TWO-WAY CLUSTERING WITH NON-EXCHANGEABLE DATA

KOEN JOCHMANS*

TOULOUSE SCHOOL OF ECONOMICS

This version: January 20, 2026

Abstract

Inference procedures for dyadic data based on two-way clustering rely on the data being exchangeable and dissociated. In particular, observations must be independent if they have no index in common. In an effort to relax this we consider, instead, data where Y_{ij} and Y_{pq} can be dependent for all index pairs, with the dependence vanishing as the distance between the indices grows large. We establish limit theory for the sample mean and propose analytical and bootstrap procedures to perform inference.

Keywords: bootstrap, clustering, dependence, dyadic data, inference, serial correlation

1 Introduction

Consider dyadic data Y_{ij} arising from the pairwise interaction between units $1 \leq i < j \leq n$. Such data arise naturally in the analysis of networks. A concern that arises when dealing with such data is how to take into account the potential dependence between observations when performing statistical inference. Independence across all dyads, i.e., all pairs of units, will often feel too strong to impose. An early contribution where this issue was raised is Fafchamps and Gubert (2007). They proposed what has become known as a two-way

*Address: Toulouse School of Economics, 1 esplanade de l'Université, 31080 Toulouse, France. E-mail: koen.jochmans@tse-fr.eu.
Funded by the European Union (ERC-NETWORK-101044319) and by the French Government and the French National Research Agency under the Investissements d' Avenir program (ANR-17-EURE-0010). Views and opinions expressed are however those of the author only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

clustering approach to constructing standard errors. The key assumption underlying the appropriateness of such a procedure is that Y_{ij} and Y_{pq} are independent unless they have an index in common.

This dependency structure is a natural generalization of the one-way error-component formulation (Moulton 1986, 1990). It holds generally for a dyadic data array provided that it is exchangeable and dissociated. In that case, by well-known results of Aldous (1981), Hoover (1979), and Kallenberg (1989), the data have a nonparametric representation as in

$$Y_{ij} = h(Z_i, Z_j, Q_{ij})$$

for some function h and (latent) random variables $\{Z_i\}$ and $\{Q_{ij}\}$ that are independent and identically distributed across units and dyads, respectively, and independent of each other. This representation is a two-way random-effect model and forms the basis for many popular models of pairwise interaction. Limit theorems and inference procedures that account for the two-way dependence in such data have been formalized in early work by Silverman (1976) and, more recently, Menzel (2021) and Davezies, D'Haultfœuille and Guyonvarch (2021).

Dissociatedness means that dependence is restricted to be within the same column and row of the data array along which, by exchangeability, it is necessarily constant. This may be too severe in many applications. This is particularly clear in a spatial context, for example, but Thompson (2011) has made essentially the same point in a panel data setting featuring random effects for both units and time. As a response to this we consider a generalized version of the representation above by allowing for $\{Z_i\}$ to be a stationary strong-mixing process. To see how this is helpful suppose that the random effect is a moving-average process of order one. Then a given observation will correlate not only with all observations in its own row and column, but also with all observations in the rows right above and below, and the column immediately to the left and right. Furthermore, along the row and diagonal, the correlation is different for the observations immediately adjacent and those further away. Other types of processes, such as autoregressive processes or processes with both an autoregressive and a moving-average component allow for richer forms of dependence.

Within this setup we first establish the limit distribution of the sample mean. The form of the asymptotic variance suggests the construction of a standard error by first clustering observations by unit and then constructing a HAC-type estimator as in the time-series literature. Such an estimator is straightforward to implement and we give conditions under which it is consistent. We also propose a bootstrap procedure that takes into account all forms of dependence. Our proposal takes the form of a (circular) resampling procedure over the units to construct a new index set to which dyadic observations are then attributed. For this procedure we give two results. The first is the consistency of the variance of the bootstrap distribution (conditional on the data) for the long-run variance of the sample mean. The second is the consistency of the bootstrap distribution for the distribution of the sample mean. The latter can be used to perform inference directly, bypassing the construction of a standard error. Some results from Monte Carlo experiments are reported on before concluding remarks discuss various generalizations and extensions. All proofs are collected in the Appendix.

2 Limit behavior of the sample mean

For $1 \leq i \neq j \leq n$ we observe the dyadic variable

$$Y_{ij} = h(Z_i, Z_j, Q_{ij}),$$

where the sequence $\{Z_i\}$ is stationary and strong mixing, the sequence $\{Q_{ij}\}$ is independent and identically distributed, and the sequences $\{Q_{ij}\}$ and $\{Z_i\}$ are independent of one another. The data are undirected and so h is symmetric in its first two arguments, but is otherwise unspecified. Because all its arguments are latent it is without any loss of generality to impose that the marginal distribution of Z_i and of Q_{ij} is uniform on the interval $[0, 1]$.

We are interested in the behavior of the sample mean,

$$\bar{Y}_n = \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{i<j} Y_{ij}.$$

We will derive its large-sample distribution under a set of regularity conditions. To state them, introduce

$$h_2(Z_i, Z_j) = \mathbb{E}(Y_{ij}|Z_i, Z_j),$$

and let

$$\alpha_\tau = \sup_{A\in\mathcal{A}_i} \sup_{B\in\mathcal{B}_{i+\tau}} |\mathbb{P}(A\cap B) - \mathbb{P}(A)\,\mathbb{P}(B)|,$$

for $\mathcal{A}_i$ and $\mathcal{B}_i$ the sigma algebras generated by the sequences $Z_i, Z_{i-1}, \ldots$ and $Z_i, Z_{i+1}, \ldots$.

Assumption 1.

(i) There is a finite constant $c > 0$ so that

$$|h_2(z', z) - h_2(z'', z)| \leq c\,|z' - z''|$$

for all z, z', z''.

(ii) There is a $\delta > 0$ and a finite constant c so that $\mathbb{E}(|Y_{ij}|^2) < c$ and

$$\mathbb{E}(|h_2(Z_i, Z_{i+\tau})|^{2+\delta}) < c$$

for all τ.

(iii) The mixing coefficients satisfy $\alpha_\tau = O(\tau^{-\rho})$ for $\rho > {}^{3\delta+5}/_{2\delta}$ with $\delta > 0$ as given in (ii).

Assumption 1 collects a Lipschitz-continuity condition along with conventional moment conditions and an associated rate of decay on the strong-mixing coefficients of the latent process $\{Z_i\}$.

The smoothness requirement concerns the kernel h_2, and not the link function h. To illustrate the difference, consider the baseline model for non-cooperative network formation

$$Y_{ij} = \{Z_i > Q_{ij}\}\,\{Z_j > Q_{ij}\},$$

where $\{\cdot\}$ denotes the indicator function. Then $h_2(z, z') = \min(z, z')$, which is Lipschitz continuous. Similarly, typical specifications for cooperative link formation are variations of

$$Y_{ij} = \{|Z_i - Z_j| < Q_{ij}\}.$$

Integrating out Q_{ij} thus leads to $h_2(z, z') = 1 - |z - z'|$, which again satisfies Assumption 1(i).

The key step in obtaining the asymptotic behavior of the sample mean is showing that

$$\sqrt{n}(\bar{Y}_n - \theta) = \frac{2}{\sqrt{n}} \sum_{i=1}^{n} X_i + o_P(1),$$

where $\theta = \mathbb{E}(h_1(Z_i))$ and $X_i = h_1(Z_i) - \theta$ for $h_1(z) = \mathbb{E}(h_2(z, Z_j))$. Under Assumption 1 the long-run variance

$$\sigma^2 = \sum_{\tau=-\infty}^{\infty} \omega_\tau, \qquad \omega_\tau = \mathbb{E}(X_i X_{i+\tau})$$

is well defined. The following theorem, whose proof is given in the Appendix, is then obtained.

