Limit Theorems for Network Dependent Random Variables

Denis Kojevnikov, Vadim Marmer, and Kyungchul Song
Vancouver School of Economics, University of British Columbia

Abstract. This paper considers a general form of network dependence where dependence between two sets of random variables becomes weaker as their distance in a network gets longer. We show that such network dependence cannot be embedded as a random field on a lattice in a Euclidean space with a fixed dimension when the maximum clique increases in size as the network grows. This paper applies Doukhan and Louhichi (1999)'s weak dependence notion to network dependence by measuring dependence strength by the covariance between nonlinearly transformed random variables. While this approach covers examples such as strong mixing random fields on a graph and conditional dependency graphs, it is most useful when dependence arises through a large functional-causal system of equations. The main results of our paper include the law of large numbers, and the central limit theorem. We also propose a heteroskedasticity-autocorrelation consistent variance estimator and prove its consistency under regularity conditions. The finite sample performance of this latter estimator is investigated through a Monte Carlo simulation study.

Key words. Network Dependence; Random Fields; Central Limit Theorem; Networks; Law of Large Numbers; Cross-Sectional Dependence; Spatial Processes
JEL Classification: C12, C21, C31

1. Introduction

Network models have been used to capture a complex form of interdependence among cross-sectional observations. These observations may represent actions by people or

Date: March 5, 2019.
Song acknowledges the financial support of the Social Sciences and Humanities Research Council of Canada.

firms, or outcomes from industry sectors, assets or products. Random fields indexed by points in a lattice in a Euclidean space have often been adopted as a model of spatial dependence in econometrics and statistics. Conley (1999) proposed using random field modeling to specify the cross-sectional dependence of observations in the context of GMM estimation. More recent contributions include Jenish and Prucha (2009) and Jenish and Prucha (2012). See Jia (2008) for an application in entry decisions in retail markets, and Boucher and Mourifie (2017) for an inference problem for a network formation model. For limit theorems for such random fields in statistics, see Comets and Janžura (1998) and references in the paper.

When the dependence ordering arises from geographic distances or their analogues, using such random fields appears natural. However, the dependence ordering often stems from pairwise relations among the sample units, which can be viewed as a form of a network. To apply the random field modeling, one would first need to transform these relations into a random field on a lattice in a Euclidean space using methods such as multidimensional scaling (see, e.g., Borg and Groenen (2005); see also footnote 16 of Conley (1999) on page 15).

However, such embedding of a network into a lattice can distort the dependence ordering. In fact, this paper shows that network dependence is not necessarily embedded as a random field indexed by a lattice in the Euclidean space with a fixed dimension, when the network has a maximum clique whose size increases as the network grows. Networks with a growing maximum clique size often arise from those with a power-law degree distribution and high clustering coefficients. These features are typically shared by social networks that are observed in practice. In this paper, we directly use a network as a model of dependence ordering, so that such an embedding is not required when dependence ordering comes from pairwise relations.

Associating a dependence pattern with a network has existed in the literature. Stein (1972) introduced a notion of dependency graphs in studying the normal approximation of a sum of random variables which are allowed to be dependent only when they are adjacent in a given network. See also Janson (1988), Baldi and Rinott (1989), Chen and Shao (2004), and Rinott and Rotar (1996) for various results for normal approximation for variables with related local dependence structures. See Leung (2016) and Song (2017) for recent applications of dependency graphs to network data. Modeling based on dependency graphs has a drawback; it requires independence between variables that are not adjacent in the network, and hence is not adequate to model dependence which becomes gradually weaker as two nodes get farther away from each other in the network.

