Peer Effects in Endogenous Social Network

$$\hat{\beta}_N^{2SLS} = \left(W_N^T Z_N (Z_N^T Z_N)^{-1} Z_N^T W_N\right)^{-1} W_N^T Z_N \left(Z_N^T Z_N\right) Z_N^T Y_N$$

where $W_N = [G_N Y_N,\ X_{1N},\ G_N X_{1N}]$

$Z_N = [X_{1N},\ G_N X_{1N},\ G_N^2 X_{1N}]$

when $E[G_N v_N] = 0$

$\Rightarrow E[v_N | Z_N] = 0$

$\Rightarrow E[v_N | X_{1N}, D_N] = 0$

unobservable characteristics that influence link formation (a_i) can also have a direct influence both link formation (D_N or G_N) and individual outcomes (Y_N)

Assumption 1

(i) (X_i, a_i, v_i) are iid for all $i = 1, \dots, N$

(ii) $\{u_{ij}\}_{i,j=1,\dots,N}$ are indep of (X_N, a_N, v_N) and i.i.d across (i,j) with cdf $\Phi(\cdot)$

(iii) $E(v_i | X_i, a_i) = E(v_i | a_i)$

(i) impies observable X_i and unobservable a_i, v_i are randomly drawn

(ii) link formation error u_{ij} is orthogonal to all other observables and unobservables in the model

(iii) assume that the dependency between X_i and v_i exists only through a_i

Lemma 1 Control function of peer Group Endogeneity

$$E[v_i \mid X_N, D_N, a_i] = E[v_i \mid a_i]$$

Assumption 2

$E[(z_i - E[z_i \mid a_i])(w_i - E[w_i \mid a_i])^T]$ has full rank

Theorem 3.1

β^0 is identified by moment condition

$$E[(z_i - E[z_i \mid a_i])(y_i - E[y_i \mid a_i] - (w_i - E[w_i \mid a_i])^T\beta] = 0$$

$$\Leftrightarrow \quad \beta = \beta^0$$

Assumption3 $X_{1i} \cap X_{2i} = \phi$

$E[v_i \mid X_N, D_N, a_i] = E[v_i \mid a_i] = E[v_i \mid X_{2i}, a_i]$

Assumption 4

$E[(z_i - E[z_i \mid X_{2i}, a_i])(w_i - E[w_i \mid X_{2i}, a_i])^T]$ has full rank

Theorem 3.2 β^0 is identified by the moment condition

$$E[(z_i - E[z_i \mid X_{2i}, a_i])((y_i - E[y_i \mid X_{2i}, a_i]) - (w_i^T - E[w_i \mid X_{2i}, a_i]))^T\beta] = 0$$

$$\Leftrightarrow \quad \beta = \beta^0$$

So far $X_{1i} \cap X_{2i} = \emptyset$, A more general case is when the regressor X_{1i} consist of two component $X_{1i} = (X_{11i}, X_{12i})$

where X_{11i} does not share any elements with X_{2i}

$X_{12i} \subset X_{2i}$

Let $\beta_2^o = (\beta_{21}^o, \beta_{22}^o)$, $\beta_3^o = (\beta_{31}^o, \beta_{32}^o)$

Properly modified rank condition of $Z_{(2)i}$ and $W_{(2)i}$ that exclude the variables associated with X_{12i}

4 Estimation

4.1 With a_i as control function

$$y_i - E[y_i | a_i] = (w_i - E[w_i | a_i])^T \beta^o + v_i - E[v_i]$$

Let
$$h(a_i) = (h^y(a_i), h^w(a_i), h^z(a_i))$$
$$:= (E[y_i | a_i], E[w_i | a_i], E[z_i | a_i])$$

