Asymptotic unidentifiability of peer effects in the linear-in-means model

Alex Hayes | PhD Defense

2024-04-04 @ 12:30 pm

Service Memorial Institute 133, Medical Sciences Center, UW-Madison

Department of Statistics, University of Wisconsin-Madison

Contagion: if my friends get sick, I am more likely to set sick

Direct effect: if I get vaccinated, I am less likely to get sick

Interference: if my friends get vaccinated, I am less likely to get sick

* Not a causal talk. But causally inspired.

$$\underbrace{Y_i}_{\text{sick?}} = \underbrace{\alpha}_{\substack{\text{base}\\\text{rate}}} + \underbrace{\beta}_{\substack{\text{contagion}\\\text{effect}}} \underbrace{\sum_{i \neq j} \frac{A_{ij}}{d_i} Y_j}_{\substack{\text{portion}\\\text{sick}\\\text{peers}}} + \underbrace{\gamma}_{\substack{\text{direct}\\\text{effect}}} \underbrace{T_i}_{\text{vaccinated?}} + \underbrace{\delta}_{\substack{\text{interference}\\\text{effect}}} \underbrace{\sum_{i \neq j} \frac{A_{ij}}{d_i} T_j}_{\substack{\text{portion}\\\text{vaccinated}\\\text{peers}}} + \underbrace{\varepsilon_i}_{\text{error}}$$

Outcome	Y_i	$\in \{0, 1\}$
Treatment	T_i	$\in \{0, 1\}$
Adjacency matrix	A	$\in \{0, 1\}^{n \times n}$
Edge $i \sim j$	A_{ij}	$\in \{0, 1\}$
Node degree	d_i	$\in \mathbb{Z}^+$

* In practice, no domain restrictions required for Y, T, A.

Let $G_{ij} = A_{ij}/d_i$ be the row-normalization of A. We express the previous model in matrix-vector notation as

$$Y = \alpha 1_n + \beta GY + \gamma T + \delta GT + \varepsilon$$

If $|\beta| < 1$, then $I - \beta G$ is invertible and there is a unique solution

$$\begin{aligned} Y &= (I - \beta G)^{-1}(\alpha 1_n + \gamma T + \delta GT + \varepsilon) \\ &= \sum_{k=0}^{\infty} \beta^k G^k (\alpha 1_n + \gamma T + \delta GT + \varepsilon) \qquad \text{since } (I - \beta G)^{-1} = \sum_{k=0}^{\infty} \beta^k G^k \end{aligned}$$

Proposition (Finite sample identification, Bramoullé et al. (2009))

Let ε be mean zero, i.i.d. noise and let

$$Y = \alpha 1_n + \beta GY + \gamma T + \delta GT + \varepsilon$$

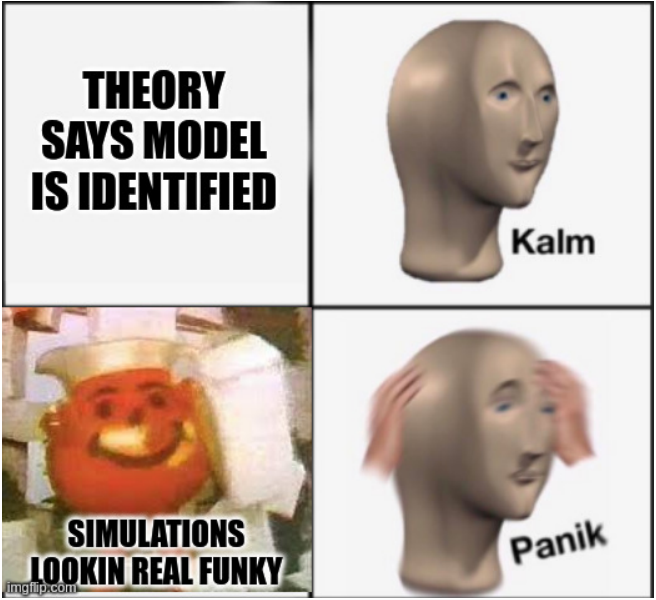
Suppose that $|\beta| < 1$ and $\gamma\beta + \delta \neq 0$. If I, G and G^2 are linearly independent, in the sense that $aI + bG + cG^2 = 0$ requires $a = b = c = 0$, then α, β, γ and δ are identified.

$\gamma\beta + \delta \neq 0$ means that there is either some interference effect, or some direct effect and some contagion effect, and if there are both, they don't cancel each other out.

Proposition

If $\gamma\beta + \delta \neq 0$ and G has three or more distinct eigenvalues, then α, β, γ and δ are identified.

Promising! If G is from a stochastic blockmodel, identification likely.


