Dyadic Double/Debiased Machine Learning for Analyzing Determinants of Free Trade Agreements

Harold Chiang Yukun Ma Joel Rodrigue Yuya Sasaki

Wisconsin-Madison Vanderbilt Vanderbilt Vanderbilt

October 28th, 2022
MEG 2022

Abstract

- For dyadic data, we develop a novel dyadic cross fitting algorithm to remove over-fitting biases under arbitrary dyadic dependence.
- **Dyadic data**, e.g.,
  - free/preferential trade agreement,
  - friendship, and
  - financial relationships, etc.
- **DML**[1] ⇒ generic method of estimation & inference for parametric, semi-parametric, high-dimensional models, etc. based on machine learning (ML).
- We illustrate an application of the general framework to high-dimensional network link formation models.
- We reconfirm that distance and the size of economics are two important determinants of FTA.

[1] Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, Newey, and Robins (2018)

Dyadic Data

Consider the sample $\{W_{ij} : 1 \leq i \leq N, 1 \leq j \leq N\}$.

- Assume the sample contains N nodes with no self link

$$i \neq j.$$

- Assume that

$$W_{ij} \perp\!\!\!\perp W_{i'j'}$$

unless $\{i, j\} \bigcap \{i', j'\} \neq \emptyset$.

$\cdots$ But if $\{i, j\} \bigcap \{i', j'\} \neq \emptyset$, then we allow for dependence.

- Notation:

$$\overline{\mathbb{N}^{+2}} := \{(i, j) \in \mathbb{N}^{+2} : i \neq j\}.$$

- An example: Free Trade Agreements

Free Trade Agreements (FTA)

Analyze the determinants of FTA,

- Consider the empirical model

$$\mathbf{E}_{P}[\boldsymbol{Y}_{ij}|\boldsymbol{D}_{ij},\boldsymbol{X}_{ij}] = \boldsymbol{\Lambda}(\boldsymbol{D}_{ij}\boldsymbol{\theta} + \boldsymbol{X}'_{ij}\boldsymbol{\beta}) \text{ for } (\boldsymbol{i},\boldsymbol{j}) \in \overline{[\boldsymbol{N}]^{2}}.$$

- Pioneering analysis of economic factors of FTA by Baier and Bergstrand (2004)
  - a greater distance between economics makes an FTA less beneficial ⇒ the population-weighted bilateral distance between i and j in kilometer.
  - larger sizes of economics make an FTA more beneficial ⇒ the sum of the logarithms of the per-capita GDP.
  - more similar economic sizes make an FTA more beneficial ⇒ the absolute difference of the logarithms of the per-capita GDP in baseline year.
  - wider relative factor endowments make an FTA more beneficial ⇒ the absolute difference of the logarithms of the capital-labor ratios in baseline year.

Double/Debiased Machine Learning (DML)

- Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, Newey, and Robins (CCDDHNR, 2018) provide a general DML toolbox of estimation & inference for parametric, semi-parametric, high-dimensional models, etc:

  DML ≈ Neyman Orthogonal Score + Cross-Fitting.

  - The former mitigates the slow convergence rates of ML-based estimates of nuisance parameters.
  - The latter removes the error induced by overfitting.
- i.i.d. sampling is crucial for cross-fitting.
  - Our dyadic sampling $\neq$ i.i.d.

Objective of the Paper

- We propose a novel dyadic cross-fitting algorithm and theories for estimation and inference using machine learning of nuisance parameters when data are dyadic.
- This objective is motivated by:
  - empirical applications that use dyadic data are lacking theoretical support (determinants of FTA).
  - recently growing interest in use of double/debiased machine learning methods of estimation and inference for high-dimensional models in today's big data environments.

Relations to the Literature

- Dyadic cluster robust variance formulas:
  - Fafchamps and Gubert (2007) propose dyadic cluster robust variance estimators for the OLS and logit.
  - Cameron and Miller (2014) generalize the dyadic cluster robust variance estimator for GMM and M-estimation.
- Asymptotic behavior:
  - Davezies, D'Haultfoeuille, and Guyonvarch (2019) study the asymptotic behavior of empirical process and their bootstrap counterparts of dyadic data.
  - Chiang, Kato and Sasaki (2020) develop methods of inference for high-dimensional parameters.
- Determinants of FTAs:
  - Baier and Bergstrand (2004) identify a parsimonious set of key economic determinants for the formation of free trade agreements: trade costs, the market size of the free trade zone, and the similarity of trading partners in terms of economic development and/or factor-endowments.

Table of Contents

Overview

Theory

Application to Dyadic Link Formation Models

Simulation

Empirical Application

Conclusions

Setup

- Assume

$$\mathbf{E}_{P}[\psi(W_{ij};\theta_0,\eta_0)] = 0.$$

- The nuisance parameter η may be finite-, high-, or infinite-dimensional. Its true value is denoted by $\eta_0 \in T$.
- Object of interest: the true value $\theta_0 \in \Theta$ of θ.
- Consider a linear score

$$\psi(w;\theta,\eta) = \psi^a(w;\eta)\theta + \psi^b(w;\eta)$$

  with

  - low-dimensional parameter vector $\theta \in \Theta \subset \mathbb{R}^{d_\theta}$.
  - nuisance parameter $\eta \in T$ for a convex set T.

