Econometric Theory, **0**, 2025, 1–22.
doi:10.1017/S0266466625100273

DOUBLE/DEBIASED MACHINE LEARNING FOR DYADIC DATA

HAROLD D. CHIANG
University of Wisconsin-Madison

YUKUN MA
University of Rochester

JOEL B. RODRIGUE
Vanderbilt University

YUYA SASAKI
Vanderbilt University

This article presents novel methods and theories for estimation and inference about parameters in statistical models using machine learning for nuisance parameter estimation when data are dyadic. We propose a dyadic cross-fitting method to remove over-fitting biases under arbitrary dyadic dependence. Together with the use of Neyman orthogonal scores, this novel cross-fitting method enables root-n consistent estimation and inference robustly against dyadic dependence. We demonstrate its versatility by applying it to high-dimensional network formation models and reexamine the determinants of free trade agreements.

1. INTRODUCTION

Empirical research investigating network formation and outcomes spans the literature covering labor economics, industrial organization, macroeconomics, development, and international trade, among others. In a large fraction of these studies, the process through which entities are determined—e.g., free/preferential trade agreements, friendships, and financial relationships—concerns dyadic/network link formation models. While recent methodological advances have increased the popularity of empirical network formation models (see Graham (2020) and Graham and de Paula (2020) for surveys), the large majority of existing studies focus on parsimonious (low-dimensional) model specifications. The increasing availability of "big data" should, in principle, allow researchers to consider increasingly rich specifications and uncover new insights into the nature of network

We benefited from valuable comments by Peter C. B. Phillips (Editor), Ivan Fernandez-Val (Co-Editor), two referees, and participants in AMES 2021, ESAM 2021, ESEM 2021, IAAE 2021, NASMES 2021, and New York Camp Econometrics XV. Y.S. thanks Brian and Charlotte Grove Chair for research support. H.D.C. was supported by the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin-Madison with funding from the Wisconsin Alumni Research Foundation. Address correspondence to Yukun Ma, Department of Economics, University of Rochester, Rochester, United States, e-mail: yma69@ur.rochester.edu.

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0266466625100273 Published online by Cambridge University Press

formation. Rather, researchers are hamstrung by two competing limitations. On the one hand, there is no existing statistical methodology that allows for both dyadic robust inference and high-dimensional link formation specifications. In the presence of significant data heterogeneity in the policy variable of interest, one needs to control a large number of covariates. However, simply estimating a rich specification across a wide set of covariates is not yet possible. Moreover, we document that common off-the-shelf approaches, such as the conventional double/debiased machine learning (DML) for i.i.d. sampling, are likely to result in both biased estimates and misleading standard errors under dyadic dependence. On the other hand, the most common empirical approach, reducing the dimensionality of the network formation model by assumption, risks model misspecification, is a likely source of estimation bias and/or misleading inference, and reduces the set of questions researchers can investigate. In light of the increasing availability of big data today, we advance state-of-the-art approaches to estimating network models by developing novel methods for root-n consistent estimation and inference for high-dimensional dyadic regressions and high-dimensional dyadic/network link formation models.

Our work builds on the literature which suggests using (near) Neyman orthogonal scores to accommodate possibly slow convergence rates of general machine learners (e.g., Belloni et al. 2015, Belloni et al., 2018, Chernozhukov et al. 2018, 2022) in high-dimensional regression estimation (and more generally semiparametric Z-estimation). In particular, our approach extends the branch of research which employs cross-fitting, in conjunction with orthogonal scores, to mitigate over-fitting biases. This combined method is referred to as DML (Chernozhukov et al. 2018).

Standard DML requires independent sampling. In a recent work, Chiang et al. (2022) develop a multiway cross-fitting algorithm to extend the DML to multiway cluster dependent data with linear scores. Although conceptually related, the methodology in this article differs fundamentally from Chiang et al. (2022) in two respects. First, Chiang et al. (2022) does not permit the estimation of network models in the presence of dyadic dependence. Dyadic dependence is different from (multiway) cluster dependence,[1] and hence the existing cross-fitting algorithms are not guaranteed to work for dyadic regressions and dyadic/network link formation models. In this light, we propose a novel cross-fitting algorithm that is effective under dyadic dependence. Second, Chiang et al. (2022) do not cover estimators defined by nonlinear scores, such as Logit. Coping with such nonlinear structures presents extra theoretical challenges as it demands tight control of the stochastic magnitude of various empirical processes under dyadic dependence. Our dyadic machine learning approach and asymptotic theories cover a large class of general

[1] Specifically, while multiway clustered data are represented by separately exchangeable arrays, dyadic data are represented by jointly exchangeable arrays (see Section 7 in Kallenberg (2006)). The mathematical characterizations of these two types of exchangeable arrays are different, and the separate exchangeability does not imply the joint exchangeability.

https://doi.org/10.1017/S0266466625100273 Published online by Cambridge University Press

nonlinear estimation problems and guarantee that estimation and inference based on this method are robust against arbitrary dyadic dependence. See Appendix I of the Supplementary Material for further details.

Relation to the Literature: Due to space constraints, a comprehensive literature review is provided in Appendix H of the Supplementary Material.

2. THE DYADIC DOUBLE/DEBIASED MACHINE LEARNING

In this section, we introduce the model and present our proposed method of dyadic machine learning. A theoretical guarantee that this method works will be presented in Section 3. We start by fixing notations to be used to describe dyadic data. Assume that a dyadic sample contains N nodes with no self-links. For any $r \in \mathbb{N}$, let $[r] = \{1, \dots, r\}$. For any finite set I with $I \subset [N]$, we use the "bar-square" notation to denote the set of two-tuples of I without repetition, i.e., $\overline{I^2} = \{(i,j) \in I \times I : i \neq j\}$. A dyadic observation is written as W_{ij} for $(i,j) \in \overline{\mathbb{N}^{+2}}$, where $\mathbb{N}^+$ represents the set of positive integers, and we assume that this random vector W_{ij} is Borel measurable throughout.[2] We assume the identical distribution of W_{ij}. Let $\{\mathcal{P}_N\}_N$ be a sequence of sets of probability laws of $\{W_{ij}\}_{ij}$, let $P = P_N \in \mathcal{P}_N$ denote the law with the sample size N, and let E_P denote the expectation with respect to P. We write the dyadic sample expectation operator by $\mathbb{E}_{N C_2}[\cdot] = \frac{1}{N(N-1)} \sum_{(i,j) \in \overline{[N]^2}} [\cdot]$. For any finite set I with $I \subset [N]$, we let $|I|$ denote the cardinality of I, and let I^c denote the complement of I, namely, $I^c = [N] \setminus I$. We write the dyadic subsample expectation operator by $\mathbb{E}_{|I| C_2}[\cdot] := \frac{1}{|I|(|I|-1)} \sum_{(i,j) \in \overline{I^2}} [\cdot]$.

2.1. The Model

The economic model is assumed to satisfy the moment restriction

$$\mathrm{E}_P[\psi(W_{ij}; \theta_0, \eta_0)] = 0 \tag{2.1}$$

for some score function ψ that depends on a low-dimensional parameter vector $\theta \in \Theta \subset \mathbb{R}^{d_\theta}$ and a nuisance parameter $\eta \in T$ for a convex set T. The nuisance parameter η may be finite-, high-, or infinite-dimensional. The true values of θ and η are denoted by $\theta_0 \in \Theta$ and $\eta_0 \in T$, respectively.

Let $\widetilde{T} = \{\eta - \eta_0 : \eta \in T\}$, and define the Gâteaux derivative map $D_r : \widetilde{T} \to \mathbb{R}^{d_\theta}$ by $D_r[\eta - \eta_0] := \partial_r \left\{ \mathrm{E}_P[\psi(W_{ij}; \theta_0, \eta_0 + r(\eta - \eta_0))] \right\}$ for all $r \in [0,1)$. Let its limit, as $r \to 0$, be denoted by $\partial_\eta \mathrm{E}_P \psi(W_{ij}; \theta_0, \eta_0)[\eta - \eta_0] := D_0[\eta - \eta_0]$. We say that the Neyman orthogonality condition holds at (θ_0, η_0) with respect to a nuisance realization set $\mathcal{T}_N \subset T$ if the score ψ satisfies (2.1), the pathwise derivative $D_r[\eta - \eta_0]$ exists for all $r \in [0,1)$ and $\eta \in \mathcal{T}_N$, and the orthogonality equation

[2] Although we focus on a balanced array with N nodes, the method extends to an unbalanced setting. See Remark 2 ahead.

https://doi.org/10.1017/S0266466625100273 Published online by Cambridge University Press

$$\partial_\eta \mathrm{E}_P \psi(W_{ij}; \theta_0, \eta_0)[\eta - \eta_0] = 0 \tag{2.2}$$

holds for all $\eta \in \mathcal{T}_N$. Furthermore, we also say that the λ_N Neyman near-orthogonality condition holds at (θ_0, η_0) with respect to a nuisance realization set $\mathcal{T}_N \subset T$ if the score ψ satisfies (2.1), the pathwise derivative $D_r[\eta - \eta_0]$ exists for all $r \in [0,1)$ and $\eta \in \mathcal{T}_N$, and the orthogonality equation

$$\sup_{\eta \in \mathcal{T}_N} \left\| \partial_\eta \mathrm{E}_P \psi(W_{ij}; \theta_0, \eta_0)[\eta - \eta_0] \right\| \leq \lambda_N \tag{2.3}$$

holds for all $\eta \in \mathcal{T}_N$ for some positive sequence $\{\lambda_N\}_N$ such that $\lambda_N = o(N^{-1/2})$. We refer readers to Chernozhukov et al. (2018) for detailed discussions of the Neyman (near) orthogonality, including its intuitions and a general procedure to construct scores ψ satisfying it.

