A Semiparametric Network Formation Model with Multiple Linear Fixed Effects

Luis E. Candelaria
University of Warwick

January 7, 2018

Overview

This paper is the first to analyze a static network formation model with two main features:

Overview

This paper is the first to analyze a static network formation model with two main features:

1. *Multiple and unobserved agent-specific components*.

Overview

This paper is the first to analyze a static network formation model with two main features:

1. *Multiple and unobserved agent-specific components*.
2. *Semiparametric approach*.

Overview

This paper is the first to analyze a static network formation model with two main features:

1. *Multiple and unobserved agent-specific components*.
2. *Semiparametric approach*.

Motivation: Friendships network.

- Homophily and unobserved agent-specific heterogeneity.

Overview

This paper is the first to analyze a static network formation model with two main features:

1. *Multiple and unobserved agent-specific components*.
2. *Semiparametric approach*.

Motivation: Friendships network.

- Homophily and unobserved agent-specific heterogeneity.

This paper:

- One large network is observed.

Overview

This paper is the first to analyze a static network formation model with two main features:

1. *Multiple and unobserved agent-specific components*.
2. *Semiparametric approach*.

Motivation: Friendships network.

- Homophily and unobserved agent-specific heterogeneity.

This paper:

- One large network is observed.

- Unrestricted dependence.

Overview

This paper is the first to analyze a static network formation model with two main features:

1. *Multiple and unobserved agent-specific components*.
2. *Semiparametric approach*.

Motivation: Friendships network.

- Homophily and unobserved agent-specific heterogeneity.

This paper:

- One large network is observed.
- Unrestricted dependence.
- No distributional assumptions on the unobserved components.

Introduction

Network formation models study the creation of relationships.

- e.g. friendships, partnerships, scientific collaborations.

Introduction

Network formation models study the creation of relationships.

- e.g. friendships, partnerships, scientific collaborations.

Why are they important?

1. **Peer effects:** network endogeneity.
    - Goldsmith-Pinkham and Imbens 2013.
2. **Policy:** social programs.
    - Banerjee et al. 2013.
3. **Social meaning:** homophily.
    - McPherson et al. 2001.

Introduction

Network formation with unobserved agent-specific attributes.

Introduction

Network formation with unobserved agent-specific attributes.

Challenges:

- Arbitrarily correlation with the observed attributes.
- Semiparametric framework: identification?

Introduction

Network formation with unobserved agent-specific attributes.

Challenges:

- Arbitrarily correlation with the observed attributes.
- Semiparametric framework: identification?

Implications:

- Biased and inconsistent results if these attributes are omitted.

Friendship network

Definition (Network)

A network is an ordered pair, $(\mathcal{N}_n, D^n)$, where $\mathcal{N}_n = \{1, \cdots, n\}$ is a set of nodes and $D^n = (D^n_{ij})$ is a $n \times n$ adjacency matrix.

Friendship network

Definition (Network)

A network is an ordered pair, $(\mathcal{N}_n, D^n)$, where $\mathcal{N}_n = \{1, \cdots, n\}$ is a set of nodes and $D^n = (D^n_{ij})$ is a $n \times n$ adjacency matrix.

Assume the network is

- **Undirected:** $D^n_{ij} = D^n_{ji}$ for any $i, j \in \mathcal{N}_n$.
- **Unweighted:** $D^n_{ij} \in \{0, 1\}$ for any $i, j \in \mathcal{N}_n$.

Normalize $D^n_{ii} = 0$ for any $i \in \mathcal{N}_n$.

Example: Friendship network

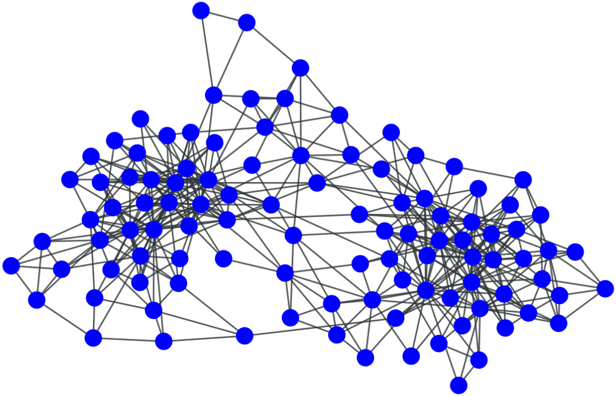


Figure: Undirected Network

Example: Friendship network

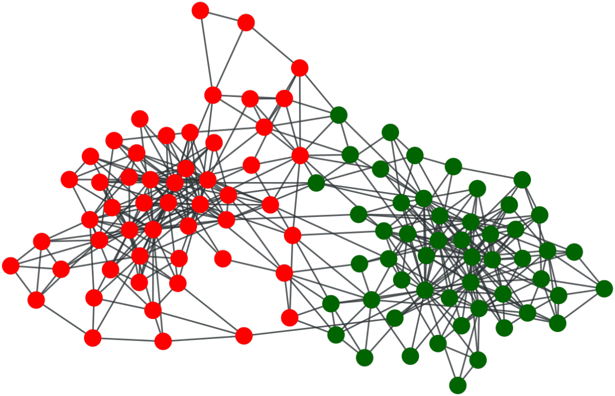


Figure: Homophily on Observed Characteristics

Example: Friendship network

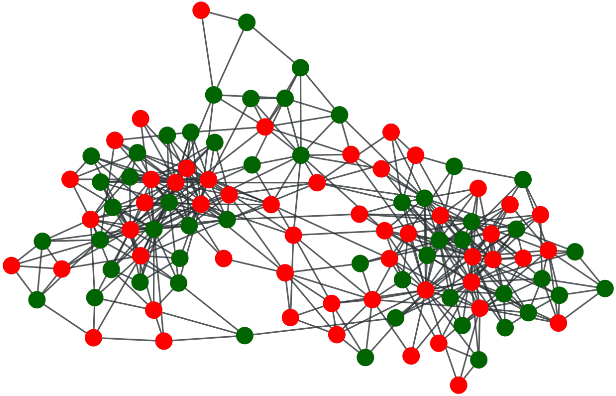


Figure: Unobserved Agent-Specific Heterogeneity

Model of Interest

Agents $i, j \in \mathcal{N}_n$ form an undirected link according to the equation:

$$D_{ij}^n = \mathbf{1}\left[X_{ij}^{n'}\beta_0 + \mu_i + \mu_j - \varepsilon_{ij}^n \geq 0\right], \tag{NF}$$

for $i \neq j$.

Model of Interest

Agents $i, j \in \mathcal{N}_n$ form an undirected link according to the equation:

$$D_{ij}^n = \mathbf{1}\left[X_{ij}^{n\prime}\beta_0 + \mu_i + \mu_j - \varepsilon_{ij}^n \geq 0\right], \qquad \text{(NF)}$$

for $i \neq j$.

- $X_{ij}^{n\prime}\beta_0$: systematic part of the net benefit.

Model of Interest

Agents $i, j \in \mathcal{N}_n$ form an undirected link according to the equation:

$$D_{ij}^n = \mathbf{1}\left[X_{ij}^{n\prime}\beta_0 + \mu_i + \mu_j - \varepsilon_{ij}^n \geq 0\right], \tag{NF}$$

for $i \neq j$.

- $X_{ij}^{n\prime}\beta_0$: systematic part of the net benefit.
- μ_i, μ_j: unobserved agent-specific factors.

Model of Interest

Agents $i, j \in \mathcal{N}_n$ form an undirected link according to the equation:

$$D_{ij}^n = \mathbf{1}\left[X_{ij}^{n\prime}\beta_0 + \mu_i + \mu_j - \varepsilon_{ij}^n \geq 0\right], \tag{NF}$$

for $i \neq j$.

- $X_{ij}^{n\prime}\beta_0$: systematic part of the net benefit.
- μ_i, μ_j: unobserved agent-specific factors.
- ε_{ij}^n: pair-specific exogenous factor.