Theorem 1. *Let Assumption 1 hold and suppose that* $\sigma^2 > 0$. *Then*

$$\sqrt{n}(\bar{Y}_n - \theta) \xrightarrow{L} N(0, 4\sigma^2)$$

as $n \to \infty$

Having $\sigma^2 > 0$ rules out the degenerate case, which would yield a different convergence rate and limit distribution; see, e.g., Leucht (2012) and Menzel (2021) for results in related settings.

3 Inference

We now move to performing inference. We consider two approaches. The first approach operationalizes the limit result in Theorem 1 by combining it with an estimator of the asymptotic variance. Here, we propose two such variance estimators, a plug-in estimator and a bootstrap estimator. The second approach uses the bootstrap to directly construct critical values or confidence intervals.

3.1 HAC variance estimator

To estimate σ^2 one option is to use a truncation-based estimator in the spirit of Newey and West (1987), i.e.,

$$\hat{\sigma}^2 = \hat{\omega}_0 + 2 \sum_{\tau=1}^{m-1} (1 - {}^{\tau}\!/\!{}_{m})\, \hat{\omega}_\tau,$$

where m is a chosen bandwidth parameter and $\hat{\omega}_\tau$ is an estimator of $\omega_\tau = \mathbb{E}(X_i X_{i+\tau})$. As the X_i are unobserved we need to replace them by an estimator. We will use the estimator

$$\hat{\omega}_\tau = \frac{1}{n-\tau} \sum_{i=1}^{n-\tau} (\tilde{Y}_i - \bar{Y}_n)(\tilde{Y}_{i+\tau} - \bar{Y}_n),$$

which amounts to using the centered sample mean $(\tilde{Y}_i - \bar{Y}_n) = {}^{1}\!/\!{}_{(n-1)} \sum_{j \neq i} (Y_{ij} - \bar{Y}_n)$ as an estimator of X_i.

To show consistency of this variance estimator we impose a further moment condition.

Assumption 2. *There is an $r > 2$ so that $\mathbb{E}(|X_i|^{2r}) < \infty$ and $\rho \geq {}^{2r}\!/\!{}_{(r-2)}$.*

The next theorem states the result.

Theorem 2. *Let Assumption 1 and Assumption 2 hold and suppose that $m \to \infty$ with n so that $mn^{-1/4} \to 0$. Then*

$$\hat{\sigma}^2 \xrightarrow{P} \sigma^2$$

as $n \to \infty$.

Selection of the bandwidth parameter in the construction of this estimator can be done by following the procedure in Newey and West (1994).

3.2 Bootstrap

The dyadic nature and complex dependency structure of the data imply that there is no bootstrap scheme available in the literature that is applicable to our problem. We take inspiration from Green and Shalizi (2022) (see also Owen 2007 for a closely-related proposal) and devise a circular version of the moving block bootstrap (Politis and Romano 1992) for dyadic data.

To state the bootstrap scheme we introduce the notation

$$i \bmod n = i - n \left\lfloor {}^{(i-1)}\!/_{n} \right\rfloor .$$

For a chosen integer b, let $\varpi_1, \ldots \varpi_b$ be independent draws from the uniform distribution on the index set $\{1, 2, \ldots, n\}$. For each $1 \leq u \leq b$, we then construct the block of m consecutive indices starting at ϖ_u,

$$\{\varpi_u, (\varpi_u + 1) \bmod n, \ldots, (\varpi_u + m - 1) \bmod n\},$$

and concatenate the b blocks to form a reshuffled index set. We will maintain for notational simplicity that $bm = n$ and note that, here, we re-use the notation m to indicate the block length. We do so as it plays a similar role to the bandwidth parameter in the previous subsection. Thus, the approach is to randomly draw b blocks of length m from the index set $\{1, 2, \ldots, n\}$, wrapping around in a circle. Each $1 \leq i \leq n$ in the bootstrapped index set maps to the original index set via

$$\phi_i = \left(\varpi_{\lceil i/m \rceil} + i - \left(\lceil i/m \rceil - 1\right) m - 2\right) \bmod n + 1.$$

We then generate a bootstrap sample as

$$Y_{ij}^* = \begin{cases} Y_{\phi_i \phi_j} & \text{if } \phi_i \neq \phi_j \\ 0 & \text{if } \phi_i = \phi_j \end{cases} .$$

from which the bootstrap sample mean,

$$\bar{Y}_n^* = \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{i<j} Y_{ij}^*,$$

may then be constructed.

Setting $Y_{ij}^* = 0$ when $\phi_i = \phi_j$ is done because there is no corresponding observation in the original data. In the Appendix, Theorems 3 and 4 below are shown to hold when, instead of setting the bootstrap observation to zero, one sets it to a random variable D_i

for which $\mathbb{E}(D_i^2) < \infty$. The use of zero effectively means that the observation does not contribute to $\sum_{i=1}^n \sum_{i<j} Y_{\phi_i \phi_j}$, and so seems a natural choice. We show in the Appendix that

$$\mathbb{E}^*(\bar{Y}_n^*) = \left(\frac{n-m}{n}\right) \bar{Y}_n + O_P\left(\frac{m-1}{n}\right) \tag{3.1}$$

Here, the first right-hand side term features a shrinkage factor that is due to setting the artificial observations equal to zero. The remainder term, whose precise form is given in the Appendix, appears only when $m > 1$ and does not depend on how what value we choose for the artificial observations. Given this expression it is possible to adjust $\bar{Y}_n^*$ to make it (conditionally) unbiased for $\bar{Y}_n$. For the case where $m = 1$ this is immediate by a simple rescaling by ${}^{n}/_{(n-m)}$.

The intuition for why our bootstrap scheme works is that, as we show in the Appendix,

$$\sqrt{n}(\bar{Y}_n^* - \bar{Y}_n) = \frac{2}{\sqrt{n}} \sum_{i=1}^n X_i^* + o_{P^*}(1),$$

where $X_i^* = X_{\phi_i}$ and $A_n^* = o_{P^*}(1)$ means that $\mathbb{P}(\mathbb{P}^*(|A_n^*| > \epsilon^*) > \epsilon) = o(1)$ for all $\epsilon^* > 0$ and $\epsilon > 0$. The dominant term on the right-hand side is a (scaled) sample mean of a stationary mixing process, for which the validity of the traditional circular block bootstrap is well established.

A bootstrap estimator of σ^2 is

$$n \operatorname{var}^*(\bar{Y}_n^*) = n\, \mathbb{E}^*((\bar{Y}_n^* - \mathbb{E}^*(\bar{Y}_n^*))^2),$$

that is, the variance of the bootstrap distribution, scaled by n. This estimator, too, is consistent for σ^2. We show this under the following strengthening of the conditions in Assumption 2.

Assumption 3. *There is an $r > 2$ so that $\mathbb{E}(|X_i|^{3r}) < \infty$ and $\rho \geq {}^{3r}/_{(r-2)}$.*

The result is as follows.

Theorem 3. *Let Assumption 1 and Assumption 3 hold and suppose that $m \to \infty$ with n so that $mn^{-1} \to 0$. Then*

$$n \operatorname{var}^*(\bar{Y}_n^*) \xrightarrow{P} \sigma^2$$

as $n \to \infty$.

The choice of block length m in practice can be guided by taking the approach of Politis and White (2004).

Theorems 2 and 3 are qualitatively similar. In conjunction with Theorem 1, either can be used to construct test statistics that yield asymptotically size-correct inference. While our argument for the bootstrap variance requires stronger moment conditions, it allows for the block size to grow considerably more quickly than the bandwidth parameter in the HAC estimator.