Throughout, we will consider structural models satisfying the moment restriction (2.1) and either form of the Neyman orthogonality conditions, (2.2) or (2.3). As a concrete motivating example of the above abstract formulation, Example 1 below presents the logit dyadic link formation models, which we consider for our main application in this article. In addition, we also present a couple of simpler examples with the linear regression models and the linear IV regression models in Appendix A.1 of the Supplementary Material.

Example 1 (Logit Dyadic Link Formation Models). Let $W_{ij} = (Y_{ij}, D_{ij}, X_{ij}')'$, where Y_{ij} is a binary indicator for a link formed between nodes i and j, D_{ij} is an explanatory variable of interest, and X_{ij} is a high-dimensional random vector of controls. Consider the logistic dyadic link formation model

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

where θ is a scalar parameter of interest,[3] β is a nuisance parameter, and $\Lambda(t) = \exp(t)/(1+\exp(t))$ for all $t \in \mathbb{R}$. A nonlinear Neyman orthogonal score ψ in this model is given by

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

where $\eta = (\beta', \gamma')'$ with γ denoting the coefficients of a weighted projection of D_{ij} on X_{ij}. See Section 4 for more details about this example as well as our general theory applied to this example.

2.2. Dyadic Cross-Fitting

For the class of models introduced in Section 2.1, we now propose a novel dyadic cross-fitting procedure for estimation of θ_0. With a fixed positive integer K, randomly partition the set $[N] = \{1, \ldots, N\}$ of indices of N nodes into K parts $\{I_1, \ldots, I_K\}$. For each $k \in [K] = \{1, \ldots, K\}$, obtain an estimate

[3] Though θ is a scalar in this particular example, we treat θ as a generic vector in the rest of the article.

https://doi.org/10.1017/S0266466625100273 Published online by Cambridge University Press

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

of the nuisance parameter η by a machine learning method (e.g., lasso, post-lasso, elastic nets, ridge, deep neural networks, and boosted trees) using only the subsample of those observations with dyadic indices (i,j) in $\overline{([N]\setminus I_k)^2}$. In turn, we define the dyadic machine learning estimator $\widetilde{\theta}$ for θ_0 as the solution to

$$\frac{1}{K}\sum_{k\in[K]} \mathbb{E}_{|I_k|C_2}[\psi(W;\widetilde{\theta},\widehat{\eta}_k)] = 0, \tag{2.4}$$

where we recall that $\mathbb{E}_{|I_k|C_2}[f(W)] = \frac{1}{|I_k|(|I_k|-1)}\sum_{(i,j)\in\overline{I_k^2}} f(W_{ij})$ denotes the subsample empirical expectation using only those observations with dyadic indices (i,j) in $\overline{I_k^2}$. If an achievement of the exact 0 is not possible as in (2.4), then one may define the estimator $\widetilde{\theta}$ as an approximate ϵ_N-solution:

$$\left\|\frac{1}{K}\sum_{k\in[K]} \mathbb{E}_{|I_k|C_2}[\psi(W;\widetilde{\theta},\widehat{\eta}_k)]\right\| \leq \inf_{\theta\in\Theta}\left\|\frac{1}{K}\sum_{k\in[K]} \mathbb{E}_{|I_k|C_2}[\psi(W;\theta,\widehat{\eta}_k)]\right\| + \epsilon_N, \tag{2.5}$$

where $\epsilon_N = o(\delta_N N^{-1/2})$ and restrictions on the sequence $\{\delta_N\}_N$ will be formally discussed in Section 3. Under the assumptions to be introduced in Section 3, this dyadic machine learning estimator $\widetilde{\theta}$ enjoys the root-N asymptotic normality $\sqrt{N}\sigma^{-1}(\widetilde{\theta}-\theta_0) \rightsquigarrow N(0, I_{d_\theta})$, where a concrete expression for the asymptotic variance σ^2 will be presented in Theorem 1 ahead.

Note that, for each $k \in [K]$, the nuisance parameter estimator $\widehat{\eta}_k$ is computed using the subsample of those observations with dyadic indices $(i,j) \in \overline{([N]\setminus I_k)^2}$, and in turn, the average score $\mathbb{E}_{|I_k|C_2}[\psi(W;\cdot,\widehat{\eta}_k)]$ is computed using the subsample of those observations with dyadic indices $(i,j) \in \overline{I_k^2}$. This two-step computation is repeated K times. We call this cross-fitting procedure the K-fold dyadic cross-fitting. It differs from and complements the cross-fitting procedure of Chernozhukov et al. (2018) for i.i.d. data and the cross-fitting procedure of Chiang et al. (2022) for multiway clustered data—neither of these existing cross-fitting methods will work under dyadic dependence. If the dyadic sample $\{W_{ij}\}_{ij}$ would reduce to the monadic sample $\{W_i\}_i$, then the K-fold dyadic cross-fitting would accordingly boil down to the K-fold cross-fitting procedure of Chernozhukov et al. (2018).

Figure 1 illustrates the dyadic cross-fitting for the case of $K = 2$. We let $N = 8$ for simplicity of illustration, although actual sample sizes should be much larger. Suppose that a random partition of $[8] = \{1,\ldots,8\}$ entails two folds with one consisting of $I_1 = \{1,\ldots,4\}$ and the other consisting of $I_2 = \{5,\ldots,8\}$. In the left panel, we compute $\widehat{\eta}_1$ with a machine learner applied to the subsample $\overline{(I_1^c)^2}$ marked by "Nuisance" in the gray shades at the bottom right quarter of the grid, and then evaluate the subsample mean $\mathbb{E}_{|I_1|C_2}[\psi(W;\widetilde{\theta},\widehat{\eta}_1)]$ of the score using the subsample $\overline{I_1^2}$ marked by "Score" in the gray shades at the top left quarter of the grid. In the right panel, in turn, we compute $\widehat{\eta}_2$ with a machine learner applied

https://doi.org/10.1017/S0266466625100273 Published online by Cambridge University Press

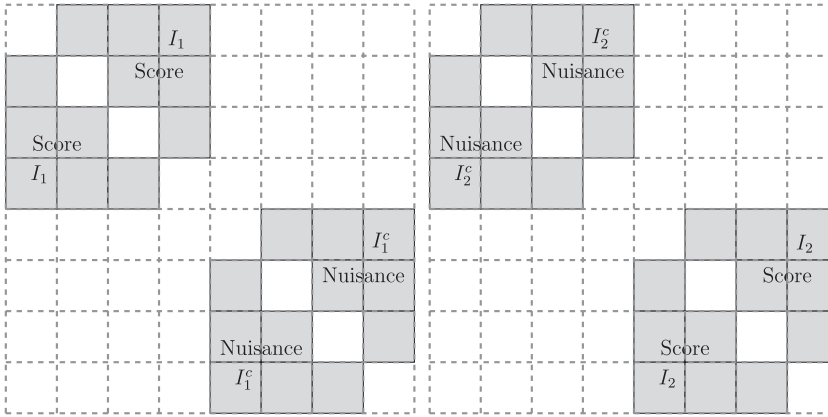


FIGURE 1. An illustration of the 2-fold dyadic cross-fitting.

to the subsample $\overline{(I_2^c)^2}$ marked by "Nuisance" in the gray shades at the top left quarter of the grid, and then evaluate the subsample mean $\mathbb{E}_{|I_2|C_2}[\psi(W;\widetilde{\theta},\widehat{\eta}_2)]$ of the score using the subsample $\overline{I_2^2}$ marked by "Score" in the gray shades at the bottom right quarter of the grid. Although the figure may apparently suggest as if we were discarding the information in the off-diagonal half of the dyad, our dyadic machine learning method uses the full information of N nodes for estimation of θ. Specifically, this means that Equation (2.4) evaluates the score with all of the N nodes (see Equation (3.3) ahead for a precise mathematical expression in support of this argument). The white off-diagonal blocks reflect that the $(K-1)/K$ fraction of data are used to estimate the nuisance parameter η_k for each $k \in [K]$, but this will not affect the asymptotic behavior of the estimator $\widetilde{\theta}$ of interest due to the Neyman orthogonality, as is the case with the conventional DML of Chernozhukov et al. (2018). As a matter of fact, the next section shows that our estimator $\widetilde{\theta}$ achieves the root-N asymptotic normality with the asymptotic variance being the same as the one in the case where an oracle would let us know the true nuisance parameter η.

