






Inference for High-Dimensional Exchangeable Arrays

Harold D. Chiang, Kengo Kato & Yuya Sasaki


To cite this article: Harold D. Chiang, Kengo Kato & Yuya Sasaki (2021): Inference for High-Dimensional Exchangeable Arrays, Journal of the American Statistical Association, DOI: [10.1080/01621459.2021.2000868](https://doi.org/10.1080/01621459.2021.2000868)

To link to this article: <https://doi.org/10.1080/01621459.2021.2000868>

 View supplementary material 

 Published online: 21 Dec 2021.

 Submit your article to this journal 


 Article views: 398

 View related articles 

 View Crossmark data 



Inference for High-Dimensional Exchangeable Arrays

Harold D. Chiang^a , Kengo Kato^b, and Yuya Sasaki^c

^aDepartment of Economics, University of Wisconsin-Madison, Madison, WI; ^bDepartment of Statistics and Data Science, Cornell University, Ithaca, NY;

^cDepartment of Economics, Vanderbilt University, Nashville, TN

ABSTRACT

We consider inference for high-dimensional separately and jointly exchangeable arrays where the dimensions may be much larger than the sample sizes. For both exchangeable arrays, we first derive high-dimensional central limit theorems over the rectangles and subsequently develop novel multiplier bootstraps with theoretical guarantees. These theoretical results rely on new technical tools such as Hoeffding-type decomposition and maximal inequalities for the degenerate components in the Hoeffding-type decomposition for the exchangeable arrays. We exhibit applications of our methods to uniform confidence bands for density estimation under joint exchangeability and penalty choice for ℓ_1 -penalized regression under separate exchangeability. Extensive simulations demonstrate precise uniform coverage rates. We illustrate by constructing uniform confidence bands for international trade network densities.

ARTICLE HISTORY

Received February 2021

Accepted October 2021

KEYWORDS

Bootstrap; Exchangeable array; High-dimensional CLT; Network data

1. Introduction

Many recent statistical problems involve non-independent observations indexed by multiple interlocking sets of entities. Examples include dyadic/polyadic networks, bipartite networks, and multiway clustering. When the sets of entities that form each of these indices are different, as is the case with market-product data and book-reader data, a natural stochastic framework is *separate exchangeability* (MacKinnon, Nielsen, and Webb 2021). Separately exchangeable arrays include row-column exchangeable models (McCullagh 2000), additive cross random effect models (Owen 2007; Owen and Eckles 2012), and multiway clustering (Cameron, Gelbach, and Miller 2011). Meanwhile, when all indices belong to a common set of entities, as is the case with friendship network data, the underlying structure is well-captured by *joint exchangeability* (Bickel and Chen 2009). Joint exchangeability covers nonparametric random graph models of Bickel and Chen (2009) for dyadic networks, which contain widely used models in the statistical network analysis literature such as stochastic block models.

Analysis of these types of data requires accounting for the underlying complex dependence structures induced by these exchangeability notions. Thus, developing valid inference methods for exchangeable arrays is challenging. The literature has witnessed some research on statistical inference that focuses on exchangeable arrays with low or fixed dimensions. For modern statistical learning methods, it is crucial to allow the dimension of data to increase with sample size. However, the existing literature has been silent about statistical inference for such high-dimensional exchangeable arrays.

This article is concerned with the problem of inference for separately or jointly exchangeable high-dimensional arrays. We

develop new high-dimensional central limit theorems (CLTs) over the rectangles for the sample mean under both exchangeability notions. Building on the high-dimensional CLTs, we propose new multiplier bootstrap methods tailored to separate and jointly exchangeable arrays and derive their nonasymptotic error bounds. Such nonasymptotic results can be translated into asymptotic results that hold uniformly over a large set of distributions, which is crucial in many high-dimensional statistical applications.

To derive these theoretical results, we develop several new technical tools, which are of independent interest and would be useful for other analyses of exchangeable arrays. Specifically, we develop novel Hoeffding-type decompositions for both separately and jointly exchangeable arrays and establish novel maximal inequalities for Hoeffding-type projections in both cases. Such maximal inequalities lead to sharp rates for degenerate components in Hoeffding-type decompositions in both cases and play a crucial role in establishing the high-dimensional CLTs and the validity of the bootstrap methods. The proofs of these technical results are highly nontrivial. For example, the proof of the symmetrization inequality for exchangeable arrays involves a careful induction argument (see Lemma B.2 in the appendix) combined with a repeated conditioning argument. Furthermore, the proof of the maximal inequality for jointly exchangeable arrays involves a delicate conditioning argument combined with the decoupling inequalities for U -statistics with index-dependent kernels (see de la Peña and Giné 1999).

We illustrate applications of the bootstrap methods to a couple of concrete statistical problems. Specifically, (i) we develop a method to construct simultaneous or uniform confidence bands for density functions with jointly exchangeable dyadic arrays, and (ii) we develop a method to choose a penalty level for ℓ_1 -

penalized regression (Lasso) and establish error bounds for the Lasso with separately exchangeable arrays. These applications are also new in the literature.

We conduct extensive simulation studies, which demonstrate precise uniform coverage across various designs and under both notions of exchangeability, thereby supporting our theoretical results. Finally, we apply our bootstrap method to international trade network data to draw uniform confidence bands for trade flow volumes in 1990, 1995, 2000, and 2005. The results indicate that there have been increasing numbers of bilateral trading pairs with high flow volumes as time progresses.

1.1. Relation to the Literature

There is now a large literature on high-dimensional CLTs and bootstraps with the “ $p \gg n$ ” regime; see Chernozhukov, Chetverikov, and Kato (2013, 2014b, 2015, 2016, 2017), Deng and Zhang (2020), Chernozhukov, Chetverikov, and Kato (2019), Kuchibhotla, Mukherjee, and Banerjee (2020), and Fang and Koike (2020) for the independent case, Chen (2018), Chen and Kato (2020, 2019) for U -statistics and processes, and Zhang and Wu (2017), Zhang and Cheng (2018), Chernozhukov, Chetverikov, and Kato (2019), Koike (2019) for time series dependence. However, none of the above references covers extensions to exchangeable arrays. The present article builds on and contributes to this literature by developing high-dimensional CLTs and bootstrap methods for exchangeable arrays.

Early applications of exchangeable arrays in statistics include Arnold (1979), Bowman and George (1995), and Andrews (2005), to name a few. For reviews, see, for example, Goldenberg et al. (2010), Orbanz and Roy (2014), and Kuchibhotla (2020). Analysis of exchangeable random graphs has been an active research area in the recent statistics literature; see, for example, Diaconis and Janson (2008), Bickel, Chen, and Levina (2011), Lloyd et al. (2012), Choi and Wolfe (2014), Caron and Fox (2017), Choi (2017), Zhang, Levina, and Zhu (2017), and Crane and Dempsey (2018). Limit theorems for jointly exchangeable arrays (in the fixed dimensional case) date back to Silverman (1976) and Eagleson and Weber (1978). Fafchamps and Gubert (2007) and Cameron, Gelbach, and Miller (2011) derive standard error formulas for jointly exchangeable dyadic arrays and separately exchangeable arrays, respectively; see also Cameron and Miller (2014, 2015), Aronow, Samii, and Assenova (2015), and Tabord-Meehan (2019) for further development. Menzel (2021) studies inference for separately exchangeable arrays, covering both degenerate and nondegenerate cases. Davezies, D’Haultfœuille, and Guyonvarch (2021) developed functional limit theorems for Donsker classes under separate and joint exchangeability. To the best of our knowledge, however, no existing work in this literature permits high-dimensional inference. We note that Davezies, D’Haultfœuille, and Guyonvarch (2021) developed symmetrization inequalities different from ours. Specifically, symmetrization inequalities developed in Davezies, D’Haultfœuille, and Guyonvarch (2021) are applied to the whole empirical process and do not lead to correct orders for degenerate components in Hoeffding-type decompositions (indeed, Davezies, D’Haultfœuille, and Guyonvarch (2021) did not derive Hoeffding-type decompositions), thereby not

powerful enough to derive our results; see Remarks 2 and 3 in the appendix for details.

Methodologically, this article is also related to the recent literature on high-dimensional U -statistics, such as Chen (2018), Chen and Kato (2019, 2020), among others. Under suitable assumptions, the data of our interest can be written as U -statistic-like latent structure (in distribution) via the Aldous-Hoover-Kallenberg representation (Aldous 1981; Hoover 1979; Kallenberg 2006), that is, the data can be written as a kernel function of some latent independent random variables. However, unlike in U -statistics, neither the kernel nor the latent independent random variables is known to us. In addition, we need to cope with the existence of extra higher-order shocks in the latent structure. Both aspects present extra challenges.

Regarding our bootstraps, McCullagh (2000) showed that no resampling scheme for the raw data is consistent for variance of a sample mean under separate exchangeability. A Pigeonhole bootstrap is subsequently proposed by Owen (2007) and its different variants are further investigated in Owen and Eckles (2012), Davezies, D’Haultfœuille, and Guyonvarch (2021), and Menzel (2021). Whether the pigeonhole bootstrap works for increasing or high-dimensional test statistics remains unknown to us. We therefore develop a novel bootstrap method in this article which we argue works for high-dimensional data.

1.2. Notations and Organization

Let \mathbb{N} denote the set of positive integers. We use $\|\cdot\|$, $\|\cdot\|_0$, $\|\cdot\|_1$, and $\|\cdot\|_\infty$ to denote the Euclidean, ℓ_0 , ℓ_1 , and ℓ_∞ -norms for vectors, respectively (precisely, $\|\cdot\|_0$ is not a norm but a seminorm). For two real vectors $\mathbf{a} = (a_1, \dots, a_p)^T$ and $\mathbf{b} = (b_1, \dots, b_p)^T$, the notation $\mathbf{a} \leq \mathbf{b}$ means that $a_j \leq b_j$ for all $1 \leq j \leq p$. Let $\text{supp}(\mathbf{a})$ denote the support of $\mathbf{a} = (a_1, \dots, a_p)^T$, that is, $\text{supp}(\mathbf{a}) = \{j : a_j \neq 0\}$. We denote by \odot the Hadamard (element-wise) product, that is, for $\mathbf{i} = (i_1, \dots, i_K)$ and $\mathbf{j} = (j_1, \dots, j_K)$, $\mathbf{i} \odot \mathbf{j} = (i_1 j_1, \dots, i_K j_K)$. For any $a, b \in \mathbb{R}$, let $a \vee b = \max\{a, b\}$. For $0 < \beta < \infty$, let ψ_β be the function on $[0, \infty)$ defined by $\psi_\beta(x) = e^{x^\beta} - 1$. Let $\|\cdot\|_{\psi_\beta}$ denote the associated Orlicz norm, that is, $\|\xi\|_{\psi_\beta} = \inf\{C > 0 : \mathbb{E}[\psi_\beta(|\xi|/C)] \leq 1\}$ for a real-valued random variable ξ . “Constants” refer to nonstochastic and finite positive numbers.

The rest of the article is organized as follows. In Section 2, we develop a high-dimensional CLT (over the rectangles) and a bootstrap method for separately exchangeable arrays. In Section 3, we develop analogous results to jointly exchangeable arrays. We illustrate two applications in Section 4, present simulation results in Section 5, and demonstrate an empirical application in Section 6. We defer all the technical proofs to the appendix.