Model of Interest

Agents $i, j \in \mathcal{N}_n$ form an undirected link according to the equation:

$$D_{ij}^n = \mathbf{1}\left[X_{ij}^{n\prime}\beta_0 + \mu_i + \mu_j - \varepsilon_{ij}^n \geq 0\right], \tag{NF}$$

for $i \neq j$.

- $X_{ij}^{n\prime}\beta_0$: systematic part of the net benefit.
- μ_i, μ_j: unobserved agent-specific factors.
- ε_{ij}^n: pair-specific exogenous factor.
- $\beta_0 \in \mathbb{R}^K$: unknown parameter.

Overview

This paper is the first to analyze a static network formation model

$$D_{ij}^n = \mathbf{1}\left[X_{ij}^{n\prime}\beta_0 + \mu_i + \mu_j - \varepsilon_{ij}^n \geq 0\right].$$

with two main features:

1. Multiple and unobserved agent-specific fixed effects:

$$\mu_i + \mu_j.$$

2. Semiparametric approach:

$$F_{\varepsilon_{ij}^n|\mathbf{x},\mu} \quad \text{and} \quad F_{\mu|\mathbf{x}} \quad \text{are unrestricted.}$$

Objective: Identification and estimation of β_0.

Main Results

1. New identification strategy.

    - Point identification of β_0.

    - Identified set and bounds on each element of β_0.

Main Results

1. New identification strategy.
   - Point identification of β_0.
   - Identified set and bounds on each element of β_0.
2. Semiparametric pairwise estimator.
   - Computationally tractable.
   - Asymptotics: growing number of agents.

Main Results

1. New identification strategy.
   - Point identification of β_0.
   - Identified set and bounds on each element of β_0.
2. Semiparametric pairwise estimator.
   - Computationally tractable.
   - Asymptotics: growing number of agents.
3. Empirical application.
   - Friendship network: Add Health dataset.
   - Evidence for homophily on age, Hispanic, and father's education.

Literature Review

I. Network Formation.

- **Observed Heterogeneity:** Brock and Durlauf (2005), Christakis, Fowler, Imbens, and Kalyanaraman (2010), Sheng (2012), Boucher and Mourifié (2013), Chandrasekhar and Jackson (2014), Souza (2014), Leung (2015a,b, 2016), Menzel (2015), Ridder and Sheng (2015), Hsieh and Lee (2016), de Paula, Richards-Shubik, and Tamer (2017), and Mele (2017).
- **Unobserved Heterogeneity:** Goldsmith-Pinkham and Imbens (2013), Charbonneau (2014), Auerbach (2016), Dzemski (2017), Graham (2017) and Jochmans (2017).

II. Semiparametric Methods.

- **Identification:** Andersen (1973), Manski (1985, 1987) and Vytlacil and Yildiz (2007).
- **Maximum Rank:** Han (1987) and Abrevaya (1999).
- **Inference:** Andrews and Schafgans (1998), Newey (1990), Chamberlain (2010), Khan and Tamer (2010) and Khan and Nekipelov (2015).

Outline

1. Network Formation Model

2. Identification

3. Inference

4. Simulations

5. Application

6. Conclusions and Extensions

Outline

1. Network Formation Model

2. Identification

3. Inference

4. Simulations

5. Application

6. Conclusions and Extensions

Model - Framework

1. Triangular array of random networks

$$\{(\mathcal{N}_n, D^n) : n \in \mathbb{N}\}.$$

Model - Framework

1. Triangular array of random networks

$$\{(\mathcal{N}_n, D^n) : n \in \mathbb{N}\}.$$

2. A dyad is a pair (i, j) of agents with $i, j \in \mathcal{N}_n$ and $i \neq j$.

Model - Framework

1. Triangular array of random networks

$$\{(\mathcal{N}_n, D^n) : n \in \mathbb{N}\}.$$

2. A dyad is a pair (i, j) of agents with $i, j \in \mathcal{N}_n$ and $i \neq j$.

   - Unique dyads: $\mathcal{N}_n^{(2)} \equiv \{(1,2), (1,3), \cdots\cdots, (n-1, n)\}$.

Model - Framework

1. Triangular array of random networks

$$\{(\mathcal{N}_n, D^n) : n \in \mathbb{N}\}.$$

2. A dyad is a pair (i, j) of agents with $i, j \in \mathcal{N}_n$ and $i \neq j$.

   - Unique dyads: $\mathcal{N}_n^{(2)} \equiv \{(1,2), (1,3), \cdots\cdots, (n-1, n)\}$.
   - Cardinality: $N \equiv \#\mathcal{N}_n^{(2)} = O(n^2)$.

Model - Framework

1. Triangular array of random networks

$$\{(\mathcal{N}_n, D^n) : n \in \mathbb{N}\}.$$

2. A dyad is a pair (i, j) of agents with $i, j \in \mathcal{N}_n$ and $i \neq j$.

   - Unique dyads: $\mathcal{N}_n^{(2)} \equiv \{(1,2), (1,3), \cdots\cdots, (n-1, n)\}$.

   - Cardinality: $N \equiv \#\mathcal{N}_n^{(2)} = O(n^2)$.

   - Each $(i, j) \in \mathcal{N}_n^{(2)}$ is endowed with X_{ij}^n, and let

$$\mathbf{X}^n \equiv \left(X_{12}^n, \cdots, X_{n-1,n}^n\right).$$

Model - Preferences

Agent i's latent marginal benefit of adding the link $\{ij\}$ to D^n is

$$V_{ij}(\mathbf{X}^n, \eta_{ij}; \beta_0) = u_{ij}(\mathbf{X}^n; \beta_0) + \eta_{ij}.$$

Model - Preferences

Agent i's latent marginal benefit of adding the link $\{ij\}$ to D^n is

$$V_{ij}(\mathbf{X}^n, \eta_{ij}; \beta_0) = u_{ij}(\mathbf{X}^n; \beta_0) + \eta_{ij}.$$

- $u_{ij}(\mathbf{X}^n; \beta_0)$ denotes the systematic part:

$$u_{ij}(\mathbf{X}^n; \beta_0) \equiv \frac{1}{2} X_{ij}^{\prime} \beta_0.$$

Model - Preferences

Agent i's latent marginal benefit of adding the link $\{ij\}$ to D^n is

$$V_{ij}(\mathbf{X}^n, \eta_{ij}; \beta_0) = u_{ij}(\mathbf{X}^n; \beta_0) + \eta_{ij}.$$

- $u_{ij}(\mathbf{X}^n; \beta_0)$ denotes the systematic part:

$$u_{ij}(\mathbf{X}^n; \beta_0) \equiv \frac{1}{2} X_{ij}^{'} \beta_0.$$

- η_{ij} denotes the unobserved valuation component:

$$\eta_{ij} \equiv \mu_j - \frac{1}{2} \varepsilon_{ij}.$$

Model - Preferences

Agent i's latent marginal benefit of adding the link $\{ij\}$ to D^n is

$$V_{ij}(\mathbf{X}^n, \eta_{ij}; \beta_0) = u_{ij}(\mathbf{X}^n; \beta_0) + \eta_{ij}.$$

- $u_{ij}(\mathbf{X}^n; \beta_0)$ denotes the systematic part:

$$u_{ij}(\mathbf{X}^n; \beta_0) \equiv \frac{1}{2} X_{ij}^{'} \beta_0.$$

- η_{ij} denotes the unobserved valuation component:

$$\eta_{ij} \equiv \mu_j - \frac{1}{2} \varepsilon_{ij}.$$

Remarks:

Rules out:

- Network externalities: $u_{ij}(\mathbf{X}^n, D^n; \beta_0)$.
- Unobserved complementarity: $g(\mu_i, \mu_j)$ as in Candelaria (2016).