$\widetilde{W}_N = (w_1 - h^w(a_1), \cdots, w_N - h^w(a_N))^T$, similarly define $\tilde{Z}_N, \tilde{y}_N$

$$\tilde{\beta}^{inf}_{2SLS} = (\widetilde{W}_N^T \tilde{Z}_N (\tilde{Z}_N^T Z_N)^{-1} \tilde{Z}_N^T \widetilde{W}_N)^{-1} \widetilde{W}_N^T \tilde{Z}_N (\tilde{Z}_N^T \tilde{Z}_N)^{-1} \tilde{Z}_N^T \tilde{y}_N$$

$* \; E[G_N v_N] = 0$

$\Rightarrow E[v_N | Z_N] = 0 \Rightarrow E[v_N | X_{1N}, P_N] = 0$

$\Rightarrow$ with a_i as control function

$E[v_i | X_N, P_N, a_i] = E[v_i | a_i]$

as a_i is not observed and the functions $h(\cdot)$ are not known. need to replace $h(a_i)$ in $\widetilde{W}_N, \widetilde{Z}_N$ and $\widetilde{y}_N$, ~~say~~ $\hat{h}(a_i)$ if a_i is observed

Sieve estimation method

$h^\ell(a)$ is ℓ^{th} element in $h(a)$ for $\ell = 1, \dots, L$ L is dimension of $(y_i, w_i^T, z_i^T)^T$

each function $h^\ell(a)$ is well approximated by a linear combination of base functions $(q_1(a), \dots, q_{k_N}(a))$

$$h^\ell(a) \cong \sum_{k=1}^{k_N} q_k(a)\, \alpha_k^\ell$$

Let $q^k(a) = (q_1(a), \dots, q_{k_N}(a))^T$, $Q_N := Q_N(a_N) = (q^k(a_1), \dots, q^k(a_N))^T$

$h^\ell(a_N) = (h^\ell(a_1), \dots, h^\ell(a_N))$, $\alpha_k^\ell = (\alpha_1^\ell, \dots, \alpha_{k_N}^\ell)^T$

b_i^ℓ be the ℓ^{th} element in $(y_i, w_i^T, z_i^T)^T$

$b_N^\ell = (b_1^\ell, \dots, b_N^\ell)$

If $a_N = (a_1, \dots, a_N)^T$ is observed unknown function $h^\ell(a_N)$ by OLS of b_i^ℓ on $q^k(a_i)$ $\ell = 1, \dots, L$

$$\hat{h}^\ell(a_N) = P_{Q_N} b_N^\ell$$

where $P_{Q_N} = Q_N (Q_N^T Q_N)^- Q_N^T$

Suppose $\hat{a}_N = (\hat{a}_1, \dots, \hat{a}_N)$ is an estimator of $a_N = (a_1, \dots, a_N)^T$

denote $\hat{Q}_N := Q_N(\hat{a}_N) = (q^{k_N}(\hat{a}_1), \dots, q^{k_N}(\hat{a}_N))^T$

then $$\hat{h}^\ell := \hat{h}^\ell(\hat{a}_N) = P_{\hat{Q}_N} b_N^\ell$$

$$\hat{\beta}_{2SLS} = (W_N^T M_{\hat{Q}_N} Z_N (Z_N^T M_{\hat{Q}_N} Z_N)^{-1} Z_N^T M_{\hat{Q}_N} W_N)^{-1} \times W_N^T M_{\hat{Q}_N} Z_N (Z_N^T M_{\hat{Q}_N} Z_N)^{-1} Z_N^T M_{\hat{Q}_N} y_N$$

where $M_{\hat{Q}_N} = I_N - P_{\hat{Q}_N}$

Assumption 5 we assume that we can estimate a_i with $\hat{a}_i$, such that $\max_i |\hat{a}_i - a_i| = O_p(\zeta_n(N)^{-1})$, where $\zeta_n(N) \to \infty$ as $N \to \infty$