A closely related strand of literature has studied various models of Markov random fields and spatial autoregressive models. Markov random fields constitute an alternative class of models of dependence which imposes conditional independence restrictions based on the network structure (see, e.g., Lauritzen (1996) and Pearl (2009); see also Chapter 19 of Murphy (2012) for applications in the literature of machine learning). Recently Lee and Song (2018) established a central limit theorem using a more general local dependence notion that encompasses both dependency graphs and a class of Markov random fields. Spatial autoregressive models specify cross-sectional dependence through the weight matrix in linear simultaneous equations, and have been extensively studied in econometrics. See, among others, Lee (2004) and Lee et al. (2010) and references therein. Also see Gaetan and Guyon (2010) for an extensive review of spatial modeling and limit theorems.

In contrast to the dependency graph modeling, our network dependence permits dependence between random variables that are only indirectly linked through intermediary variables. In fact random variables having a graph as a dependency graph can be viewed as a special case of our network dependence modeling. The approach in this paper is also distinct from the approach of Markov random fields modeling. Markov random fields are based on conditional independence restrictions among the variables. While the limit theorems on Markov random fields rely on independence restrictions that come from conditioning on certain random variables, our modeling expresses the degree of stochastic dependence in terms of the distance in the network.

A major distinction of our paper is that in modeling dependence and associating it with a given network, we adopt the approach of ψ-dependence proposed by Doukhan and Louhichi (1999) and extend the notion to accommodate common shocks. The notion of ψ-dependence is simple and intuitive. Roughly speaking ψ-dependence measures the strength of dependence between two sets of random variables in terms of the covariance between nonlinear functions of random variables.

A primary benefit of modeling dependence through ψ-dependence comes when dependence among the variables is produced through a system of causal equations in which sharing of exogenous shocks creates cross-sectional dependence among the variables of interest. In this paper, we give four broad classes of such examples, including those where the random variables are generated from primitive random variables through a nonlinear transform. These classes cover many sub-examples that are used in statistics and econometrics. In such examples, a traditional approach of modeling through various mixing properties is cumbersome, because it is hard to find primitive conditions that guarantee the mixing properties for the variables of interest. Nevertheless, one can write the covariance bounds of those variables in terms of the primitive exogenous shocks using the causal equations.

This flexibility of the ψ-dependence notion, however, carries a cost. The ψ-dependence of a nonlinearly transformed ψ-dependent random variables is not necessarily ensured, if the nonlinear transform does not belong to the class in the original definition. This paper provides various auxiliary results which could be helpful in such situations. (They are found in the appendix of the paper.)

The main results of this paper are three-fold: the Law of the Large Numbers, the Central Limit Theorem, and the consistency of HAC estimators. The results show the trade-off between the extensiveness of the cross-sectional dependence and the quality of the asymptotic approximation in the statistics. The extensiveness is characterized by the density properties of the underlying graph. We also investigate the performance of our HAC estimators through various simulation designs. As expected, the performance becomes better if the cross-sectional dependence is less extensive.

The paper is organized as follows. In the next section, we define network dependence of stochastic processes and provide examples. In Section 3, we present the main results of the law of the large numbers and the central limit theorem. Section 4 focuses on HAC estimation and investigates its finite sample performance using Monte Carlo simulations. In Section 5, we conclude. Some auxiliary results and the proofs of the results are provided in the appendix.

2. Network Dependence and Examples

2.1. Network Topology and a Lattice in a Euclidean Space

Let $N_n = \{1, 2, \ldots, n\}$ be the set of cross-sectional unit indices. Modeling cross-sectional weak dependence or spatial dependence usually assumes a certain metric on this set N_n. For some examples, this distance can be motivated by geographic distances or economic distances measured in terms of economic outcomes. This paper focuses on the pattern of cross-sectional dependence that is shaped along a given network. More specifically, suppose that we observe an *undirected* network G_n on $N_n = \{1, 2, ..., n\}$, where $G_n = (N_n, E_n)$, and $E_n \subset \{ij : i, j \in N_n, i \neq j\}$ denotes the set of links. For $i, j \in N_n$, we define $d_n(i, j)$ to be the distance between i and j in G_n, i.e., the length of the shortest path between nodes i and j given G_n. The distance d_n defines a metric on the set N_n. We refer to *network dependence* as a stochastic dependence pattern of random variables governed by the distance d_n on G_n.