Neyman Orthogonality Condition

- Path-wise derivative map D_r

$$D_r[\eta - \eta_0] := \partial_r \left\{ \mathbf{E}_P \left[\psi(W; \theta_0, \eta_0 + r(\eta - \eta_0)) \right] \right\} \quad \eta \in T$$

- Notation of $D_r[\eta - \eta_0]$ evaluated at $r = 0$:

$$\partial_\eta \mathbf{E}_P \left[\psi(W; \theta_0, \eta_0) \right] [\eta - \eta_0] := D_0 [\eta - \eta_0] \qquad \eta \in T$$

- The Neyman orthogonality condition holds at (θ_0, η_0) with respect to a nuisance realization set $\mathcal{T}_n \subset T$ if

$$\partial_\eta \mathbf{E}_P \psi(W; \theta_0, \eta_0)[\eta - \eta_0] = 0$$

  holds for all $\eta \in \mathcal{T}_n$.

- Can be generalized to near orthogonality.

Review of DML (CCDDHNR) under i.i.d Sampling

- Randomly partition $\{1, ..., N\}$ into K parts $\{I_1, ..., I_K\}$.

- For each $k \in \{1, ..., K\}$, obtain an estimate

$$\widehat{\eta}_k = \widehat{\eta}\left((W_i)_{i \in \{1,...,N\} \setminus I_k}\right)$$

  of the nuisance parameter η by some machine learner using only the subsample with $i \in \{1, ..., N\} \setminus I_k$.

- Define $\widetilde{\theta}$, the double/debiased machine learning (DML) estimator for θ_0, as the solution to

$$\frac{1}{K}\sum_{k=1}^{K} \mathbb{E}_{n,k}[\psi(W; \widetilde{\theta}, \widehat{\eta}_k)] = 0,$$

  where $\mathbb{E}_{n,k}[f(W)] = \frac{1}{|I_k|}\sum_{i \in I_k} f(W_i)$ denotes the subsample empirical mean using only data with $i \in I_k$.

DML (CCDDHNR) under i.i.d Sampling, Continued

Figure: An illustration of **2**-fold cross-fitting.

$\mathbb{E}_{n,1}[\psi(W;\widetilde{\theta},\widehat{\eta}_1)]$ $\qquad$ $\mathbb{E}_{n,2}[\psi(W;\widetilde{\theta},\widehat{\eta}_2)]$

$$\Downarrow$$

The DML estimator $\widetilde{\theta}$ is obtained by solving

$$\mathbb{E}_{n,1}[\psi(W;\widetilde{\theta},\widehat{\eta}_1)] + \mathbb{E}_{n,2}[\psi(W;\widetilde{\theta},\widehat{\eta}_2)] = 0$$

- If i.i.d. is violated (as in dyadic sampling), then **blue** and **red** subsamples are no longer independent.

Dyadic Cross Fitting

- Notations:
  - $[r] := \{1, ..., r\}$ for any $r \in \mathbb{N}$.
  - For any finite set I with $I \subset [N]$, $|I|$ denote the cardinality of I, and I^c denote the complement of I.
  - $\overline{\mathbb{N}^{+2}} = \{(i, j) \in \mathbb{N}^{+2} : i \neq j\}$ denote the set of two-tuple of $\mathbb{N}^+$ without repetition.

Dyadic Cross Fitting

- Randomly partition $[N]$ into K parts $\{I_1, ..., I_K\}$.

- For each $k \in [K]$, obtain an estimate

$$\widehat{\eta}_k = \widehat{\eta}\left((W_{ij})_{(i,j)\in\overline{([N]\setminus I_k)^2}}\right)$$

  of the nuisance parameter η by some machine learner using only the subsample with $(i,j) \in \overline{([N] \setminus I_k)^2}$.

- Define $\widetilde{\theta}$, the dyadic machine learning estimator for θ_0, as the solution to

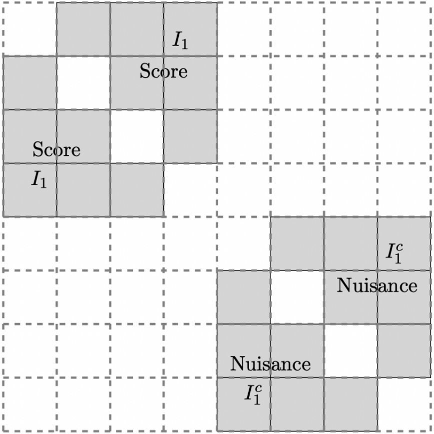
$$\frac{1}{K}\sum_{k\in[K]} \mathbb{E}_{|I_k|}[\psi(W;\widetilde{\theta},\widehat{\eta}_k)] = 0,$$

  where $\mathbb{E}_{|I_k|}[f(W)] = \frac{1}{|I_k|(|I_k|-1)}\sum_{(i,j)\in\overline{I_k^2}} f(W_{ij})$ denotes the subsample empirical mean using only data with $(i,j) \in \overline{I_k^2}$.