2. Separately Exchangeable Arrays

In this section, we consider separately exchangeable arrays that correspond to multiway clustered data. Pick any $K \in \mathbb{N}$. With $\mathbf{i} = (i_1, \dots, i_K) \in \mathbb{N}^K$, we consider a K -array $(\mathbf{X}_i)_{i \in \mathbb{N}^K}$ consisting of random vectors in \mathbb{R}^p with $p \geq 2$. We denote by X_i^j the j -th coordinate of \mathbf{X}_i : $\mathbf{X}_i = (X_i^1, \dots, X_i^p)^T$. We say that the array $(\mathbf{X}_i)_{i \in \mathbb{N}^K}$ is *separately exchangeable* if the following condition is satisfied (cf. Kallenberg 2006, Section 3.1).

Definition 1 (Separate exchangeability). A K -array $(\mathbf{X}_i)_{i \in \mathbb{N}^K}$ is called separately exchangeable if for any K permutations π_1, \dots, π_K of \mathbb{N} , the arrays $(\mathbf{X}_i)_{i \in \mathbb{N}^K}$ and $(\mathbf{X}_{(\pi_1(i_1), \dots, \pi_K(i_K))})_{i \in \mathbb{N}^K}$ are identically distributed in the sense that their finite dimensional distributions agree.

See Appendix I in the supplementary material for more details, discussions, and examples. From the Aldous-Hoover-Kallenberg representation (see Kallenberg 2006, corol. 7.23), any separately exchangeable array $(\mathbf{X}_i)_{i \in \mathbb{N}^K}$ is generated by the structure

$$\begin{aligned} \mathbf{X}_i &= f((U_{i \odot \mathbf{e}})_{\mathbf{e} \in \{0,1\}^K}), \\ i \in \mathbb{N}^K, \quad \{U_{i \odot \mathbf{e}} : i \in \mathbb{N}^K, \mathbf{e} \in \{0,1\}^K\} &\stackrel{\text{iid}}{\sim} U[0,1] \end{aligned}$$

for some Borel measurable map $f : [0,1]^{2^K} \rightarrow \mathbb{R}^p$.

The latent variable U_0 appears commonly in all \mathbf{X}_i 's. In the present article, as in Andrews (2005) and Menzel (2021), we consider inference conditional on U_0 and treat it as fixed. In the rest of Section 2, we will assume (without further mentioning) that the array $(\mathbf{X}_i)_{i \in \mathbb{N}^K}$ has mean zero (conditional on U_0) and is generated by the structure

$$\mathbf{X}_i = \mathfrak{g}((U_{i \odot \mathbf{e}})_{\mathbf{e} \in \{0,1\}^K \setminus \{0\}}), \quad i \in \mathbb{N}^K, \quad (1)$$

where \mathfrak{g} is now a map from $[0,1]^{2^K-1}$ into \mathbb{R}^p .

Suppose that we observe $\{\mathbf{X}_i : i \in [N]\}$ with $N = (N_1, \dots, N_K)$ and $[N] = \prod_{k=1}^K \{1, \dots, N_k\}$. We are interested in approximating the distribution of the sample mean

$$\mathbf{S}_N = \frac{1}{\prod_{k=1}^K N_k} \sum_{i \in [N]} \mathbf{X}_i$$

in the high-dimensional setting where the dimension p is allowed to entail $p \gg \min\{N_1, \dots, N_K\}$.

Example 1 (Empirical process indexed by function class with increasing cardinality). Our setting covers the following situation: let $\{Y_i : i \in \mathbb{N}^K\}$ be random variables taking values in an abstract measurable space (S, \mathcal{S}) , and suppose that they are generated as $Y_i = \mathfrak{g}((U_{i \odot \mathbf{e}})_{\mathbf{e} \in \{0,1\}^K \setminus \{0\}})$. Let $f_j : S \rightarrow \mathbb{R}$ for $1 \leq j \leq p$ be measurable functions, and define $X_i^j = f_j(Y_i) - \mathbb{E}[f_j(Y_i)]$. In this case, the sample mean \mathbf{S}_N can be regarded as the empirical process $f \mapsto (\prod_{k=1}^K N_k)^{-1} \sum_{i \in [N]} (f(Y_i) - \mathbb{E}[f(Y_i)])$ indexed by the function class $\mathcal{F} = \{f_1, \dots, f_p\}$. Allowing $p \rightarrow \infty$ as $\min_{1 \leq k \leq K} N_k \rightarrow \infty$ enables us to cover empirical processes indexed by function classes with increasing cardinality.

For later convenience, we fix some additional notations. Let $n = \min_{1 \leq k \leq K} N_k$ and $\bar{N} = \max_{1 \leq k \leq K} N_k$ denote the minimum and maximum cluster sizes, respectively. For $1 \leq k \leq K$, denote by $\mathcal{E}_k = \{\mathbf{e} = (e_1, \dots, e_K) \in \{0,1\}^K : \sum_{k=1}^K e_k = k\}$ the set of vectors in $\{0,1\}^K$ whose support has cardinality k . Let $\mathbf{e}_k \in \mathbb{R}^K$ denote the vector such that the k -th coordinate of \mathbf{e}_k is 1 and the other coordinates are 0. For a given $\mathbf{e} \in \{0,1\}^K$, define

$$I_{\mathbf{e}}([N]) = \{i \odot \mathbf{e} : i \in [N]\} \subset \mathbb{N}_0^K \quad \text{with } \mathbb{N}_0 = \mathbb{N} \cup \{0\}.$$

The following decomposition of the sample mean \mathbf{S}_N will play a fundamental role in our analysis, which is reminiscent of the Hoeffding decomposition for U -statistics (Lee 1990; de la Peña and Giné 1999).

Lemma 1 (Hoeffding decomposition of separately exchangeable array). For any $i \in \mathbb{N}^K$, define recursively $\hat{\mathbf{X}}_{i \odot \mathbf{e}_k} = \mathbb{E}[\mathbf{X}_i | U_{i \odot \mathbf{e}_k}]$ for $k = 1, \dots, K$ and $\hat{\mathbf{X}}_{i \odot \mathbf{e}} = \mathbb{E}[\mathbf{X}_i | (U_{i \odot \mathbf{e}'})_{\mathbf{e}' \leq \mathbf{e}}] - \sum_{\substack{\mathbf{e}' \leq \mathbf{e} \\ \mathbf{e}' \neq \mathbf{e}}} \hat{\mathbf{X}}_{i \odot \mathbf{e}'}$

for $\mathbf{e} \in \bigcup_{k=2}^K \mathcal{E}_k$. Then, we have $\mathbf{X}_i = \sum_{\mathbf{e} \in \{0,1\}^K \setminus \{0\}} \hat{\mathbf{X}}_{i \odot \mathbf{e}}$. Consequently, we can decompose the sample mean $\mathbf{S}_N = (\prod_{k=1}^K N_k)^{-1} \sum_{i \in [N]} \mathbf{X}_i$ as

$$\mathbf{S}_N = \sum_{k=1}^K \sum_{\mathbf{e} \in \mathcal{E}_k} \frac{1}{\prod_{k' \in \text{supp}(\mathbf{e})} N_{k'}} \sum_{i \in I_{\mathbf{e}}([N])} \hat{\mathbf{X}}_i. \quad (2)$$

The proof of this lemma can be found in Appendix C.1.

Remark 1 (Hoeffding decomposition). The reason that we call (2) the Hoeffding decomposition comes from the fact that if the dimension p is fixed, for each fixed $k = 1, \dots, K$ and $\mathbf{e} \in \mathcal{E}_k$, the component $(\prod_{k' \in \text{supp}(\mathbf{e})} N_{k'})^{-1} \sum_{i \in I_{\mathbf{e}}([N])} \hat{\mathbf{X}}_i$ scales as $(\prod_{k' \in \text{supp}(\mathbf{e})} N_{k'})^{-1/2} = O(n^{-k/2})$ with $n = \min_{1 \leq k' \leq K} N_{k'}$ under moment conditions. See Corollary B.1 in Appendix B. This is completely analogous to the Hoeffding decomposition of U -statistics and from this analogy we shall call (2) the Hoeffding decomposition.

The leading term in the decomposition (2) is

$$\begin{aligned} \sum_{\mathbf{e} \in \mathcal{E}_1} \frac{1}{\prod_{k' \in \text{supp}(\mathbf{e})} N_{k'}} \sum_{i \in I_{\mathbf{e}}([N])} \hat{\mathbf{X}}_i \\ = \sum_{k=1}^K N_k^{-1} \sum_{i_k=1}^{N_k} \mathbb{E}[\mathbf{X}_i | U_{(0, \dots, 0, i_k, 0, \dots, 0)}], \end{aligned}$$

which we call the Hájek projection of \mathbf{S}_N . With this in mind, define $\mathbf{W}_{k,i_k} = \mathbb{E}[\mathbf{X}_i | U_{(0, \dots, 0, i_k, 0, \dots, 0)}]$ for $k = 1, \dots, K$.

2.1. High-Dimensional CLT for Separately Exchangeable Arrays

We first establish a high-dimensional CLT for \mathbf{S}_N over the class of rectangles, $\mathcal{R} = \{\prod_{j=1}^p [a_j, b_j] : -\infty \leq a_j \leq b_j \leq \infty, 1 \leq j \leq p\}$. This high-dimensional CLT will be a building block for establishing the validity of the multiplier bootstrap (cf. Section 2.2).

We start with discussing regularity conditions. Denote by $\mathbf{1} = (1, \dots, 1)$ the vector of ones. Let $D_N \geq 1$ be a given constant that may depend on the cluster sizes N (and p ; when we consider asymptotics we have in mind that p is a function of N or n so we omit the dependence of D_N on p), and let $\underline{\sigma} > 0$ be another given constant independent of the cluster sizes N . We will assume *either* of the following moment conditions:

$$\max_{1 \leq j \leq p} \|X_1^j\|_{\psi_1} \leq D_N, \quad \text{or} \quad (3)$$

$$\mathbb{E}[\|\mathbf{X}_1\|_{\infty}^q] \leq D_N^q \quad \text{for some } q \in (4, \infty). \quad (4)$$

We will also assume the following condition.

$$\begin{aligned} \max_{1 \leq j \leq p; 1 \leq k \leq K} \mathbb{E}[|W_{k,1}^j|^{2+\kappa}] \leq D_N^{\kappa}, \quad \kappa = 1, 2, \quad \text{and} \\ \min_{1 \leq j \leq p; 1 \leq k \leq K} \mathbb{E}[|W_{k,1}^j|^2] \geq \underline{\sigma}^2. \end{aligned} \quad (5)$$

Condition (3) requires that each coordinate of \mathbf{X}_1 is sub-exponential. By Jensen's inequality, Condition (3) implies that $\max_{1 \leq j \leq p; 1 \leq k \leq K} \|W_{k,1}^j\|_{\psi_1} \leq D_N$. Condition (4) is an alternative moment condition on \mathbf{X}_1 . Condition (4) is satisfied for example under the following situation: Suppose that \mathbf{X}_i is given by $\mathbf{X}_i = \varepsilon_i \mathbf{Z}_i$ where ε_i is a scalar "error" variable while \mathbf{Z} is a vector of "covariates." If each coordinate of \mathbf{Z}_i is bounded by a constant \bar{D} and ε_i has finite q th moment, then $\mathbb{E}[\|\mathbf{X}_i\|_\infty^q] \leq \bar{D}^q \mathbb{E}[|\varepsilon_i|^q]$. Also Condition (4) is satisfied if, in the discretized empirical process application (cf. Example 1), the function class possesses an envelope function with finite q th moment. Again, by Jensen's inequality, Condition (4) implies that $\max_{1 \leq k \leq K} \mathbb{E}[\|\mathbf{W}_{k,1}\|_\infty^q] \leq D_N^q$. The restriction $q > 4$ is needed to guarantee that Condition (7) appearing in Theorem 2 to be nonvoid.