The next theorem shows that the bootstrap can also be used to approximate the entire distribution of the sample mean.

Theorem 4. *Let Assumption 1 and Assumption 3 hold and suppose that* $m \to \infty$ *with* n *so that* $mn^{-1} \to 0$*. Then*

$$\sup_a |\mathbb{P}^*(\sqrt{n}(\bar{Y}_n^* - \bar{Y}_n) \leq a) - \mathbb{P}(\sqrt{n}(\bar{Y}_n - \theta) \leq a)| \xrightarrow{P} 0$$

as $n \to \infty$.

There are results on selecting the optimal block length for estimation of a distribution in the time-series literature; see Hall, Horowitz and Jing (1995). These results concern optimality at a point or in a mean-squared error sense. The optimal choices differ depending on the objective.

By van der Vaart (2000, Lemma 23.3), Theorem 4 implies that inference based on the reverse-percentile bootstrap is asymptotically justified. For any $\alpha \in (0, 1)$, denote by $\hat{Q}_\alpha$ the α-quantile of the distribution of $\sqrt{n}(\bar{Y}_n^* - \bar{Y}_n)$, conditional on the original data. Then,

$$\mathbb{P}(\sqrt{n}(\bar{Y}_n - \theta) \leq \hat{Q}_\alpha) \xrightarrow{n \to \infty} \alpha.$$

As an example, a two-sided confidence set for θ with nominal coverage $1 - \alpha$ is the interval

$$\left\{ \vartheta : \bar{Y}_n - \frac{\hat{Q}_{1-\alpha/2}}{\sqrt{n}} \leq \vartheta \leq \bar{Y}_n - \frac{\hat{Q}_{\alpha/2}}{\sqrt{n}} \right\}$$

for any $\alpha < 1/2$. Critical values and p-values for specific null hypotheses follow from the same argument.

4 Numerical illustration

A simulation experiment was conducted to evaluate the performance of the various inference techniques proposed above. Here we will restrict attention to data generating process where $\{X_i\}$ is an AR(1) process with autoregressive parameter set to either $1/4$ or $1/2$, and innovations drawn from the standard-normal distribution. Simulation results for other designs and other distributions gave very similar results and are omitted for brevity. Table 1 contains simulations results for $n = 250$ and various choices for the bandwidth parameter for the HAC estimator (m_{HAC}) and the block length for the circular bootstrap (m_{CB}).

For each of the values for m_{HAC} considered the table provides the average (over the Monte Carlo simulations) of the HAC standard error, $\hat{\sigma}$, along with the coverage of the associated 95% confidence interval for θ constructed via the normal approximation implied by Theorem 1. The same quantities are reported for the bootstrap (CB) for each of the values of m_{CB}. Coverage rates for two intervals based on the reverse-percentile method are equally reported. The first, CB%, uses the quantiles of the bootstrap distribution centered at $\bar{Y}_n$, following directly Theorem 4. The second, CCB%, instead, centers around $\mathbb{E}^*(\bar{Y}_n^*)$; this is done by using the exact formula in the Appendix. We do this as such recentering is known to lead to improvements in related situations (Lahiri 1991). Yet another possibility to obtain refinements, one that we do not explore here, would be to use the double bootstrap (Beran 1988).

First, for $m_{\text{HAC}} = m_{\text{CB}} = 1$, both the analytical and the bootstrap approach mistake the data array for being dissociated by not taking into account serial dependence. This leads to standard errors that underestimate the Monte Carlo standard deviation of $\bar{Y}_n$ (reported in the table for each design). The standard deviation is underestimated by more than 20% when $\rho = 1/4$ and by more than 40% when $\rho = 1/4$. This downward bias decreases as the respective bandwidth parameters increase and additional covariance terms are taken into account.

The coverage intervals constructed using either the HAC or bootstrap-based variance estimator follow suit. The severe undercoverage observed at $m_{\text{HAC}} = m_{\text{CB}} = 1$ diminishes

Table 1: Simulation results for the AR(1) process

$\rho = 1/4$		Std. error (0.1690)		Coverage rate (95%)			
m_{HAC}	m_{CB}	HAC	CB	HAC	CB	CB%	CCB%
1	1	0.1298	0.1295	0.8616	0.8608	0.8576	0.8588
2	2	0.1446	0.1441	0.9002	0.8988	0.8954	0.8960
3	5	0.1513	0.1559	0.9144	0.9206	0.9170	0.9180
5	10	0.1566	0.1581	0.9238	0.9206	0.9188	0.9194
10	20	0.1591	0.1657	0.9216	0.9264	0.9450	0.9452
$\rho = 1/2$		Std. error (0.2530)		Coverage rate (95%)			
m_{HAC}	m_{CB}	HAC	CB	HAC	CB	CB%	CCB%
1	1	0.1446	0.1442	0.7370	0.7370	0.7324	0.7346
2	2	0.1765	0.1758	0.8244	0.8208	0.8208	0.8226
3	5	0.1946	0.2121	0.8592	0.8886	0.8872	0.8884
5	10	0.2133	0.2258	0.8908	0.9076	0.9050	0.9056
10	20	0.2276	0.2421	0.9080	0.9186	0.9360	0.9364

Results are based on 5000 Monte Carlo replications and 999 bootstrap replications.

as the bandwidth and block length increase. The percentile-based bootstrap confidence intervals both perform similar to the normal approximation. Lastly, we can observe an improvement of CCB% relative to CB% for all values of m_{CB} and in both tables, although it is very modest throughout.

5 Concluding remarks

In this paper we have looked at the problem of inference with dyadic data. The conventional setting implies that dependence between Y_{ij} and Y_{pq} is present only when $i < j$ and $p < q$ have an index in common, and that their covariance in such a case is independent of the indices. Such a setup can be too restrictive in a variety of practical applications. Our framework allows for dependence between Y_{ij} and Y_{pq} for all values of the indices. The dependency structure between non-overlapping index pairs is ergodic, and we stated conditions under which this ensures the sample mean to converge to a Gaussian process at the usual $n^{-1/2}$ convergence rate. To perform inference, we gave a plug-in variance estimator that generalizes the usual two-way clustering approach and devised a new bootstrap scheme that can be used to estimate both the asymptotic variance and the quantiles of the limit distribution.

Extensions in various directions are possible. First, a directed version of our setup can be set up by having a pair of variables $Z_i = (Z_i^o, Z_i^{oo})$, and then asymmetric dyadic variables

$$Y_{ij} = h(Z_i^o, Z_j^{oo}, Q_{ij}).$$

Here, subject to suitable regularity conditions we will have a representation for the sample mean as

$$\sqrt{n}(\bar{Y}_n - \theta) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i^o + X_i^{oo}) + o_P(1),$$

where, now, $X_i^o = h_1^o(Z_i) - \theta$ and $X_i^{oo} = h_1^{oo}(Z_i) - \theta$ for functions $h_1^o(z) = \mathbb{E}(h(z, Z_j, Q_{ij}))$ and $h_1^{oo}(z) = \mathbb{E}(h(Z_i, z, Q_{ij}))$. The asymptotic variance of the sample mean will then be equal to the long-run variance of $X_i^o + X_i^{oo}$, for which a HAC or bootstrap estimator can again be constructed.