2-Fold Cross-Fitting under Dyadic Sampling

Figure: An illustration of **2**-fold cross fitting

$i, j \in \{1, 2, 3, 4\}$

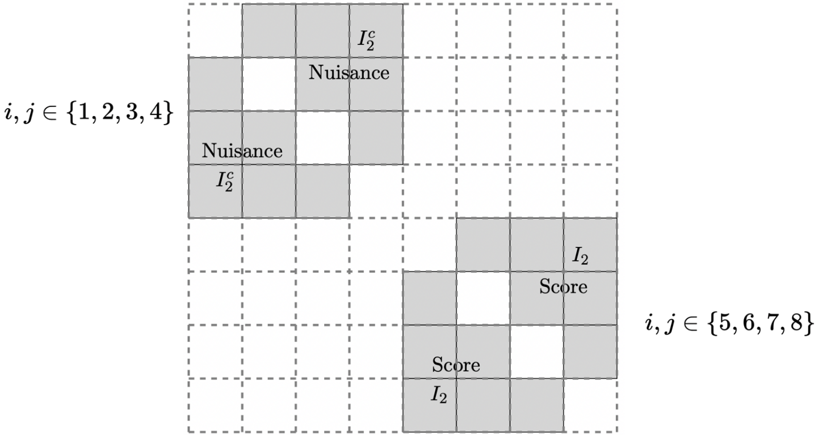


$i, j \in \{5, 6, 7, 8\}$

$$\mathbb{E}_{|I_1|}[\psi(W; \widetilde{\theta}, \widehat{\eta}_1)]$$

2-Fold Cross-Fitting under Dyadic Sampling

Figure: An illustration of **2**-fold cross fitting



$$\mathbb{E}_{|I_2|}[\psi(W; \widetilde{\theta}, \widehat{\eta}_2)]$$

Dyadic Cross Fitting

- We call this procedure **K-fold dyadic cross-fitting**.
- For each $k \in [K]$,
  - The nuisance parameter $\widehat{\eta}_k$ is computed using the subsample with $(i, j) \in ([N] \setminus I_k)^2$.
  - The score $\mathbb{E}_{|I_k|}[\psi(W; \cdot, \cdot)]$ is computed using the subsample with $(i, j) \in I_k^2$.
- This two-step computation is repeated K times for every partitioning pair $k \in [K]$.

Inference

- We propose to estimate the asymptotic variance of $\sqrt{N}(\widetilde{\theta} - \theta_0)$ by $\widehat{\sigma}^2 = \widehat{J}^{-1}\widehat{\Gamma}(\widehat{J}^{-1})'$, where

$$\widehat{J} = \frac{1}{K}\sum_{k\in[K]} \mathbb{E}_{|I_k|}[\psi^a(W;\widehat{\eta}_k)],$$

$$\widehat{\Gamma} = \frac{1}{K}\sum_{k\in[K]} \frac{|I_k|-1}{(|I_k|(|I_k|-1))^2}\Big[\sum_{i\in I_k}\sum_{\substack{j,j'\in I_k\\ j,j'\neq i}} \psi(W_{ij};\widetilde{\theta},\widehat{\eta}_k)\psi(W_{ij'};\widetilde{\theta},\widehat{\eta}_k)'$$

$$+\sum_{j\in I_k}\sum_{\substack{i,i'\in I_k\\ i,i'\neq j}} \psi(W_{ij};\widetilde{\theta},\widehat{\eta}_k)\psi(W_{i'j};\widetilde{\theta},\widehat{\eta}_k)'$$

$$+\sum_{i\in I_k}\sum_{\substack{j,j'\in I_k\\ j,j'\neq i}} \psi(W_{ij};\widetilde{\theta},\widehat{\eta}_k)\psi(W_{j'i};\widetilde{\theta},\widehat{\eta}_k)'$$

$$+\sum_{j\in I_k}\sum_{\substack{i,i'\in I_k\\ i,i'\neq j}} \psi(W_{ij};\widetilde{\theta},\widehat{\eta}_k)\psi(W_{ji'};\widetilde{\theta},\widehat{\eta}_k)'\Big].$$

- For a d_θ-dimensional vector r, the $(1-a)$ confidence interval for the linear functional $r'\theta_0$ can be constructed by

$$\mathrm{CI}_a := [r'\widetilde{\theta} \pm \Phi^{-1}(1-a/2)\sqrt{r'\widehat{\sigma}^2 r/N}].$$

Table of Contents

Overview

Theory

Application to Dyadic Link Formation Models

Simulation

Empirical Application

Conclusions

Notations

- Let $\boldsymbol{c_0 > 0}$, $\boldsymbol{c_1 > 0}$, $\boldsymbol{s > 0}$, $\boldsymbol{q \geq 4}$ be finite constants with $\boldsymbol{c_0 \leq c_1}$.

- Let $\boldsymbol{\{\delta_N\}_{N\geq 1}}$ (estimation errors), $\boldsymbol{\{\Delta_N\}_{N\geq 1}}$ (probability bounds) and $\boldsymbol{\{\tau_N\}_{N\geq 1}}$ be sequences of positive constants that converge to zero such that $\boldsymbol{\delta_N \geq N^{-1/2}}$.