Condition (5) requires the maximum of third (respectively, fourth) moment across coordinates to be increasing at speed no faster than the first (respectively, second) power of D_N . By Jensen's inequality, the first part of Condition (5) is satisfied if $\max_{1 \leq j \leq p} \mathbb{E}[|X_1^j|^{2+\kappa}] \leq D_N^\kappa$ for $\kappa = 1, 2$. The second part of Condition (5) guarantees that the Hájek projection is nondegenerate.

Let $\gamma = N(\mathbf{0}, \Sigma)$ with $\Sigma = \sum_{k=1}^K (n/N_k) \Sigma_{W_k}$ and $\Sigma_{W_k} = \mathbb{E}[\mathbf{W}_{k,1} \mathbf{W}_{k,1}^T]$ for $k = 1, \dots, K$.

Theorem 1 (High-dimensional CLT for separately exchangeable arrays). Suppose that either Condition (3) or (4) holds, and further that Condition (5) holds. Then, there exists a constant C such that

$$\sup_{R \in \mathcal{R}} |\mathbb{P}(\sqrt{n} \mathbf{S}_N \in R) - \gamma_\Sigma(R)| \leq \begin{cases} C \left(\frac{D_N^2 \log^7(p\bar{N})}{n} \right)^{1/6} & \text{if (3) holds,} \\ C \left[\left(\frac{D_N^2 \log^7(p\bar{N})}{n} \right)^{1/6} + \left(\frac{D_N^2 \log^3(p\bar{N})}{n^{1-2/q}} \right)^{1/3} \right] & \text{if (4) holds,} \end{cases}$$

where the constant C depends only on $\underline{\sigma}$ and K if Condition (3) holds, while C depends only on $q, \underline{\sigma}$, and K if Condition (4) holds.

Remark 2 (Refinement under subgaussianity). The recent article of Chernozhukov, Chetverikov, and Kato (2019) provided some improvements on convergence rate of Gaussian approximation under the subgaussian tail assumption for the sample mean of independent random vectors. With this new technique, if we strengthen Condition (3) by replacing the ψ_1 -norm $\|\cdot\|_{\psi_1}$ with the ψ_2 -norm $\|\cdot\|_{\psi_2}$ (i.e., each coordinate \mathbf{X}_1 is sub-Gaussian), the bound $C (n^{-1} D_N^2 \log^7(p\bar{N}))^{1/6}$ in Theorem 1 can be improved to $C (n^{-1} D_N^2 \log^5(p\bar{N}))^{1/4}$.

2.2. Multiplier Bootstrap for Separately Exchangeable Arrays

Let $\{\xi_{1,i_1}\}_{i_1=1}^{N_1}, \dots, \{\xi_{K,i_k}\}_{i_k=1}^{N_K}$ be independent $N(0, 1)$ random variables independent of the data. Ideally, we want to make use of the bootstrap statistic $\sum_{k=1}^K N_k^{-1} \sum_{i_k=1}^{N_k} \xi_{k,i_k} (\mathbf{W}_{k,i_k} - \mathbf{S}_N)$. However, this bootstrap is infeasible as $\mathbf{W}_{k,i_k} =$

$\mathbb{E}[\mathbf{X}_i | U_{(0, \dots, i_k, \dots, 0)}]$ are unknown to us. Estimation of \mathbf{W}_{k,i_k} is nontrivial as $U_{(0, \dots, i_k, \dots, 0)}$ is a latent variable. We propose to estimate each \mathbf{W}_{k,i_k} by

$$\bar{\mathbf{X}}_{k,i_k} = \frac{1}{\prod_{k' \neq k} N_{k'}} \sum_{i_1, \dots, i_{k-1}, i_{k+1}, \dots, i_K} \mathbf{X}_i, \\ i_k = 1, \dots, N_k; k = 1, \dots, K,$$

that is, the sample mean taken over all indices but i_k . Then, we apply the multiplier bootstrap to $\bar{\mathbf{X}}_{k,i_k}$ in place of \mathbf{W}_{k,i_k}

$$\mathbf{S}_N^{MB} = \sum_{k=1}^K N_k^{-1} \sum_{i_k=1}^{N_k} \xi_{k,i_k} (\bar{\mathbf{X}}_{k,i_k} - \mathbf{S}_N).$$

To the best of our knowledge, this multiplier bootstrap for separately exchangeable arrays is new in the literature. We will formally study the validity of this multiplier bootstrap for high-dimensional separately exchangeable arrays with $p \gg n$.

We are now in position to establish the validity of the proposed multiplier bootstrap for separately exchangeable arrays. Let $\mathbb{P}_{|\mathbf{X}_{[N]}}$ denote the law conditional on the data $\mathbf{X}_{[N]} = (\mathbf{X}_i)_{i \in [N]}$ and $\bar{\sigma} = \max_{1 \leq j \leq p; 1 \leq k \leq K} \sqrt{\mathbb{E}[|W_{k,1}^j|^2]}$.

Theorem 2 (Validity of multiplier bootstrap for separately exchangeable arrays). Consider the following two cases:

- (i). Assume that Conditions (3) and (5) hold, and further there exist constants C_1 and $\zeta \in (0, 1)$ such that

$$\frac{\bar{\sigma}^2 D_N^2 \log^7 p}{n} \sqrt{\frac{D_N^2 (\log^2 n) \log^5(p\bar{N})}{n}} \leq C_1 n^{-\zeta}. \quad (6)$$

- (ii). Assume that Conditions (4) and (5) hold, and further there exist constants C_1 and $\zeta \in (2/q, 1)$ such that

$$\frac{\bar{\sigma}^2 D_N^2 \log^5(pn)}{n} \sqrt{\left(\frac{D_N^2 \log^3 p}{n^{1-4/q}} \right)^2} \leq C_1 n^{-\zeta}. \quad (7)$$

Then, under Case (i), for any $\nu \in (1/\zeta, \infty)$, there exists a constant C depending only on ν, σ, K , and C_1 such that $\sup_{R \in \mathcal{R}} |\mathbb{P}_{|\mathbf{X}_{[N]}}(\sqrt{n} \mathbf{S}_N^{MB} \in R) - \gamma_\Sigma(R)| \leq C n^{-(\zeta-1/\nu)/4}$ with probability at least $1 - Cn^{-1}$. Under Case (ii), the same conclusion holds with $n^{-(\zeta-1/\nu)/4}$ replaced by $n^{-(\zeta-2/q)/4}$, while the constant C depends only on $q, \underline{\sigma}, K$, and C_1 .

Remark 3 (Discussion on Conditions (6) and (7)). Conditions (6) and (7) are placed to guarantee that the error bound for our multiplier bootstrap decreases at a polynomial rate in n . If we are to show a weaker result, namely,

$$\sup_{R \in \mathcal{R}} |\mathbb{P}_{|\mathbf{X}_{[N]}}(\sqrt{n} \mathbf{S}_N^{MB} \in R) - \gamma_\Sigma(R)| = o_P(1) \quad (8)$$

as $n \rightarrow \infty$ (with the understanding that $p, \bar{\sigma}, D_N$, and \bar{N} are functions of n), then Conditions (6) and (7) can be weakened to $(\bar{\sigma}^2 D_N^2 \log^7 p) \vee D_N^2 \log^5(p\bar{N}) = o(n)$ and $(n^{-1} \bar{\sigma}^2 D_N^2 \log^5(pn)) \vee (n^{-(1-2/q)} D_N^2 \log^3 p) = o(1)$, respectively. (The critical case $q = 4$ is allowed for Equation (8); note that the high-dimensional CLT (Theorem 1) also holds with $q = 4$.)

Remark 4 (Normalized sample mean). In practice, we often normalize the coordinates of the sample mean by estimates of the standard deviations, so that each coordinate is approximately distributed as $N(0, 1)$. We can estimate the variance of the j -th coordinate of $\sqrt{n}\mathbf{S}_N$ by the conditional variance of the j th coordinate of $\sqrt{n}\mathbf{S}_N^{MB}$. The validity of the multiplier bootstrap to the normalized sample mean follows similarly to the preceding theorem; see Appendix A.1 for details. A similar comment applies to the joint exchangeable case; see Appendix A.2 for details.

3. Jointly Exchangeable Arrays

In this section, we consider another class of exchangeable arrays, namely, jointly exchangeable arrays. The notations in the current section are independent from those in Section 2 unless otherwise noted. Joint exchangeability induces a more complex dependence structure on arrays than separate exchangeability, but still we are able to develop analogous results to the preceding section for jointly exchangeable arrays as well. It should be noted, however, that we do require a different bootstrap and technical tools (see Appendix D) to accommodate a specific dependence structure induced from joint exchangeability.

Pick any $K \in \mathbb{N}$. For a given positive integer $n \geq K$, let $I_{n,K} = \{(i_1, \dots, i_K) : 1 \leq i_1, \dots, i_K \leq n \text{ and } i_1, \dots, i_K \text{ are distinct}\}$. Also let $I_{\infty,K} = \bigcup_{n=K}^{\infty} I_{n,K}$. For any $\mathbf{i} = (i_1, \dots, i_K) \in \mathbb{N}^K$, let $\{\mathbf{i}\}^+$ denote the set of distinct nonzero elements of (i_1, \dots, i_K) .

In this section, we consider a K -array $(\mathbf{X}_{\mathbf{i}})_{\mathbf{i} \in I_{\infty,K}}$ consisting of random vectors in \mathbb{R}^p with $p \geq 2$. We say that the array $(\mathbf{X}_{\mathbf{i}})_{\mathbf{i} \in I_{\infty,K}}$ is *jointly exchangeable* if the following condition is satisfied (cf. Kallenberg 2006, sec. 3.1).

Definition 2 (Joint exchangeability). A K -array $(\mathbf{X}_{\mathbf{i}})_{\mathbf{i} \in I_{\infty,K}}$ is called jointly exchangeable if for any permutation π of \mathbb{N} , the arrays $(\mathbf{X}_{\mathbf{i}})_{\mathbf{i} \in I_{\infty,K}}$ and $(\mathbf{X}_{(\pi(i_1), \dots, \pi(i_K))})_{\mathbf{i} \in I_{\infty,K}}$ are identically distributed.