We introduce two major network properties that we use to characterize the conditions for the network later. Let $N_n(i;s)$ denote the set of the nodes that are within the distance s from node i, and $N_n^{\partial}(i;s)$ denote the set of the nodes that are exactly at the distance s from node i. That is,

$$N_n(i;s) = \{j \in N_n : d_n(i,j) \leq s\} \quad \text{and} \quad N_n^{\partial}(i;s) = \{j \in N_n : d_n(i,j) = s\}.$$

Define

$$\delta_n(s;k) = \frac{1}{n}\sum_{i \in N_n} |N_n(i;s)|^k, \qquad \delta_n^{\partial}(s;k) = \frac{1}{n}\sum_{i \in N_n} |N_n^{\partial}(i;s)|^k,$$
$$D_n(s) = \max_{i \in N_n} |N_n(i;s)|, \quad \text{and} \quad D_n^{\partial}(s) = \max_{i \in N_n} |N_n^{\partial}(i;s)|.$$

When $k = 1$, we simply write $\delta_n(s;1) = \delta_n(s)$ and $\delta_n^{\partial}(s;1) = \delta_n^{\partial}(s)$. These quantities measure the denseness of a network. For example, the so-called small world phenomenon in social network data is reflected by rapidly growing $\delta_n(s;k)$ as we increase s.

Our first focus is on the relation between modeling dependence through network topology and that through random fields indexed by the elements of a finite subset of a metric space $(\mathcal{X}, d_{\mathcal{X}})$. We denote the equilateral dimension of $\mathcal{X}$, i.e. the maximum number of equidistant points in $\mathcal{X}$ with respect to the distance $d_{\mathcal{X}}$, as $e(\mathcal{X})$. The main question here is whether any given connected network[1] is embeddable in $\mathcal{X}$. The following definition makes precise the notion of embedding.

Definition 2.1. An isometric embedding of a network $G_n = (N_n, E_n)$ into a metric space $(\mathcal{X}, d_{\mathcal{X}})$ is an injective map $b : N_n \to \mathcal{X}$ such that for all $i, j \in N_n$

$$d_{\mathcal{X}}(b(i), b(j)) = d_n(i,j). \tag{2.1}$$

When such an isometry exists, it means that modeling cross-sectional dependence using a network topology can be viewed as a special case of modeling a random field on a finite subset of $\mathcal{X}$. The following result shows that this is not always possible when the clique number $\omega(G_n)$ of G_n, i.e. the number of nodes in a maximum clique[2] in G_n, is large enough.

Proposition 2.1. *A connected network G_n is isometrically embeddable into a metric space $(\mathcal{X}, d_{\mathcal{X}})$ only if $\omega(G_n) \leq e(\mathcal{X})$.*

Proof. Suppose that C is a maximum clique of G_n. It is obvious that there is no isometry between C and $\mathcal{X}$ when $|C| > e(\mathcal{X})$. ∎

[1] A network/graph is connected if there is a path between every pair of nodes.

[2] A clique of a graph G is a subset of nodes such that every two distinct nodes are adjacent.