- Let $\boldsymbol{K \geq 2}$ be a fixed integer.

Assumption Summary

- Linear Score
  - Sampling
  - Linear Neyman Orthogonal Score
  - Score Regularity and Nuisance Parameter Estimators
- Nonlinear and Nonseparable Score
  - Sampling
  - Nonlinear Moment Condition Problem with Approximate Neyman Orthogonality
  - Score Regularity and Nuisance Parameter Estimators

Assumption: Dyadic Sampling

Suppose $N \to \infty$. The following conditions hold.

(i) $(W_{ij})_{(i,j)\in\overline{\mathbb{N}^2}}$ is an infinite sequence of jointly exchangeable p-dimensional random vectors. That is, for any permutation π of $\mathbb{N}$, we have

$$(W_{ij})_{(i,j)\in\overline{\mathbb{N}^2}} \overset{d}{=} (W_{\pi(i)\pi(j)})_{(i,j)\in\overline{\mathbb{N}^2}}.$$

(ii) $(W_{ij})_{(i,j)\in\overline{\mathbb{N}^2}}$ is dissociated. That is, for any disjoint subsets A, B of $\mathbb{N}^+$, with $\min(|A|, |B|) \geq 2$, $(W_{ij})_{(i,j)\in\overline{A^2}}$ is independent of $(W_{ij})_{(i,j)\in\overline{B^2}}$.

Aldous-Hoover-Kallenberg representation

- Assumption 1 (i) & (ii) together imply the Aldous-Hoover-Kallenberg representation (e.g. Kallenberg, 2006; Corollary 7.35), which states that there exists an unknown (to the researcher) Borel measurable function $\boldsymbol{\tau_n}$ such that

$$\boldsymbol{W_{ij} \overset{d}{=} \tau_n(U_i, U_j, U_{\{i,j\}})},$$

  where $\boldsymbol{\{U_i, U_{\{i,j\}} : i, j \in [N], i \neq j\}}$ are some i.i.d. latent shocks that can be taken to be Unif$[\mathbf{0}, \mathbf{1}]$ without loss of generality – see Aldous (1981).
- We can exploit the independence over them.

Assumption: Linear Neyman Orthogonal Score

For $N \geq 4$ and $P \in \mathcal{P}_N$, the following conditions hold.

(i) The map $\eta \mapsto \mathrm{E}_P[\psi(W_{12};\theta,\eta)]$ is twice continuously Gateaux differentiable on T.

(ii) ψ satisfies either the Neyman orthogonality condition or the Neyman near orthogonality condition.

(iii) The identification condition holds; namely, the singular values of the matrix $J_0 := \mathrm{E}_P[\psi^a(W_{12};\eta_0)]$ are between c_0 and c_1.

Assumption: Score Regularity and Nuisance Parameter Estimators

For all $N \geq 4$ and $P \in \mathcal{P}_N$, the following conditions hold.

(i) Given random subsets $I \subset [N]$ such that $|I| = \lfloor N/K \rfloor$, the nuisance parameter estimator $\widehat{\eta} = \widehat{\eta}((W_{ij})_{(i,j) \in \overline{([N] \setminus I)^2}})$, belongs to $\mathcal{T}_N$ with probability $1 - \Delta_N$, where $\mathcal{T}_N$ contains η_0.

(ii) All eigenvalues of the matrix

$$\begin{aligned}\Gamma = {} & \mathrm{E}_P[\psi(W_{12};\theta_0,\eta_0)\psi(W_{13};\theta_0,\eta_0)] + \mathrm{E}_P[\psi(W_{12};\theta_0,\eta_0)\psi(W_{31};\theta_0,\eta_0)] \\ & + \mathrm{E}_P[\psi(W_{21};\theta_0,\eta_0)\psi(W_{13};\theta_0,\eta_0)] + \mathrm{E}_P[\psi(W_{21};\theta_0,\eta_0)\psi(W_{31};\theta_0,\eta_0)]\end{aligned}$$

are bounded from below by c_0.

Main Result

Suppose that the above assumptions are satisfied. If $\delta_N \geq N^{-1/2}$ for all $N \geq 4$, then

$$\sqrt{N}\sigma^{-1}(\widetilde{\theta} - \theta_0) = \frac{\sqrt{N}}{K} \sum_{k \in [K]} \mathbb{E}_{|I_k|} \bar{\psi}(W_{ij}) + O_{P_N}(\rho_N) \rightsquigarrow N(0, I_{d_\theta})$$

holds uniformly over $P \in \mathcal{P}_N$, where the size of the remainder terms follows

$$\rho_N := N^{-1/2} + r_N + r'_N + \underbrace{(N^{1/2}\lambda_N)}_{\substack{\text{Neyman Near} \\ \text{Orthogonal}}} + N^{1/2}\lambda'_N \lesssim \delta_N,$$

the influence function takes the form $\bar{\psi}(\cdot) := -\sigma^{-1} J_0^{-1} \psi(\cdot; \theta_0, \eta_0)$, and the approximate variance is given by

$$\sigma^2 := J_0^{-1} \Gamma (J_0^{-1})'.$$

Variance Estimation

Under the same set of assumptions as above,

$$\widehat{\boldsymbol{\sigma}}^{\mathbf{2}} = \boldsymbol{\sigma}^{\mathbf{2}} + O_{\mathbf{P}}(\boldsymbol{\rho}_{\boldsymbol{n}}).$$

Furthermore, the statement of the theorem in the previous slide holds true with $\widehat{\boldsymbol{\sigma}}^{\mathbf{2}}$ in place of $\boldsymbol{\sigma}^{\mathbf{2}}$.