Remark 1. In the context of dyadic data, the first-order asymptotic distribution is governed by information at the individual (node) level rather than at the dyad (pair) level. This is consistent with the convergence rate in dyadic asymptotics being root-N, as opposed to root-$N(N-1)$, reflecting that the effective sample size is determined by the number of nodes. Although estimation utilizes dyadic pairs, the leading influence function terms depend on node-specific information. Consequently, the fact that some dyadic pairs are excluded from particular folds during cross-fitting does not result in a first-order efficiency loss.

Remark 2 (Unbalanced dyads). We now consider the case of unbalanced dyads. For simplicity, let us focus on the case of undirected dyadic data, that is, dyads indexed by pairs with $i < j$. Suppose the data are unbalanced in the sense that we

https://doi.org/10.1017/S0266466625100273 Published online by Cambridge University Press

only observe dyads (i,j) for $i = 1,\dots,N$ and $j = 1,\dots,M$, where $N < M$, and both N and M diverge proportionally. The proposed method continues to apply in this setting, with the only required modification being an appropriate adjustment to the effective sample size. In such cases, we recommend using all available dyads rather than restricting attention to a balanced subset of $\binom{N}{2}$ observations. Retaining the full set of dyads yields an asymptotic variance that lies between $\frac{1}{4}$ and 1 times the variance in the balanced design, and hence can lead to efficiency gains.

3. THEORY

In this section, we present a formal theory that guarantees that the dyadic machine learning estimator introduced in Section 2.2 works. For convenience, we fix additional notations. Let $\{a_N\}_{N\geq 1}$, $\{v_N\}_{N\geq 1}$, $\{D_N\}_{N\geq 1}$, $\{B_{1N}\}_{N\geq 1}$, and $\{B_{2N}\}_{N\geq 1}$ be some sequences of positive constants, possibly growing to infinity, where $a_N \geq N \vee D_N$ and $v_N \geq 1$, $B_{1N} \geq 1$, $B_{2N} \geq 1$ for all $N \geq 1$. Let $\{\delta_N\}_{N\geq 1}$, $\{\Delta_N\}_{N\geq 1}$, and $\{\tau_N\}_{N\geq 1}$ be sequences of positive constants that converge to zero such that $\delta_N \geq N^{-1/2}$. Let c_0, c_1, C_0, and $q' > 2$ be strictly positive and finite constants. For sequences $a = a_N$, $b = b_N$, we write $a \lesssim b$ to mean that there exists a universal constant $c > 0$, independent of N, such that $a_N \leq c b_N$ holds for all sufficiently large N. Similarly, $a \gtrsim b$ denotes $a_N \geq c b_N$. We use the notations $a \vee b = \max\{a,b\}$ and $a \wedge b = \min\{a,b\}$. For any vector δ, we define the l_1-norm by $\|\delta\|_1$, l_2-norm by $\|\delta\|$, l_∞-norm by $\|\delta\|_\infty$, and l_0-seminorm (the number of nonzero components of δ) by $\|\delta\|_0$. For any function $f \in L^q(P)$, we define $\|f\|_{P,q} = (\int |f(w)|^q dP(w))^{1/q}$. We use $\|x_{ij}'\delta\|_{2,N}$ to denote the prediction norm of δ, namely, $\|x_{ij}'\delta\|_{2,N} = \sqrt{\mathbb{E}_{N C_2}[(x_{ij}'\delta)^2]}$.

The following assumption formally states the dyadic sampling under our consideration.

Assumption 1 (Sampling). Suppose that $N \to \infty$ and 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 $(W_{ij})_{(i,j)\in\overline{B^2}}$.

(iii) For each N, a statistician observes $(W_{ij})_{(i,j)\in\overline{[N]^2}}$.

Part (i) requires the identical distribution of W_{ij} while we do allow for dyadic dependence. Part (ii) requires that two groups of links in the dyad are independent whenever they do not share a common node. On the other hand, we do allow for arbitrary statistical dependence between any pair of links whenever they share a common node. Assumption 1 replaces the i.i.d. sampling assumption of Chernozhukov et al. (2018), and is a dyadic counterpart of Assumption 1 in Chiang et al. (2022) for multiway clustered data. We note that our exchangeability

https://doi.org/10.1017/S0266466625100273 Published online by Cambridge University Press

assumption excludes models with individual fixed effects, as the distribution of W_{ij} would then systematically depend on node identities. Assumption 1(i) and (ii) together imply the Aldous–Hoover–Kallenberg representation (e.g., Kallenberg, 2006, Cor. 7.35), which states that there exists an unknown (to the researcher) Borel measurable function κ_n such that

$$W_{ij} \overset{d}{=} \kappa_n(U_i, U_j, U_{\{i,j\}}), \tag{3.1}$$

where $\{U_i, U_{\{i,j\}} : i,j \in [N], i \neq j\}$ are some i.i.d. latent shocks that can be taken to be Unif$[0,1]$ without loss of generality (see Aldous, 1981). This is a mathematically different structural representation (Aldous–Hoover–Kallenberg representation; see, e.g., Kallenberg, 2006, Thm. 7.22 and Cor. 7.35) from that in either of these preceding papers and thus none of the existing DML methods apply here (see Chiang et al. (2023, Appx. I) for a detailed comparison and discussion of the differences). Finally, we remark that monadic variables (i.e., coordinates of W_{ij} that do not vary with i or j) are not ruled out from our sampling framework of Assumption 1.

Remark 3. While our methodology is designed for directional dyadic data, it can be adapted for undirectional data. Specifically, given undirectional data $(W_{ij})_{i<j}$, define $\widetilde{W}_{ij} = (W_{ij})_{i<j}$ and $\widetilde{W}_{ij} = (W_{ji})_{i>j}$. Then, we effectively convert the undirectional data into a form that can be treated as directional, allowing us to apply our existing methodology without any loss of generality. We can verify that $\widetilde{W}_{ij}$ satisfies the jointly exchangeable condition, as the order of i and j does not matter—only the set of indices does. This ensures that our existing algorithm remains applicable and that the methodology holds for undirectional data.

We further state the following two conditions.

Assumption 2 (Nonlinear Moment Condition Problem with Approximate Neyman Orthogonality). The following conditions hold for all $P \in \mathcal{P}_N$ for all $N \geq 4$.

(i) θ_0 is in the interior of Θ, where Θ is a convex set.
(ii) The map $(\theta, \eta) \mapsto \mathrm{E}_P[\psi(W_{12}; \theta, \eta)]$ is twice continuously Gâteaux differentiable on $\Theta \times \mathcal{T}_N$.
(iii) For all $\theta \in \Theta$, the identification relation

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

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.
(iv) The score ψ satisfies the Neyman orthogonality or, more generally, the Neyman near-orthogonality with $\lambda_N = \delta_N N^{-1/2}$ for the set $\mathcal{T}_N \subset T$.

https://doi.org/10.1017/S0266466625100273 Published online by Cambridge University Press

Assumption 3 (Score Regularity and Requirement on the Quality of Estimation of Nuisance Parameters). Let K be a fixed integer. The following conditions hold for all $P \in \mathcal{P}_N$ for all $N \geq 4$.

(i) The function class $\mathcal{F}_1 = \{\psi(\cdot;\theta,\eta) : \theta \in \Theta, \eta \in \mathcal{T}_N\}$ is pointwise measurable[4] and its uniform entropy numbers[5] satisfy

$$\sup_{Q \in \mathcal{Q}} \log N(\mathcal{F}_1, \|\cdot\|_{Q,2}, \varepsilon \|F_1\|_{Q,2}) \leq v_N \log(a_N/\varepsilon), \text{ for all } 0 < \varepsilon \leq 1, \quad \textbf{(3.2)}$$

where F_1 is a measurable envelope for $\mathcal{F}_1$ satisfying $\|F_1\|_{P,q'} \leq D_N$, and $\mathcal{Q}$ denotes the set of probability measures Q such that $QF_1^2 < \infty$.

(ii) For all $f \in \mathcal{F}_1$, we have $c_0 \leq \|f\|_{P,2} \leq C_0$.

(iii) 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 the realization set $\mathcal{T}_N$ with probability at least $1 - \Delta_N$, where $\mathcal{T}_N$ contains η_0.