See Appendix I in the supplementary material for more details, discussions, and examples. From the Aldous-Hoover-Kallenberg representation (see Kallenberg 2006, theor. 7.22), any jointly exchangeable array $(\mathbf{X}_{\mathbf{i}})_{\mathbf{i} \in I_{\infty,K}}$ is generated by the structure

$$\mathbf{X}_{\mathbf{i}} = \mathbf{f}((U_{\{\mathbf{i} \odot \mathbf{e}\}^+})_{\mathbf{e} \in \{0,1\}^K}), \quad \mathbf{i} \in I_{\infty,K},$$

$$\{U_{\{\mathbf{i} \odot \mathbf{e}\}^+} : \mathbf{i} \in I_{\infty,K}, \mathbf{e} \in \{0,1\}^K\} \stackrel{\text{iid}}{\sim} U[0,1]$$

for some Borel measurable map $\mathbf{f} : [0,1]^{2^K} \rightarrow \mathbb{R}^p$. Here the coordinates of the vector $(U_{\{\mathbf{i} \odot \mathbf{e}\}^+})_{\mathbf{e} \in \{0,1\}^K}$ are understood to be properly ordered, so that, for example, when $K = 2$, $\mathbf{X}_{(i_1, i_2)} = \mathbf{f}(U_{\emptyset}, U_{i_1}, U_{i_2}, U_{\{i_1, i_2\}})$ and $\mathbf{X}_{(i_2, i_1)} = \mathbf{f}(U_{\emptyset}, U_{i_2}, U_{i_1}, U_{\{i_1, i_2\}})$ differ (although they have the identical distribution).

As in the separately exchangeable case, we consider inference conditional on U_{\emptyset} , and in what follows, we will assume that the array $(\mathbf{X}_{\mathbf{i}})_{\mathbf{i} \in I_{\infty,K}}$ has mean zero (conditional on U_{\emptyset}) and is generated by the structure

$$\mathbf{X}_{\mathbf{i}} = \mathbf{g}((U_{\{\mathbf{i} \odot \mathbf{e}\}^+})_{\mathbf{e} \in \{0,1\}^K \setminus \{0\}}), \quad \mathbf{i} \in I_{\infty,K}, \quad (9)$$

where \mathbf{g} is now a map from $[0,1]^{2^K-1}$ into \mathbb{R}^p .

Suppose that we observe $\{\mathbf{X}_{\mathbf{i}} : \mathbf{i} \in I_{n,K}\}$ with $n \geq K$ and are interested in distributional approximation of the polyadic sample mean

$$\mathbf{S}_n := \frac{(n-K)!}{n!} \sum_{\mathbf{i} \in I_{n,K}} \mathbf{X}_{\mathbf{i}}$$

in the high-dimensional setting where the dimension p is allowed to entail $p \gg n$.

As in Section 2, define $\mathcal{E}_k = \{\mathbf{e} = (e_1, \dots, e_K) \in \{0,1\}^K : \sum_{k=1}^K e_k = k\}$ for $1 \leq k \leq K$. The analysis of the jointly exchangeable array relies on the following decomposition:

$$\begin{aligned} \mathbf{S}_n &= \frac{1}{n} \sum_{j=1}^n \mathbb{E} \left[\frac{(n-K)!}{(n-1)!} \sum_{k=1}^K \sum_{\mathbf{i} \in I_{n,K} : i_k = j} \mathbf{X}_{\mathbf{i}} \mid U_j \right] \\ &\quad + \frac{(n-K)!}{n!} \sum_{\mathbf{i} \in I_{n,K}} \left(\mathbb{E}[\mathbf{X}_{\mathbf{i}} | U_{i_1}, \dots, U_{i_K}] - \sum_{k=1}^K \mathbb{E}[\mathbf{X}_{\mathbf{i}} | U_{i_k}] \right) \\ &\quad + \sum_{k=2}^K \frac{(n-K)!}{n!} \sum_{\mathbf{i} \in I_{n,K}} \left(\mathbb{E}[\mathbf{X}_{\mathbf{i}} | (U_{\{\mathbf{i} \odot \mathbf{e}\}^+})_{\mathbf{e} \in \cup_{r=1}^k \mathcal{E}_r}] \right. \\ &\quad \left. - \mathbb{E}[\mathbf{X}_{\mathbf{i}} | (U_{\{\mathbf{i} \odot \mathbf{e}\}^+})_{\mathbf{e} \in \cup_{r=1}^{k-1} \mathcal{E}_r}] \right). \end{aligned} \quad (10)$$

It turns out that the first term on the right-hand side, which we call the Hájek projection of \mathbf{S}_n , is a dominant term. Defining $h_k(u) = \mathbb{E}[\mathbf{X}_{(1, \dots, k)} | U_k = u]$ for $k = 1, \dots, K$, we can simplify the Hájek projection into $n^{-1} \sum_{i=1}^n \mathbf{W}_j$ where $\mathbf{W}_j = \sum_{k=1}^K h_k(U_j)$.

3.1. High-Dimensional CLT for Jointly Exchangeable Arrays

We consider to approximate the distribution of $\sqrt{n}\mathbf{S}_n$ by a Gaussian distribution on the set of rectangles \mathcal{R} as defined in Section 2.

Let $D_n \geq 1$ be a given constant that may depend on n , and $\underline{\sigma} > 0$ be another given constant independent of n . We will assume either of the following moment conditions:

$$\max_{1 \leq \ell \leq p} \|X_{(1, \dots, K)}^{\ell}\|_{\psi_1} \leq D_n, \quad \text{or} \quad (11)$$

$$\mathbb{E}[\|\mathbf{X}_{(1, \dots, K)}\|_{\infty}^q] \leq D_n^q \quad \text{for some } q \in (4, \infty). \quad (12)$$

We will also assume the following condition:

$$\begin{aligned} \max_{1 \leq \ell \leq p} \mathbb{E}[|W_1^{\ell}|^{2+\kappa}] &\leq D_n^{\kappa}, \quad \kappa = 1, 2, \quad \text{and} \\ \min_{1 \leq \ell \leq p} \mathbb{E}[|W_1^{\ell}|^2] &\geq \underline{\sigma}^2 \end{aligned} \quad (13)$$

The conditions required here are similar to those in the case of separate exchangeability in Section 2. The main difference is that Condition (13) is now imposed on \mathbf{W}_1 .

Let $\gamma_{\Sigma} = N(\mathbf{0}, \Sigma)$ with $\Sigma = \mathbb{E}[\mathbf{W}_1 \mathbf{W}_1^T]$.

Theorem 3 (High-dimensional CLT for jointly exchangeable arrays). Suppose that either Condition (11) or (12) holds, and further Condition (13) holds. Then, there exists a constant C such that

$$\begin{aligned} \sup_{R \in \mathcal{R}} |\mathbb{P}(\sqrt{n}\mathbf{S}_n \in R) - \gamma_{\Sigma}(R)| \\ \leq \begin{cases} C \left(\frac{D_n^2 \log^7(pn)}{n} \right)^{1/6} & \text{if (11) holds,} \\ C \left[\left(\frac{D_n^2 \log^7(pn)}{n} \right)^{1/6} + \left(\frac{D_n^2 \log^3(pn)}{n^{1-2/q}} \right)^{1/3} \right] & \text{if (12) holds,} \end{cases} \end{aligned}$$

where the constant C depends only on $\underline{\sigma}$ and K if Condition (11) holds, while C depends only on $q, \underline{\sigma}$, and K if Condition (12) holds.

Remark 5 (Comparison with Silverman 1976). Theorem 3 is a high-dimensional extension of (Silverman 1976, theor. A) that established a CLT for jointly exchangeable arrays with fixed p . The covariance matrix of the limiting Gaussian distribution in Silverman (1976) had a different expression than our Σ , but we will verify below that two expressions are indeed the same. The covariance matrix given in Corollary to Silverman (1976, theor. A) read as follows: Let $\check{X}_{(i_1, \dots, i_K)}$ be the symmetrized version of $X_{(i_1, \dots, i_K)}$, that is, $\check{X}_{(i_1, \dots, i_K)} = (K!)^{-1} \sum_{(i'_1, \dots, i'_K)} X_{(i'_1, \dots, i'_K)}$ where the summation is taken over all permutations of (i_1, \dots, i_K) . The covariance matrix given in Silverman (1976) is $\Sigma_S = K^2 \mathbb{E}[\check{X}_{(1, \dots, K)} \check{X}_{(1, K+1, \dots, 2K)}]$. On the other hand, $\sum_{k=1}^K \mathbb{E}[X_{(1, \dots, K)} | U_k = u] = \sum_{k=1}^K \mathbb{E}[\check{X}_{(1, \dots, K)} | U_k = u] = K \mathbb{E}[\check{X}_{(1, \dots, K)} | U_1 = u]$, so that $\Sigma = K^2 \mathbb{E}[\mathbb{E}[\check{X}_{(1, \dots, K)} | U_1] \mathbb{E}[\check{X}_{(1, \dots, K)} | U_1]'] = K^2 \mathbb{E}[\check{X}_{(1, \dots, K)} \check{X}_{(1, K+1, \dots, 2K)}] = \Sigma_S$.

3.2. Multiplier Bootstrap for Jointly Exchangeable Arrays

Let $\{\xi_j\}_{j=1}^n$ be independent $N(0, 1)$ random variables independent of the data. Ideally, we want to make use of the multiplier bootstrap statistic $n^{-1} \sum_{j=1}^n \xi_j (\mathbf{W}_j - K \mathbf{S}_n)$. This is infeasible, however, as the projections \mathbf{W}_j are unknown. As an alternative, we replace each \mathbf{W}_j by its estimate

$$\hat{\mathbf{W}}_j = \frac{(n-K)!}{(n-1)!} \sum_{k=1}^K \sum_{i \in I_{n,K}: i_k=j} X_i,$$

and apply the multiplier bootstrap to $\hat{\mathbf{W}}_j$, that is,

$$\mathbf{S}_n^{MB} := \frac{1}{n} \sum_{j=1}^n \xi_j (\hat{\mathbf{W}}_j - K \mathbf{S}_n)$$

When $K = 2$ (dyadic), this multiplier bootstrap coincides with the multiplier bootstrap statistic considered in Davezies, D'Haultfœuille, and Guyonvarch (2021). However, Davezies, D'Haultfœuille, and Guyonvarch (2021, sec. 3.2) did not consider the extension to general K arrays, and focus on the empirical process indexed by a Donsker class, which excludes the high-dimensional sample mean. We will study the validity of this multiplier bootstrap for jointly exchangeable arrays.

Let $\mathbb{P}_{|X_{I_{n,K}}}$ denote the law conditional on the data $(X_i)_{i \in I_{n,K}}$ and $\bar{\sigma} = \max_{1 \leq \ell \leq p} \sqrt{\mathbb{E}[|W_1^\ell|^2]}$.

Theorem 4 (Validity of multiplier bootstrap for jointly exchangeable arrays). Consider the following two cases.