Assumption: Nonlinear and Nonseparable Scores

For $N \geq 4$ and $P \in \mathcal{P}_N$, the following conditions hold.

(i) Θ contains a ball of radius $c_1 N^{-1/2} \log N$ centred at θ_0.

(ii) ψ satisfies either the Neyman orthogonality condition or the Neyman near orthogonality condition.

(iii) For all $\theta \in \Theta$, the identification relation

$$2||\mathrm{E}_P[\psi(W_{12}; \theta, \eta_0)]|| \geqslant ||J_0(\theta - \theta_0)|| \wedge c_0$$

is satisfied with the Jacobian matrix

$$J_0 := \partial_{\theta'} \{\mathrm{E}_P[\psi(W; \theta, \eta_0)]\}|_{\theta=\theta_0}$$

having singular values between c_0 and c_1.

Assumption: Score Regularity and Nuisance Parameter Estimators

For all $N \geq 4$ and $P \in \mathcal{P}_N$, the following conditions hold.

(i) The function class
$\mathcal{F}_1 = \{\psi_j(\cdot;\theta,\eta) : j = 1, ..., d_\theta, \theta \in \Theta, \eta \in \mathcal{T}_N\}$ is suitably measurable and its uniform entropy numbers satisfy

$$\sup_Q \log N(\mathcal{F}_1, \|\cdot\|_{Q,2}, \varepsilon\|F_1\|_{Q,2}) \leqslant v_N \log(a_N/\varepsilon), \text{ for all } 0 < \varepsilon \leq 1,$$

where F_1 is a measurable envelope for $\mathcal{F}_1$ that satisfies $\|F_1\|_{P,q} \leqslant K_N$.

(ii) All eigenvalues of the matrix

$$\Gamma = \mathrm{E}_P[\psi(W_{12};\theta_0,\eta_0)\psi(W_{13};\theta_0,\eta_0)] + \mathrm{E}_P[\psi(W_{12};\theta_0,\eta_0)\psi(W_{31};\theta_0,\eta_0)] + \mathrm{E}_P[\psi(W_{21};\theta_0,\eta_0)\psi(W_{13};\theta_0,\eta_0)] + \mathrm{E}_P[\psi(W_{21};\theta_0,\eta_0)\psi(W_{31};\theta_0,\eta_0)]$$

are bounded from below by c_0.

Main Result

Suppose that the above assumptions are satisfied. If $\delta_N \geq N^{-1/2+1/q} \log N$ and $N^{-1/2} \log N \leq \tau_N \leq \delta_N$ for all $N \geq 4$, then

$$\sqrt{N}\sigma^{-1}(\widetilde{\theta} - \theta_0) = \frac{\sqrt{N}}{K} \sum_{k \in [K]} \mathbb{E}_{|I_k|} \bar{\psi}(W_{ij}) + O_{P_N}(\rho_N) \rightsquigarrow N(0, I_{d_\theta})$$

holds uniformly over $P \in \mathcal{P}_N$, where the size of the remainder terms follows

$$\rho_N := N^{-1/2+1/q} + r'_N \log^{1/2}(1/r'_N) + N^{1/2}\lambda_N + N^{1/2}\lambda'_N \lesssim \delta_N,$$

the influence function takes the form $\bar{\psi}(\cdot) := -\sigma^{-1} J_0^{-1} \psi(\cdot; \theta_0, \eta_0)$, and the approximate variance is given by

$$\sigma^2 := J_0^{-1} \Gamma (J_0^{-1})'.$$

Variance Estimation

Under the same set of assumptions as above,

$$\widehat{\boldsymbol{\sigma}}^{\mathbf{2}} = \boldsymbol{\sigma}^{\mathbf{2}} + O_{\mathbf{P}}(\boldsymbol{\rho_n}),$$

for $\widehat{\boldsymbol{\sigma}}^{\mathbf{2}} = \widehat{\boldsymbol{J}}^{-\mathbf{1}}\widehat{\boldsymbol{\Gamma}}(\widehat{\boldsymbol{J}})^{-\mathbf{1}}$, where

$$\widehat{\boldsymbol{J}} := \frac{\mathbf{1}}{\boldsymbol{K}} \sum_{\boldsymbol{k} \in [\boldsymbol{K}]} \mathbb{E}_{|\boldsymbol{I_k}|}[\boldsymbol{\partial_\theta \psi(W; \theta, \widehat{\eta}_k)}|_{\boldsymbol{\theta = \widetilde{\theta}}}].$$

Our Approach to the Variance Formula

- The *Aldous-Hoover-Kallenberg representation*

  Under our sampling assumption, there exists a measurable function $\boldsymbol{\tau_n}$ such that

  $$\boldsymbol{W_{ij} = \tau_n(U_i, U_j, U_{\{i,j\}})} \text{ a.s.,}$$

  where $\boldsymbol{U}$'s are independent uniform $[\mathbf{0}, \mathbf{1}]$ random variables.