Model - Stability Condition

A network D^n is stable with transfers if for each $(i, j) \in \mathcal{N}_n^{(2)}$:

1. for all $D_{ij} = 1$, $V_{ij}(\mathbf{X}^n, \eta_{ij}; \beta_0) + V_{ji}(\mathbf{X}^n, \eta_{ji}; \beta_0) \geq 0$;
2. for all $D_{ij} = 0$, $V_{ij}(\mathbf{X}^n, \eta_{ij}; \beta_0) + V_{ji}(\mathbf{X}^n, \eta_{ji}; \beta_0) < 0$.

Model - Stability Condition

A network D^n is stable with transfers if for each $(i, j) \in \mathcal{N}_n^{(2)}$:

1. for all $D_{ij} = 1$, $V_{ij}(\mathbf{X}^n, \eta_{ij}; \beta_0) + V_{ji}(\mathbf{X}^n, \eta_{ji}; \beta_0) \geq 0$;
2. for all $D_{ij} = 0$, $V_{ij}(\mathbf{X}^n, \eta_{ij}; \beta_0) + V_{ji}(\mathbf{X}^n, \eta_{ji}; \beta_0) < 0$.

Equivalently, the network D^n is stable with transfers if:

$$D_{ij} = \mathbf{1}\left[X_{ij}'\beta_0 + \mu_i + \mu_j - \varepsilon_{ij} \geq 0\right], \quad \forall (i, j) \in \mathcal{N}_n^{(2)}. \qquad \text{(NF)}$$

Model-Assumptions

Assumption (A1)

The following hold for any n.

1. *For any distinct* $(i,k),(j,l) \in \mathcal{N}_n^{(2)}$*:*

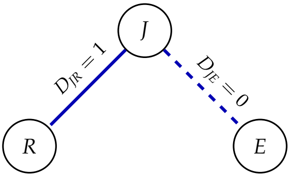
$$\varepsilon_{ik} \perp\!\!\!\perp \varepsilon_{jl} \mid \mathbf{X^n} = \mathbf{x}, \mu^n = \mu, \text{ and } F_{\varepsilon_{ik}|\mathbf{x},\mu} = F_{\varepsilon_{jl}|\mathbf{x},\mu}.$$

2. *The pdf* $f_{\varepsilon_{i1}|\mathbf{x},\mu}$ *is positive everywhere for all* $(\mathbf{x}, \mu)$*.*

- A1 used in Graham (2017) and Menzel (2015).
- Agnostic about $F_{\varepsilon_{i1}|\mathbf{x},\mu}$ and $F_{\mu|\mathbf{x}}$.

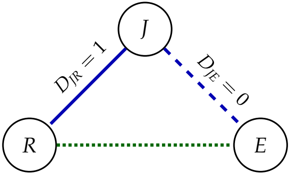
Identification Strategy

Consider the subnetwork given by $J, R, E \in \mathcal{N}_n$.



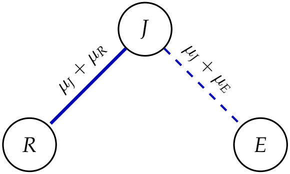
Identification Strategy

Consider the subnetwork given by $J, R, E \in \mathcal{N}_n$.



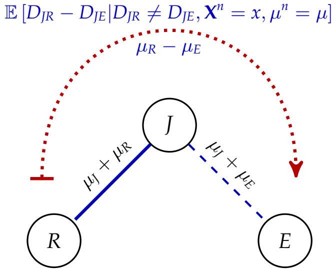
Identification Strategy

Conditional on $\{\mathbf{X}^n = \mathbf{x}, \mu^n = \mu\}$:



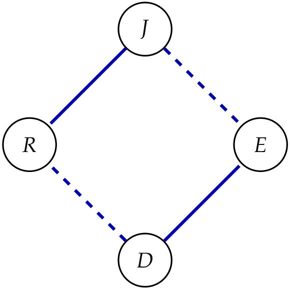
Identification Strategy

Conditional on $\{\mathbf{X}^n = \mathbf{x}, \mu^n = \mu\}$:



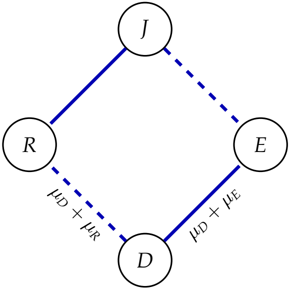
Identification Strategy

Consider the tetrad given by $\{J, R, E, D\}$.



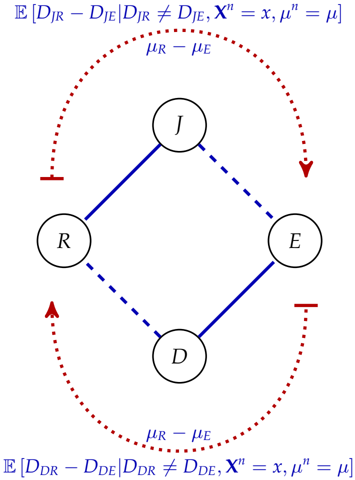
Identification Strategy

Conditional on $\{\mathbf{X}^n = \mathbf{x}, \mu^n = \mu\}$:



Identification Strategy

Conditional on $\{\mathbf{X}^n = \mathbf{x}, \mu^n = \mu\}$:



Assumptions

Let

$$\Delta_{kl} X_i \equiv X_{ik} - X_{il} \text{ for any distinct } (i,k),(i,l) \in \mathcal{N}_n^{(2)};$$

Assumptions

Let

$$\Delta_{kl}X_i \equiv X_{ik} - X_{il} \;\text{ for any distinct } (i,k),(i,l) \in \mathcal{N}_n^{(2)};$$
$$\Delta_{kl}X_i = (\Delta_{kl}X_i^{(1)}, \Delta_{kl}X_i^{(-1)}).$$

Assumptions

Let

$$\Delta_{kl} X_i \equiv X_{ik} - X_{il} \text{ for any distinct } (i,k),(i,l) \in \mathcal{N}_n^{(2)};$$

$$\Delta_{kl} X_i = (\Delta_{kl} X_i^{(1)}, \Delta_{kl} X_i^{(-1)}).$$

Assumption (A2)

The following hold for any n, and any distinct $(i,k),(i,l) \in \mathcal{N}_n^{(2)}$.

1. $\Delta_{kl} X_i$ *is not contained in a proper subspace of* $\mathbb{R}^K$.

Assumptions

Let

$$\Delta_{kl} X_i \equiv X_{ik} - X_{il} \text{ for any distinct } (i,k),(i,l) \in \mathcal{N}_n^{(2)};$$
$$\Delta_{kl} X_i = (\Delta_{kl} X_i^{(1)}, \Delta_{kl} X_i^{(-1)}).$$

Assumption (A2)

The following hold for any n, and any distinct $(i,k),(i,l) \in \mathcal{N}_n^{(2)}$.

1. $\Delta_{kl} X_i$ *is not contained in a proper subspace of* $\mathbb{R}^K$.
2. *Exists* $\Delta_{kl} X_i^{(1)}$ *with* $\beta_0^{(1)} \neq 0$ *s.t. the cond. density of* $\Delta_{kl} X_i^{(1)}$ *is positive everywhere for any* $\Delta_{kl} x_i^{(-1)}$.

Assumptions

Let

$$\Delta_{kl}X_i \equiv X_{ik} - X_{il} \;\; \text{for any distinct } (i,k),(i,l) \in \mathcal{N}_n^{(2)};$$

$$\Delta_{kl}X_i = (\Delta_{kl}X_i^{(1)}, \Delta_{kl}X_i^{(-1)}).$$

Assumption (A2)

The following hold for any n, and any distinct $(i,k),(i,l) \in \mathcal{N}_n^{(2)}$.

1. *$\Delta_{kl}X_i$ is not contained in a proper subspace of $\mathbb{R}^K$.*
2. *Exists $\Delta_{kl}X_i^{(1)}$ with $\beta_0^{(1)} \neq 0$ s.t. the cond. density of $\Delta_{kl}X_i^{(1)}$ is positive everywhere for any $\Delta_{kl}x_i^{(-1)}$.*

- Sign of $\beta_0^{(1)}$ is identified, and scale is normalized: $\mid \beta_0^{(1)} \mid = 1$.
- A2 used in Manski(1985,1987), Han (1987) and Abrevaya (1999).