Second, there is no apparent reason why our results could not be extended to multi-adic data, where we observe $Y_{\boldsymbol{i}}$ for tuples $\boldsymbol{i} = (i_1, i_2, \ldots, i_d)$ of size $d > 2$. Such data are studied in Owen and Eckles (2012) and Davezies, D'Haultfœuille and Guyonvarch (2021), where they are assumed to be exchangeable and dissociated. To build in additional dependence, the same intuition as here applies. In particular, subject to regularity conditions, we should have

$$\sqrt{n}(\bar{Y}_n - \theta) + \frac{d}{\sqrt{n}} \sum_{i=1}^{n} X_i + o_P(1),$$

where, now, the sample mean is computed over all unique d-tuples $\boldsymbol{i}$, $X_i = h_1(Z_i) - \theta$ as before with h_1 now corresponding to the function obtained on integrating out all random variables but Z_i from the Kallenberg (1989) representation of the data array; see, e.g., Equation (2.1) in Davezies, D'Haultfœuille and Guyonvarch (2021) for this representation.

Appendix

Proof of Theorem 1

We begin by defining

$$V_{ij} = Y_{ij} - \theta - X_i - X_j$$

where, recall, $X_i = h_1(Z_i) - \theta$ for $h_1(z) = \mathbb{E}(h_2(z, Z_j))$ with $h_2(z, z') = \mathbb{E}(h(z, z', Q_{ij}))$. Then

$$Y_{ij} = \theta + X_i + X_j + V_{ij}.$$

Averaging over all observations yields

$$\hat{\theta} = \theta + \frac{2}{n} \sum_{i=1}^{n} X_i + \bar{V}_n$$

for $\bar{V}_n = {}^{2}\!/_{n(n-1)} \sum_{i=1}^{n} \sum_{i<j} V_{ij}$. The proof of the theorem proceeds in two steps. The first step is to show that $\bar{V}_n = o_p(n^{-1/2})$. The second step is to show that, as n diverges, ${}^{1}\!/_{\sqrt{n}} \sum_{i=1}^{n} X_i$ converges in distribution to a normal random variable with mean zero and variance $\sigma^2 > 0$. From this

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{L} N(0, 4\sigma^2)$$

follows, thereby completing the proof of the theorem.

Remainder term. To show that $\bar{V}_n = o_P(n^{-1/2})$ it is convenient to introduce the variables

$$W_{ij} = Y_{ij} - h_2(Z_i, Z_j), \qquad U_{ij} = h_2(Z_i, Z_j) - h_1(Z_i) - h_1(Z_j) + \theta.$$

Then $\bar{V}_n = \bar{W}_n + \bar{U}_n$ for $\bar{W}_n = {}^2/_{n(n-1)} \sum_{i=1}^n \sum_{i<j} W_{ij}$ and $\bar{U}_n = {}^2/_{n(n-1)} \sum_{i=1}^n \sum_{i<j} U_{ij}$. We handle each of these two terms in turn.

Conditional on $Z_1, \ldots, Z_n$, the variables W_{ij} are independent across dyads. They have (conditional) mean zero and variance $\mathbb{E}(W_{ij}^2 | Z_i, Z_j)$. We therefore see that $\mathbb{E}(\bar{W}_n) = 0$ and

$$\mathbb{E}(\bar{W}_n^2) = 4\, \mathbb{E}\left(\operatorname{var}\left(\left. \frac{\sum_{i=1}^n \sum_{i<j} W_{ij}}{n(n-1)} \right| Z_1, \ldots, Z_n \right) \right) = 4\, \frac{\sum_{i=1}^n \sum_{i<j} \mathbb{E}(\mathbb{E}(W_{ij}^2 | Z_i, Z_j))}{n^2(n-1)^2},$$

where we have used the law of total variance in the first step. Notice that, because we have that

$$\mathbb{E}(W_{ij}^2 | Z_i, Z_j) = \mathbb{E}((Y_{ij} - \theta)^2 | Z_i, Z_j) - (h_2(Z_i, Z_j) - \theta)^2 \leq \mathbb{E}((Y_{ij} - \theta)^2 | Z_i, Z_j),$$

it holds that $\mathbb{E}(\mathbb{E}(W_{ij}^2 | Z_i, Z_j)) = \mathbb{E}(W_{ij}^2) \leq \mathbb{E}((Y_{ij} - \theta)^2)$, which is uniformly bounded by Assumption 1. Therefore,

$$\mathbb{E}(\bar{W}_n^2) \lesssim \frac{2}{n(n-1)}$$

where, here and later, we use the notation $A \lesssim B$ to indicate that $A \leq c\, B$ for some finite constant $c > 0$. Consequently,

$$\bar{W}_n = O_P(n^{-1})$$

follows by Markov's inequality.

Moving on, observe that $\bar{U}_n$ is a degenerate U-statistic in the variables $Z_1, \ldots, Z_n$. Hence, $\mathbb{E}(\bar{U}_n) = 0$. We now proceed by deriving an upper bound on its variance, that is, on

$$\mathbb{E}(\bar{U}_n^2) = \frac{4}{n^2(n-1)^2} \sum_{i_1=1}^n \sum_{i_1<j_1} \sum_{i_2=1}^n \sum_{i_2<j_2} \mathbb{E}(U_{i_1 j_1} U_{i_2 j_2}).$$

Under the conditions stated in Assumption 1, and given the rate of decay of the mixing coefficients α_τ being at most of order $\tau^{-\rho}$, we have, letting $\varepsilon = \rho\, {}^{2\delta}/_{(3\delta+5)}$, the upper bound

$$|\mathbb{E}(U_{i_1 j_1} U_{i_2 j_2})| \lesssim \alpha_{\max\{j_1 - i_1, j_2 - i_2\}}^{2\delta/(3\delta+5)} \lesssim \max\{j_1 - i_1, j_2 - i_2\}^{-\varepsilon}$$

by an application of Lemma 3.3 in Dehling and Wendler (2010). Introduce the shorthands

$$m_1 = j_1 - i_1 \geq 1, \qquad m_2 = j_2 - i_2 \geq 1.$$

For fixed m_1 there are $n - m_1$ choices of i_1, and for fixed m_2 there are $n - m_2$ choices of i_2. Thus,

$$\mathbb{E}(\bar{U}_n^2) \lesssim \frac{1}{n^2(n-1)^2} \sum_{m_1=1}^{n-1} \sum_{m_2=1}^{n-1} (n - m_1)(n - m_2) \max\{m_1, m_2\}^{-\varepsilon}.$$

By symmetry of the summand in the pair (m_1, m_2) we are free to presume that $m_1 \geq m_2$ and multiply through by two to obtain

$$\mathbb{E}(\bar{U}_n^2) \lesssim \frac{1}{n^2(n-1)^2} \sum_{m_1=1}^{n-1} \sum_{m_2=1}^{m_1} (n - m_1)(n - m_2)\, m_1^{-\varepsilon}.$$

Note that $n - m_1 \leq n$ and $n - m_2 \leq n$, and there are at most m_1 terms in the inner sum. Hence, $(n - m_1) \sum_{m_2=1}^{m_1} (n - m_2) \leq n^2 m_1$ and

$$\mathbb{E}(\bar{U}_n^2) \lesssim \frac{1}{(n-1)^2} \sum_{m_1=1}^{n-1} m_1^{1-\varepsilon} \lesssim n^{-\varepsilon},$$

using that $\sum_{m_1=1}^{n-1} m_1^{1-\varepsilon} \lesssim n^{2-\varepsilon}$. Observing that $\varepsilon > 1$ we thus have $\mathbb{E}(\bar{U}_n^2) = o(n^{-1})$ and so

$$\bar{U}_n = o_P(n^{-1/2})$$

follows by another application of Markov's inequality. We have, therefore, shown that $\bar{V}_n = o_P(n^{-1/2})$.