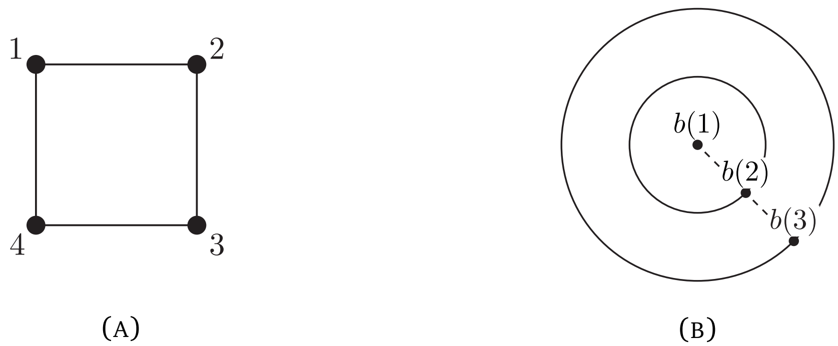

FIGURE 1. An example of a network with the maximum clique size of two in panel (A) that cannot be embedded into $\mathbf{R}^2$ equipped with the Euclidean distance (with the equilateral dimension of three), as shown in panel (B). Node 2 has distance one from nodes 1 and 3, and node 3 has distance two from node 1. Their maps $b(1)$, $b(2)$, and $b(3)$ must be on the same line. If one maps node 4 to preserve its distance of one from nodes 1 and 3, $b(4)$ would have zero distance from $b(2)$.

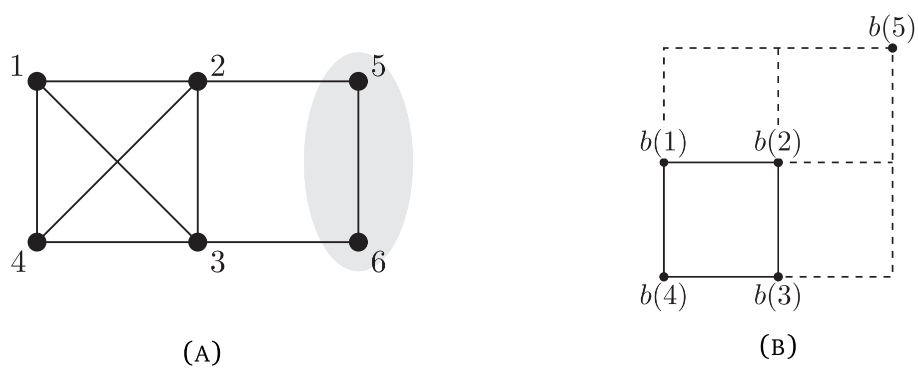

FIGURE 2. An example of a network with the maximum clique size of four in panel (A) that cannot be embedded into $\mathbf{R}^2$ equipped with the L_∞ distance (with the equilateral dimension of 4), as shown in panel (B). Node 5 has distance one from node 2, and distance two from nodes 1, 3, and 4, which uniquely determine its map $b(5)$. Similarly, the distances between node 6 and nodes 1, 2, 3, and 4 uniquely determines $b(6)$, however, it would be inconsistent with distance one between nodes 5 and 6.

Proposition 2.1 gives only a necessary condition for isometric embedding. Consider, for example, $\mathbf{R}^k$ equipped with the Euclidean distance, which has the equilateral dimension of $k+1$. Figure 1 provides an example of a network with the maximum clique size of two that cannot be embedded into the Euclidean $\mathbf{R}^2$ space, which has the equilateral distance of three. Figure 2 provides an example with a non-Euclidean space. It shows a network with the maximal clique size of four that cannot be imbedded into $\mathbf{R}^2$ equipped with the L_∞ distance, which has the equilateral dimension of four.

An important consequence of Proposition 2.1 is that when the size of the maximum cliques in the network G_n grows to infinity as $n \to \infty$, the sequence of networks cannot be embedded into a metric space having finite equilateral dimension. Examples of such spaces include a k-dimensional normed space M^k and a sphere $\mathbb{S}^k$ equipped with the usual distance because $e(M^k) \leq 2^k$ (see Petty, 1971, Theorem 4) and $e(\mathbb{S}^k) = k + 2$. As a consequence, the random field models used in Conley (1999) with the Euclidean distance and in Jenish and Prucha (2009) with the Chebychev distance cannot include a network dependence model when the maximum clique size of the networks increases with the sample size. Indeed, there are random graphs whose degree distribution takes the form of a power law and the size of the maximum cliques grows to infinity as $n \to \infty$ (see Bläsius et al., 2017). Such models accommodate both dense and sparse graphs, and are often motivated as a model of many real networks that we observe in practice.

