Estimation and inference in network data

Haoyuan Xu

KU Leuven

September 2, 2024

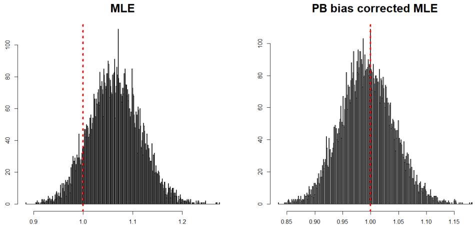
Motivation

- Networks are increasingly important as the world becomes more interconnected
  - *Systemic risk* in finance (e.g. 2007-2008 credit crunch and subsequent sovereign debt crisis) (Elliott et al., 2014).
  - *International trade* flows along trade links (Silva and Tenreyro, 2006).
  - *Production* network effect firms' performance (Bernard et al., 2019).
  - *Peer effects* run through networks of peers (Manski, 1993).
  - *R&D spillovers* across networked firms (Bloom et al., 2013).
  - *Risk-sharing* in developing countries (Ambrus et al., 2014).

- Why study network formation?
  - Example: inefficient production networks cause welfare losses.
  - Policy interventions may lead to more effective production networks.
  - But how? What determines the network formation? Network formation analysis is an important recent research area.

Network formation models

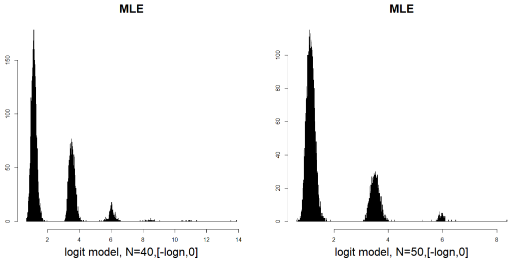
- *Degree heterogeneity* (few agents have many links; most agents have few links) is a common feature of social/economic networks. However, many determinants of link formation are unobserved.
- The *fixed effects* approach (Arellano and Bonhomme, 2011) can be used to capture unobserved heterogeneity.
- It leads to asymptotic bias and incorrect confidence interval due to the *incidental parameter problem* (Neyman and Scott, 1948).



- One novel solution (Part 1 of my proposal): the *parametric bootstrap* method (Higgins and Jochmans, 2024).

Sparse networks

- *Many economic networks are sparse*: most agents have only a few links with other agents.
- Estimators in sparse models often have *non-standard asymptotic behavior* (Matsushita and Otsu, 2023).
- In sparse network formation models, the distribution of maximum likelihood estimators exhibits *multimodality*.



- Incorporating *Unobserved heterogeneity* (Part 2 of my proposal) can explain degree heterogeneity and sparsity, but makes this even worse/more challenging.

Dyadic regression with dependence across dyads

- Dyadic regression has been widely used in the modern trade literature (Silva and Tenreyro, 2006).

$$y_{ij} = \mathbf{x}_{ij}'\beta + u_{ij}, \quad i \neq j \text{ and } i, j = 1, \ldots, G$$

Exporter	Importer $i = 1$	$i = 2$	$i = 3$	$\ldots$
$i = 1$	-	y_{12}	y_{13}	
$i = 2$	y_{21}	-	y_{23}	
$i = 3$	y_{31}	y_{32}	-	
$\vdots$				

- However, current methods only allow dyadic dependence, i.e $\mathrm{Cov}(u_{ij}, u_{kl}|\mathbf{x}) = 0$ unless $i = k$ or $i = l$ or $j = k$ or $j = l$ (Tabord-Meehan, 2019).
- This assumption is unrealistic. Example: The trade war between China and the US also influences trade between the EU and the rest of the world.

Beyond dyadic dependence

- Ignoring potential error dependencies induced through a network structure between dyads will underestimate the variance of estimators and lead to incorrect confidence intervals.
- Example: variance of the mean of error terms.

$$\mathrm{Var}\left(\frac{1}{G(G-1)}\sum_{i\neq j} u_{ij}\right)$$

$$=\frac{1}{G^2(G-1)^2}\left(\underbrace{\sum_{\{i,j\}\cap\{k,l\}\neq\emptyset}\mathrm{Cov}(u_{ij},u_{kl})}_{\text{dyadic dependence}}+\underbrace{\sum_{\{i,j\}\cap\{k,l\}=\emptyset}\mathrm{Cov}(u_{ij},u_{kl})}_{\text{network induced dependence}}\right)$$

- Part 3 of my proposal: develop weak dependence notions for dyadic/network data and provide valid inference methods.