Social Interactions with Endogeneous Group Formation
by Shuyang Sheng and Xiaoting Sun

Angelo Mele
Johns Hopkins University - Carey Business School

Interactions 2023 - Northwestern University

September 22-23, 2023

Model

- Classical linear-in-means model

$$y_i = \sum_{j=1}^{n} w_{ij} y_j \gamma_1 + \sum_{j=1}^{n} w_{ij} x_j' \gamma_2 + x_i' \gamma_3 + \epsilon_i \tag{1}$$

- workhorse in the literature on peer effects
- $\boldsymbol{w}$: exogenous or endogenous
- $\boldsymbol{w}$: group structure observed or unobserved

This paper

Groups **observed** and **endogenously** formed via two-sided many-to-one matching without transfers

- COMMENT: **groups do not interact**, $w_{ij} = 0$ if $g_i \neq g_j$ how restrictive? network formation with block structure?
- COMMENT: individuals join only one group (restrictive?)

Group formation

Two-sided many-to-one matching w/o transfers (Azevedo and Leshno (2016))

$$u_{ig} = z_i'\delta_g^u + \xi_{ig} \quad \text{and} \quad v_{gi} = z_i'\delta_g^v + \eta_{gi} \tag{2}$$

COMMENT: No peer effects in group formation

EQUILIBRIUM: Group cutoffs $p = (p_1, ..., p_G)$

$$\mathbf{1}\{g_i = g\} =$$

$$\mathbf{1}\{\underbrace{\eta_{gi} \geq p_g - z_i'\delta_g^v}_{i \text{ qualifies for } g}\} \prod_{h \neq g} \mathbf{1}\{\underbrace{\xi_{ih} - \xi_{ig} < z_i'(\delta_g^u - \delta_h^u)}_{i \text{ prefers } g \text{ to } h} \quad \text{OR} \quad \underbrace{\eta_{hi} < p_h - z_i'\delta_h^v}_{i \text{ doesn't qualify for } h}\}$$

Essentially a multinomial choice problem with endogenous choice set

Equlibrium p clears demand and supply

Selection Bias

We assume a completely nonparametric distribution $f(\epsilon_i, \xi_i, \eta_i)$

$$\mathbb{E}\left[\epsilon_i | \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{g}\left(\boldsymbol{z}, \boldsymbol{\xi}, \boldsymbol{\eta}; p(\boldsymbol{z}, \boldsymbol{\xi}, \boldsymbol{\eta})\right)\right] \neq 0 \tag{3}$$

1. Because ϵ_i is allowed to be correlated with unobservables ξ_i and η_i
2. Because ϵ_i is correlated with $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ since $p(\boldsymbol{z}, \boldsymbol{\xi}, \boldsymbol{\eta})$

This is a very high-dimensional object

Approximation

The equilibrium cutoffs for a fixed n are

$$p_n = (p_{n,1}, ..., p_{n,G}) \tag{4}$$

Azevedo and Leshno (2016) show that as $n \longrightarrow \infty$

$$p_n \longrightarrow p^* = (p_1^*, ..., p_G^*) \tag{5}$$

Note: p^* is **unique** and **deterministic**
By continuous mapping we get

$$\mathbb{E}\left[\epsilon_i | \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{g}\left(\boldsymbol{z}, \boldsymbol{\xi}, \boldsymbol{\eta}; p_n\right)\right] \longrightarrow \mathbb{E}\left[\epsilon_i | x_i, z_i, g_i\left(z_i, \xi_i, \eta_i; p^*\right)\right] \tag{6}$$

which is now lower dimension

Selection function

There exist a selection function (a la Heckman)

$$\mathbb{E}\left[\epsilon_i | x_i, z_i, g_i = g\right] = \lambda_g(\tau_i) \tag{7}$$

where $\tau_i = (z_i'\delta_1^u, ..., z_i'\delta_G^u, z_i'\delta_1^v, ..., z_i'\delta_G^v)$

Note: The bias correction function $\lambda_g(\tau_i)$ is **group-specific**

$$y_i = \boldsymbol{w}_i\boldsymbol{y}\gamma_1 + \boldsymbol{w}_i\boldsymbol{x}\gamma_2 + \boldsymbol{x}_i'\gamma_3 + \lambda_{g_i}(\tau_i) + \nu_i \tag{8}$$

Identification problem?: if $w_{ij} = 1/n_{g_i}$,

$$\boldsymbol{w}_i\boldsymbol{y} = \sum_{j=1}^{n} w_{ij}y_j \quad \text{and} \quad \boldsymbol{w}_i\boldsymbol{x} = \sum_{j=1}^{n} w_{ij}x_j \tag{9}$$

are also **group-specific** averages

Exchangeability

So assume **fixed effect** in the preferences

$$u_{ig} = \alpha_g + z_i' \delta_g^u + \xi_{ig} \quad \text{and} \quad v_{gi} = z_i' \delta_g^v + \eta_{gi} \tag{10}$$

and define extended indices $\tau_{ig}^e = \left(\alpha_g + z_i' \delta_g^u, p_g - z_i' \delta_g^v\right)$.