(iv) The following conditions hold for all $r \in [0,1)$, $\theta \in \Theta$, and $\eta \in \mathcal{T}_N$:
  (a) $\mathrm{E}_P\big[\|\psi(W_{12};\theta,\eta) - \psi(W_{12};\theta_0,\eta_0)\|^2\big] \lesssim (\|\theta - \theta_0\| \vee \|\eta - \eta_0\|)^2$;
  (b) $\big\|\partial_r \mathrm{E}_P\big[\psi\big(W_{12};\theta,\eta_0 + r(\eta - \eta_0)\big)\big]\big\| \leq B_{1N}\|\eta - \eta_0\|$;
  (c) $\big\|\partial_r^2 \mathrm{E}_P\big[\psi\big(W_{12};\theta_0 + r(\theta - \theta_0),\eta + r(\eta - \eta_0)\big)\big]\big\| \leq B_{2N}(\|\theta - \theta_0\|^2 \vee \|\eta - \eta_0\|^2)$.

(v) For all $\eta \in \mathcal{T}_N$, $\|\eta - \eta_0\| \leq \tau_N$.

(vi) The matrix

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

is positive definite.

(vii) a_N and v_N satisfy:
  (a) $N^{-1/2}(v_N \log a_N)^{1/2} \leq C_0\tau_N$;
  (b) $N^{-1/2+1/q} v_N D_N \log a_N \leq C_0\delta_N$;
  (c) $N^{1/2} B_{1N}^2 B_{2N} \tau_N^2 \leq C_0\delta_N$.

Assumption 3(i) is a high-level condition adapted from Assumption 3.4(b) in Chernozhukov et al. (2018), controlling the complexity of the score function class through uniform entropy numbers. This assumption is formulated analogously in our dyadic context to ensure the empirical-process maximal inequalities required under dyadic dependence. The last two assumptions are counterparts of Assumptions 2.1 and 2.2 in Belloni et al. (2018), except that the data in this article are allowed to exhibit dyadic dependence. One observation is that although there are multiple N-dependent sequences of constants appearing in these assumptions that might be, at first glimpse, interacting in complicated ways, their rates are pinned

[4] See van der Vaart and Wellner (1996, pp. 110) for its definition.

[5] See van der Vaart and Wellner (1996, Chap. 2.6) for its definition.

https://doi.org/10.1017/S0266466625100273 Published online by Cambridge University Press

down by δ_N in Assumption 3(vii), analogously to the high level conditions in Belloni et al. (2018) and Chernozhukov et al. (2018). Although these conditions are high-level statements, we provide explicit lower-level primitive conditions that imply these assumptions in the context of the logit dyadic link formation model (Example 1) in Section 4.2. In particular, Assumption 3(vi) rules out degeneracy by ensuring that the long-run variance matrix is positive definite.

In Assumption 3, v_N scales the logarithm term, determining the growth rate of the uniform entropy bound, while a_N normalizes the complexity based on the sample size. In the application to the high-dimensional logit dyadic link formation models, for instance, we set $v_N = cs_N$ and $a_N = p \vee N$, with s_N denoting the number of nonzero components in the nuisance parameters. Second, τ_N represents the convergence rates of the nuisance parameter estimators. In the case of using the Post-Lasso OLS/Logit model to estimate the parameters, for instance, we can set $\tau_N = C(s_N \log a_N/N)^{1/2}$ (see Appendix D of the Supplementary Material). Finally, B_{1N} and B_{2N} constrain the change in first and second derivatives relative to the difference between η and η_0.

While W denotes a generic dyadic observation, W_{0i} and W_{i0} specifically represent dyadic observations involving node i, paired with a generic other node. This notation highlights the node-specific indexing structure inherent in dyadic data. With these notations, the following theorem establishes the root-N asymptotic normality of the dyadic DML estimator $\widehat{\theta}$ along with its consistent asymptotic variance estimation.

THEOREM 1 (Dyadic DML for Nonlinear Scores). *Suppose that Assumptions 1–3 hold. Then,*

$$\begin{aligned}\sqrt{N}\sigma^{-1}(\widetilde{\theta} - \theta_0) &= \frac{\sqrt{N}}{K}\sum_{k\in[K]} \mathbb{E}_{|I_k|C_2}\bar{\psi}(W) + O_{P_N}(\rho_N) \\ &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left\{\mathrm{E}_P[\bar{\psi}(W_{i0}) \mid U_i] + \mathrm{E}_P[\bar{\psi}(W_{0i}) \mid U_i]\right\} + O_{P_N}(\rho_N) \rightsquigarrow N(0, I_{d_\theta})\end{aligned} \tag{3.3}$$

holds uniformly over $P \in \mathcal{P}_N$, where the order of remainder follows $\rho_N \lesssim \delta_N$, the influence function takes the form $\bar{\psi}(\cdot) := -\sigma^{-1}J_0^{-1}\psi(\cdot;\theta_0,\eta_0)$, and the variance is given by

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

Moreover, σ^2 can be replaced by $\widehat{\sigma}^2 = \widehat{J}^{-1}\widehat{\Gamma}(\widehat{J}^{-1})'$, where

$$\widehat{J} := \frac{1}{K}\sum_{k\in[K]} \mathbb{E}_{|I_k|C_2}[\partial_\theta \psi(W;\theta,\widehat{\eta}_k)]|_{\theta=\widetilde{\theta}}$$

https://doi.org/10.1017/S0266466625100273 Published online by Cambridge University Press

and

$$\widehat{\Gamma} := \frac{1}{K}\sum_{k\in[K]}\frac{|I_k|-1}{(|I_k|(|I_k|-1))^2}\Big[\sum_{i\in I_k}\sum_{\substack{j,j'\in I_k\\ j,j'\neq i}}\psi(W_{ij};\widetilde{\theta},\widehat{\eta}_k)\psi(W_{ij'};\widetilde{\theta},\widehat{\eta}_k)' \tag{3.4}$$

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

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

The asymptotic distribution is different from both that of the prototypical DML (Chernozhukov et al., 2018) for i.i.d. data and that of the multi-way cluster-robust DML (Chiang et al., 2022). In addition to the differences in dependence structures, a yet major departure from Theorem 1 in Chiang et al. (2022) is that this result allows a large class of nonlinear score, whereas only linear scores are permitted in Chiang et al. To handle the complications resulted from the nonlinearity, it is necessary to handle empirical processes that consist of dyadic random elements. Coping with these technical issues requires delicate work and hence Theorem 1 is not a straightforward extension of results in Chiang et al. (2022).

Remark 4. While our results establish $\sqrt{N}$-consistency and asymptotic normality for dyadic DML estimators, constructing efficient orthogonal moments under dyadic dependence remains an open question. Chernozhukov et al. (2018) propose an approach for building locally efficient orthogonal scores under i.i.d. sampling through projections onto tangent spaces orthogonal to nuisance parameters. Extending their efficiency theory to the dyadic context would require deriving appropriate semiparametric efficiency bounds that explicitly account for two-way dependence, a task beyond the scope of the current article but an important topic for future research.

4. APPLICATION TO DYADIC LINK FORMATION MODELS

4.1. The High-Dimensional Logit Dyadic Link Formation Models

In this section, we demonstrate an application of the general theory of dyadic machine learning from Section 3 to high-dimensional dyadic/network link formation models. Consider a class of logit dyadic link formation models where the binary outcome Y_{ij}, which indicates a link formed between nodes i and j, is chosen based on an explanatory variable D_{ij} and p-dimensional control variables X_{ij}. Specifically, we write the model

https://doi.org/10.1017/S0266466625100273 Published online by Cambridge University Press

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

where $\Lambda(t) = \exp(t)/(1+\exp(t))$ for all $t \in \mathbb{R}$. Suppose that we are interested in the parameter θ, while $x \mapsto x'\beta_0$ is treated as a nuisance function η.

Under i.i.d. sampling environments, Belloni, Chernozhukov, and Wei (2016) propose a method of inference for θ_0 in high-dimensional logit models. We demonstrate that a modification of their suggested procedure with a flavor of our proposed dyadic cross-fitting enables root-N consistent estimation and asymptotically valid inference for θ_0 under dyadic sampling environments by virtue of Theorem 1. In Section 4.2, we will supplement this procedure with a formal discussion of lower-level primitive conditions for the high-level statements in Assumptions 2 and 3 that are invoked in Theorem 1.

As in Belloni et al. (2016), our goal is to construct a generated random variable $Z_{ij} = Z(D_{ij}, X_{ij})$ such that the following three conditions are satisfied:

$$\mathrm{E}_P\left[\left\{Y_{ij} - \Lambda\left(D_{ij}\theta_0 + X_{ij}'\beta_0\right)\right\} Z_{ij}\right] = 0, \tag{4.1}$$

$$\frac{\partial}{\partial\theta}\mathrm{E}_P\left[\left\{Y_{ij} - \Lambda\left(D_{ij}\theta + X_{ij}'\beta_0\right)\right\} Z_{ij}\right]\Big|_{\theta=\theta_0} \neq 0, \qquad \text{and} \tag{4.2}$$

$$\frac{\partial}{\partial\beta}\mathrm{E}_P\left[\left\{Y_{ij} - \Lambda\left(D_{ij}\theta_0 + X_{ij}'\beta\right)\right\} Z_{ij}\right]\Big|_{\beta=\beta_0} = 0. \tag{4.3}$$

Equations (4.1) and (4.2) are used for consistent estimation of θ_0, while Equation (4.3) obeys the Neyman orthogonality condition (see Section 2.1).