4.2 with (X_{2i}, a_i) as control function

Let $\widehat{deg}_i$ be the degree of node i scaled by the network size

$$\widehat{deg}_i := \frac{1}{N-1} \sum_{j=1, \neq i}^{N} d_{ij,N}$$

$$\widehat{deg}_i := \frac{1}{N-1} \sum_{j=1, \neq i}^{N} I\left(g(t(X_{2i}, X_{2j}), a_i, a_j) - u_{ij} \geq 0\right)$$

$$\xrightarrow{P} \int \Phi\left(g(t(X_{2i}, X_{2j}), a_i, a_j\right) \pi(X_2, a) \, dX_2 \, da$$

$$=: deg(X_{2i}, a_i)$$

$$=: deg_i$$

(X_{2i}, a_i) and (X_{2i}, deg_i) have an one-to-one relation

$$h_*^{\ell}(X_{2i}, a_i) := E[b_i^{\ell} | X_{2i}, a_i] = E[b_i^{\ell} | X_2, deg_i] =: h_{**}^{\ell}(X_{2i}, deg_i)$$

$h_{**}^{\ell}(X_{2i}, deg_i)$, $\ell = 1, \ldots, L$ is well approximated by a linear combination of base function $(r_1(X_2, deg_i), \ldots, r_K(X_2, deg_i))$

$$h_{**}^{\ell}(X_{2i}, deg_i) \cong \sum_{k=1}^{K_N} r_k(X_2, deg_i) \gamma_k^{\ell}$$

Let $Deg_N = (deg_1, \ldots, deg_N)^T$, $r^k(X_{2i}, deg_i) = (r_1(X_{2i}, deg_i), \ldots, r_K(X_{2i}, deg_i))^T$

$$R_N := R_N(X_{2N}, Deg_N) = (r^k(X_{2i}, deg_i), \ldots, r^k(X_{2N}, deg_N))^T$$

$$\gamma^{\ell} = (\gamma_1^{\ell}, \ldots, \gamma_{K_N}^{\ell}), \quad b_N^{\ell} = (b_1^{\ell}, \ldots, b_N^{\ell})$$

where (X_{2i}, deg_i) are observed we can estimate

$$h_{**}^{\ell}(X_{2N}, Deg_N) = (h_{**}^{\ell}(X_{21}, deg_1), \ldots, h_{**}^{\ell}(X_{2N}, deg_N)) \quad \text{for } \ell = 1, \ldots, L$$

$$\widehat{h}^{\ell}_{**}(X_{2N}, Deg_N) := P_{R_N} b_N^{\ell}$$

where $P_{R_N} = R_N (R_N^T R_N)^- R_N^T$

natural estimator of deg_i is $\widehat{deg}_i$, we estimate $h^{\ell}_{**}(X_{2i}, deg_i)$ by using $\widehat{deg}_i$ in place of deg_i

Suppose that $\widehat{Deg}_N = (\widehat{deg}_1, \cdots, \widehat{deg}_N)$, Denote $\widehat{R}_N := R_N(X_{2N}, \widehat{Deg}_N) = (r^k(X_{21}, \widehat{deg}_1), \cdots, r^k(X_{2N}, \widehat{deg}_N))^T$

estimator of $h^{\ell}_{*}(X_{2i}, a_i) = h^{\ell}_{**}(X_{2i}, deg_i)$ is defined by ith element of

$$\widehat{h}^{\ell}_{*}(X_{2N}, a_N) := h^{\ell}_{**}(X_{2N}, \widehat{Deg}_N) = P_{\widehat{R}_N} b_N^{\ell}$$

$$\bar{B}_{2SLS} := (W_N^T M_{\widehat{R}_N} Z_N (Z_N^T M_{\widehat{R}_N} Z_N)^{-1} Z_N^T M_{\widehat{R}_N} W_N)^{-1}$$
$$\times W_N^T M_{\widehat{R}_N} Z_N (Z_N^T M_{\widehat{R}_N} Z_N)^{-1} Z_N^T M_{\widehat{R}_N} y_N$$

where $M_{\widehat{R}_N} = I_N - P_{\widehat{R}_N}$

5 Limit Distribution And Standard Error

5.1 $\widehat{\beta}_{2SLS}$

define $\eta_i^y := y_i - h^y(a_i)$, $\eta_i^r := r_i - h^r(a_i)$, $\eta_i^w := w_i - h^w(a_i)$, $\eta_i^z = z_i - h^z(a_i)$