Leading term. Having established that

$$\sqrt{n}(\hat{\theta} - \theta) = \frac{2}{\sqrt{n}} \sum_{i=1}^{n} X_i + o_P(1)$$

we set out to show asymptotic normality. In light of Assumption 1, the sequence $\{X_i\}$ is zero mean, stationary, and strong mixing. Moreover, we have that $\mathbb{E}(|X_i|^{2+\delta}) < \infty$ and that the mixing coefficients are of size $-\rho < -{(2+\delta)}/{\delta}$. These conclusions on the sequence

$\{X_i\}$ follow from an application of standard results (see, e.g., White 2001, Theorem 3.49). Therefore, provided that the long-run variance of the process $\{X_i\}$ is positive, that is, that, as assumed in the theorem,

$$\sigma^2 = \mathbb{E}(X_i^2) + 2\sum_{\tau=1}^{\infty}\mathbb{E}(X_i X_{i+\tau}) > 0,$$

we readily obtain that ${}^1\!/\!\sqrt{n}\sum_{i=1}^{n} X_i \xrightarrow{L} N(0,\sigma^2)$ (see, e.g., White 2001, Theorem 5.20), thereby completing the proof. $\square$

Proof of Theorem 2

Introduce the infeasible estimator $\breve{\sigma}^2 = \breve{\omega}_0 + 2\sum_{\tau=1}^{m-1}(1 - {}^{\tau}\!/\!{}_{m})\,\breve{\omega}_\tau$ that uses the covariance estimators

$$\breve{\omega}_\tau = \frac{1}{n-\tau}\sum_{i=1}^{n-\tau} X_i X_{i+\tau}.$$

By virtue of Assumption 2 and the requirement that ${}^{m^4}\!/\!{}_{n} \to 0$ as $n \to \infty$, Theorem 6.20 in White (2001) can be applied to obtain $\breve{\sigma}^2 \xrightarrow{P} \sigma^2$ as $n \to \infty$. We then need to show that $|\breve{\sigma}^2 - \sigma^2| \xrightarrow{P} 0$ to complete the proof.

To do so we establish that (i) $|\hat{\omega}_0 - \breve{\omega}_0| = o_P(1)$ and that (ii) $\sum_{\tau=1}^{m-1}|\hat{\omega}_\tau - \breve{\omega}_\tau| = o_P(1)$, which suffices in light of the fact that $0 \leq (1 - {}^{\tau}\!/\!{}_{m}) \leq 1$ for all relevant τ. To proceed, we begin by recalling, from the proof of Theorem 1, the decomposition of the observations as

$$Y_{ij} = \theta + X_i + X_j + V_{ij}.$$

Then, by averaging and recentering, we obtain $(\tilde{Y}_i - \bar{Y}_n) = (X_i - \bar{X}_n) + (\tilde{X}_i - \bar{X}_n) + (\tilde{V}_i - \bar{V}_n)$, for

$$\tilde{X}_i = \frac{1}{n-1}\sum_{j\neq i} X_j, \qquad \tilde{V}_i = \frac{1}{n-1}\sum_{j\neq i} V_{ij},$$

and $\bar{X}_n = {}^1\!/\!{}_n\sum_{i=1}^{n}\tilde{X}_i$ and $\bar{V}_n = {}^1\!/\!{}_n\sum_{i=1}^{n}\tilde{V}_i$. The sampling error in $(\bar{Y}_i - \bar{Y}_n)$ as an estimator of X_i thus equals

$$A_i = (\tilde{Y}_i - \bar{Y}_n) - X_i = -\bar{X}_n + (\tilde{X}_i - \bar{X}_n) + (\tilde{V}_i - \bar{V}_n).$$

For any given τ, by an application of the Cauchy-Schwarz inequality and using stationarity,

$$\mathbb{E}(|\hat{\omega}_\tau - \tilde{\omega}_\tau|) \leq \mathbb{E}(|X_i A_{i+\tau}|) + \mathbb{E}(|X_{i+\tau} A_i|) + \mathbb{E}(|A_i A_{i+\tau}|)) \leq 2\sqrt{\mathbb{E}(X_i^2)\,\mathbb{E}(A_i^2)} + \mathbb{E}(A_i^2).$$

By assumption, $\mathbb{E}(X_i^2) = O(1)$. Below we show that $\mathbb{E}(A_i^2) = O(n^{-1})$. Then, by Markov's inequality, for any $\epsilon > 0$,

$$\mathbb{P}\left(|(\hat{\omega}_0 - \tilde{\omega}_0)| > \epsilon\right) \leq \frac{\mathbb{E}\left(|(\hat{\omega}_0 - \tilde{\omega}_0)|\right)}{\epsilon} = O(n^{-1/2}),$$

which establishes (i), and, similarly,

$$\mathbb{P}\left(\sum_{\tau=1}^{m-1} |(\hat{\omega}_\tau - \tilde{\omega}_\tau)| > \epsilon\right) \leq \frac{\sum_{\tau=1}^{m-1} \mathbb{E}\left(|(\hat{\omega}_\tau - \tilde{\omega}_\tau)|\right)}{\epsilon} = O(mn^{-1/2}),$$

from which (ii) follows given that $m^2 n^{-1} \to 0$ as $n \to \infty$.

To bound $\mathbb{E}(A_i^2)$ it suffices to obtain bounds on the second moment of each of the components of A_i. First, as,

$$n^2\,\mathbb{E}(\bar{X}_n^2) = \sum_{i=1}^n \sum_{j=1}^n \mathbb{E}(X_i X_j) = n\omega_0 + 2\sum_{\tau=1}^{n-1} (n-\tau)\,\omega_\tau$$

and $|\omega_\tau| \lesssim \alpha_\tau^{1/2 - 1/2r} = O(\tau^{-\rho/2\,(r-1)/r})$ by Corollary 6.17 in White (2001) and $\alpha_\tau = O(\tau^{-\rho})$, it follows that

$$\mathbb{E}(\bar{X}_n^2) \lesssim n^{-2} \sum_{\tau=1}^{n-1} (n-\tau)\tau^{-\rho\,(r-1)/2r} = O(n^{-1})$$

on noting that $\rho\,{}^{(r-1)}/_{2r} > 1$ by Assumption 2.

Moving on to $\mathbb{E}(\tilde{X}_i^2)$ we first expand the sum to write

$$(n-1)^2\,\mathbb{E}(\tilde{X}_i^2) = \sum_{j=1}^n \sum_{k=1}^n \mathbb{E}(X_j X_k) - \sum_{k \neq i} \mathbb{E}(X_i X_k) - \mathbb{E}(X_i^2).$$

We note that the first term is equal to $n^2\,\mathbb{E}(\bar{X}_n^2)$, which has already been shown to be $O(n)$, while the third term is $\mathbb{E}(X_i^2) = \omega_0 = O(1)$. The middle term, finally, behaves like $\sum_{\tau=1}^{n-1} \tau^{-\rho\,(r-1)/2r}$, which remains bounded as n grows large given our rate condition on the mixing coefficients. Therefore, $\mathbb{E}(\bar{X}_n^2) = O(n^{-1})$ holds.