To this end, following Belloni et al. (2016), consider the weighted regression of D_{ij} on X_{ij}:

$$f_{ij}D_{ij} = f_{ij}X_{ij}'\gamma_0 + V_{ij}, \qquad \text{with} \qquad \mathrm{E}_P\left[f_{ij}V_{ij}X_{ij}\right] = 0, \tag{4.4}$$

where $f_{ij} := w_{ij}/\sigma_{ij}$, $\sigma_{ij}^2 := \mathrm{V}(Y_{ij}|D_{ij},X_{ij})$, $w_{ij} := \Lambda^{(1)}(D_{ij}\theta_0 + X_{ij}'\beta_0)$, and $\Lambda^{(1)}(t) = \frac{\partial}{\partial t}\Lambda(t)$. Under the logit link Λ, these objects are specifically given by $f_{ij}^2 = w_{ij}$ and $\sigma_{ij}^2 = w_{ij} = \Lambda(D_{ij}\theta_0 + X_{ij}'\beta_0)\{1 - \Lambda(D_{ij}\theta_0 + X_{ij}'\beta_0)\}$. We let $Z_{ij} = D_{ij} - X_{ij}'\gamma_0$.

From the log-likelihood function of the logit model $L(\theta,\beta) = \mathbb{E}_{N C_2}\left[L(W_{ij};\theta,\beta)\right]$, where $L(W_{ij};\theta,\beta) = \log\{1+\exp(D_{ij}\theta + X_{ij}'\beta)\} - Y_{ij}(D_{ij}\theta + X_{ij}'\beta)$, one can derive the following Neyman orthogonal score:

$$\psi(W_{ij};\theta,\eta) = \{Y_{ij} - \Lambda(D_{ij}\theta + X_{ij}'\beta)\}(D_{ij} - X_{ij}'\gamma), \tag{4.5}$$

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

With this procedure, applying Theorem 1 yields the asymptotic normality result

$$\sqrt{N}\sigma^{-1}(\widetilde{\theta} - \theta_0) \rightsquigarrow N(0,1), \tag{4.6}$$

https://doi.org/10.1017/S0266466625100273 Published online by Cambridge University Press

where the components of the variance $\sigma^2 := J_0^{-1}\Gamma(J_0^{-1})'$ take the forms of

$$J_0 = \mathrm{E}_P[-D_{ij}\Lambda(D_{ij}\theta_0 + X'_{ij}\beta_0)\{1-\Lambda(D_{ij}\theta_0 + X'_{ij}\beta_0)\}(D_{ij} - X'_{ij}\gamma_0)] \quad \text{and} \tag{4.7}$$

$$\begin{aligned}\Gamma = {} & \mathrm{E}_P[\{Y_{12} - \Lambda(D_{12}\theta_0 + X'_{12}\beta_0)\}(D_{12} - X'_{12}\gamma_0)\{Y_{13} - \Lambda(D_{13}\theta_0 + X'_{13}\beta_0)\}(D_{13} - X'_{13}\gamma_0)] \\ & + \mathrm{E}_P[\{Y_{12} - \Lambda(D_{12}\theta_0 + X'_{12}\beta_0)\}(D_{12} - X'_{12}\gamma_0)\{Y_{31} - \Lambda(D_{31}\theta_0 + X'_{31}\beta_0)\}(D_{31} - X'_{31}\gamma_0)] \\ & + \mathrm{E}_P[\{Y_{21} - \Lambda(D_{21}\theta_0 + X'_{21}\beta_0)\}(D_{21} - X'_{21}\gamma_0)\{Y_{13} - \Lambda(D_{13}\theta_0 + X'_{13}\beta_0)\}(D_{13} - X'_{13}\gamma_0)] \\ & + \mathrm{E}_P[\{Y_{21} - \Lambda(D_{21}\theta_0 + X'_{21}\beta_0)\}(D_{21} - X'_{21}\gamma_0)\{Y_{31} - \Lambda(D_{31}\theta_0 + X'_{31}\beta_0)\}(D_{31} - X'_{31}\gamma_0)].\end{aligned} \tag{4.8}$$

Section 4.2 will present formal theories to guarantee that this application of Theorem 1 is valid, based on lower-level sufficient conditions tailored to the current specific model.

Summarizing the above procedure, we provide a step-by-step algorithm below for dyadic machine learning estimation and inference about θ_0. We remind readers of the notation for the dyadic subsample expectation operator $\mathbb{E}_{|I|C_2}[\cdot] := \frac{1}{|I|(|I|-1)}\sum_{(i,j)\in\overline{I^2}}[\cdot]$.

Algorithm 1. *K*-fold Dyadic DML for High-Dimensional Logit Dyadic Link Formation Models

1. Randomly partition $[N]$ into K parts $\{I_1, ..., I_K\}$.
2. For each $k \in [K]$: obtain a 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 $\overline{([N]\setminus I_k)^2}$,

$$(\widehat{\theta}_k, \widehat{\beta}_k) \in \arg\min_{\theta,\beta} \mathbb{E}_{|I_k^c|C_2}\left[L(W_{ij};\theta,\beta)\right] + \frac{\lambda_1}{|I_k^c|}\|(\theta,\beta)\|_1,$$

$$(\widetilde{\theta}_k, \widetilde{\beta}_k) \in \arg\min_{\theta,\beta} \mathbb{E}_{|I_k^c|C_2}\left[L(W_{ij};\theta,\beta)\right] : \mathrm{support}(\theta,\beta) \subseteq \mathrm{support}(\widehat{\theta}_k, \widehat{\beta}_k).$$

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

$$\widehat{\gamma}_k \in \arg\min_{\gamma} \mathbb{E}_{|I_k^c|C_2}\left[\widehat{f}^2_{ij,k}\left(D_{ij} - X'_{ij}\gamma\right)^2\right] + \frac{\lambda_2}{|I_k^c|}\|\gamma\|_1,$$

$$\widetilde{\gamma}_k \in \arg\min_{\gamma} \mathbb{E}_{|I_k^c|C_2}\left[\widehat{f}^2_{ij,k}\left(D_{ij} - X'_{ij}\gamma\right)^2\right] : \quad \mathrm{support}(\gamma) \subseteq \mathrm{support}(\widehat{\gamma}_k).$$

5. Solve the equation $\frac{1}{K}\sum_{k\in[K]} \mathbb{E}_{|I_k|C_2}[\psi(W;\theta,\widetilde{\eta}_k)] = 0$ for θ to obtain the dyadic machine learning estimate $\check{\theta}$, where $\psi(W;\theta,\widetilde{\eta}_k)$ takes the form in Equation (4.5) with $\widetilde{\eta}_k = (\widetilde{\beta}'_k, \widetilde{\gamma}'_k)'$ and $(i,j) \in \overline{I_k^2}$.

https://doi.org/10.1017/S0266466625100273 Published online by Cambridge University Press

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

$$\widehat{J} = -\frac{1}{K}\sum_{k\in[K]} \mathbb{E}_{|I_k|C_2}\big[\Lambda(D_{ij}\check{\theta} + X_{ij}'\widetilde{\beta}_k)\{1-\Lambda(D_{ij}\check{\theta} + X_{ij}'\widetilde{\beta}_k)\}(D_{ij} - X_{ij}'\widetilde{\gamma}_k)D_{ij}\big],$$

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

7. Report the estimate $\check{\theta}$, its standard error $\sqrt{\widehat{\sigma}^2/N}$, and/or the $(1-a)$ confidence interval $\mathrm{CI}_a := [\check{\theta} \pm \Phi^{-1}(1-a/2)\sqrt{\widehat{\sigma}^2/N}]$.

Remark 5. There are a couple of technical and practical reasons why we opted for a logit model, instead of probit model in this study. First, the probit model presents greater challenges for Neyman orthogonalization, particularly in the context of high-dimensional binary regression. This complexity has led to a lack of papers in the literature utilizing the probit model in such settings. Second, mainstream software packages for regularized estimation of high-dimensional models, such as glmnet in R, do not support the probit model; they are primarily designed for the logit model. As a result, interpreting the probit model in high-dimensional binary regression is more difficult, which is likely why current high-dimensional packages focus exclusively on the logit model.