THEORY SAYS MODEL IS IDENTIFIED
Kalm
SIMULATIONS LOOKIN REAL FUNKY
imgflip.com
Panik

Disappearing asymptotic identification: a simple case (Li and Wager, 2022)

$$\underbrace{Y_i}_{\text{sick?}} = \underbrace{\alpha}_{\substack{\text{base}\\\text{rate}}} + \underbrace{\gamma}_{\substack{\text{direct}\\\text{effect}}} \underbrace{T_i}_{\text{vaccinated?}} + \underbrace{\delta}_{\substack{\text{interference}\\\text{effect}}} \underbrace{\sum_{i \neq j} \frac{A_{ij}}{d_i} T_j}_{\substack{\text{portion}\\\text{vaccinated}\\\text{peers}}} + \underbrace{\varepsilon_i}_{\text{error}}$$

Problem: *GT* term converges to a constant

$$\sum_{i \neq j} \frac{A_{ij}}{d_i} T_j \to \pi$$

GY can also be a problem

Suppose no node is isolated (otherwise the intercept differs for isolated and connected nodes in reduced form) then the reduced form of Y is given by

$$Y = \frac{\alpha}{1-\beta}1_n + \gamma T + (\gamma\beta + \delta)\sum_{k=0}^{\infty}\beta^k G^{k+1} T + \sum_{k=0}^{\infty}\beta^k G^k \varepsilon$$

then

$$GY = \frac{\alpha}{1-\beta}1_n + \underbrace{\gamma G T}_{\substack{\text{neighborhood}\\ \text{average}}} + \underbrace{(\gamma\beta + \delta)\sum_{k=0}^{\infty}\beta^k G^{k+2} T}_{\substack{\text{repeated}\\ \text{neighborhood}\\ \text{averages}}} + \underbrace{\sum_{k=0}^{\infty}\beta^k G^{k+1} \varepsilon}_{\substack{\text{neighborhood}\\ \text{repeated}\\ \text{averages}}}$$

$$GY = \frac{\alpha}{1-\beta}1_n + \underbrace{\gamma GT}_{\substack{\text{neighborhood}\\\text{average}}} + \underbrace{(\gamma\beta + \delta)\sum_{k=0}^{\infty} \beta^k G^{k+2} T}_{\substack{\text{repeated}\\\text{neighborhood}\\\text{averages}}} + \underbrace{\sum_{k=0}^{\infty} \beta^k G^{k+1} \varepsilon}_{\substack{\text{neighborhood}\\\text{repeated}\\\text{averages}}}$$

If neighborhood averages GT are converging, repeated neighborhood averages converge as well[1], because $G1_n = 1_n$ (neighborhood average of constants is constant)

[1] Possibly under suitable conditions

Roughly, if

1. neighborhoods d_i are all getting larger in n, and
2. there is no CLT/LLN breaking strong dependence between network A and nodal covariate T

Then

$$\lim_{n \to \infty} GT \to C \qquad C \in \mathbb{R}$$
$$\lim_{n \to \infty} GY \to C' \qquad C' \in \mathbb{R}$$

Then contagion and interference effects are the same for everyone, and indistinguishable from base rates

The direct effect γ is still identified.

Theorem (Identification failure when T independent of A)

Let ε be mean zero, i.i.d. noise and let

$$Y = \alpha 1_n + \beta GY + \gamma T + \delta GT + \varepsilon$$

Suppose $|\beta| < 1$. If T is sub-gamma and independent of the network A, and $\min_{i \in [n]} d_i = \omega(\log n)$, then there exists constants $C, C' \in \mathbb{R}$ such that

$$\lim_{n \to \infty} \sup_{i \in [n]} |GT - 1_n C| = o_p(1), \qquad \textit{and}$$

$$\lim_{n \to \infty} \sup_{i \in [n]} \left| GY - 1_n C' \right| = o_p(1).$$

* This covers fixed networks and conditioning on the network

Theorem (Identification failure in random dot product graphs)

Let ε be mean zero, i.i.d. noise and let

$$Y = \alpha 1_n + \beta GY + \gamma T + \delta GT + \varepsilon$$

Suppose $|\beta| < 1$. Let $(A, X) \sim RDPG(F, n)$ and let $T = X \in \mathbb{R}^{n \times d}$. Then there exists constants $C, C' \in \mathbb{R}^d$ such that

$$\lim_{n\to\infty} \sup_{i\in[n]} |GT - 1_n C| = o_p(1), \qquad \text{and}$$

$$\lim_{n\to\infty} \sup_{i\in[n]} \left|GY - 1_n C'\right| = o_p(1).$$

Why this matters

1. Shalizi and Thomas (2011); McFowland and Shalizi (2021): homophily and contagion are non-parametrically confounded. Must make parametric assumptions to estimate contagion effects.
2. Bramoullé et al. (2020): most parametric contagion identification theory considers fixed n setting but not the $n \to \infty$ limit. Many results won't hold under stochastic blockmodels.

Open questions

- Do *GT* and *GY* ever not converge to constants?
- Are longitudinal contagion models also affected?
- What happens in sparse networks asymptotically?

Thank you! Questions?

Stay in touch

- @alexpghayes
- alex.hayes@wisc.edu
- *https://www.alexpghayes.com*
- *https://github.com/alexpghayes*

References

Bramoullé, Y., H. Djebbari, and B. Fortin (2009, May). Identification of peer effects through social networks. *Journal of Econometrics 150*(1), 41–55.

Bramoullé, Y., H. Djebbari, and B. Fortin (2020). Peer Effects in Networks: A Survey. *Annual Review of Economics 12*(1), 603–629.

Li, S. and S. Wager (2022, March). Random Graph Asymptotics for Treatment Effect Estimation under Network Interference. *arXiv:2007.13302 [math, stat]*.

McFowland, E. and C. R. Shalizi (2021, July). Estimating Causal Peer Influence in Homophilous Social Networks by Inferring Latent Locations. *Journal of the American Statistical Association 0*(0), 1–12.

Shalizi, C. R. and A. C. Thomas (2011, May). Homophily and Contagion Are Generically Confounded in Observational Social Network Studies. *Sociological Methods & Research 40*(2), 211–239.