Next we note that $\mathbb{E}(\bar{V}_n^2) = O(n^{-1})$ was already shown to hold in the proof of Theorem 1. The argument to show that $\mathbb{E}(\tilde{V}_i^2) = O(n^{-1})$ proceeds in the same way. Moreover, recalling the decomposition $V_{ij} = U_{ij} + W_{ij}$ from the proof of Theorem 1 we can write that

$$\tilde{V}_i = \frac{1}{n-1} \sum_{j \neq i} (U_{ij} + W_{ij}) = \tilde{U}_i + \tilde{W}_i.$$

First, because the W_{ij} are serially uncorrelated across dyads, it is immediate that we have

$$\mathbb{E}(\tilde{W}_i^2) = \frac{1}{(n-1)^2} \sum_{j \neq i} \sum_{k \neq i} \mathbb{E}(W_{ij} W_{ik}) = \frac{1}{(n-1)^2} \sum_{j \neq i} \mathbb{E}(W_{ij}^2) = O(n^{-1}).$$

Next, for the remaining part, we can again rely on Lemma 3.3 in Dehling and Wendler (2010) to arrive at

$$(n-1)^2 \mathbb{E}(\tilde{U}_i^2) = \sum_{j \neq i} \sum_{k \neq i} \mathbb{E}(U_{ij}\, U_{ik}) \lesssim \sum_{m_1=1}^{n-1} \sum_{m_2=1}^{n-1} \max\{m_1, m_2\}^{-\epsilon} = 2 \sum_{m_1=1}^{n-1} m_1^{1-\epsilon} = O(n^{2-\epsilon}),$$

for some $\epsilon > 1$, in light of our rate requirement on the mixing coefficients. Therefore, $\mathbb{E}(\tilde{U}_i^2) = O(n^{-\epsilon}) = o(n^{-1})$. Taken together, the final two rates imply that $\mathbb{E}(\tilde{V}_i^2) = O(n^{-1})$.

We have thus shown that $\mathbb{E}(A_i^2) = O(n^{-1})$. Then (i) and (ii) both follow and the proof is complete. $\square$

Proof of Equation (3.1)

We will derive the bootstrap expectation in the slightly more general setting where we set

$$\dot{Y}_{ij}^* = \dot{Y}_{\phi_i \phi_j} = \begin{cases} Y_{\phi_i \phi_j} & \text{if } \phi_i \neq \phi_j \\ D_{\phi_i} & \text{if } \phi_i = \phi_j \end{cases}$$

for chosen $D_1, \ldots, D_n$ which are non-random conditional on the data. We wish to compute $\mathbb{E}^*(\bar{Y}_n^*) = {}^1\!/_{n(n-1)} \sum_{i=1}^n \sum_{j \neq i} \mathbb{E}^*(\dot{Y}_{\phi_i \phi_j})$. It will be useful to make the bootstrap expectation explicit. We have

$$\mathbb{E}^*(\bar{Y}_n^*) = \frac{1}{n(n-1)} \sum_{u=1}^{b} \sum_{i=0}^{m-1} \sum_{v=1}^{b} \sum_{j=0}^{m-1} \mathbb{E}^*(\dot{Y}_{(\varpi_u + i) \bmod n\, (\varpi_v + j) \bmod n}) \{i \neq j \text{ if } u = v\},$$

where the indicator function prevents (undefined) entries $\dot{Y}^*_{ii}$ from arising in the summand. We next split up the right-hand sided into two terms; the first concerns observations within the same block, i.e.,$u = v$, the second observations across blocks, i.e., $u \neq v$.

The within-block contribution is

$$\frac{1}{n(n-1)} \sum_{u=1}^{b} \sum_{i=0}^{m-1} \sum_{j \neq i} \mathbb{E}^*(\dot{Y}_{(\varpi_u+i) \bmod n\, (\varpi_u+j) \bmod n}).$$

Notice that this term is exactly equal to zero when $m = 1$. Thus take $m > 1$. As ϖ_u is uniformly distributed on $\{1, 2, \dots, n\}$, this is equal to

$$\frac{b}{n(n-1)} \sum_{i=0}^{m-1} \sum_{j \neq i} \frac{\sum_{k=1}^{n} \dot{Y}_{(k+i) \bmod n\, (k+j) \bmod n}}{n}.$$

Choose a pair of indices (i, j) with $j - i = d > 0$. By the circularity of the bootstrap design,

$$\frac{\sum_{k=1}^{n} \dot{Y}_{(k+i) \bmod n\, (k+j) \bmod n}}{n} = \frac{\sum_{k=1}^{n} \dot{Y}_{(k) \bmod n\, (k+d) \bmod n}}{n} = \frac{\sum_{i=1}^{n-d} Y_{i(i+d)} + \sum_{i=1}^{d} Y_{(n-d+i)i}}{n},$$

which is the average of the (extended) dth off-diagonal of the data matrix, say $\bar{Y}_{(d)}$. This only depends on d. Hence, exploiting symmetry of the data, the within-block contribution to $\mathbb{E}^*(\bar{Y}^*_n)$ equals

$$2 \frac{b}{n(n-1)} \sum_{d=1}^{m-1} (m-d)\, \bar{Y}_{(d)} = O_P((m-1)(n-1)^{-1});$$

here, the rate holds because $\bar{Y}_{(d)} = O_P(1)$ uniformly in d and $\sum_{d=1}^{m-1}(m-d) = {}^{m(m-1)}/_{2}$, and we have used that $bm = n$.

The between-block contribution is

$$\frac{1}{n(n-1)} \sum_{u=1}^{b} \sum_{i=0}^{m-1} \sum_{v \neq u} \sum_{j=0}^{m-1} \mathbb{E}^*(\dot{Y}_{(\varpi_u+i) \bmod n\, (\varpi_v+j) \bmod n}),$$

observing that the indicator function will always take the value one for observations across different blocks. Here, ϖ_u and ϖ_v are independent draws from the uniform distribution on $\{1, 2, \dots, n$, and so

$$\mathbb{E}^*(\dot{Y}_{(\varpi_u+i) \bmod n\, (\varpi_v+j) \bmod n}) = \frac{\sum_{k_1=1}^{n} \sum_{k_2=1}^{n} \dot{Y}_{(k_1+i) \bmod n\, (k_2+j) \bmod n}}{n^2} = \frac{\sum_{k_1=1}^{n} \sum_{k_2=1}^{n} \dot{Y}_{k_1\, k_2}}{n^2}$$

for each pair of blocks $u \neq v$. The independence of this expectation on the pair (i, j) again follows from the circularity of the bootstrap scheme. The between-block contribution, therefore, is

$$\frac{b(b-1)m^2}{n(n-1)} \frac{\sum_{k_1=1}^{n} \sum_{k_2=1}^{n} \dot{Y}_{k_1 k_2}}{n^2} = \frac{n-m}{n-1} \left(\frac{n-1}{n} \bar{Y}_n + \frac{1}{n} \bar{D}_n \right) = \bar{Y}_n + O_P(mn^{-1})$$

as long as $\bar{D}_n = {}^1/_n \sum_{i=1}^{n} D_i = O_P(1)$. In the case that $D_i = 0$ for all $1 \leq i \leq n$ we arrive at

$$\mathbb{E}^*(\bar{Y}_n^*) = \left(1 - \frac{m}{n}\right) \bar{Y}_n + O_P\left(\frac{m-1}{n}\right),$$

which is Equation (3.1). □

Proof of Theorems 3 and 4

As in the derivation of Equation (3.1) we will work under the more general setting where we set

$$\dot{Y}_{ij}^* = \dot{Y}_{\phi_i \phi_j} = \begin{cases} Y_{\phi_i \phi_j} & \text{if } \phi_i \neq \phi_j \\ D_{\phi_i} & \text{if } \phi_i = \phi_j \end{cases}$$

for $D_1, \ldots, D_n$ that are non-random conditional on the data. We will only require that $\mathbb{E}(D_i^2) < \infty$. As in the proof of Theorem 1 we begin by setting up a Hoeffding-type decomposition for the bootstrap data. This can be done by defining

$$\dot{W}_{ij}^* = \dot{W}_{\phi_i \phi_j} = \begin{cases} Y_{\phi_i \phi_j} - h_2(Z_{\phi_i}, Z_{\phi_j}) & \text{if } \phi_i \neq \phi_j \\ D_{\phi_i} \quad - h_2(Z_{\phi_i}, Z_{\phi_i}) & \text{if } \phi_i = \phi_j \end{cases},$$

and $\dot{U}_{ij}^* = \dot{U}_{\phi_i \phi_j} = h_2(Z_{\phi_i, Z_{\phi_j}}) - h_1(Z_{\phi_i}) - h_1(Z_{\phi_j}) + \theta$. Here, the D_i only appear in the construction of the artificial 'diagonal' variables $\dot{W}_{ii}^*$. Let $\bar{W}_n^*$ and $\bar{U}_n^*$ be the corresponding averages over all observations. Then,

$$\bar{Y}_n^* = \theta + \frac{2}{n} \sum_{i=1}^{n} X_i^* + \bar{U}_n^* + \bar{W}_n^*.$$

We next recenter this expression by using the decomposition for $\bar{Y}_n$ from before and multiply through by $\sqrt{n}$ to arrive at

$$\sqrt{n}(\bar{Y}_n^* - \bar{Y}_n) = \frac{2}{\sqrt{n}} \sum_{i=1}^{n} (X_i^* - X_i) + \sqrt{n}(\bar{U}_n^* - \bar{U}_n) + \sqrt{n}(\bar{W}_n^* - \bar{W}_n),$$

which is the starting point of our analysis.