4.2. Lower-Level Sufficient Conditions

In this section, we provide lower-level sufficient conditions that guarantee the root-N asymptotic normality (4.6) in the application to the high-dimensional logit dyadic link formation models. For convenience of stating such conditions, we first introduce additional notations. We continue to use $Z_{ij} = D_{ij} - X_{ij}'\gamma_0$ from the previous section. Let $a_N = p \vee N$. Let q, c_1, and C_1 be strictly positive and finite constants with $q > 4$, M_N be a positive sequence of constants such that $M_N \geq \{\mathrm{E}_P[(D_{12} \vee \|X_{12}\|_\infty)^{2q}]\}^{1/2q}$. For any $T \subset [p+1]$, $\delta = (\delta_1, \ldots, \delta_{p+1})' \in \mathbb{R}^{p+1}$, denote $\delta_T = (\delta_{T,1}, \ldots, \delta_{T,p+1})'$ with $\delta_{T,j} = \delta_j$ if $j \in T$ and $\delta_{T,j} = 0$ if $j \notin T$. Define the minimum and maximum sparse eigenvalues by

$$\phi_{\min}(m) = \inf_{\|\delta\|_0 \leq m} \frac{\|\sqrt{w_{ij}}(D_{ij}, X_{ij}')'\delta\|_{2,N}}{\|\delta_T\|_1}, \quad \phi_{\max}(m) = \sup_{\|\delta\|_0 \leq m} \frac{\|\sqrt{w_{ij}}(D_{ij}, X_{ij}')'\delta\|_{2,N}}{\|\delta_T\|_1}.$$

https://doi.org/10.1017/S0266466625100273 Published online by Cambridge University Press

The following assumptions constitute the lower level sufficient conditions. Let $s_N \geq 1$ be a sequence of integers.

Assumption 4. (Sparse eigenvalue conditions). The sparse eigenvalue conditions hold with probability $1 - o(1)$, namely, for some $\ell_N \to \infty$ slow enough, we have $1 \lesssim \phi_{\min}(\ell_N s_N) \leq \phi_{\max}(\ell_N s_N) \lesssim 1$.

Assumption 5. (Sparsity). $\|\beta_0\|_0 + \|\gamma_0\|_0 \leqslant s_N$.

Assumption 6. (Parameters). (i) $\|\beta_0\| + \|\gamma_0\| \leq C_1$. (ii) $\sup_{\theta\in\Theta} |\theta| \leq C_1$.

Assumption 7. (Covariates). The following inequalities hold for some $q > 4$:

Here, U_j refers to the individual-level latent variable appearing in the Aldous–Hoover–Kallenberg representation introduced in Equation (3.1),

(i) $\max_{j=1,2} \inf_{\|\xi\|=1} \mathrm{E}_P\big[\big(\mathrm{E}_P[f_{12}\,(D_{12}, X'_{12}) \mid U_j]\xi\big)^2\big] \geqslant c_1$.
(ii) $\max_{j=1,2} \min_{l\in[p]} \big(\mathrm{E}_P\big[\big(\mathrm{E}_P[f_{12}^2 Z_{12} X_{12,l} \mid U_j]\big)^2\big] \wedge \mathrm{E}_P\big[\big(\mathrm{E}_P[f_{12}^2 D_{12} X_{12,l} \mid U_j]\big)^2\big]\big) \geqslant c_1$.
(iii) $\sup_{\|\xi\|=1} \mathrm{E}_P[\{(D_{12}, X'_{12})\xi\}^4] \leqslant C_1$.
(iv) $N^{-1/2+2/q} M_N^2 s_N \log^2 a_N \leq \delta_N$.

Assumption 8. All eigenvalues of the matrix Γ defined in (4.8) are bounded from below by c_0.

Assumption 4 is a sparse eigenvalue condition similar to the one in Bickel, Ritov, and Tsybakov (2009) under i.i.d. setting. Sufficient conditions for the sparse eigenvalue conditions under the related multiway clustering is provided by Proposition 3 in Chiang et al. (2023). Assumption 5 imposes a restriction on the number of nonzero components in the nuisance parameter vectors by s_N, where a constraint on the growth rate of s_N in turn will be introduced in Assumption 7(iv). While our current theory imposes exact sparsity, it is conceptually feasible to extend the results to accommodate approximate sparsity by allowing the nuisance parameters to be well-approximated by sparse vectors. Formalizing such an extension would require additional regularity conditions to control the approximation error, and we leave this as an important direction for future research. Assumption 6 is standard and requires the parameter space Θ to be bounded as well as the true nuisance parameter vectors β_0 and γ_0 to have bounded ℓ_2-norms. It permits $\|\beta_0\|_1$ and $\|\gamma_0\|_1$ to be growing as N increases. Assumption 7(i) and (ii) impose lower bounds on eigenvalues and variances of some population conditional moments. Note that the U_j's here come from the Aldous–Hoover–Kallenberg representation introduced in Equation (3.1). We emphasize that, although this representation is not unique, we focus on a fixed representation for the purpose of defining and evaluating conditional expectations, and all assumptions and derivations are understood with respect to this fixed representation. Assumption 7(iii) requires the fourth moments of the covariates to be bounded in a uniform manner. Finally, Assumption 7(iv) constrains the rates at which the sparsity index s_N, the dimensionality p, as well as the bound for the $2q$-th moments of the variable of interest and the maximal

https://doi.org/10.1017/S0266466625100273 Published online by Cambridge University Press

components of the covariates vector can grow. It also implicitly requires these $2q$-th moments to exist (albeit the bounds can be diverging with N). As an illustration, suppose that $\|(D_{ij}, X_{ij}')\|_\infty = 1$. Then, q can be taken arbitrarily large and $M_N = 1$. In this case, it essentially requires $N^{-1}s_N^2 \log^2 a_N = o(1)$. These rate requirements are comparable to those assumed in state-of-the-art results in the literature under i.i.d. sampling (e.g., Belloni et al., 2018).

These assumptions are sufficient for the high-level conditions invoked in Theorem 1, as formally stated in the following lemma.

LEMMA 1. *Assumptions 4–8 imply Assumptions 2 and 3.*

As a consequence of this lemma and the general result in Theorem 1, we obtain the following result of the root-N asymptotic normality of the dyadic machine learning estimator $\check{\theta}$ in the high-dimensional logit dyadic link formation models.

THEOREM 2. *(Inference about Regression Coefficients in the High-Dimensional Logit Dyadic Model) Suppose that Assumptions 1 and 4–8 hold. Then, the dyadic machine learning estimator $\check{\theta}$ constructed using score (4.5) follows*

$$\sqrt{N}\sigma^{-1}(\check{\theta} - \theta_0) \rightsquigarrow N(0,1),$$

where $\sigma^2 = J_0^{-1}\Gamma(J_0^{-1})'$, J_0 and Γ are defined in Equations (4.7) and (4.8) separately. Moreover, this result continues to hold when σ^2 is replaced by $\widehat{\sigma}^2$ defined in Algorithm 1.

5. MONTE CARLO SIMULATION STUDIES

Focusing on the application to the high-dimensional dyadic link formation model introduced in Section 4, we study and present the finite sample performance of the proposed method of dyadic machine learning in this section through Monte Carlo simulations.

We design the data generating process for our Monte Carlo simulation studies as follows. For each $(i,j) \in \overline{[N]^2}$, generate the random vector $(D_{ij}, X_{ij}', \varepsilon_{ij})'$ according to

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

where $\{(\widetilde{D}_i, \widetilde{X}_i')'\}_{i\in[N]}$, $\{(\widetilde{D}_{ij}, \widetilde{X}_{ij}')'\}_{(i,j)\in\overline{[N]^2}}$, $\{\widetilde{\varepsilon}_i\}_{i\in[N]}$, and $\{\widetilde{\varepsilon}_{ij}\}_{(i,j)\in\overline{[N]^2}}$ are independently drawn with the laws

$$(\widetilde{D}_i, \widetilde{X}_i')' \sim \text{Normal}(0, \Sigma), \qquad (\widetilde{D}_{ij}, \widetilde{X}_{ij}')' \sim \text{Normal}(0, \Sigma),$$
$$\widetilde{\varepsilon}_i \sim \text{Normal}(0,1), \qquad \widetilde{\varepsilon}_{ij} \sim \text{Normal}(0,1),$$

https://doi.org/10.1017/S0266466625100273 Published online by Cambridge University Press

and Σ is the variance–covariance matrix whose (r,c)-th entry given by $\Sigma_{rc} = 5^{-|r-c|}$. While D_{ij} is designed as a scalar random variable, we will vary the dimension of the random vector X_{ij} across sets of simulations.