Assume **exchangeability** for unobservables ξ_i and η_i

$$f\left(\epsilon_i, \xi_{i1}, ...\xi_{iG}, \eta_{1i}, ..., \eta_{Gi}\right) = f\left(\epsilon_i, \xi_{ik_1}, ...\xi_{ik_G}, \eta_{k_1 i}, ..., \eta_{k_G i}\right) \tag{11}$$

for any permutation $(k_1, ..., k_G$ of the groups.

Then the selection bias is **group-invariant**

$$\mathbb{E}\left[\epsilon_i | x_i, z_i, g_i = g\right] = \lambda^e(\tau_{ig}^e, \tau_{i,-g}^e) \tag{12}$$

IDENTIFICATION p^*

He et al 2022 show how to identify the parameters $\boldsymbol{\delta}$

To identify p^* (and $\boldsymbol{\alpha}$) compute conditional probabilities

$$\begin{aligned} \sigma_{h|g}\left(p_g - z_i'\delta_g^v\right) &= P\left(g_i = h|p_g - z_i'\delta_g^v\right) = P\left(g_i = h|z_i'\delta_g^v\right) && (13) \\ \sigma_{h|1}\left(p_1 - z_j'\delta_1^v\right) &= P\left(g_j = h|p_1 - z_j'\delta_1^v\right) = P\left(g_j = h|z_j'\delta_1^v\right) && (14) \end{aligned}$$

because of **exchangeability**

$$\sigma_{h|g}\left(\cdot\right) = \sigma_{h|1}\left(\cdot\right) = \sigma_h\left(\cdot\right) \tag{15}$$

Assuming $\sigma_h\left(\cdot\right)$ **strictly monotone**, if you can find i and j such that

$$\sigma_{h|g}\left(p_g - z_i'\delta_g^v\right) = \sigma_{h|1}\left(p_1 - z_j'\delta_1^v\right) \Rightarrow p_g - z_i'\delta_g^v = p_1 - z_j'\delta_1^v \tag{16}$$

Then identification of p_g follow from **normalization** $p_1 = 0$.

IDENTIFICATION γ

$$y_i = \boldsymbol{w}_i\boldsymbol{y}\gamma_1 + \boldsymbol{w}_i\boldsymbol{x}\gamma_2 + \boldsymbol{x}_i'\gamma_3 + \lambda^e(\tau^e_{ig}, \tau^e_{i,-g}) + \nu_i \tag{17}$$

Take $\mathbb{E}(\cdot|\tau^e_{ig}, \tau^e_{i,-g})$ and subtract to obtain

$$y_i = \widetilde{\boldsymbol{w}_i\boldsymbol{y}}\gamma_1 + \widetilde{\boldsymbol{w}_i\boldsymbol{x}}\gamma_2 + \widetilde{\boldsymbol{x}}_i'\gamma_3 + \lambda^e(\tau^e_{ig}, \tau^e_{i,-g}) + \nu_i \tag{18}$$

IDEA: Use z_i as instrument for $\boldsymbol{w}_i\boldsymbol{y}$

$$X_i = \left(\boldsymbol{w}_i\boldsymbol{y}, \boldsymbol{w}_i\boldsymbol{x}, \boldsymbol{x}_i'\right) \quad \text{and} \quad Z_i = \left(z_i, \boldsymbol{w}_i\boldsymbol{x}, \boldsymbol{x}_i'\right) \tag{19}$$

and you can build moment like this

$$\mathbb{E}\left[Z_i\left(\widetilde{y}_i - \widetilde{X}_i'\gamma\right)\right] = 0 \tag{20}$$

So identification follows from rank conditions on $\mathbb{E}\left[Z_iZ_i'\right]$ and $\mathbb{E}\left[Z_i\widetilde{X}_i'\right]$

Summary

- Great paper
- Some clever tricks and methodology
- Needs a killer application to become **memorable**