Let $\eta_N^r = (\eta_1^r, \cdots, \eta_N^r)$ and $H_N^r(a_N) = (h^r(a_1), \cdots, h^r(a_N))^T$

Let $\hat{h}^r(a_i)$, $\hat{h}^w(a_i)$ and $\hat{h}^z(a_i)$ denote the sieve estimators of $h^r(a_i), h^w(a_i), h^z(a_i)$

First show that the sampling error caused by the use of $\hat{a}_N$ instead of a_N is asymptotically negligible

Next, we control the error introduced by the non parametric estimation of $h^\ell(a_i)$, where $\ell \in \{r, w, z\}$

under the regularity condtions, the estimation error in $\hat{h}^\ell(a_i)$ vanishes at a suitable rate

Combining these two, we deduce

$$\sqrt{N}(\hat{\beta}_{2SLS} - \hat{\beta}_{2SLS}^{inf}) = o_p(1)$$

Last step is to derive the limiting distribution of the infeasible estimator

$$\sqrt{N}(\hat{\beta}_{2SLS}^{inf} - \beta^0)$$

$$\frac{1}{N}\sum_{i=1}^{N}(w_i - h^w(a_i))(z_i - h^z(a_i))^T \xrightarrow{p} S^{wz}$$

$$\frac{1}{N}\sum_{i=1}^{N}(z_i - h^z(a_i))(z_i - h^z(a_i))^T \xrightarrow{p} S^{zz}$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(z_i - h^z(a_i))\eta_i^r \Rightarrow N(0, S^{zz\sigma})$$

$$\sqrt{N}(\hat{\beta}_{2SLS} - \beta^{\circ}) \Rightarrow N(0, \Omega)$$

where

$$\Omega = (S^{wz}(S^{zz})^{-1}(S^{wz})^{T})^{-1}(S^{wz}(S^{zz})^{-1} S^{zz\delta}(S^{zz})^{-1}(S^{wz})^{T})(S^{wz}(S^{zz})^{-1}(S^{wz})^{T})^{-1}$$

The asymptotic variance can be consistently estimated by

$$\hat{\Omega} = (\hat{S}^{wz}(\hat{S}^{zz})^{-1}(\hat{S}^{wz})^{T})^{-1}(\hat{S}^{wz}(\hat{S}^{zz})^{-1} \hat{S}^{zz\delta}(\hat{S}^{zz})^{-1}(\hat{S}^{wz})^{T})(\hat{S}^{wz}(\hat{S}^{zz})^{-1}(\hat{S}^{wz})^{T})^{-1}$$

$$\hat{S}^{wz} = \frac{1}{N}\sum_{i=1}^{N}(w_i - \hat{h}^{w}(a_i))(z_i - \hat{h}^{z}(a_i))^{T}$$

$$\hat{S}^{zz} = \frac{1}{N}\sum_{i=1}^{N}(z_i - \hat{h}^{z}(a_i))(z_i - \hat{h}^{z}(a_i))^{T}$$

$$\hat{S}^{zz\delta^2} = \frac{1}{N}\sum_{i=1}^{N}(z_i - \hat{h}^{z}(a_i))(z_i - \hat{h}^{z}(a_i))^{T}(\hat{\eta}_i^{v})^2$$

$$\hat{\eta}_i^{v} = y_i - \hat{h}^{y}(a_i) - (w_i - \hat{h}^{w}(a_i))^{T}\hat{\beta}_{2SLS}$$