Remainder term. We first set out to show that

$$\sqrt{n}(\bar{U}_n^* - \bar{U}_n) = o_{P^*}(1), \qquad \sqrt{n}(\bar{W}_n^* - \bar{W}_n) = o_{P^*}(1),$$

where, for a random variable A_n^*, $A_n^* = o_{P^*}(1)$ means that $\mathbb{P}(\mathbb{P}^*(|A_n^*| > \epsilon^*) > \epsilon) = o(1)$ for all $\epsilon^* > 0$ and $\epsilon > 0$. It was already established in the proof of Theorem 1 that $\bar{U}_n = o_P(n^{-1/2})$ and that $\bar{W}_n = o_P(n^{-1/2})$. Further, by Lemma 3.7 in Dehling and Wendler (2010) we have that $n\,\mathbb{E}(\mathbb{E}^*(\bar{U}_n^{*2})) = o(1)$. From this $\bar{U}_n^* = o_{P^*}(n^{-1/2})$ follows by repeated application of Markov's inequality. It then only remains to verify that $\bar{W}_n^* = o_{P^*}(n^{-1/2})$. We turn to doing so next.

As in the proof of Equation (3.1), we begin by writing out the sample average in full as

$$\bar{W}_n^* = \frac{1}{n(n-1)} \sum_{u=1}^{b} \sum_{v=1}^{b} \sum_{i=0}^{m-1} \sum_{j=0}^{m-1} \dot{W}_{(\varpi_u+i) \bmod n (\varpi_v+j) \bmod n} \{i \neq j \text{ if } u = v\}.$$

Hence, $\mathbb{E}^*(\bar{W}_n^{*2})$ is the (normalized) sum of

$$\mathbb{E}^*(\dot{W}_{(\varpi_{u_1}+i_1) \bmod n (\varpi_{v_1}+j_1) \bmod n} \dot{W}_{(\varpi_{u_2}+i_2) \bmod n (\varpi_{v_2}+j_2) \bmod n}) \quad \begin{matrix} \{i_1 \neq j_1 \text{ if } u_1 = v_1\} \\ \{i_2 \neq j_2 \text{ if } u_2 = v_2\} \end{matrix}$$

over all b^4 combinations of blocks u_1, u_2, v_1, v_2 and all m^4 combinations of indices i_1, j_1, i_2, j_2. The form of this bootstrap expectation depends on whether some of the indices coincide and, if so, which ones.

First consider the situation where all four blocks u_1, v_1, u_2, v_2 are distinct. Then the $\varpi_{u_1}, \varpi_{v_1}, \varpi_{u_2}, \varpi_{v_2}$ are all independent uniform random variables on $\{1, 2, \ldots, n\}$ and so (again exploiting cyclicality of the bootstrap scheme) the bootstrap expectation is equal to

$$\left(\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \dot{W}_{ij}\right)^2 = \left(\frac{n-1}{n}\bar{W}_n + \frac{1}{n}\bar{H}_n\right)^2 = \frac{(n-1)^2}{n^2}\bar{W}_n^2 + 2\frac{(n-1)}{n^2}\bar{W}_n\,\bar{D}_n + \frac{1}{n^2}\bar{D}_n^2,$$

where (relative to the proof of Equation (3.1)) we redefine $\bar{D}_n = {}^1\!/\!{}_n \sum_{i=1}^n D_i - h_2(Z_i, Z_i)$. This does not depend on any of the indices under consideration. Furthermore, from the proof of Theorem 1 we know that $\mathbb{E}(\bar{W}_n^2) = O(n^{-2})$. Provided that $\mathbb{E}(D_i^2) < \infty$ we also have that $\mathbb{E}(\bar{D}_n^2) = O(1)$ because $\mathbb{E}(h_2(Z_i, Z_i)^2) < \infty$ by Assumption 1. We thus have that

$$\mathbb{E}\left(\left(\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\dot{W}_{ij}\right)^2\right) = O(n^{-2}).$$

There are $b!/(b-4)!$ combinations of four distinct blocks and m^4 different combinations of indices i_1, j_1, i_2, j_2. The total contribution of distinct-block terms to $\mathbb{E}(\mathbb{E}^*(\bar{W}_n^{*2}))$ is then at most of order n^{-2}.

Next consider the opposite extreme where all observations are from the same block, i.e., $u_1 = v_1 = u_2 = v_2$. For such terms to contribute we need that $i_1 \neq j_1$ and $i_2 \neq j_2$. Now the bootstrap expectation of a representative term of this configuration has as expectation

$$\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}(W_{(k+i_1) \bmod n\,(k+j_1) \bmod n}\,W_{(k+i_2) \bmod n\,(k+j_2) \bmod n})$$

which is non-zero only when $i_1 = i_2$ and $j_1 = j_2$, in which case it equals $\mathbb{E}(W_{i_1 j_1}^2) < \infty$. Therefore, for this second case, a contribution to $\mathbb{E}(\mathbb{E}^*(\bar{W}_n^{*2}))$ comes only from $bm(m-1)$ terms. Normalized by $n^2(n-1)^2$ this then gives an order of magnitude of $mn^{-3} = o(n^{-2})$ given that $mn^{-1} \to 0$.

Then there are a variety of different configurations where some but not all observations belong to the same block. These configurations feature either (i) two pair of observations with pairs in different blocks but observations within a pair in the same block, or (ii) a triplet of observations in the same block and the fourth observation in a different block. The different cases for each configuration have the same structure and so can be handled in the same way.

For the first of these two configurations, consider the case where $u_1 = v_1$ and $u_2 = v_2$, but $u_1 \neq v_2$. There are $b(b-1)$ such group pairings. A representative configuration of i_1, j_1, i_2, j_2 has $i_1 \neq j_1$ and $i_2 \neq j_2$ and, therefore, the bootstrap expectation here is equal to

$$\frac{1}{n^2}\sum_{k_1=1}^{n}\sum_{k_2=1}^{n}W_{(k_1+i_1) \bmod n\,(k_1+j_1) \bmod n}W_{(k_2+i_2) \bmod n\,(k_2+j_2) \bmod n}.$$

The summand here has zero expectation unless (i) $k_1+i_1=k_2+i_2$ and $k_1+j_1=k_2+j_2$ or (by symmetry) (ii) $k_1+i_1=k_2+j_2$ and $k_1+j_1=k_2+i_2$. Take (i), which is equivalent to having $k_1=k_2+d$ for each admissible value $d=i_2-i_1=j_2-j_1$. The bootstrap expectation equals ${}^1\!/\!{}_{n^2}\sum_{k=1}^n W_{k,k+d}^2$, whose expectation is $O(n^{-1})$ for any d. Ranging over all d yields $O(m^4)$ index combinations. Hence, the contribution of this case to $\mathbb{E}(\mathbb{E}^*(\bar{W}_n^{*2}))$ is of the order $b^2m^4n^{-5}=m^2n^{-3}=o(n^{-1})$. By symmetry, the same conclusion holds for (ii).