Assumption (A3)

For any $i \in \mathcal{N}_n$*,*

$$\text{supp}(\mu_i \mid \mathbf{X}^n = \mathbf{x}) \subseteq [-B, B],$$

for any $\mathbf{x} \in \text{supp}(\mathbf{X}^n)$*, and some* $B < \infty$*.*

Assumption (A3)

For any $i \in \mathcal{N}_n$,

$$\text{supp}(\mu_i \mid \mathbf{X}^n = \mathbf{x}) \subseteq [-B, B],$$

for any $\mathbf{x} \in \text{supp}(\mathbf{X}^n)$*, and some* $B < \infty$.

- Allows for continuous or a discrete fixed effects.
- Intuitively:

$$\text{supp}(\mu_k - \mu_l \mid \mathbf{X}^n = \mathbf{x}) \subset \text{supp}(\Delta_{kl} X_i' \beta_0)$$

Assumption (A3)

For any $i \in \mathcal{N}_n$,

$$\text{supp}(\mu_i \mid \mathbf{X}^n = \mathbf{x}) \subseteq [-B, B],$$

for any $\mathbf{x} \in \text{supp}(\mathbf{X}^n)$*, and some* $B < \infty$*.*

- Allows for continuous or a discrete fixed effects.
- Intuitively:

$$\text{supp}(\mu_k - \mu_l \mid \mathbf{X}^n = \mathbf{x}) \subset \text{supp}(\Delta_{kl} X_i^{\prime} \beta_0)$$

- Let:

$$\mathcal{X}_B = \{\mathbf{x} \in \mathbf{X}^n : \text{ for any } \; i, j, k, l \in \mathcal{N}_n, \mid \Delta_{kl} x_i \beta_0 \mid \geq 2B, \text{ and } \\ \text{sign}\{\Delta_{kl} x_i \beta_0\} \neq \text{sign}\{\Delta_{kl} x_j \beta_0\}\}$$

Point Identification

For any distinct $i, j, l, k \in \mathcal{N}_n$, let:

$$Y_{kl}^{(s)} \equiv (D_{sk} - D_{sl}) \quad \text{for} \quad s = i, j,$$

Point Identification

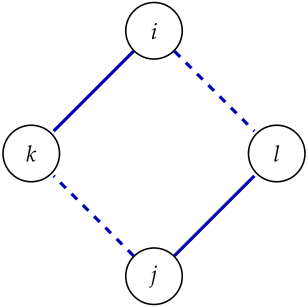
For any distinct $i, j, l, k \in \mathcal{N}_n$, let:

$$Y_{kl}^{(s)} \equiv (D_{sk} - D_{sl}) \quad \text{for} \quad s = i, j,$$
$$\Omega(ijlk) \equiv \{ \ D_{ik} \neq D_{il}, \ D_{jl} \neq D_{jk} \quad D_{ik} \neq D_{jk} \ \}$$

Point Identification

For any distinct $i, j, l, k \in \mathcal{N}_n$, let:

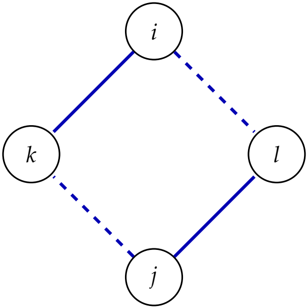
$$Y_{kl}^{(s)} \equiv (D_{sk} - D_{sl}) \quad \text{for} \quad s = i, j,$$
$$\Omega(ijlk) \equiv \{ \underbrace{D_{ik} \neq D_{il}, \;\; D_{jl} \neq D_{jk}}_{\text{Within-ind}} \;\; D_{ik} \neq D_{jk} \}$$



Point Identification

For any distinct $i, j, l, k \in \mathcal{N}_n$, let:

$$\begin{aligned} Y_{kl}^{(s)} &\equiv (D_{sk} - D_{sl}) \quad \text{for} \quad s = i, j, \\ \Omega(ijlk) &\equiv \{ \;\; D_{ik} \neq D_{il}, \;\; D_{jl} \neq D_{jk} \;\; \underbrace{D_{ik} \neq D_{jk}}_{\text{Across-inds}} \;\; \} \end{aligned}$$



Point Identification

Theorem

1. *Let assumptions 1-3 hold. Then, for any n, and any* $i, j, k, l \in \mathcal{N}_n$*:*

$$\text{Med}\left[Y_{kl}^{(i)} - Y_{kl}^{(j)} | \mathbf{X}^n = \mathbf{x}, \Omega(ijlk)\right] = 2 \times \text{sign}\left\{\left[\Delta_{kl} x_i - \Delta_{kl} x_j\right]' \beta_0\right\}, \tag{MC}$$

   where $\mathbf{x} \in \mathcal{X}_B$*.*

2. *Let assumptions 1-3 hold. Then* β_0 *is point identified.*

Point Identification

$\Omega(ijlk)$ contains the subnetworks with enough variation to identify β_0.

Point Identification

$\Omega(ijlk)$ contains the subnetworks with enough variation to identify β_0.

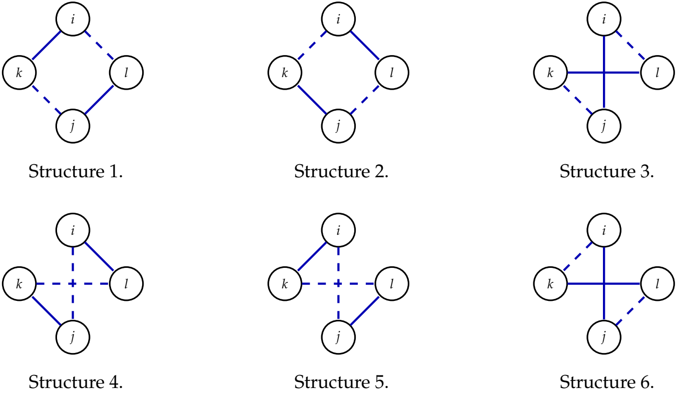


Figure: Subnetwork by the tetrad (i, j, k, l) in $\Omega(ijlk)$.

Identification Failure

I. Thin Set

Let

$$\Omega_n \equiv \{\Omega(ijlk) : \text{ for any distinct } i, j, k, l \in \mathcal{N}_n\}.$$

Identification Failure

I. Thin Set

Let

$$\Omega_n \equiv \{\Omega(ijlk) : \text{ for any distinct } i, j, k, l \in \mathcal{N}_n\}.$$

Theorem (Thin Set)

Under Ass. 1-3. If the class Ω_n has probability zero, then:

Identification Failure

Let

$$\Omega_n \equiv \{\Omega(ijlk) : \text{ for any distinct } i, j, k, l \in \mathcal{N}_n\} .$$

Theorem (Thin Set)

Under Ass. 1-3. If the class Ω_n has probability zero, then:

$\text{Med}\left\{ Y_{kl}^{(i)} - Y_{kl}^{(j)} \,\middle|\, \mathbf{X}^n = \mathbf{x} \right\}$ *does not have identification power.*

Identification Failure

I. Thin Set

Let

$$\Omega_n \equiv \{\Omega(ijlk) : \text{ for any distinct } i, j, k, l \in \mathcal{N}_n\}.$$

Theorem (Thin Set)

Under Ass. 1-3. If the class Ω_n has probability zero, then:

$\text{Med}\left\{ Y_{kl}^{(i)} - Y_{kl}^{(j)} \middle| \mathbf{X}^n = \mathbf{x} \right\}$ *does not have identification power.*

- In the empirical application: $P(\Omega_n) = 2.24\%$.
- "Thin set identification" as in Khan and Tamer (2010).