- (i). Assume that Conditions (11) and (13) hold, and further there exist constants C_1 and $\zeta \in (0, 1)$ such that

$$\frac{\bar{\sigma}^2 D_n^2 \log^7 p}{n} \sqrt{\frac{D_n^2 (\log^2 n) \log^5(pn)}{n}} \leq C_1 n^{-\zeta}. \quad (14)$$

- (ii). Assume that Conditions (12) and (13) hold, and further there exist constants C_1 and $\zeta \in (2/q, 1)$ such that

$$\frac{\bar{\sigma}^2 D_n^2 \log^5(pn)}{n} \sqrt{\left(\frac{D_n^2 \log^3 p}{n^{1-4/q}}\right)^2} \leq C_1 n^{-\zeta}. \quad (15)$$

Then, under Case (i), for any $\nu \in (1/\zeta, \infty)$, there exists a constant C depending only on $\nu, \underline{\sigma}, K$, and C_1 such that $\sup_{R \in \mathcal{R}} \left| \mathbb{P}_{|X_{I_{n,K}}}(\sqrt{n} \mathbf{S}_n^{MB} \in R) - \gamma_\Sigma(R) \right| \leq C n^{-(\zeta-1/\nu)/4}$ with probability at least $1 - Cn^{-1}$. Under Case (ii), the same conclusion holds with $n^{-(\zeta-1/\nu)/4}$ replaced by $n^{-(\zeta-2/q)/4}$, while the constant C depends only on $q, \underline{\sigma}, K$, and C_1 .

Remark 6 (Discussion on Conditions (14) and (15)). Similar to Remark 3, if one is interested only in bootstrap consistency, Conditions (14) and (15) can be weakened to $(\bar{\sigma}^2 D_n^2 \log^7 p) \vee (D_n^2 \log^5(pn)) = o(n)$ and $(n^{-1} \bar{\sigma}^2 D_n^2 \log^5(pn)) \vee (n^{-(1-2/q)} D_n^2 \log^3 p) = o(1)$, respectively.

4. Applications

In this section, we illustrate a couple of applications of our bootstrap methods. Section 4.1 is concerned with construction of confidence bands for densities of flows in dyadic data. Section 4.2 is concerned with penalty choice for the Lasso and the performance of the corresponding estimate.

4.1. Confidence Bands for Densities of Flows in Dyadic Data

Researchers are often interested in “the densities of migration across states, trade across nations, liabilities across banks, or minutes of telephone conversation among individuals” (Graham, Niu, and Powell 2019). Densities of these flow measures use dyadic data. We illustrate an application of our method in Section 3 to constructing confidence bands for such density functions. We refer the reader to Bickel and Rosenblatt (1973), Claeskens and van Keilegom (2003), and Chernozhukov, Chetverikov, and Kato (2014a) as references on confidence bands for density estimation with iid data.

Following Graham, Niu, and Powell (2019), suppose that we observe the dyadic data $\{Y_{ij} : 1 \leq i \neq j \leq n\}$ that admits the structure

$$Y_{ij} = \mathfrak{g}(U_i, U_j, U_{\{i,j\}}) \quad (16)$$

where \mathfrak{g} is symmetric in the first two arguments and hence $Y_{ij} = Y_{ji}$. We are interested in inference on the density of Y_{ij} . However, in certain empirical applications, such as international trade (see Head and Mayer 2014), a proportion of the variable of interest is zero. Hence we assume that Y_{ij} has a probability mass at zero, i.e. Y_{ij} is such that $\mathbb{P}(Y_{ij} \neq 0) = a \in (0, 1]$, and $Y_{ij} \sim f$ when $Y_{ij} \neq 0$, where f is a density function on \mathbb{R} . Let $b(y) = af(y)$ denote the scaled density. We may estimate $f(\cdot) = b(\cdot)/a$ by $\hat{f}(\cdot) = \hat{b}(\cdot)/\hat{a}$, where $\hat{a} = \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} \mathbb{1}(Y_{ij} \neq 0)$ and $\hat{b}(y) = \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n} K_h(y - Y_{ij}) \mathbb{1}(Y_{ij} \neq 0)$. Here $K : \mathbb{R} \rightarrow \mathbb{R}$ is a kernel function (a function that integrates to one), $K_h(\cdot) := h^{-1}K(\cdot/h)$, and $h = h_n \rightarrow 0$ is a bandwidth.

We consider to construct simultaneous confidence intervals (bands) for f over the set of design points y_1, \dots, y_p , where $p = p_n \rightarrow \infty$ is allowed. Define

$$\tilde{X}_{ij}^\ell = \left\{ \frac{K_h(y_\ell - Y_{ij})}{\hat{a}} - \frac{\hat{b}(y_\ell)}{\hat{a}^2} \right\} \mathbb{1}(Y_{ij} \neq 0), \quad 1 \leq i < j \leq n,$$

$$\tilde{X}_{ij}^\ell = \tilde{X}_{ji}^\ell, \quad 1 \leq j < i \leq n,$$

for $\ell = 1, \dots, p$. Then, the multiplier bootstrap statistic is given by

$$\tilde{S}_n^{MB} = \frac{1}{n} \sum_{i=1}^n \xi_i (\tilde{W}_i - 2\tilde{S}_n), \text{ with}$$

$$\tilde{S}_n = \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} \tilde{X}_{ij} \text{ and } \tilde{W}_i = \frac{1}{n-1} \sum_{j \neq i} 2\tilde{X}_{ij},$$

where $\sum_{j \neq i} = \sum_{j \in \{1, \dots, n\} \setminus \{i\}}$. For a given $\alpha \in (0, 1)$, consider the $(1 - \alpha)$ -simultaneous confidence intervals defined by

$$\mathcal{I}(1 - \alpha) := \prod_{\ell=1}^p \left[\hat{f}(y_\ell) \pm \frac{\tilde{c}(1 - \alpha)}{\sqrt{n}} \right] \text{ and}$$

$$\mathcal{I}^N(1 - \alpha) := \prod_{\ell=1}^p \left[\hat{f}(y_\ell) \pm \frac{\tilde{\sigma}_\ell \tilde{c}^N(1 - \alpha)}{\sqrt{n}} \right],$$

where $\tilde{\sigma}_\ell^2 = n^{-1} \sum_{i=1}^n (\tilde{W}_i^\ell - 2\tilde{S}_n^\ell)^2$, $\tilde{\Lambda} = \text{diag}(\tilde{\sigma}_1^2, \dots, \tilde{\sigma}_p^2)$, $\tilde{c}(1 - \alpha)$ is the conditional $(1 - \alpha)$ -quantile of $\|\sqrt{n}\tilde{S}_n^{MB}\|_\infty$, and $\tilde{c}^N(1 - \alpha)$ is the conditional $(1 - \alpha)$ -quantile of $\|\sqrt{n}\hat{\Lambda}^{-1/2}\tilde{S}_n^{MB}\|_\infty$. The first method $\mathcal{I}(1 - \alpha)$ is a constant-length confidence band, while the second method $\mathcal{I}^N(1 - \alpha)$ is a variable-length confidence band based on Studentization.

The following proposition establishes asymptotic validity of the confidence bands. We will assume that there exists a conditional density of Y_{ij} given U_i and $Y_{ij} \neq 0$, denoted by $f_{Y_{12}|U_1, Y_{12} \neq 0}(y|u)$ (more formally, we assume that the conditional distribution of Y_{ij} given U_i is $\mathbb{P}(Y_{ij} \in dy|U_i) = \mathbb{P}(Y_{ij} = 0|U_i)\delta_0(dy) + \mathbb{P}(Y_{ij} \neq 0|U_i)f_{Y_{12}|U_1, Y_{12} \neq 0}(y|U_i)dy$, where δ_0 is the Dirac delta at 0). Let $\bar{f}_h(y) = \int K_h(y - z)f(z)dz$ and $\bar{f}_h(y|u) = \int K_h(y - z)f_{Y_{12}|U_1, Y_{12} \neq 0}(z|u)dz$ denote the surrogate density and conditional density, respectively. Recall that a kernel K is an r th-order kernel for some $r \geq 2$ if $\int y^t K(y)dy = 0$ for $t = 1, \dots, r - 1$ and $\int |y^r K(y)|dy < \infty$. Let M, h_0, σ_0 , and $a \in (0, 1]$ be given positive constants independent of n .

Proposition 1. Suppose that: (i) the data is generated following Equation (16) with point mass at zero, $\mathbb{P}(Y_{ij} \neq 0) = a$ and $Y_{ij} \sim f$ with probability a ; (ii) $\|f\|_\infty \leq M$ and $\sup_{y \in \mathbb{R}, u \in [0, 1]} |f_{Y_{12}|U_1, Y_{12} \neq 0}(y|u)| \leq M$; (iii) for the set of non-zero design points $\{y_1, \dots, y_p\} \subset \mathbb{R}$ and $h \leq h_0$, $\text{Var}(\bar{f}_h(y_\ell|U_1) \cdot \mathbb{P}(Y_{12} \neq 0|U_1)) \geq \sigma_0^2$; (iv) the kernel K is a bounded r th-order kernel for some $r \geq 2$; (v) the bandwidth satisfies $h \rightarrow 0, nh^2 \rightarrow \infty$ as $n \rightarrow \infty$ and $\log^7(pn) = o(nh^2)$. Then we have

$$\mathbb{P}\left(\left(\bar{f}_h(y_\ell)\right)_{\ell=1}^p \in \mathcal{I}(1 - \alpha)\right) \rightarrow (1 - \alpha) \text{ and}$$

$$\mathbb{P}\left(\left(\bar{f}_h(y_\ell)\right)_{\ell=1}^p \in \mathcal{I}^N(1 - \alpha)\right) \rightarrow (1 - \alpha).$$

In addition, if f is r -continuously differentiable, $\|f^{(r)}\|_\infty < \infty$, and $nh^{2r} \log p = o(1)$, then

$$\mathbb{P}\left(\left(f(y_\ell)\right)_{\ell=1}^p \in \mathcal{I}(1 - \alpha)\right) \rightarrow (1 - \alpha) \text{ and}$$

$$\mathbb{P}\left(\left(f(y_\ell)\right)_{\ell=1}^p \in \mathcal{I}^N(1 - \alpha)\right) \rightarrow (1 - \alpha).$$

Some comments on the proposition are in order.

Remark 7. (i) The assumption that \mathbf{g} in Equation (16) is symmetric in its first two arguments can in fact be relaxed. In such case, the conclusions in Proposition 1 continue to hold under a few minor modifications to the regularity conditions. Also, when $a = 1$ and $r = 2$, the proposed dyadic kernel density estimator reduces to the estimator of Graham, Niu, and Powell (2020). The proposition complements Graham, Niu, and Powell (2020) by providing valid simultaneous confidence intervals for their dyadic kernel density estimator. (ii) In some applications, such as in our empirical illustration in Section 6, the object of interest is $b(\cdot)$. For such case, one can simply omit the estimation of a by setting $\hat{a} = 1$ while keeping $\hat{b}(\cdot)$ unaltered. The conclusions in Proposition 1 continue to hold with this modification. (iii) The proof of Proposition 1 does not follow directly from the results of Section 3, as we have to handle the estimation errors of \hat{a} and $\hat{b}(\cdot)$, which involves additional substantial work.