From these primitive variables, construct the dyadic binary decision Y_{ij} in turn by the logistic threshold-crossing model $Y_{ij} = \mathbb{1}\{D_{ij}\theta_0 + X_{ij}'\beta_0 \geq \varepsilon\}$. Note that this data-generating model implies the logistic regression $\mathrm{E}[Y_{ij}|D_{ij}, X_{ij}] = \Lambda(D_{ij}\theta_0 + X_{ij}'\beta_0)$. We thus obtain a dyadic sample $\{(Y_{ij}, D_{ij}, X_{ij}')'\}_{(i,j)\in\overline{[N]^2}}$, which we assume is observed by a researcher. For the true parameter values, set $\theta_0 = 1$, and let the c-th coordinate of β_0 be $2(-2)^{-c}$ for $c \leq \lfloor\sqrt{N}\rfloor$ and 0 otherwise. We will experiment with various sample sizes N as well as various dimensions of X_{ij} across sets of simulations. The number of Monte Carlo iterations is set to 2,500 for each set of simulations.

Table 1 summarizes Monte Carlo simulation results. The top panel of the table shows results for the conventional DML without the dyadic cross-fitting or the dyadic robust standard errors. The bottom panel of the table shows results for our proposed dyadic DML with the dyadic cross-fitting and the dyadic robust standard errors. Displayed are simulation results that vary with the sample size N, the dimension $\dim(X)$ of X_{ij}, and the number K of folds for the dyadic cross-fitting. Displayed statistics include the simulation bias of $\check{\theta}$, the average estimated standard error of $\check{\theta}$ across replications, the simulation standard deviation of $\check{\theta}$, the simulation root mean squared error of $\check{\theta}$, the simulation inter-quartile range of $\check{\theta}$, and the simulation coverage frequencies for nominal sizes of 90% and 95%. First, compare the top panel and bottom panel of the table in terms of simulation statistics of the estimates. Notice that the conventional DML (without accounting for dyadic data) produces biased estimates of $\check{\theta}$, while the dyadic DML produces less biased estimates. Recall that the conventional cross-fitting works under independent sampling, while dyadic data are not independent. As such, for dyadic data, conventional cross-fitting may not be able to remove overfitting biases. On the other hand, the simulation results support the claim that dyadic cross-fitting accounts for dyadic dependence and is successful in removing overfitting biases.

Second, compare the top panel and bottom panel of the table now in terms of simulation coverage frequencies. Notice that the conventional DML (without accounting for dyadic data) suffers from severe under-coverage especially for larger sample sizes, while the dyadic DML maintains correct coverage. This difference may be imputed to two sources. One source of this difference is that estimates by the conventional DML are biased while those by the dyadic DML are less biased as discussed in the previous paragraph. As the sample size increases, it becomes less likely that shorter confidence intervals centered around the biased estimates $\check{\theta}$ contain the true parameter value θ_0. Another source of the difference is that the conventional standard error without accounting for dyadic sampling is simply too small compared to the dyadic robust standard error.

In summary, the dyadic DML produces less biased estimates and asymptotically accurate coverage. Failure to account for dyadic dependence in cross-fitting results

https://doi.org/10.1017/S0266466625100273 Published online by Cambridge University Press

https://doi.org/10.1017/S0266466625100273 Published online by Cambridge University Press

TABLE 1. Simulation results based on 2,500 Monte Carlo iterations

Method	N	dim(X)	K	True	Bias	Avg. SE	SD	RMSE	Q25	Q50	Q75	90%	95%
Conventional DML	50	25	5	1.000	0.144	0.100	0.316	0.347	0.928	1.139	1.354	0.363	0.422
Conventional DML	100	50	5	1.000	0.148	0.049	0.220	0.265	0.999	1.141	1.293	0.233	0.277
Conventional DML	50	25	10	1.000	0.145	0.100	0.316	0.348	0.929	1.139	1.356	0.360	0.421
Conventional DML	100	50	10	1.000	0.149	0.049	0.220	0.265	0.998	1.143	1.294	0.234	0.274
Conventional DML	50	50	5	1.000	0.253	0.116	0.316	0.405	1.044	1.253	1.464	0.332	0.392
Conventional DML	100	100	5	1.000	0.252	0.056	0.222	0.336	1.100	1.248	1.404	0.170	0.209
Conventional DML	50	50	10	1.000	0.256	0.116	0.316	0.407	1.044	1.257	1.465	0.324	0.388
Conventional DML	100	100	10	1.000	0.254	0.056	0.222	0.338	1.102	1.249	1.406	0.171	0.206
Method	N	dim(X)	K	True	Bias	Avg. SE	SD	RMSE	Q25	Q50	Q75	90%	95%
Dyadic DML	50	25	5	1.000	0.059	0.477	0.470	0.474	0.746	1.052	1.369	0.908	0.950
Dyadic DML	100	50	5	1.000	0.045	0.288	0.288	0.292	0.848	1.040	1.236	0.901	0.946
Dyadic DML	50	25	10	1.000	0.047	0.561	0.518	0.520	0.688	1.034	1.394	0.922	0.965
Dyadic DML	100	50	10	1.000	0.039	0.319	0.303	0.305	0.825	1.042	1.239	0.916	0.957
Dyadic DML	50	50	5	1.000	0.113	0.526	0.522	0.534	0.736	1.111	1.460	0.897	0.953
Dyadic DML	100	100	5	1.000	0.105	0.313	0.304	0.322	0.903	1.105	1.314	0.883	0.940
Dyadic DML	50	50	10	1.000	0.114	0.613	0.582	0.593	0.715	1.101	1.497	0.908	0.964
Dyadic DML	100	100	10	1.000	0.095	0.346	0.320	0.334	0.880	1.093	1.316	0.908	0.958

Note: Displayed are the sample size N, dimension dim(X) of X_{ij}, number K of folds for the dyadic cross-fitting, simulation bias of $\breve{\theta}$, average estimated standard error ("Avg. SE") of $\breve{\theta}$, simulation standard deviation of $\breve{\theta}$, simulation root mean square error of $\breve{\theta}$, simulation inter quartile range of $\breve{\theta}$, and coverage frequency for the nominal sizes of 90% and 95%. The top panel of the table displays results for the conventional DML without dyadic cross-fitting or dyadic robust standard errors, and the bottom panel of the table displays results for our proposed method of dyadic DML with dyadic cross-fitting and dyadic robust standard errors.

in biased estimates, while failure to account for dyadic dependence in cross-fitting and/or standard errors results in severe under-coverage.

6. EMPIRICAL ILLUSTRATION: THE DETERMINANTS OF FTAS

In this section, we consider the classic network setting: the formation of free trade agreements (FTAs) across countries. We compare our proposed method of dyadic DML to traditional methods for analyzing determinants of FTAs using both parsimonious and rich empirical models. Baier and Bergstrand (2004) pioneered empirical studies in this field, and a large number of subsequent studies have followed since that time based on their foundational specification. We aim to strengthen our understanding of the conclusions produced in this literature by conducting a new empirical analysis with the following econometric details: 1. we control for high-dimensional covariates in order to mitigate unobserved endogeneity or unobserved confoundedness and 2. we account for the dyadic dependence inherent in FTA data and, for the first time, report standard errors robust to this source of misleading inference.

Consider the empirical model $\mathrm{E}_P[Y_{ij}|D_{ij}, X_{ij}] = \Lambda(D_{ij}\theta + X_{ij}'\beta)$ for $(i,j) \in \overline{[N]^2}$, where Y_{ij} is the dummy variable if there is a bilateral FTA between i and j as of year 2000. The parsimonious specification includes four classic explanatory variables $(D_{ij}, X_{ij}')'$: (A) the logarithm of the population-weighted bilateral distance between i and j in kilometers, (B) the sum of the logarithms of the per-capita GDPs of i and j in the baseline year, (C) the absolute difference of the logarithms of the per-capita GDPs between i and j in the baseline year, and (D) the absolute difference of the logarithms of the capital-labor ratios between i and j in the baseline year. Following Baier and Bergstrand (2004) we complete our most parsimonious specification by including the square of the difference of the logarithms of the capital-labor ratio.

Baier and Bergstrand (2004) state several hypotheses: 1. "[t]he net gain from an FTA between two countries increases as the distance between them decreases", motivating (A), 2. "[t]he net welfare gain from an FTA between a pair of countries increases the larger are their economic sizes (i.e., average real GDPs)", motivating (B), 3. "[t]he net welfare gain from an FTA between a pair of countries increases the more similar are their economic sizes (i.e., real GDPs)", motivating (C), and 4."[t]he net welfare gain from an FTA between a pair of countries increases with wider relative factor endowments, but might eventually decline due to increased specialization", motivating (D).