Thin Set

Lemma (Sufficient Conditions)

For any n, the class Ω_n has probability zero if for any $(i,j) \in \mathcal{N}_n^{(2)}$:

1. *D^n is empty, i.e.,*

$$\operatorname{supp}\left(X_{ij}'\beta_0 \mid \mu^n = \mu, \varepsilon_{ij} = e\right) = \left(-\infty,\ \mu_i + \mu_j - e\right]$$

Thin Set

Lemma (Sufficient Conditions)

For any n, the class Ω_n has probability zero if for any $(i,j) \in \mathcal{N}_n^{(2)}$:

1. *D^n is empty, i.e.,*

$$\operatorname{supp}\left(X_{ij}'\beta_0 \mid \mu^n = \mu, \varepsilon_{ij} = e\right) = \left(-\infty,\ \mu_i + \mu_j - e\right]$$

2. *D^n is dense, i.e.,*

$$\operatorname{supp}\left(X_{ij}'\beta_0 \mid \mu^n = \mu, \varepsilon_{ij} = e\right) = \left[\mu_i + \mu_j - e,\ \infty\right)$$

Thin Set

Lemma (Sufficient Conditions)

For any n, the class Ω_n has probability zero if for any $(i,j) \in \mathcal{N}_n^{(2)}$:

1. *D^n is empty, i.e.,*

$$\text{supp}\left(X_{ij}'\beta_0 \mid \mu^n = \mu, \varepsilon_{ij} = e\right) = \left(-\infty,\ \mu_i + \mu_j - e\right]$$

2. *D^n is dense, i.e.,*

$$\text{supp}\left(X_{ij}'\beta_0 \mid \mu^n = \mu, \varepsilon_{ij} = e\right) = \left[\mu_i + \mu_j - e,\ \infty\right)$$

3. *D^n is homogeneous, i.e.,*

$$\text{supp}\left(\mu_i + \mu_j \mid X_{ij} = x, \varepsilon_{ij} = e\right) = \left[\, e - x'\beta_0,\ \infty\right)$$

Additional Identification Results

Is the large support assumption necessary for identification?

Additional Identification Results

Is the large support assumption necessary for identification?

- Suppose all the covariates have bounded support, and

Additional Identification Results

Is the large support assumption necessary for identification?

- Suppose all the covariates have bounded support, and

1. A2′: there exists at least one continuous variable with

$$\text{supp}\left(\Delta_{kl}X_i\beta_0 \mid \Delta_{kl}X_i^{(-1)} = \Delta_{kl}x_i^{(-1)}\right) = [-\delta, \delta]$$

$\Rightarrow \beta_0$ is still identified

Additional Identification Results

Is the large support assumption necessary for identification?

- Suppose all the covariates have bounded support, and

1. A2′: there exists at least one continuous variable with

$$\text{supp}\left(\Delta_{kl}X_i\beta_0 \mid \Delta_{kl}X_i^{(-1)} = \Delta_{kl}x_i^{(-1)}\right) = [-\delta, \delta]$$

   $\Rightarrow \beta_0$ is still identified

2. A2″: they are all discrete variables.

   $\Rightarrow$ Bounds for each element of β_0 are obtained.

Outline

1. Network Formation Model
2. Identification
3. Inference
4. Simulations
5. Application
6. Conclusions and Extensions

Inference

The identification condition in (MC) suggests an estimator for β_0.

Limiting objective function:

$$Q(b) \equiv 2\mathbb{E}\left[S(\mathcal{X}_B) \times \text{sign}\left\{\left[\Delta_{kl}X_i - \Delta_{kl}X_j\right]' b\right\} \times \left(Y_{kl}^{(i)} - Y_{kl}^{(j)}\right) \mid \Omega(ijlk)\right],$$

where, $S(\mathcal{X}_B) = 1$ if $\mathbf{x} \in \mathcal{X}_B$, and 0 otherwise.

Inference

The identification condition in (MC) suggests an estimator for β_0.

Limiting objective function:

$$Q(b) \equiv 2\mathbb{E}\left[S(\mathcal{X}_B) \times \text{sign}\left\{\left[\Delta_{kl}X_i - \Delta_{kl}X_j\right]' b\right\} \times \left(Y_{kl}^{(i)} - Y_{kl}^{(j)}\right) \mid \Omega(ijlk)\right],$$

where, $S(\mathcal{X}_B) = 1$ if $\mathbf{x} \in \mathcal{X}_B$, and 0 otherwise.

- $Q(b)$ is uniquely maximized at $b = \beta_0$.

Inference

The identification condition in (MC) suggests an estimator for β_0.

Limiting objective function:

$$Q(b) \equiv 2\mathbb{E}\left[S(\mathcal{X}_B) \times \text{sign}\left\{\left[\Delta_{kl}X_i - \Delta_{kl}X_j\right]' b\right\} \times \left(Y_{kl}^{(i)} - Y_{kl}^{(j)}\right) \mid \Omega(ijlk)\right],$$

where, $S(\mathcal{X}_B) = 1$ if $\mathbf{x} \in \mathcal{X}_B$, and 0 otherwise.

- $Q(b)$ is uniquely maximized at $b = \beta_0$.

The semiparametric pairwise difference estimator is

$$\hat{\beta}_n = \arg\max_{b \in \tilde{\mathcal{B}}} Q_n(b)$$

Inference

Given a random sample of n agents, let $\left\{z_{ij}^{n}\right\}_{(i,j)\in\mathcal{N}_n^{(2)}} = \left\{D_{ij}^{n}, x_{ij}\right\}_{(i,j)\in\mathcal{N}_n^{(2)}}$.

Inference

Given a random sample of n agents, let $\left\{z_{ij}^n\right\}_{(i,j)\in\mathcal{N}_n^{(2)}} = \left\{D_{ij}^n, x_{ij}\right\}_{(i,j)\in\mathcal{N}_n^{(2)}}$.

The sample analog of $Q(b)$:

$$Q_n(b) \equiv \binom{n}{4}^{-1} \sum_{C_{n,4}} h(z_{i_{1,3}}, z_{i_{1,4}}, z_{i_{2,3}}, z_{i_{2,4}}, b), \qquad \text{(Qn)}$$

where $C_{n,4}$ indexes all the unique tetrads in $\{1, 2, \cdots, n\}$.

Inference

Given a random sample of n agents, let $\left\{z_{ij}^n\right\}_{(i,j)\in\mathcal{N}_n^{(2)}} = \left\{D_{ij}^n, x_{ij}\right\}_{(i,j)\in\mathcal{N}_n^{(2)}}$.

The sample analog of $Q(b)$:

$$Q_n(b) \equiv \binom{n}{4}^{-1} \sum_{C_{n,4}} h(z_{i_{1,3}}, z_{i_{1,4}}, z_{i_{2,3}}, z_{i_{2,4}}, b), \tag{Qn}$$

where $C_{n,4}$ indexes all the unique tetrads in $\{1, 2, \cdots, n\}$.

Kernel function:

$$h(z_{i_{1,3}}, z_{i_{1,4}}, z_{i_{2,3}}, z_{i_{2,4}}, b) \equiv \frac{2}{4!} \sum_{P_4} \{\text{sign}\left\{[\Delta_{3,4}x_1 - \Delta_{3,4}x_2]' \, b\right\}$$
$$\times \, (y_{3,4}^{(1)} - y_{3,4}^{(2)}) \times \mathbf{1}\left\{|y_{3,4}^{(1)} - y_{3,4}^{(2)}| = 2\right\} \times S(x_{i_{1,3}}, x_{i_{1,4}}, x_{i_{2,3}}, x_{i_{2,4}}, B)\right\},$$

where P_4 denotes the 4! permutations of $\{i_{1,3}, i_{1,4}, i_{2,3}, i_{2,4}\}$.