4.2. Penalty Choice for Lasso Under Separate Exchangeability

Consider a regression model

$$Y_i = f(\mathbf{Z}_i) + \varepsilon_i, \quad \mathbb{E}[\varepsilon_i|\mathbf{Z}_i] = 0,$$

$$\mathbf{i} = (i_1, \dots, i_K) \in [N] = \prod_{k=1}^K \{1, \dots, N_k\},$$

where Y_i is a scalar outcome variable, $\mathbf{Z}_i \in \mathbb{R}^d$ is a d -dimensional vector of covariates, $f : \mathbb{R}^d \rightarrow \mathbb{R}$ is an unknown regression function of interest, and ε_i is an error term. We approximate f by a linear combination of technical controls $\mathbf{X}_i = P(\mathbf{Z}_i)$ for some transformation $P : \mathbb{R}^d \rightarrow \mathbb{R}^p$, that is, $f(\mathbf{Z}_i) = \mathbf{X}_i^T \beta_0 + r_i$, $\mathbf{i} \in [N]$, where r_i is a bias term. The dimension p can be much larger than the cluster sizes N , but we assume that the vector $\beta_0 \in \mathbb{R}^p$ is sparse in the sense that $\|\beta_0\|_0 = s \ll n$ with $n = \min_{1 \leq k \leq K} N_k$. Suppose that the array $((Y_i, \mathbf{Z}_i^T)^T)_{i \in \mathbb{N}^K}$ is separately exchangeable and generated as

$$(Y_i, \mathbf{Z}_i^T)^T = \mathbf{g}((U_{i \circ \mathbf{e}})_{\mathbf{e} \in \{0, 1\}^K \setminus \{0\}}), \quad \mathbf{i} \in \mathbb{N}^K,$$

$$\{U_{i \circ \mathbf{e}} : \mathbf{i} \in \mathbb{N}^K, \mathbf{e} \in \{0, 1\}^K \setminus \{0\}\} \stackrel{i.i.d.}{\sim} U[0, 1],$$

for some Borel measurable map $\mathbf{g} : [0, 1]^{2^K - 1} \rightarrow \mathbb{R}^{1+d}$.

Arguably, one of the most popular estimation methods for such a high-dimensional regression problem is the Lasso (Tibshirani 1996); we refer to Bühlmann and van de Geer (2011), Giraud (2015); Wainwright (2019) as standard references on high-dimensional statistics. Let $N = \prod_{k=1}^K N_k$ denote the total sample size. The Lasso estimate for β_0 is defined by

$$\hat{\beta}^\lambda = \arg \min_{\beta \in \mathbb{R}^p} \left\{ \frac{1}{N} \sum_{i \in [N]} (Y_i - \mathbf{X}_i^T \beta)^2 + \lambda \|\beta\|_1 \right\},$$

where $\lambda > 0$ is a penalty level. We estimate the vector $\mathbf{f} = (f_i)_{i \in [N]} = (f(\mathbf{Z}_i))_{i \in [N]}$ by $\hat{\mathbf{f}}^\lambda = (\mathbf{X}_i^T \hat{\beta}^\lambda)_{i \in [N]}$. Let $\|\mathbf{t}\|_{N,2}^2 = N^{-1} \sum_{i \in [N]} t_i^2$ for $\mathbf{t} = (t_i)_{i \in [N]}$.

In what follows, we discuss the statistical performance of the Lasso estimate. Following Bickel, Ritov, and Tsybakov (2009), we say that Condition RE(s, c_0) holds (RE refers to ‘‘restricted eigenvalue’’) if, for a given positive constant $c_0 \geq 1$, the inequality

$$\kappa(s, c_0) = \min_{\substack{J \subset \{1, \dots, p\} \\ 1 \leq |J| \leq s}} \inf_{\substack{\theta \in \mathbb{R}^p, \theta \neq 0 \\ \|\theta_J\|_1 \leq c_0 \|\theta\|_1}} \frac{\sqrt{sN^{-1} \sum_{i \in [N]} (\theta^T \mathbf{X}_i)^2}}{\|\theta_J\|_1} > 0$$

holds with $J^c = \{1, \dots, p\} \setminus J$. Here for $\theta = (\theta_1, \dots, \theta_p)^T$ and $J \subset \{1, \dots, p\}$, $\theta_J = (\theta_j)_{j \in J}$.

In addition, to guarantee fast rates for the Lasso, it is important to choose the penalty level λ in such a way that $\lambda \geq 2c\|\mathbf{S}_N\|_\infty$ with $\mathbf{S}_N = N^{-1} \sum_{i \in [N]} \varepsilon_i \mathbf{X}_i$ for some $c > 1$ (Bickel, Ritov, and Tsybakov 2009; Belloni and Chernozhukov 2013). To this end, we shall estimate the $(1 - \eta)$ -quantile of $2c\|\mathbf{S}_N\|_\infty$ for some small $\eta > 0$. We first estimate the error terms ε_i by pre-estimating β_0 by the preliminary Lasso estimate $\hat{\beta} = \hat{\beta}^{\lambda^0}$ with penalty $\lambda^0 = \tau_n(n^{-1} \log p)^{1/2}$ for some slowing growing sequence $\tau_n \rightarrow \infty$. In the following, we take $\tau_n = \log n$ for the sake of simplicity but other choices also work. We apply the multiplier bootstrap to $\tilde{\mathbf{S}}_N = N^{-1} \sum_{i \in [N]} \tilde{\varepsilon}_i \mathbf{X}_i$ instead of \mathbf{S}_N .

The Hájek projection to \mathbf{S}_N is given by $\sum_{k=1}^K N_k^{-1} \sum_{k=1}^{N_k} \mathbf{V}_{k,i_k}$, where \mathbf{V}_{k,i_k} is given by $\mathbf{V}_{k,i_k} = \mathbb{E}[\varepsilon_{(1, \dots, 1, i_k, 1, \dots, 1)} \mathbf{X}_{(1, \dots, 1, i_k, 1, \dots, 1)} | U_{(0, \dots, 0, i_k, 0, \dots, 0)}]$. We estimate \mathbf{V}_{k,i_k} by $\tilde{\mathbf{V}}_{k,i_k} = (\prod_{k' \neq k} N_{k'})^{-1} \sum_{i_1, \dots, i_{k-1}, i_{k+1}, \dots, i_K} \tilde{\varepsilon}_i \mathbf{X}_i$. Let $\{\xi_{1,i_1}\}_{i_1=1}^{N_1}, \dots, \{\xi_{K,i_K}\}_{i_K=1}^{N_K}$ be iid $N(0, 1)$ variables independent of the data, and consider

$$\Lambda_N^\xi = \left\| \sum_{k=1}^K \frac{1}{N_k} \sum_{i_k=1}^{N_k} \xi_{k,i_k} (\tilde{\mathbf{V}}_{k,i_k} - \tilde{\mathbf{S}}_N) \right\|_\infty.$$

We propose to choose λ as $\lambda = \lambda(\eta) = 2c\Lambda_N^\xi(1 - \eta)$, where $\Lambda_N^\xi(1 - \eta)$ denotes the conditional $(1 - \eta)$ -quantile of Λ_N^ξ . We allow η to decrease with n , i.e., $\eta = \eta_n \rightarrow 0$.

The following proposition establishes the asymptotic validity of our choice of λ (as $n \rightarrow \infty$) under separate exchangeability. In what follows, we understand that s, p, N, η are functions of n while other parameters such as $c, q, \underline{\kappa}$ are independent of n .

Proposition 2 (Penalty choice for the Lasso under separate exchangeability). Suppose that: (i) there exist some constants $q \in [4, \infty)$ independent of n and D_N that may depend on N (and thus on n) such that $\mathbb{E}[|\varepsilon_1|^{2q}] \vee \mathbb{E}[\|\mathbf{X}_1\|_\infty^{2q}] \leq D_N^q$ and $\max_{1 \leq j \leq p} \max_{1 \leq k \leq K} \mathbb{E}[|V_{k,1}^j|^{2+\ell}] \leq D_N^\ell$ for $\ell = 1, 2$; (ii) $\mathbb{E}[|V_{k,1}^j|^2]$ is bounded and bounded away from zero uniformly in $1 \leq j \leq p$ and $1 \leq k \leq K$; (iii) there exists a positive constant $\underline{\kappa}$ independent of n such that $\kappa(s, c_0) \geq \underline{\kappa}$ with probability $1 - o(1)$; and (iv) as $n \rightarrow \infty$, $\|\mathbf{r}\|_{N,2} = O(\sqrt{(s \log p)/n})$ and $\frac{sN^{1/q} D_N^3 \log^7(pN)}{n} \vee \frac{D_N^2 \log^5(pn)}{n^{1-2/q}} = o(1)$. Then, we have $\lambda \geq 2c\|\mathbf{S}_N\|_\infty$ with probability $1 - \eta - o(1)$. Further, we have $\|\hat{\mathbf{f}}^\lambda - \mathbf{f}\|_{N,2} = O_p\left(\sqrt{\frac{s \log p}{n}} \vee \sqrt{\frac{s \log(1/\eta)}{n}}\right)$.

The proof of Proposition 2 does not follow directly from the results of Section 2, as we have to take care of the estimation error of the preliminary Lasso estimate $\tilde{\beta}$, which requires extra work.

Condition (iii) in the preceding proposition is a high-level condition on the sample gram matrix. The following proposition provides primitive sufficient conditions for Condition (iii) to hold for the case of $K = 2$.

Proposition 3 (RE condition under $K = 2$). Consider $K = 2$ and let $B_N = \sqrt{\mathbb{E}[\max_{i \in [N]} \|\mathbf{X}_i\|_\infty^2]}$. Suppose that the eigenvalues of $\mathbb{E}[\mathbf{X}_1 \mathbf{X}_1^T]$ are bounded and bounded away from zero, and $sB_N^2 \log^4(pN) = o(n)$. Then, there exists a positive constant $\underline{\kappa}$ independent of n such that $\kappa(s, c_0) \geq \underline{\kappa}$ with probability $1 - o(1)$.

Under Condition (i) of Proposition 2, $B_N \leq \bar{N}^{1/q} D_N$, so that $sB_N^2 \log^4(pN) = o(n)$ reduces to $s\bar{N}^{1/q} D_N \log^4(pN) = o(n)$, which is implied by Condition (iv) of Proposition 2.

5. Simulation Studies

In this section, we present simulation studies to evaluate the finite sample performance of the proposed multiplier bootstrap methods.

We first describe the simulation design for separately exchangeable arrays. With Σ_Z denoting the $p \times p$ covariance matrix consisting of elements of the form $4^{-|r-c|}$ in its (r, c) th position, separately exchangeable data with $K = 2$ indices are generated according to $\mathbf{X}_i = \frac{1}{4}(\mathbf{Z}_{(i,0)} + \mathbf{Z}_{(0,i_2)}) + \frac{1}{2}\mathbf{Z}_{(i_1,i_2)}$, where $\mathbf{Z}_{i \circ \mathbf{e}} \sim BN(\mathbf{0}, \Sigma_Z) + (1 - B)N(\mathbf{0}, 2\Sigma_Z)$ and $B \sim \text{Bernoulli}(0.5)$ independently for $\mathbf{i} \in \{(i_1, i_2) \in \mathbb{N}^2 : 1 \leq i_1 \leq N_1, 1 \leq i_2 \leq N_2\}$ and $\mathbf{e} \in \{0, 1\}^2$. For this data generating design, we run 2,500 Monte Carlo iterations to compute the uniform coverage frequencies of $\mathbb{E}[\mathbf{X}_i]$ for the nominal probabilities of 90% and 95% using our proposed multiplier bootstrap for separately exchangeable arrays with 2500 bootstrap iterations.