To investigate the impact of restricting attention to a parsimonious specification, we add a wide set of additional controls to our benchmark empirical model (see Appendix K.3 of the Supplementary Material for details of how we construct each covariate). In total, there are more than 140 explanatory variables in the rich specifications that we consider, and these dimensions are high relative to the effective sample size N of the dyad which consists of $N = 229$ economies. Following Baier and Bergstrand (2004), we set year 1960 as the baseline year for

https://doi.org/10.1017/S0266466625100273 Published online by Cambridge University Press

Table 2. Estimation and inference results based on 50 iterations of resampled cross-fitting

Dependent variable:	Logit		Full logit	Conventional DML		Dyadic DML	
free trade agreement	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)
(A) Distance	−1.690	−1.690	−1.358	−1.662	−1.660	−1.515	−1.762
	(0.046)	(0.058)	(0.075)	(0.081)	(0.079)	(0.111)	(0.115)
(B) Size (Sum of log GDP)	0.236	0.236	0.343	0.359	0.360	0.263	0.244
	(0.013)	(0.011)	(0.020)	(0.008)	(0.007)	(0.043)	(0.035)
(C) Similarity (Δ log GDP)	−0.003	−0.003	−0.004	−0.004	−0.004	−0.001	0.001
	(0.015)	(0.001)	(0.018)	(0.014)	(0.014)	(0.002)	(0.002)
(D) Rel. Factor endowments	0.231	0.231	0.187	−0.460	−0.460	−0.432	−0.396
(Δ log K/L)	(0.060)	(0.077)	(0.072)	(0.143)	(0.143)	(0.326)	(0.362)
Effective sample size	13,027	229	13,027	13,027	13,027	229	229
Dimension dim$(D', X')'$	5	5	141	141	141	141	141
Number K of folds	N/A	N/A	N/A	5	10	5	10

Note: Displayed are estimates and standard errors for the coefficients of (A) the logarithm of population-weighted distance, (B) the sum of the logarithms of GDPs in the baseline year, (C) the absolute difference of the logarithms GDPs in the baseline year, and (D) the difference of the logarithms of capital-labor ratios. The first column (I) reports results based on the simple, parsimonious logistic regression without accounting for dyadic dependence, and the second column (II) reports results based on the same parsimonious logistic regression, but accounting for dyadic dependence. Column (III) reports results based on the simple logistic regression with all the covariates without accounting for dyadic dependence. The next two columns (IV) and (V) report results based on conventional DML with $K = 5$ and 10 of folds, respectively. The last two columns (VI) and (VII) report results based on our proposed dyadic DML with $K = 5$ and 10 of folds, respectively.

variable measurement, which is roughly the year when FTAs started to be signed across global trading partners.

Table 2 summarizes estimation and inference results based on 50 iterations of resampled cross-fitting. We observe the following sharp results.

First, our simple logit model confirms the Baier and Bergstrand (2004) finding, with all covariates significant at the conventional level. However, our model suggests that the economic similarity coefficient, which is a significant determinant of FTAs in their analysis, is small and not significantly different from zero. Data differences, including a focus on FTAs signed by 2000 and expanding the analysis from 54 to 167 countries, may explain this discrepancy.

Second, rows (A), (B), and (D) each highlight a separate, but equally important, feature of dyadic machine learning in this classic setting. Row (A) shows all coefficients are negative and significantly different from zero, with larger standard errors in columns (VI) and (VII). Our findings support the hypothesis that decreasing distance and trade costs between countries increases FTA net gain. The dyadic ML estimates again confirm that this common empirical conclusion is robust to the

https://doi.org/10.1017/S0266466625100273 Published online by Cambridge University Press

inclusion of high-dimensional covariates and controlling for dyadic dependence among FTA formation data.

In row (B), dyadic ML shows that the coefficient of economic size (sum of log GDP) is positive and significant, aligning with all of the basic logit, full logit, and conventional ML results. This supports the idea that larger economic size between countries increases FTA net welfare gain. Columns (VI) and (VII), which control dyadic dependence, have 27%–32% smaller coefficients on the size variable than columns (IV) and (V). These results align with our Monte Carlo simulations, indicating that conventional ML may possibly produce biased estimates under dyadic dependence.

Row (D) highlights a different bias source. The simple logit shows the coefficient on relative factor endowments (difference of log capital-labor ratio) is positive and significant in columns (I)–(III), consistent with existing results. However, when using machine learning for variable selection with high-dimensional controls (columns (IV) and (V)), the coefficients become negative. This suggests that benchmark analyses that do not account for high-dimensional covariate confounding may yield misleading results. When comparing the results for Conventional DML and Dyadic DML, we observe that, while the positive estimated coefficient on relative factor endowments persists after controlling for dyadic dependence, the standard error more than doubles. This challenges the classic result, indicating that relative factor endowments may not be crucial determinants of FTA formation when a researcher accounts for the dyadic dependence.

7. SUMMARY AND DISCUSSION

This article presents new methods and theories for the use of machine learning when data are dyadic. A novel dyadic cross-fitting algorithm is proposed to remove over-fitting biases under arbitrary dyadic dependence. Together with the use of Neyman orthogonal scores, this dyadic cross-fitting method enables DML for dyadic data. Applying the general results (Theorem 1), we demonstrate how to estimate and conduct inference for high-dimensional logit dyadic link formation models.

Motivated by the application to the dyadic link formation model for the analysis of FTA determinants, we focus our statistical methodology on unconditional moment restrictions in this article. In other applications, such as the instrumental variable regression in dyadic data, conditional moment restrictions are more relevant (cf. Ai and Chen, 2003, 2007; Chen, Linton, and Van Keilegom, 2003; Chen and Pouzo, 2015). We conjecture that our proposed method with dyadic cross-fitting will extend to estimation and inference for models based on conditional moment restrictions. Formal theoretical investigation in this important direction is left for future research, given the length of the current article.

https://doi.org/10.1017/S0266466625100273 Published online by Cambridge University Press

SUPPLEMENTARY MATERIAL

Harold D. Chiang, Yukun Ma, Joel B. Rodrigue, and Yuya Sasaki (2025): Supplement to "Double/Debiased Machine Learning for Dyadic Data," Econometric Theory, Supplementary Material. To view, please visit: https://doi.org/10.1017/S0266466625100273.

REFERENCES

Ai, C., & Chen, X. (2003). Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica*, 71, 1795–1843.

Ai, C., & Chen, X. (2007). Estimation of possibly misspecified semiparametric conditional moment restriction models with different conditioning variables. *Journal of Econometrics*, 141, 5–43.

Aldous, D. J. (1981). Representations for partially exchangeable arrays of random variables. *Journal of Multivariate Analysis*, 11, 581–598.

Baier, S. L., & Bergstrand, J. H. (2004). Economic determinants of free trade agreements. *Journal of International Economics*, 64, 29–63.

Belloni, A., Chernozhukov, V., Chetverikov, D., & Wei, Y. (2018). Uniformly valid post-regularization confidence regions for many functional parameters in z-estimation framework. *Annals of Statistics*, 46, 3643–3675.

Belloni, A., Chernozhukov, V., & Kato, K. (2015). Uniform post-selection inference for least absolute deviation regression and other Z-estimation problems. *Biometrika*, 102, 77–94.

Belloni, A., Chernozhukov, V., & Wei, Y. (2016). Post-selection inference for generalized linear models with many controls. *Journal of Business & Economic Statistics*, 34, 606–619.

Bickel, P. J., Ritov, Y., & Tsybakov, A. B. (2009). Simultaneous analysis of lasso and Dantzig selector. *Annals of Statistics*, 37, 1705–1732.

Chen, X., Linton, O., & Van Keilegom, I. (2003). Estimation of semiparametric models when the criterion function is not smooth. *Econometrica*, 71, 1591–1608.

Chen, X., & Pouzo, D. (2015). Sieve Wald and QLR inferences on semi/nonparametric conditional moment models. *Econometrica*, 83, 1013–1079.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21, C1–C68.

Chernozhukov, V., Escanciano, J. C., Ichimura, H., Newey, W. K., & Robins, J. M. (2022). Locally robust semiparametric estimation. *Econometrica*, 90, 1501–1535. Wiley Online Library.

Chiang, H. D., Kato, K., Ma, Y., & Sasaki, Y. (2022). Multiway cluster robust double/debiased machine learning. *Journal of Business & Economic Statistics*. 40, 1046–1056. Taylor & Francis.

Chiang, H. D., Kato, K., & Sasaki, Y. (2023). Inference for high-dimensional exchangeable arrays. *Journal of the American Statistical Association*. 118, 1595–1605. Taylor & Francis.

Graham, B., & de Paula, Á. (2020). *The econometric analysis of network data*. Academic Press.

Graham, B. S. (2020). Dyadic regression. In B.S. Graham, & A. de Paula (Eds.), *The econometric analysis of network data*. Academic Press, Amsterdam, (pp. 23–40).

Kallenberg, O. (2006). *Probabilistic symmetries and invariance principles*. Springer Science & Business Media.

van der Vaart, A., & Wellner, J. A. (1996). *Weak convergence and empirical processes: With applications to statistics*. Springer Science & Business Media.

https://doi.org/10.1017/S0266466625100273 Published online by Cambridge University Press