Choice of B

1. If B is known

$$\mathcal{X}_B = \{\mathbf{x} \in \mathbf{X}^n : \text{ for any } \ i, j, k, l \in \mathcal{N}_n, \ | \Delta_{kl} x_i b | \geq 2B, \text{ and } \\ \text{sign}\{\Delta_{kl} x_i b\} \neq \text{sign}\{\Delta_{kl} x_j b\}\}$$

Choice of B

1. If B is known

$$\mathcal{X}_B = \{\mathbf{x} \in \mathbf{X}^n : \text{ for any } i,j,k,l \in \mathcal{N}_n, \mid \Delta_{kl} x_i b \mid \geq 2B, \text{ and } \\ \text{sign}\{\Delta_{kl} x_i b\} \neq \text{sign}\{\Delta_{kl} x_j b\}\}$$

2. Trimming

$$\mathcal{X}_B(\gamma_n) = \{\mathbf{x} \in \mathbf{X}^n : \text{ for any } i,j,k,l \in \mathcal{N}_n, \mid \Delta_{kl} x_i b \mid \geq \gamma_n, \text{ and } \\ \text{sign}\{\Delta_{kl} x_i b\} \neq \text{sign}\{\Delta_{kl} x_j b\}\},$$

with $\gamma_n \to \infty$.

Choice of B

1. If B is known

$$\mathcal{X}_B = \{\mathbf{x} \in \mathbf{X}^n : \text{ for any } \; i, j, k, l \in \mathcal{N}_n, \; | \Delta_{kl} x_i b | \geq 2B, \text{ and } \\ \text{sign}\{\Delta_{kl} x_i b\} \neq \text{sign}\{\Delta_{kl} x_j b\}\}$$

2. Trimming

$$\mathcal{X}_B(\gamma_n) = \{\mathbf{x} \in \mathbf{X}^n : \text{ for any } \; i, j, k, l \in \mathcal{N}_n, \; | \Delta_{kl} x_i b | \geq \gamma_n, \text{ and } \\ \text{sign}\{\Delta_{kl} x_i b\} \neq \text{sign}\{\Delta_{kl} x_j b\}\},$$

with $\gamma_n \to \infty$.

$$\sup_{\gamma_n \in \Gamma} \sup_{b \in \tilde{\mathcal{B}}} ||Q_n(b; \gamma_n) - \mathbb{E}[Q_n(b; \gamma_n)]|| \xrightarrow{p} 0.$$

Assumptions

Assumption (B1)

The researcher observes a random sample of n agents. For each dyad in $\mathcal{N}_n^{(2)}$, the researcher observes the link status and dyad-level attributes.

$$\left\{D_{ij}, \mathbf{x}_{ij}\right\}_{(i,j)\in\mathcal{N}_n^{(2)}} .$$

- Used in Graham (2017), Leung (2015b) and Menzel (2015).

Assumptions

Assumption (B1)

The researcher observes a random sample of n agents. For each dyad in $\mathcal{N}_n^{(2)}$, the researcher observes the link status and dyad-level attributes.

$$\left\{D_{ij}, \mathbf{x}_{ij}\right\}_{(i,j)\in\mathcal{N}_n^{(2)}} .$$

- Used in Graham (2017), Leung (2015b) and Menzel (2015).

Assumption (B2)

The parameter space $\tilde{\mathcal{B}}$ is compact and β_0 is an interior point of $\tilde{\mathcal{B}}$.

- Used in Han (1987), Sherman (1993, 1994) and Abrevaya (1999).

Assumptions

Assumption (B3)

Let $p_n \equiv \mathbb{P}(\Omega_n)$, *where*

1. $p_n \to p_0 \geq 0$, *as* $n \to \infty$.
2. $\sqrt{N} p_n \to \infty$, *as* $n \to \infty$.

- The probability p_n is allowed to decay as $n \to \infty$.

Consistency

Theorem (Consistency)

Let assumptions A1, A2, B1-B3 hold. Then,

$$\hat{\beta}_n - \beta_0 \xrightarrow{p} 0$$

as $n \to \infty$.

Theorem (Asymptotic Normality)

If assumptions A1, A2, B1-B4. hold, then:

$$p_n \sqrt{N}(\hat{\beta}_n - \beta_0) \xrightarrow{d} \mathcal{N}(0, V^{-1} \Delta V^{-1}) \tag{AN}$$

with

$$\begin{aligned} 4V &= \mathbb{E}\left[\nabla_2 \tau_2(\cdot, \beta_0) \mid \Omega_n\right], \\ \Delta &= \mathbb{E}\left[\nabla_1 \tau(\cdot, \beta_0)\right]\left[\nabla_1 \tau(\cdot, \beta_0)\right]'. \end{aligned}$$

Recall that $N = O(n^2)$.

Convergence Rate

The convergence rate depends on the limit of $p_n \equiv \mathbb{P}(\Omega_n)$.

1. Regular Estimator: $p_n \to \bar{p} > 0$, as $n \to \infty$.

Convergence Rate

The convergence rate depends on the limit of $p_n \equiv \mathbb{P}(\Omega_n)$.

1. Regular Estimator: $p_n \to \bar{p} > 0$, as $n \to \infty$.
2. Irregular Estimator: $p_n \to 0$, as as $n \to \infty$.
   (Newey 1990, Andrews and Schafgans 1998 and Khan and Tamer 2010).

Theorem (Information bound)

In model given by equation (NF), *under assumptions A1, A2, B1-B4.*

*If $p_n \to 0$, then the information bound for β_0 is **zero**.*

Adaptive Rate Inference

Consider the "studentized" estimator, as in Andrews and Schafgans (1998) and Khan and Tamer (2010),

$$\hat{\Sigma}_n^{-1/2}\sqrt{N}(\hat{\beta}_n - \beta_0) \xrightarrow{d} \mathcal{N}(\mathbf{0}, I), \quad \text{as } n \to \infty$$

where $\hat{\Sigma}_n$,

$$\hat{\Sigma}_n = \hat{S}_n / \hat{p}_n^2,$$

and $\hat{S}_n$ is the Bootstrap estimate of

$$S = V^{-1} \Delta V^{-1}.$$

Subbotin (2007): Bootstrap validity for Maximum Rank estimators.

Outline

Computation

The objective function $Q_n(b)$ is a 4th order U-statistic.

- $O(n^4)$ operations.

Proposition

The estimator $\hat{\beta}_n$ can be equivalently computed from:

$$\tilde{Q}_n(b) \equiv \frac{1}{n(n-1)(n-2)} \sum_{i \neq j \neq k} S(B) Rank_{j,k} \left[(x_{ik} - x_{il})' b \right] y_{k,l}^{(i)}$$

- *$\tilde{Q}_n(b)$ can be computed in $O(n^3 log(n))$ operations.*

Simulations

Consider the following model:

$$D_{ij} = \mathbf{1}\left[X_{ij}'\beta_0 + \mu_i + \mu_j - \varepsilon_{ij} \geq 0\right], \quad \text{for } (i,j) \in \mathcal{N}_n^{(2)}.$$

1. Dyad-specific attributes, X_{ij} for $(i,j) \in \mathcal{N}_n^{(2)}$:

$$X_{ij} = \left[z_{i1}z_{j1},\ z_{i2}z_{j2},\ z_{i3}z_{j3}\right].$$

   where the individual-specific attributes are drawn as:

$$\begin{aligned} z_{i1} &\sim \text{Normal}(0,3), \\ z_{i2} &\sim \text{Uniform}\{-1,0,1\} \text{ with } p_k = 1/3, \\ z_{i3} &\sim \text{Uniform}(-2,2). \end{aligned}$$

Simulations

2. Fixed effects:

$$\alpha_i = \lambda \left(z_{i1} + z_{i2} + z_{i3} \right) /3 + (1 - \lambda)\text{Normal}(0, 1),$$

where $\lambda \in \{1/4, 1/2, 3/4\}$ measures the degree of dependence.

$$\mu_i = \begin{cases} -B & \text{if} \quad \alpha_i < -B \\ \alpha_i & \text{if} \quad -B \leq \alpha_i \leq B \\ B & \text{if} \quad B < \alpha_i \end{cases},$$

with $B = 1$.