- The Hájek projection of $\mathbb{G}_{\boldsymbol{n}}\boldsymbol{f} = \frac{\sqrt{n}}{n(n-1)} \sum_{(i,j)\in\overline{[N]^2}} \boldsymbol{f(W_{ij})}$ on functions of each single $\boldsymbol{(U_l)_{l=1}^n}$ can be written as $\boldsymbol{H_n f} = \frac{\sqrt{n}}{n(n-1)} \sum_{l\in[n]} \{\sum_{j\neq l} \mathbf{E}_P[f(W_{lj})|U_l] + \sum_{i\neq l} \mathbf{E}_P[f(W_{il})|U_l]\}.$

Table of Contents

Overview

Theory

Application to Dyadic Link Formation Models

Simulation

Empirical Application

Conclusions

High-dimensional Logit Dyadic Link Formation Models

- Logit model

$$\mathbf{E}_P[Y_{ij}|D_{ij}, X_{ij}] = \Lambda(D_{ij}\theta_0 + X'_{ij}\beta_0) \text{ for } (i,j) \in \overline{[N]^2}$$

  where $\Lambda(t) = \frac{\exp(t)}{1+\exp(t)}$ for all $t \in \mathbb{R}$.

- The goal is to construct a generated random variable $Z_{ij} = Z(D_{ij}, X_{ij})$ such that

$$\mathbf{E}_P\left[\left\{Y_{ij} - \Lambda\left(D_{ij}\theta_0 + X'_{ij}\beta_0\right)\right\} Z_{ij}\right] = 0,$$

$$\frac{\partial}{\partial\theta}\mathbf{E}_P\left[\left\{Y_{ij} - \Lambda\left(D_{ij}\theta + X'_{ij}\beta_0\right)\right\} Z_{ij}\right]\Big|_{\theta=\theta_0} \neq 0,$$

$$\frac{\partial}{\partial\beta}\mathbf{E}_P\left[\left\{Y_{ij} - \Lambda\left(D_{ij}\theta_0 + X'_{ij}\beta\right)\right\} Z_{ij}\right]\Big|_{\beta=\beta_0} = 0.$$

High-dimensional Logit Dyadic Link Formation Models

- Consider the weighted regression of D_{ij} on X_{ij}: (Belloni, Chernozhukov, Wei (2016))

$$f_{ij} D_{ij} = f_{ij} X_{ij}' \gamma_0 + V_{ij}, \quad \text{with} \quad \mathrm{E}_P \left[f_{ij} V_{ij} X_{ij} \right] = 0,$$

where

$$f_{ij} := w_{ij} / \sigma_{ij}, \quad \sigma_{ij}^2 := \mathrm{Var} \left(Y_{ij} | D_{ij}, X_{ij} \right),$$

$$w_{ij} := \Lambda^{(1)} \left(D_{ij} \theta_0 + X_{ij}' \beta_0 \right), \text{ and } \Lambda^{(1)}(t) = \frac{\partial}{\partial t} \Lambda(t).$$

- The optimal generated random variable is given by

$$Z_{ij} := V_{ij} / \sigma_{ij}.$$

- Under the logit link Λ, f_{ij}, σ_{ij}^2, w_{ij} and Z_{ij} are given by

$$f_{ij}^2 = w_{ij},$$
$$\sigma_{ij}^2 = w_{ij} = \Lambda \left(D_{ij} \theta_0 + X_{ij}' \beta_0 \right) \left\{ 1 - \Lambda \left(D_{ij} \theta_0 + X_{ij}' \beta_0 \right) \right\}, \text{ and}$$
$$Z_{ij,0} = D_{ij} - X_{ij}' \gamma_0.$$

- The Neyman orthogonal score

$$\psi(W_{ij}; \theta, \eta) = \{ Y_{ij} - \Lambda(D_{ij} \theta + X_{ij}' \beta) \} (D_{ij} - X_{ij}' \gamma),$$

where $\eta = (\beta', \gamma')'$ denotes the nuisance parameters.