We next describe the simulation design for jointly exchangeable arrays. We shall focus on the most common case in practice, the dyadic data, that is, $K = 2$. With Σ_Z denoting the $p \times p$ covariance matrix consisting of elements of the form $4^{-|r-c|}$ in its (r, c) th position, dyadic samples are generated symmetrically in i and j according to $\mathbf{X}_{i,j} = \frac{1}{4}(\mathbf{Z}_{(i,0)} + \mathbf{Z}_{(j,0)}) + \frac{1}{2}\mathbf{Z}_{(i,j)}$, where $\mathbf{Z}_{i \circ \mathbf{e}} \sim BN(\mathbf{0}, \Sigma_Z) + (1 - B)N(\mathbf{0}, 2\Sigma_Z)$ and $B \sim \text{Bernoulli}(0.5)$ independently for $\mathbf{i} \in \{(i, j) \in \mathbb{N}^2 : 1 \leq i < j \leq n\}$ and $\mathbf{e} \in \{1\} \times \{0, 1\}$. We run 2,500 Monte Carlo iterations to compute the uniform coverage frequencies of \mathbf{S}_n for the nominal probabilities of 90% and 95% using our proposed multiplier bootstrap with 2500 bootstrap iterations.

Table 1 summarizes simulation results under the separate exchangeability. The columns consist of the dimension p of \mathbf{X} and the two-way sample size (N_1, N_2) . The displayed numbers indicate the simulated uniform coverage frequencies for the nominal probabilities of 90% and 95%. For each dimension $p \in \{25, 50, 100\}$, sample sizes vary as $(N_1, N_2) \in \{(25, 25), (50, 50), (100, 100)\}$. Table 2 summarizes simulation results under the joint exchangeability. The columns consist of the dimension p of \mathbf{X} , and the dyadic sample size N . The displayed numbers indicate the simulated uniform coverage

Table 1. Simulation results for separately exchangeable data with $K = 2$ indices.

Normalization	No								
Dimension of $X_i; p$	25	25	25	50	50	50	100	100	100
Sample Sizes: N_1, N_2	25	50	100	25	50	100	25	50	100
90% Coverage	0.927	0.908	0.905	0.942	0.931	0.919	0.943	0.910	0.917
95% Coverage	0.967	0.954	0.956	0.976	0.968	0.960	0.973	0.957	0.962
Normalization	Yes								
Dimension of $X_i; p$	25	25	25	50	50	50	100	100	100
Sample Sizes: N_1, N_2	25	50	100	25	50	100	25	50	100
90% Coverage	0.884	0.892	0.905	0.885	0.885	0.900	0.857	0.878	0.901
95% Coverage	0.936	0.938	0.949	0.930	0.938	0.942	0.921	0.936	0.952

NOTE: Displayed are the dimension p of X , the two-way sample size (N_1, N_2) with $N_1 = N_2$, and the simulated uniform coverage frequencies for the nominal probabilities of 90% and 95%.

Table 2. Simulation results for dyadic data.

Normalization	No								
Dimension of $X_{ij}; p$	25	25	25	50	50	50	100	100	100
Sample Size: n	50	100	200	50	100	200	50	100	200
90% Coverage	0.902	0.896	0.891	0.912	0.914	0.908	0.904	0.915	0.893
95% Coverage	0.960	0.953	0.945	0.956	0.963	0.951	0.953	0.961	0.952
Normalization	Yes								
Dimension of $X_{ij}; p$	25	25	25	50	50	50	100	100	100
Sample Size: n	50	100	200	50	100	200	50	100	200
90% Coverage	0.851	0.854	0.887	0.819	0.865	0.884	0.802	0.870	0.864
95% Coverage	0.921	0.924	0.938	0.890	0.936	0.943	0.882	0.927	0.925

NOTE: Displayed are the dimension p of X , the dyadic sample size n , and the simulated uniform coverage frequencies for the nominal probabilities of 90% and 95%.

frequencies for the nominal probabilities of 90% and 95%. For each dimension $p \in \{25, 50, 100\}$, sample sizes vary as $n \in \{50, 100, 200\}$.

Observe that, for each simulation design and for each nominal probability, the uniform coverage frequencies approach the nominal probability as the sample size increases. These results support the theoretical property of our multiplier bootstrap method. We ran many other sets of simulations with various designs and sample sizes not presented here, but this observed pattern to support our theory remains invariant across all the different sets of simulations—see Appendix H.1. In Appendix H.2, we further experiment with the separate exchangeability with $K = 3$ indices.

6. Real Data Analysis

In this section, we present an empirical application of the method proposed in Section 4.1 to constructing uniform confidence bands for the density functions of bilateral trade volumes in the international trade, with a similar motivation to that stated in Graham, Niu, and Powell (2019, 2020). Recall that our method extends those by Graham, Niu, and Powell (2019) in that we can draw uniform confidence bands as opposed to point-wise confidence intervals. From this analysis, we can learn about the evolution of the distributions of international trade volumes over time.

We employ the international trade data used in Head and Mayer (2014), that come from the Direction of Trade Statistics (DoTS). This dataset contains information about bilateral trade flows among 208 economies for 59 years from 1948 to 2006. In

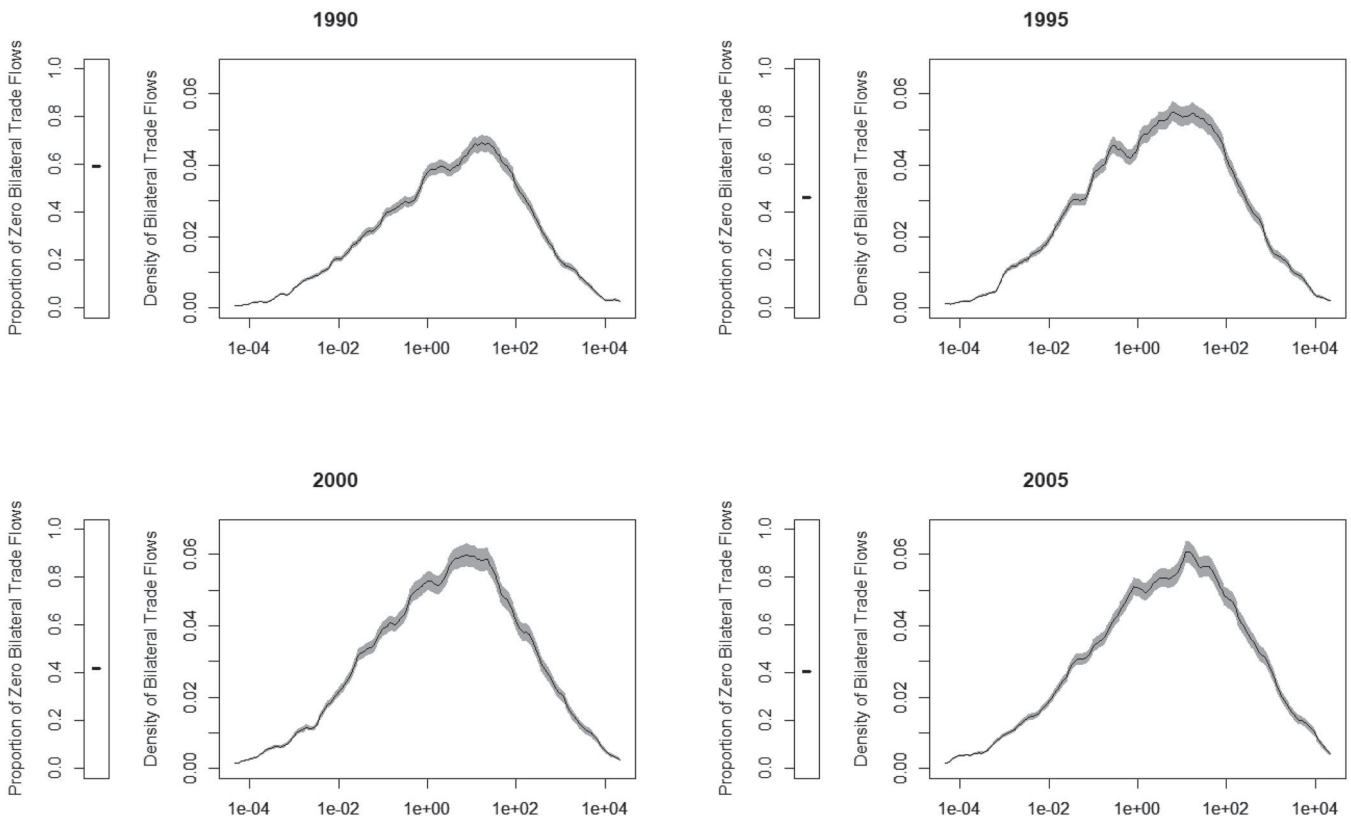


Figure 1. The kernel density estimates (solid curve) and the 95% uniform confidence bands (gray shade) of the bilateral trade volumes in 1990, 1995, 2000, and 2005.

this analysis, we will focus on the relatively recent years, 1990, 1995, 2000, and 2005. Our measure of the bilateral trade volume Y_{ij} is defined as the logarithm of the sum of the trade flow from economy i to economy j and the trade flow from economy j to economy i . We perform simulation studies on confidence bands for densities in Appendix H.3, confirm that the method works as desired, and thus use the same software code here to draw confidence bands of the probability density function of Y_{ij} . Since there is a probability mass at zero in the international trade volumes, what we estimate is precisely the Lebesgue-Radon-Nikodym derivative of the continuous part of the distribution, rather than the probability density function. Specifically, we use $\hat{b}(y)$ defined in Section 4.1 for estimation, and confidence bands are constructed by setting $\hat{\alpha} = 1$. That said, we shall call it a density for conciseness.

Figure 1 illustrates estimates and confidence bands of the density functions of Y_{ij} in each of the years 1990, 1995, 2000, and 2005. Each panel of the figure displays the kernel density estimates in a solid curve and the 95% uniform confidence bands in a gray shade. In addition, we also display the proportion of zero bilateral trade volumes to the left of the kernel density plots so we can get an idea of the complementary proportion that consists the density of the continuously distributed part of the distribution. Although we treat Y_{ij} as the logarithm of the bilateral trade volumes in estimation and inference, we use the original scale on the horizontal axis for ease of reading the graphs.

Observe that the proportion of the zero trade volume is decreasing over time, and the density function is accordingly moving upward over time. Despite this pattern of the changes over time, the shapes of the density functions are rather similar across time in the middle of the distribution. This observation entails a high level of confidence given the reasonably tight confidence bands. On the other hand, notice that the right tail of the distribution becomes fatter as time progresses, implying that there is an increasing number of bilateral trading pairs with very large trade volumes.

7. Summary

In this article, we have developed methods and theories for inference about high-dimensional parameters with separately/jointly exchangeable arrays. Building on the high-dimensional CLTs over the rectangles, we have proposed bootstrap methods and established their finite sample validity for both notions of exchangeability. Simulation studies support the theoretical properties of the methods. We have illustrated a couple of applications of the bootstrap methods. First, extending Graham, Niu, and Powell (2019), we have applied our method to construction of uniform confidence bands for density functions of dyadic data. Second, we have demonstrated an application of our method to penalty choice for ℓ_1 -penalized regression under the separate exchangeability. As such, the results in the present article pave the way for a variety of applications to analyses of separately and jointly exchangeable arrays.

Funding

Harold D. Chiang 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. Kengo Kato supported by NSF (grants nos. DMS-1952306 and DMS-2014636).