For the second of the configurations we can take $u_1=v_1=u_2=u$ and $v_2=v$ for some $u\neq v$. There are again $b(b-1)$ such permissible group pairs. In this case the indicator function ensures that we will only face terms with $i_1\neq j_1$. The bootstrap expectation for a given collection of permissible indices i_1,j_1,i_2,j_2 for this type of configuration then equals

$$\frac{1}{n^2}\sum_{k_1=1}^n\sum_{k_2=1}^n W_{(k_1+i_1)\,\mathrm{mod}\,n\,(k_1+j_1)\,\mathrm{mod}\,n}\,\dot{W}_{(k_1+i_2)\,\mathrm{mod}\,n\,(k_2+j_2)\,\mathrm{mod}\,n}.$$

We distinguish two cases. The first case has $(k_1+i_2)\,\mathrm{mod}\,n\neq(k_2+j_2)\,\mathrm{mod}\,n$. Then the average does not feature any artificial diagonal entries and the expectation of the summand is zero unless (i) $i_1=i_2=i$ and $(k_1+j_1)\,\mathrm{mod}\,n=(k_2+j_2)\,\mathrm{mod}\,n$ or (ii) $(k_1+i_1)\,\mathrm{mod}\,n=(k_2+j_2)\,\mathrm{mod}\,n$ and $i_2=j_2$. For Case (i), the expectation of the above sum becomes

$$\frac{1}{n^2}\sum_{k_1=1}^n\mathbb{E}(W^2_{(k_1+i)\,\mathrm{mod}\,n\,(k_1+j_1)\,\mathrm{mod}\,n})=O(n^{-1}).$$

Because there are at most m^3 such terms for each of the $b(b-1)$ block pairs the total contribution of such configuration is of the order $b^2m^3n^{-5}=(mn^{-1})n^{-2}=o(n^{-2})$. For Case (ii), in turn, we arrive at

$$\frac{1}{n^2}\sum_{k_1=1}^n\mathbb{E}(W_{(k_1+i_1)\,\mathrm{mod}\,n\,(k_1+j_1)\,\mathrm{mod}\,n}\,(D_{(k_1+i_2)\,\mathrm{mod}\,n}-h_2(Z_{(k_1+i_2)\,\mathrm{mod}\,n},Z_{(k_1+i_2)\,\mathrm{mod}\,n})))$$

which, by Cauchy-Schwarz, is easily seen to be $O(n^{-1})$. Here, for each of the $b(b-1)$ block pairings there are m^3 such terms yielding a total contribution to $\mathbb{E}(\mathbb{E}^*(\bar{W}_n^{*2}))$ that is again $o(n^{-2})$.

Having shown that $n\,\mathbb{E}(\mathbb{E}^*(\bar{W}_n^{*2})) = o(1)$ it follows that $\bar{W}_n^* = o_{P^*}(n^{-1/2})$ by Markov's inequality. This completes our analysis of the remainder term in the bootstrap version of Hoeffding's decomposition.

Bootstrap consistency. From the previous step,

$$\sqrt{n}(\bar{Y}_n^* - \bar{Y}_n) = 2\sqrt{n}(\bar{X}_n^* - \bar{X}_n) + o_{P^*}(1),$$

with $\bar{X}_n^* = {}^1\!/\!{}_n \sum_{i=1}^n X_i^*$, and it suffices to show that

(i) $\mathrm{var}^*(\sqrt{n}\bar{X}_n^*) = n\mathbb{E}^*((\bar{X}_n^* - \bar{X}_n)^2) \xrightarrow{P} \sigma^2$,

(ii) $\sup_a |\mathbb{P}^*(\sqrt{n}(\bar{X}_n^* - \bar{X}_n) \leq a) - \mathbb{P}(\sqrt{n}\bar{X}_n \leq a)| \xrightarrow{P} 0$ as $n \to \infty$.

As $\{X_i\}$ is a stationary mixing process both (i) and (ii) follow readily from the bootstrap literature on time series. Here, we use Theorem 1 in Politis and Romano (1992). This completes the proof of both Theorem 3 and Theorem 4. $\square$

References

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

Beran, R. (1988). Prepivoting test statistics: A bootstrap view of asymptotic refinements. *Journal of the American Statistical Association 83*, 687–697.

Davezies, L., X. D'Haultfœuille, and Y. Guyonvarch (2021). Empirical process results for exchangeable arrays. *Annals of Statistics 49*, 845–862.

Dehling, H. and M. Wendler (2010). Central limit theorem and the bootstrap for U-statistics on strongly mixing data. *Journal of Multivariate Analysis 101*, 126–137.

Fafchamps, M. and F. Gubert (2007). The formation of risk sharing networks. *Journal of Development Economics 83*, 326–350.

Green, A. and C. R. Shalizi (2022). Bootstrapping exchangeable random graphs. *Electronic Journal of Statistics 16*, 1058–1095.

Hall, P., J. L. Horowitz, and B.-Y. Jing (1995). On blocking rules for the bootstrap with dependent data. *Biometrika 82*, 561–574.

Hoover, D. N. (1979). Relations on probability spaces and arrays of random variables. Technical report, Institute of Advanced Study, Princeton.

Kallenberg, O. (1989). On the representation theorem for exchangeable arrays. *Journal of Multivariate Analysis 30*, 137–154.

Lahiri, S. N. (1991). Second order optimality of stationary bootstrap. *Statistic & Probability Letters 11*, 335–341.

Leucht, A. (2012). Degenerate U- and V-statistics under weak dependence: Asymptotic theory and bootstrap consistency. *Bernoulli 18*, 552–585.

Menzel, K. (2021). Bootstrap with clustering in two or more dimensions. *Econometrica 89*, 2143–2188.

Moulton, B. R. (1986). Random group effects and the precision of regression estimates. *Journal of Econometrics 32*, 385–397.

Moulton, B. R. (1990). An illustration of a pitfall in estimating the effects of aggregate variables on micro units. *Review of Economics and Statistics 72*, 334–338.

Newey, W. K. and K. D. West (1987). A simple positive semi-definite, heteroscedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*, 703–708.

Newey, W. K. and K. D. West (1994). Automatic lag selection in covariance matrix estimation. *Review of Economic Studies 61*, 631–653.

Owen, A. B. (2007). The pigeonhole bootstrap. *Annals of Applied Statistics 1*, 386–411.

Owen, A. B. and D. Eckles (2012). Bootstrapping data arrays of arbitrary order. *Annals of Applied Statistics 6*, 895–927.

Politis, D. N. and J. P. Romano (1992). A circular block resampling procedure for stationary data. In R. LePage and L. Billard (Eds.), *Exploring the Limits of the Bootstrap*, pp. 263–270.

Politis, D. N. and H. White (2004). Automatic block-length selection for the dependent bootstrap. *Econometric Reviews 23*, 53–70.

Silverman, B. W. (1976). Limit theorems for dissociated random variables. *Advances in Applied Probability 8*, 806–819.

Thompson, S. B. (2011). Simple formulas for standard errors that cluster by both firm and time. *Journal of Financial Economics 99*, 1–10.

van der Vaart, A. W. (2000). *Asymptotic Statistics.* Cambridge University Press.

White, H. (2001). *Asymptotic Theory for Econometricians.* Academic Press.