Algorithm

- Randomly partition $[N]$ into K parts $\{I_1, ..., I_K\}$.
- For each $k \in [K]$: obtain an post-lasso logistic estimate $(\widetilde{\theta}_k, \widetilde{\beta}_k)$ of the nuisance parameter by using only the subsample of those observations with dyadic indices (i, j) in $\overline{([N] \setminus I_k)^2}$,

$$(\widehat{\theta}_k, \widehat{\beta}_k) \in \arg\min_{\theta,\beta} \mathbb{E}_{|I_k^c|} [L(W_{ij}; \theta, \beta)] + \frac{\lambda_1}{|I_k^c|} \|(\theta, \beta)\|_1$$

$$(\widetilde{\theta}_k, \widetilde{\beta}_k) \in \arg\min_{\theta,\beta} \mathbb{E}_{|I_k^c|} [L(W_{ij}; \theta, \beta)] : \text{supp}(\theta, \beta) \subseteq \text{supp}(\widehat{\theta}_k, \widehat{\beta}_k).$$

- For each $k \in [K]$: calculate the weight $\widehat{f}^2_{ij,k} = \Lambda\left(D_{ij}\widetilde{\theta}_k + X'_{ij}\widetilde{\beta}_k\right)\left\{1 - \Lambda\left(D_{ij}\widetilde{\theta}_k + X'_{ij}\widetilde{\beta}_k\right)\right\}$ where $(i, j) \in \overline{([N] \setminus I_k)^2}$.

Algorithm, Continued

- For each $k \in [K]$: obtain an post-lasso OLS estimate $\widetilde{\gamma}_k$ of the nuisance parameter by using only the subsample of those observations with dyadic indices (i, j) in $\overline{([N] \setminus I_k)^2}$,

$$\widehat{\gamma}_k \in \arg\min_{\gamma} \mathbb{E}_{|I_k^c|} \left[\widehat{f}^2_{ij,k} \left(D_{ij} - X'_{ij}\gamma \right)^2 \right] + \frac{\lambda_2}{|I_k^c|} \|\gamma\|_1$$

$$\widetilde{\gamma}_k \in \arg\min_{\gamma} \mathbb{E}_{|I_k^c|} \left[\widehat{f}^2_{ij,k} \left(D_{ij} - X'_{ij}\gamma \right)^2 \right] : \quad \mathrm{supp}(\gamma) \subseteq \mathrm{supp}(\widehat{\gamma}_k).$$

- Solve the equation $\frac{1}{K} \sum_{k \in [K]} \mathbb{E}_{|I_k|}[\psi(W; \theta, \widetilde{\eta}_k)] = 0$ for θ to obtain the dyadic machine learning estimate $\check{\theta}$, with $\widetilde{\eta}_k = (\widetilde{\beta}'_k, \widetilde{\gamma}'_k)'$ and $(i, j) \in \overline{I_k^2}$.

Algorithm, Continued

- Let the dyadic Lasso DML asymtotic variance estimator be given by $\widehat{\sigma}^2 = \widehat{J}^{-1}\widehat{\Gamma}(\widehat{J}^{-1})'$ where

$$\widehat{J} = -\frac{1}{K}\sum_{k\in[K]} \mathbb{E}_{|I_k|}\Lambda(D_{ij}\check{\theta} + X_{ij}'\widetilde{\beta}_k)\{1 - \Lambda(D_{ij}\check{\theta} + X_{ij}'\widetilde{\beta}_k)\}(D_{ij} - X_{ij}'\widetilde{\gamma}_k)D_{ij},$$

$$\widehat{\Gamma} = \frac{1}{K}\sum_{k\in[K]} \frac{|I_k|-1}{(|I_k|(|I_k|-1))^2}\Big[\sum_{i\in I_k}\sum_{\substack{j,j'\in I_k\\ j,j'\neq i}} \psi(W_{ij};\check{\theta},\widetilde{\eta}_k)\psi(W_{ij'};\check{\theta},\widetilde{\eta}_k)'$$

$$+ \sum_{j\in I_k}\sum_{\substack{i,i'\in I_k\\ i,i'\neq j}} \psi(W_{ij};\check{\theta},\widetilde{\eta}_k)\psi(W_{i'j};\check{\theta},\widetilde{\eta}_k)' + \sum_{i\in I_k}\sum_{\substack{j,j'\in I_k\\ j,j'\neq i}} \psi(W_{ij};\check{\theta},\widetilde{\eta}_k)\psi(W_{j'i};\check{\theta},\widetilde{\eta}_k)'$$

$$+ \sum_{j\in I_k}\sum_{\substack{i,i'\in I_k\\ i,i'\neq j}} \psi(W_{ij};\check{\theta},\widetilde{\eta}_k)\psi(W_{ji'};\check{\theta},\widetilde{\eta}_k)'\Big].$$

- Report the estimate $\check{\theta}$, its standard error $\sqrt{\widehat{\sigma}^2/N}$, and/or the $(1-a)$ confidence interval

$$\mathrm{CI}_a := [\check{\theta} \pm \Phi^{-1}(1-a/2)\sqrt{\widehat{\sigma}^2/N}].$$

Table of Contents

Overview

Theory

Application to Dyadic Link Formation Models

Simulation

Empirical Application

Conclusions

DGP

For each $(i,j) \in \overline{[N]^2}$, generate the random vector $(D_{ij}, X'_{ij}, \varepsilon_{ij})'$ according to

$$D_{ij} = \frac{1}{3}\widetilde{D}_i + \frac{1}{3}\widetilde{D}_j + \frac{1}{3}\widetilde{D}_{ij},$$
$$X_{ij} = \frac{1}{3}\widetilde{X}_i + \frac{1}{3}\widetilde{X}_j + \frac{1}{3}\widetilde{X}_{ij},$$
$$\varepsilon_{ij} = F^{-1}_{\text{Logistic}(0,1)} \circ F_{\text{Normal}(0,1)}\left(\sqrt{\frac{1}{3}}\widetilde{\varepsilon}_i + \sqrt{\frac{1}{3}}\widetilde{\varepsilon}_j + \sqrt{\frac{1}{3}}\widetilde{\varepsilon}_{ij}\right),$$

Construct $Y_{ij} = \mathbb{1}\{D_{ij}\theta_0 + X'_{ij}\beta_0 \geq \varepsilon\}$.