Supplementary Material

Supplementary material: The supplementary material includes mathematical proofs and additional simulation studies.

ORCID

Harold D. Chiang  <http://orcid.org/0000-0002-3545-4710>

References

- Aldous, D. J. (1981), “Representations for Partially Exchangeable Arrays of Random Variables,” *Journal of Multivariate Analysis*, 11, 581–598. [2]
- Andrews, D. W. (2005), “Cross-Section Regression With Common Shocks,” *Econometrica*, 73, 1551–1585. [2,3]
- Arnold, S. F. (1979), “Linear Models With Exchangeably Distributed Errors,” *Journal of the American Statistical Association*, 74, 194–199. [2]
- Aronow, P. M., Samii, C., and Assenova, V. A. (2015), “Cluster–Robust Variance Estimation for Dyadic Data,” *Political Analysis*, 23, 564–577. [2]
- Belloni, A., and Chernozhukov, V. (2013), “Least Squares After Model Selection in High-Dimensional Sparse Models,” *Bernoulli*, 19, 521–547. [8]
- Bickel, P. J., and Chen, A. (2009), “A Nonparametric View of Network Models and Newman–Girvan and Other Modularities,” *Proceedings of the National Academy of Sciences*, 106, 21068–21073. [1]
- Bickel, P. J., Chen, A., and Levina, E. (2011), “The Method of Moments and Degree Distributions for Network Models,” *Annals of Statistics*, 39, 2280–2301. [2]
- Bickel, P. J., Ritov, Y., and Tsybakov, A. B. (2009), “Simultaneous Analysis of Lasso and Dantzig Selector,” *Annals of Statistics*, 37, 1705–1732. [8]
- Bickel, P. J., and Rosenblatt, M. (1973), “On Some Global Measures of the Deviations of Density Function Estimates,” *Annals of Statistics*, 1, 1071–1095. [6]
- Bowman, D., and George, E. O. (1995), “A Saturated Model for Analyzing Exchangeable Binary Data: Applications to Clinical and Developmental Toxicity Studies,” *Journal of the American Statistical Association*, 90, 871–879. [2]
- Bühlmann, P., and van de Geer, S. (2011), *Statistics for High-Dimensional Data*. Springer Series in Statistics. Heidelberg: Springer. [7]
- Cameron, A. C., and Miller, D. L. (2014), “Robust Inference for Dyadic Data,” Unpublished manuscript, University of California–Davis. [2]
- Cameron, C. A., Gelbach, J. B., and Miller, D. L. (2011), “Robust Inference With Multiway Clustering,” *Journal of Business & Economic Statistics*, 29, 238–249. [1,2]
- Cameron, C. A., and Miller, D. L. (2015), “A Practitioner’s Guide to Cluster–Robust Inference,” *Journal of Human Resources*, 50, 317–372. [2]
- Caron, F., and Fox, E. B. (2017), “Sparse Graphs Using Exchangeable Random Measures,” *Journal of the Royal Statistical Society, Series B*, 79, 1295. [2]
- Chen, X. (2018), “Gaussian and Bootstrap Approximations for High-Dimensional U-Statistics and Their Applications,” *Annals of Statistics*, 46, 642–678. [2]
- Chen, X., and Kato, K. (2019), “Randomized Incomplete U-Statistics in High Dimensions,” *Annals of Statistics*, 47, 3127–3156. [2]
- (2020), “Jackknife Multiplier Bootstrap: Finite Sample Approximations to The U-Process Supremum With Applications,” *Probability Theory and Related Fields*, 176, 1097–1163. [2]
- Chernozhukov, V., Chetverikov, D., and Kato, K. (2013), “Gaussian Approximations and Multiplier Bootstrap for Maxima of Sums of High-Dimensional Random Vectors,” *Annals of Statistics*, 41, 2786–2819. [2]
- (2014a), “Anti-Concentration and Honest, Adaptive Confidence Bands,” *Annals of Statistics*, 42, 1787–1818. [6]

- (2014b), “Gaussian Approximation of Suprema of Empirical Processes,” *Annals of Statistics*, 42, 1564–1597. [2]
- (2015), “Comparison and Anti-Concentration Bounds for Maxima of Gaussian Random Vectors,” *Probability Theory and Related Fields*, 162, 47–70. [2]
- (2016), “Empirical and Multiplier Bootstraps for Suprema of Empirical Processes of Increasing Complexity, and Related Gaussian Couplings,” *Stochastic Processes and their Applications*, 126, 3632–3651. [2]
- (2017), “Central Limit Theorems and Bootstrap in High Dimensions,” *Annals of Probability*, 45, 2309–2352. [2]
- (2019), “Inference on Causal and Structural Parameters Using Many Moment Inequalities,” *Review of Economic Studies*, 86, 1867–1900. [2]
- Chernozhukov, V., Chetverikov, D., Kato, K., and Koike, Y. (2019), “Improved Central Limit Theorem and Bootstrap Approximations in High Dimensions,” arXiv:1912.10529. [2,4]
- Choi, D. (2017), “Co-Clustering of Nonsmooth Graphons,” *Annals of Statistics*, 45, 1488–1515. [2]
- Choi, D., and Wolfe, P. J. (2014), “Co-Clustering Separately Exchangeable Network Data,” *The Annals of Statistics*, 42, 29–63. [2]
- Claeskens, G., and van Keilegom, I. (2003), “Bootstrap Confidence Bands for Regression Curves and Their Derivatives,” *Annals of Statistics*, 31, 1852–1884. [6]
- Crane, H., and Dempsey, W. (2018), “Edge Exchangeable Models for Interaction Networks,” *Journal of the American Statistical Association*, 113, 1311–1326. [2]
- Davezies, L., D’Haultfoeuille, X., and Guyonvarch, Y. (2021), “Empirical Process Results for Exchangeable Arrays,” *Annals of Statistics*, 49, 845–862. [2,6]
- de la Peña, V., and Giné, E. (1999), *Decoupling: From Dependence to Independence*, New York, NY: Springer. [1,3]
- Deng, H., and Zhang, C.-H. (2020), “Beyond Gaussian Approximation: Bootstrap for Maxima of Sums of Independent Random Vectors,” *Annals of Statistics*, arXiv:1705.09528. [2]
- Diaconis, P., and Janson, S. (2008), “Graph Limits and Exchangeable Random Graphs,” *Rendiconti di Matematica e delle Sue Applicazioni* (7), 28, 33–61. [2]
- Eagleson, G. K., and Weber, N. C. (1978), “Limit Theorems for Weakly Exchangeable Arrays,” *Mathematical Proceedings of the Cambridge Philosophical Society*, 84, 123–130. [2]
- Fafchamps, M., and Gubert, F. (2007), “The Formation of Risk Sharing Networks,” *Journal of Development Economics*, 83, 326–350. [2]
- Fang, X., and Koike, Y. (2020), “High-Dimensional Central Limit Theorems by Stein’s Method,” arXiv:2001.10917. [2]
- Giraud, C. (2015), *Introduction to High-Dimensional Statistics*, Volume 139 of *Monographs on Statistics and Applied Probability*, Boca Raton, FL: CRC Press. [7]
- Goldenberg, A., Zheng, A. X., Fienberg, S. E., and Airoldi, E. M. (2010), “A Survey of Statistical Network Models,” *Foundations and Trends in Machine Learning*, 2, 129–233. [2]
- Graham, B. S., Niu, F., and Powell, J. L. (2019), “Kernel Density Estimation for Undirected Dyadic Data,” arXiv:1907.13630. [6,9,10]
- (2020), “Minimax Risk and Uniform Convergence Rates for Non-parametric Dyadic Regression,” arXiv:2012.08444. [7,9]
- Head, K., and Mayer, T. (2014), “Gravity Equations: Workhorse, Toolkit, and Cookbook,” in *Handbook of International Economics*, Vol. 4, pp. 131–195. Oxford, UK: Elsevier. [6,9]
- Hoover, D. (1979), “Relations on Probability Spaces and Arrays of Random Variables,” Working paper. [2]
- Kallenberg, O. (2006), *Probabilistic Symmetries and Invariance Principles*, New York, NY: Springer Science & Business Media. [2,3,5]
- Koike, Y. (2019), “Gaussian Approximation of Maxima of Wiener Functionals and Its Application to High-Frequency Data,” *Annals of Statistics*, 47, 1663–1687. [2]
- Kuchibhotla, A. K. (2020), “Exchangeability, Conformal Prediction, and Rank Tests,” arXiv:2005.06095. [2]
- Kuchibhotla, A. K., S. Mukherjee, and D. Banerjee (2020), “High-Dimensional CLT: Improvements, Non-Uniform Extensions and Large Deviations,” *Bernoulli*, arXiv:1806.06153. [2]
- Lee, A. J. (1990), *U-Statistics: Theory and Practice*, Boca Raton, FL: CRC Press. [3]
- Lloyd, J., Orbanz, P., Ghahramani, Z., and Roy, D. M. (2012), “Random Function Priors for Exchangeable Arrays With Applications to Graphs and Relational Data,” *Advances in Neural Information Processing Systems*, 25, 998–1006. [2]
- MacKinnon, J. G., Nielsen, M. Ø., and Webb, M. D. (2021), “Wild Bootstrap and Asymptotic Inference With Multiway Clustering,” *Journal of Business and Economic Statistics*, 39, 505–519. [1]
- McCullagh, P. (2000), “Resampling and Exchangeable Arrays,” *Bernoulli*, 6, 285–301. [1,2]
- Menzel, K. (2021), “Bootstrap With Cluster-Dependence in Two or More Dimensions,” *Econometrica*, 89, 2143–2188. [2,3]
- Orbanz, P., and Roy, D. M. (2014), “Bayesian Models of Graphs, Arrays and Other Exchangeable Random Structures,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37, 437–461. [2]
- Owen, A. B. (2007), “The Pigeonhole Bootstrap,” *Annals of Applied Statistics*, 1, 386–411. [1,2]
- Owen, A. B., and Eckles, D. (2012), “Bootstrapping Data Arrays of Arbitrary Order,” *Annals of Applied Statistics*, 6, 895–927. [1,2]
- Silverman, B. W. (1976), “Limit Theorems for Dissociated Random Variables,” *Advances in Applied Probability*, 8, 806–819. [2,6]
- Tabord-Meehan, M. (2019), “Inference With Dyadic Data: Asymptotic Behavior of the Dyadic-Robust t-Statistic,” *Journal of Business & Economic Statistics*, 37, 671–680. [2]
- Tibshirani, R. (1996), “Regression Shrinkage and Selection Via The Lasso,” *Journal of the Royal Statistical Society, Series B*, 58, 267–288. [7]
- Wainwright, M. J. (2019), *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge: Cambridge University Press. [7]
- Zhang, D., and Wu, W. B. (2017), “Gaussian Approximation for High Dimensional Time Series,” *Annals of Statistics*, 45, 1895–1919. [2]
- Zhang, X., and Cheng, G. (2018), “Gaussian Approximation for High Dimensional Vector Under Physical Dependence,” *Bernoulli*, 24, 2640–2675. [2]
- Zhang, Y., Levina, E., and Zhu, J. (2017), “Estimating Network Edge Probabilities by Neighbourhood Smoothing,” *Biometrika*, 104, 771–783. [2]