mathbf{G}_N)^{-1}(\mathbf{X}_{1N}\beta_2 + \mathbf{G}_N \mathbf{X}_{1N}\beta_3 + \boldsymbol{\nu}_N))' \boldsymbol{\nu}_N] \tag{4}$$

$$= E[(\mathbf{G}_N (\mathbf{I}_N - \beta_1 \mathbf{G}_N)^{-1} \boldsymbol{\nu}_N)' \boldsymbol{\nu}_N] = \sigma_0 tr(\mathbf{G}_N (\mathbf{I}_N - \beta_1 \mathbf{G}_N)^{-1}) \neq 0 \tag{5}$$

One of the widely used estimation methods is the Instrumental Variable (IV) approach based on using $\mathbf{G}_N^2\mathbf{X}_{1N}$ as the IV of the endogenous regressor $\mathbf{G}_N\mathbf{y}_N$. Then, the natural estimator is the Two-Stage Least Squares (2SLS) estimator

$$\hat{\beta}_N^{2SLS} = (\mathbf{W}_N'\mathbf{Z}_N(\mathbf{Z}_N'\mathbf{Z}_N)^{-1}\mathbf{Z}_N'\mathbf{W}_N)\mathbf{W}_N'\mathbf{Z}_N(\mathbf{Z}_N'\mathbf{Z}_N)^{-1}\mathbf{Z}_N'\mathbf{y}_N \tag{6}$$

where $\mathbf{W}_N = [\mathbf{G}_N\mathbf{y}_N, \mathbf{X}_{1N}, \mathbf{G}_N\mathbf{X}_{1N}]$ and $\mathbf{Z}_N = [\mathbf{X}_{1N}, \mathbf{G}_N\mathbf{X}_{1N}, \mathbf{G}_N^2\mathbf{X}_{1N}]$ we asuume $\beta_2^0 \neq 0$

When the network matrix is endogenous, $E[\mathbf{G}_N|\boldsymbol{\nu}_N] \neq 0$, the procedure is no longer valid since the IV matrix $\mathbf{Z}_N = [\mathbf{X}_{1N}, \mathbf{G}_N\mathbf{X}_{1N}, \mathbf{G}_N^2\mathbf{X}_{1N}]$ is correlated with the error term $\boldsymbol{\nu}_N$ Specifically, the validity of the 2SLS estimator depends on the orthogonality condition $E[\boldsymbol{\nu}_N|\mathbf{Z}_N] = 0$, which is implied if $E[\boldsymbol{\nu}_N|\mathbf{X}_{1N}, \mathbf{D}_N] = 0$. However, it does not hold if the network $\mathbf{D}_N$ (or equivalently, the network $\mathbf{G}_N$) is correlated with $\boldsymbol{\nu}_N$.

2 Endogenous Network Formation and Identification of peer effects

2.1 Formation

Let $\mathbf{x}_{2i}$ be a vector of observable characteristics of individual i, and let $\mathbf{x}_i = \mathbf{x1}_i \cup \mathbf{x}_{2i}$. Define $\mathbf{X}_{2N}$ analogously to $\mathbf{X}_{1N}$ and let $\mathbf{X}_N = \mathbf{X}_{1N} \cup \mathbf{X2}_N$. We introduce a_i, a scalar unobserved characteristic of individual i, which is treated as an individual fixed-effect, and hence, might be correlated with $\mathbf{x}_i$. We denote the vector of individual unobserved characteristics by $\mathbf{a}_N = (a_1, a_2, \ldots, a_N)'$. Individuals are connected by an undirected network $\mathbf{D}_N$, with the $(i,j)^{th}$ element $d_{ij,N} = 1$ if i and j are directly connected and 0 otherwise. We assume the network to be undirected, $d_{ij,N} = d_{ji,N}$, and assume $d_{ii,N} = 0$ for all i, following the convention. In this case, there are $n = \binom{N}{2}$ dyads. Let $\mathbf{t}_{ij}$ denote an $l_T \times 1$ vector of dyad-specific characteristics of dyad ij, and we assume that $\mathbf{t}_{ij} = t(\mathbf{x}_{2i}, \mathbf{x}_{2j})$.

$$d_{ij,N} = I(g(t(\mathbf{x}_{2i}, \mathbf{x}_{2j}), a_i, a_j) - u_{ij} \geq 0) \tag{7}$$

$$d_{ij,N} = I(t(\mathbf{x}_{2i}, \mathbf{x}_{2j})'\lambda + a_i + a_j - u_{ij} > 0) \tag{8}$$

where $I()$ is an indicator function.