3. Link-specific disturbance term: $\varepsilon_{ij}^{(2)} \sim \text{Normal}(0, 2)$.

True DGP: $\beta_0 = [1,\ 1.5,\ -1.5]'$

MC Simulations: Normal(0,2)

	Pairwise Difference				Graham (2015)				$P(\Omega_n)$
	Median	Mean	Bias(%)	RMSE	Median	Mean	Bias(%)	RMSE	
$N = 100$									7.914%
$\beta_2/\beta_1 = 1.5$	1.630	1.585	5.715	0.727	1.651	1.665	7.454	0.437	
$\beta_3/\beta_1 = -1.5$	-1.734	-1.702	13.613	1.836	-1.735	-1.763	15.712	0.438	
$N = 250$									7.376%
$\beta_2/\beta_1 = 1.5$	1.567	1.551	5.061	0.686	1.524	1.512	4.133	0.325	
$\beta_3/\beta_1 = -1.5$	-1.677	-1.632	7.245	1.074	-1.691	-1.674	13.128	0.325	

M=500, $\lambda = 0.5$

MC Simulations: Normal(0,2)

	Pairwise Difference				Graham (2015)				$P(\Omega_n)$
	Median	Mean	Bias(%)	RMSE	Median	Mean	Bias(%)	RMSE	
$N = 100$									7.914%
$\beta_2/\beta_1 = 1.5$	1.630	1.585	5.715	0.727	1.651	1.665	7.454	0.437	
$\beta_3/\beta_1 = -1.5$	-1.734	-1.702	13.613	1.836	-1.735	-1.763	15.712	0.438	
$N = 250$									7.376%
$\beta_2/\beta_1 = 1.5$	1.567	1.551	5.061	0.686	1.524	1.512	4.133	0.325	
$\beta_3/\beta_1 = -1.5$	-1.677	-1.632	7.245	1.074	-1.691	-1.674	13.128	0.325	
$N = 500$									7.148%
$\beta_2/\beta_1 = 1.5$	1.529	1.542	4.761	0.591					
$\beta_3/\beta_1 = -1.5$	-1.572	-1.553	5.281	0.801					

M=500, $\lambda = 0.5$

Simulation: Discrete and Bounded Support:

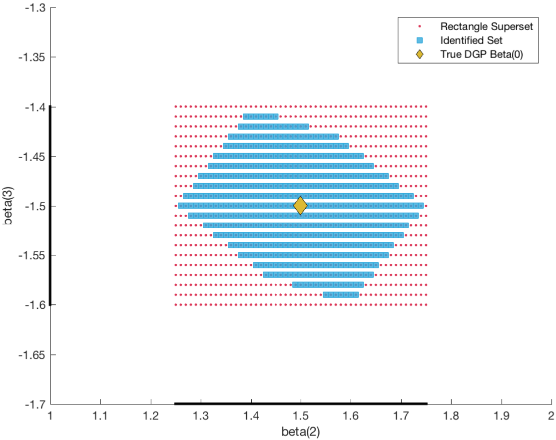
Consider the next specification for the observed covariates:

- $X_{ij}^{(1)}$ takes the values $\{0, 1, 2, 3, 4\}$.
- $X_{ij}^{(2)}$ takes the values $\{-1, 0, 1\}$.
- $X_{ij}^{(3)}$ takes the values $\{-1, 0, 1, 2\}$.

Thus, the support of X_{ij} contains 60 points.

Discrete and Bounded Support: Sharp Bounds

Figure: Bounds and Rectangular Superset



Outline

Empirical Application

This application estimates a model of friendships formation using the Add Health dataset.

Empirical Application

This application estimates a model of friendships formation using the Add Health dataset.

- Objective: Estimate the preferences for homophily.

Empirical Application

This application estimates a model of friendships formation using the Add Health dataset.

- Objective: Estimate the preferences for homophily.
- Dataset: Add Health is a longitudinal national survey.

Empirical Application

This application estimates a model of friendships formation using the Add Health dataset.

- Objective: Estimate the preferences for homophily.
- Dataset: Add Health is a longitudinal national survey.
- High school students: Grades 7-12 during the 1994-95 school years.

Empirical Application

This application estimates a model of friendships formation using the Add Health dataset.

- Objective: Estimate the preferences for homophily.
- Dataset: Add Health is a longitudinal national survey.
- High school students: Grades 7-12 during the 1994-95 school years.
- Observed network: availability of respondents' friendship network.

Empirical Application

This application estimates a model of friendships formation using the Add Health dataset.

- Objective: Estimate the preferences for homophily.
- Dataset: Add Health is a longitudinal national survey.
- High school students: Grades 7-12 during the 1994-95 school years.
- Observed network: availability of respondents' friendship network.
- Saturated high schools: each student nominates at most 5 male and 5 female friends.

Empirical Application

This application estimates a model of friendships formation using the Add Health dataset.

- Objective: Estimate the preferences for homophily.
- Dataset: Add Health is a longitudinal national survey.
- High school students: Grades 7-12 during the 1994-95 school years.
- Observed network: availability of respondents' friendship network.
- Saturated high schools: each student nominates at most 5 male and 5 female friends.
- Wave I In-home interview: One high school with 319 students.

Exogenous Covariates

Table: Descriptive Statistics

Variable	Mean	Std. Dev.	Min	Max
Household Income	51.405	29.68	4	200
Age	15.707	1.183	14	19
Female	0.441	0.497	0	1
Grade	10.255	1.085	9	12
Hispanic	0.025	0.150	0	1
White	0.942	0.233	0	1
Black	0.006	0.079	0	1
Asian	0.014	0.121	0	1
Indian	0.029	0.170	0	1
Other races	0.036	0.187	0	1
Overall GPA	2.346	0.956	0	4
Mother's Education	4.240	2.419	0	9
Father's Education	4.147	2.794	0	9
Sample size = 469.				

Estimation Results

	Logistic	Pairwise Difference	Graham (2015)
Age	−1.245***	−0.826	−1.088
Female	−1.875***	0.635**	0.032
Grade	0.764***	1.264*	0.553*
Hispanic	0.772	1.322***	1.100***
White	−3.758***	1.661**	1.544***
Black		0.382	0.085
Asian		−1.172**	−1.491**
Indian	−0.597	−0.318	−0.742
Other races	−0.461	−0.553*	−1.061
Overall GPA	−0.102***	2.436**	2.350**
Mother Education	0.276***	−0.352*	−0.615*
Father Education	0.240***	1.549***	0.748

$P(\Omega) = 2.24\%$
Average Degree = 3.62.
Number of Students = 319.
Number of dyads = 50,721.

*,**,*** represents the significant at 10%, 5%, and 1% level.

Conclusions

1. Semiparametric network model with unobserved heterogeneity.

2. Point identification and sharp bounds for each component of β_0.

3. Semiparametric pairwise difference estimator.

4. Empirical application considers a friendship network.

Thanks!

Appendix

Covariates with Bounded Support

I. At Least One Continuous Covariate

Assumption (A2′)

The following hold for any n, and any $i, l, k \in \mathcal{N}_n$, with $l \neq k$.

1. *The random vector $\Delta_{kl} X_i$ has a bounded support on $\mathbb{R}^K$.*
2. *For some $\delta > 0$, there exists an interval $I_\delta = [-\delta, \delta]$ and a set $N_\delta \in \mathbb{R}^{K-1}$ such that*
    - *N_δ is not contained in any proper linear subspace of $\mathbb{R}^{K-1}$.*
    - $\mathbb{P}\left(\Delta_{kl}\tilde{X}_i \in N_\delta\right) > 0.$
    - *For almost every $\Delta_{kl}\tilde{x} \in N_\delta$, the distribution of $\Delta_{kl}X_i'\beta_0$ conditional on $\Delta_{kl}\tilde{X}_i = \Delta_{kl}\tilde{x}_i$ has a probability density that is everywhere positive on I_δ.*

Proposition

Let Assumptions A1, A2′, and A3 hold; then β_0 is point identified.

Covariates with Bounded Support

II. Discrete Support

I obtain sharp bounds for each component in β_0 using Komarova (2013).