Monte Carlo Simulation

Table: Simulation results based on 2,500 Monte Carlo iterations

Method	N	$\dim(X)$	K	True	Mean	Bias	SD	RMSE	Q25	Q50	Q75	90%	95%
Conventional ML	50	25	5	1.000	1.144	0.144	0.316	0.347	0.928	1.139	1.354	0.363	0.422
Conventional ML	100	50	5	1.000	1.148	0.148	0.220	0.265	0.999	1.141	1.293	0.233	0.277
Conventional ML	50	25	10	1.000	1.145	0.145	0.316	0.348	0.929	1.139	1.356	0.360	0.421
Conventional ML	100	50	10	1.000	1.149	0.149	0.220	0.265	0.998	1.143	1.294	0.234	0.274
Conventional ML	50	50	5	1.000	1.253	0.253	0.316	0.405	1.044	1.253	1.464	0.332	0.392
Conventional ML	100	100	5	1.000	1.252	0.252	0.222	0.336	1.100	1.248	1.404	0.170	0.209
Conventional ML	50	50	10	1.000	1.256	0.256	0.316	0.407	1.044	1.257	1.465	0.324	0.388
Conventional ML	100	100	10	1.000	1.254	0.254	0.222	0.338	1.102	1.249	1.406	0.171	0.206

Method	N	$\dim(X)$	K	True	Mean	Bias	SD	RMSE	Q25	Q50	Q75	90%	95%
Dyadic ML	50	25	5	1.000	1.059	0.059	0.470	0.474	0.746	1.052	1.369	0.908	0.950
Dyadic ML	100	50	5	1.000	1.045	0.045	0.288	0.292	0.848	1.040	1.236	0.901	0.946
Dyadic ML	50	25	10	1.000	1.047	0.047	0.518	0.520	0.688	1.034	1.394	0.922	0.965
Dyadic ML	100	50	10	1.000	1.039	0.039	0.303	0.305	0.825	1.042	1.239	0.916	0.957
Dyadic ML	50	50	5	1.000	1.113	0.113	0.522	0.534	0.736	1.111	1.460	0.897	0.953
Dyadic ML	100	100	5	1.000	1.105	0.105	0.304	0.322	0.903	1.105	1.314	0.883	0.940
Dyadic ML	50	50	10	1.000	1.114	0.114	0.582	0.593	0.715	1.101	1.497	0.908	0.964
Dyadic ML	100	100	10	1.000	1.095	0.095	0.320	0.334	0.880	1.093	1.316	0.908	0.958

Table of Contents

Overview

Theory

Application to Dyadic Link Formation Models

Simulation

Empirical Application

Conclusions

Empirical Application: Determinants of FTA

Table: Estimation and inference
based on 50 iterations of resampled cross fitting

Dependent variable:	Logit	Full Logit	Conventional ML		Dyadic ML	
free trade agreement	(I)	(II)	(III)	(IV)	(V)	(VI)
(A) Distance	-1.690	-1.358	-1.662	-1.660	-1.515	-1.762
	(0.046)	(0.075)	(0.081)	(0.079)	(0.111)	(0.115)
(B) Size (Sum of log GDP)	0.236	0.343	0.359	0.360	0.263	0.244
	(0.013)	(0.020)	(0.008)	(0.007)	(0.043)	(0.035)
(C) Similarity (Δ log GDP)	-0.003	-0.004	-0.004	-0.004	-0.001	0.001
	(0.015)	(0.018)	(0.014)	(0.014)	(0.002)	(0.002)
(D) Rel. Factor Endowments	0.231	0.187	-0.460	-0.460	-0.432	-0.396
(Δ log K/L)	(0.060)	(0.072)	(0.143)	(0.143)	(0.326)	(0.362)
Effective sample size	13,027	13,027	13,027	13,027	229	229
Dimension $\dim(D', X')'$	5	141	141	141	141	141
Number K of folds	N/A	N/A	5	10	5	10

Table of Contents

Overview

Theory

Application to Dyadic Link Formation Models

Simulation

Empirical Application

Conclusions

Conclusions

- For dyadic data, we develop a novel cross fitting algorithm to remove over-fitting biases under arbitrary dyadic dependence.
- This novel dyadic cross fitting method enables $\sqrt{N}$ consistent estimation and inference robustly against dyadic dependence.
- We illustrate an application of the general framework to high-dimension network link formation models.
- We confirm that trade costs and market size are key determinants of FTA formation.

Thank you!