Assumption (A2″)

For any n, and any $i, k, l \in \mathcal{N}_n$, with $k \neq l$.

1. *The support of $F_{X_{ik}}$ is not contained in any proper linear space of $\mathbb{R}^K$.*
2. *The profile vector of observed attributes $\mathbf{X}^n \equiv (X_{12}, \cdots, X_{n-1,n})$ has a discrete support given by*

$$\text{supp}(\mathbf{X}^n) = \left\{\mathbf{x}^1, \cdots, \mathbf{x}^D\right\},$$

for a finite D.

Thin Set

Table: Stochastic Dominance and Sparsity

	Empty		Sparse		Dense	
	E [Degree]	$P\left[\Omega_n\right]$ (%)	E [Degree]	$P\left[\Omega_n\right]$ (%)	E [Degree]	$P\left[\Omega_n\right]$ (%)
$\lambda = 0.25$						
Log	20.30	4.32	49.53	16.71	97.15	0.06
LnN	9.34	1.01	36.98	13.73	95.88	0.11
N	19.47	3.84	49.52	18.11	98.56	0.00
Gam	19.54	3.87	49.36	19.63	87.12	1.56
T	28.59	8.30	49.45	18.25	90.54	1.03
$\lambda = 0.5$						
Log	23.56	5.71	49.44	16.95	95.48	0.21
LnN	10.58	1.28	36.62	13.72	92.34	0.47
N	22.44	5.03	49.39	18.58	98.13	0.01
Gam	23.11	5.41	49.32	21.04	76.73	4.72
T	33.90	11.29	49.30	18.84	84.53	2.71
$\lambda = 0.75$						
Log	27.81	7.88	49.30	17.14	91.75	0.86
LnN	12.38	1.74	36.06	13.64	80.39	3.52
N	26.38	6.92	49.21	18.82	96.75	0.07
Gam	27.08	7.34	49.20	22.42	54.40	11.08
T	40.51	15.00	49.26	19.29	72.11	7.27

Notes: N=100, M=500.

Thin Set

Table: Thin Set Simulations: Homogeneous Network

		$\mu = 10 * \text{Bernoulli(p)} + (-5) * (1 - \text{Bernoulli(p)})$		
N=100	E [Degree]	$P[\Omega(ijkl)]$ (%) (%)	Jaccard SI (Mean) (Mean)	Cosine SI (Mean) (Mean)
$p = 0.2$				
Log	37.66	0.38	0.55	0.70
LnN	20.52	0.83	0.35	0.53
N	36.66	0.31	0.60	0.73
Gam	31.14	0.42	0.56	0.70
T	27.30	0.34	0.57	0.70
$p = 0.8$				
Log	92.56	0.12	0.87	0.93
LnN	83.46	1.16	0.74	0.85
N	95.10	0.01	0.91	0.95
Gam	94.42	0.05	0.90	0.94
T	93.26	0.10	0.88	0.93

Notes: M=500.

Identification Failure

Proposition

1. *Let assumption 1 hold; then, for any n, and any $i, l, k \in \mathcal{N}_n$.*

$$\begin{aligned} &\text{Med}(D_{ik} - D_{il} | \mathbf{X}^n = x, D_{il} + D_{ik} = 1) \\ &\qquad = \text{sign}\left[(x_{ik} - x_{il})'\beta_0 + (\mu_k - \mu_l)\right] \end{aligned} \tag{MS}$$

2. *Let Assumptions 1 and 2 hold; then, the equation* (MS) *does not have identification power.*

References I

Abrevaya, J. (1999). Leapfrog estimation of a fixed-effects model with unknown transformation of the dependent variable. *Journal of Econometrics 93*(2), 203–228.

Andersen, E. B. (1973). Conditional inference for multiple-choice questionnaires. *British Journal of Mathematical and Statistical Psychology 26*(1), 31–44.

Andrews, D. W. and M. M. Schafgans (1998). Semiparametric estimation of the intercept of a sample selection model. *The Review of Economic Studies 65*(3), 497–517.

Auerbach, E. (2016). Identification and estimation of models with endogenous network formation. Technical report, Working Paper.

Banerjee, A., A. G. Chandrasekhar, E. Duflo, and M. O. Jackson (2013). The diffusion of microfinance. *Science 341*(6144), 1236498.

Boucher, V. and I. Mourifié (2013). My friend far far away: Asymptotic properties of pairwise stable networks. *Working Paper*.

Brock, W. A. and S. N. Durlauf (2005). Multinomial choice with social interactions. *The Economy As an Evolving Complex System, III: Current Perspectives and Future Directions*, 175.

References II

Candelaria, L. E. (2016). Network formation models with interactive fixed effects. *Working Paper*.

Chamberlain, G. (2010). Binary response models for panel data: Identification and information. *Econometrica 78*(1), 159–168.

Chandrasekhar, A. G. and M. O. Jackson (2014). Tractable and consistent random graph models. *Working Paper*.

Charbonneau, K. (2014). Multiple fixed effects in nonlinear panel data models. *Working Paper*.

Christakis, N. A., J. H. Fowler, G. W. Imbens, and K. Kalyanaraman (2010). An empirical model for strategic network formation. *Working Paper*.

de Paula, A., S. Richards-Shubik, and E. Tamer (2017). Identifying preferences in networks with bounded degree. forthcoming in. *Econometrica*.

Dzemski, A. (2017). An empirical model of dyadic link formation in a network with unobserved heterogeneity.

Goldsmith-Pinkham, P. and G. W. Imbens (2013). Social networks and the identification of peer effects. *Journal of Business & Economic Statistics 31*(3), 253–264.

Graham, B. S. (2017). An econometric model of link formation with degree heterogeneity. *Working Paper*.

References III

Han, A. K. (1987). Non-parametric analysis of a generalized regression model: the maximum rank correlation estimator. *Journal of Econometrics 35*(2), 303–316.

Hsieh, C.-S. and L. F. Lee (2016). A social interactions model with endogenous friendship formation and selectivity. *Journal of Applied Econometrics 31*(2), 301–319.

Jochmans, K. (2017). Semiparametric analysis of network formation. *Journal of Business & Economic Statistics* (just-accepted).

Khan, S. and E. Tamer (2010). Irregular identification, support conditions, and inverse weight estimation. *Econometrica 78*(6), 2021–2042.

Komarova, T. (2013). Binary choice models with discrete regressors: Identification and misspecification. *Journal of Econometrics 177*(1), 14–33.

Leung, M. (2015a). A random-field approach to inference in large models of network formation. *Available at SSRN*.

Leung, M. (2015b). Two-step estimation of network-formation models with incomplete information. *Journal of Econometrics 188*(1), 182–195.

Leung, M. (2016). A weak law for moments of pairwise-stable networks. *Working Paper*.

References IV

Manski, C. F. (1985). Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator. *Journal of Econometrics 27*(3), 313–333.

Manski, C. F. (1987). Semiparametric analysis of random effects linear models from binary panel data. *Econometrica*, 357–362.

McPherson, M., L. Smith-Lovin, and J. M. Cook (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 415–444.

Mele, A. (2017). A structural model of dense network formation. *Econometrica 85*(3), 825–850.

Menzel, K. (2015). Stategic network formation with many agents. *Working Paper*.

Newey, W. K. (1990). Semiparametric efficiency bounds. *Journal of Applied Econometrics 5*(2), 99–135.

Ridder, G. and S. Sheng (2015). Estimation of large network formation games. Technical report, Working papers, UCLA.

Sheng, S. (2012). Identification and estimation of network formation games. *Working Paper*.

Sherman, R. P. (1993). The limiting distribution of the maximum rank correlation estimator. *Econometrica*, 123–137.

References V

Sherman, R. P. (1994). Maximal inequalities for degenerate U-processes with applications to optimization estimators. *The Annals of Statistics*, 439–459.

Souza, P. (2014). Estimating network effects without network data. *LSE Working*.

Subbotin, V. (2007). Asymptotic and bootstrap properties of rank regressions. *Working Paper*.