The asymptotic results developed in this paper can accommodate network generating processes with the maximum clique size increasing with the sample size. However, our results impose certain restrictions on the rate of growth of the maximum clique size.

One may consider "approximating" the network dependence ordering by a lattice in a finite dimensional Euclidean space. Methods called multidimensional scaling (MDS) provides various ways to achieve such an approximation, see Borg and Groenen (2005). The dependence ordering obtained through MDS is itself dependent on the data, and is stochastic. Hence, it is generally different from the true dependence ordering of the data. Proposition 2.1 tells us that there is no guarantee that the approximation error of the MDS-based dependence ordering will be small with a large sample size.

2.2. Network Dependent Processes

Let us introduce the notion of network dependence that is the focus in this paper. Suppose that we are given a triangular array of $\mathbf{R}^v$-valued random vectors, $Y_{n,i}, i \in N_n$, which are laid on a network G_n. We adapt the weak dependence notion of Doukhan and Louhichi (1999) to our set up. We define $\mathbb{N} = \{1, 2, 3, \ldots\}$, and for any $v, a \in \mathbb{N}$,

$$\mathcal{L}_v = \{\mathcal{L}_{v,a} : a \in \mathbb{N}\},$$

where $\mathcal{L}_{v,a}$ denotes the collection of bounded Lipschitz real functions on $\mathbf{R}^{v \times a}$, i.e.,

$$\mathcal{L}_{v,a} = \left\{ f : \mathbf{R}^{v \times a} \to \mathbf{R} : \|f\|_\infty < \infty, \mathrm{Lip}(f) < \infty \right\},$$

with $\mathrm{Lip}(f)$ denoting the Lipschitz constant of f,[3] and $\|\cdot\|_\infty$ the sup-norm of f, i.e., $\|f\|_\infty = \sup_x |f(x)|$. For any positive integers a, b, s, consider two sets of nodes (of size a and b) with distance between each other of at least s. Let $\mathcal{P}_n(a, b; s)$ denote the collection

[3]The Lipschitz constant for a real function f on a metric space $(\mathcal{X}, d)$ is the smallest constant C such that $|f(x) - f(y)| \leq Cd(x, y)$, for all $x, y \in \mathcal{X}$.

of all such pairs:

$$\mathcal{P}_n(a,b;s) = \{(A,B) : A, B \subset N_n, |A| = a, |B| = b, \text{ and } d_n(A,B) \geq s\},$$

where

$$d_n(A,B) = \min_{i \in A} \min_{i' \in B} d_n(i,i'),$$

and $d_n(i,i')$ denotes the distance between nodes i and i' in G_n, i.e., the length of the shortest path between i and i' in G_n. For each set A of positive integers, and a given triangular array of random vectors $(Y_{n,i})_{i \in A}$, we write

$$Y_{n,A} \equiv (Y_{n,i})_{i \in A}.$$

We take $\{\mathcal{C}_n\}_{n \geq 1}$ to be a given sequence of σ-fields.

Definition 2.2. The triangular array $\{Y_{n,i}\}_{i \in N_n}, n \geq 1$, $Y_{n,i} \in \mathbf{R}^v$, is called *conditionally ψ-weakly dependent* given $\{\mathcal{C}_n\}_{n \geq 1}$, if for each $n \in \mathbb{N}$, there exists a $\mathcal{C}_n$-measurable sequence $\theta_n \equiv \{\theta_{n,s}\}_{s=1}^{\infty}$ such that $\sup_{n \geq 1} \theta_{n,s} \to_{a.s.} 0$ as $s \to \infty$, and a collection of nonrandom functions $(\psi_{a,b})_{a,b \in \mathbb{N}}$, $\psi_{a,b} : \mathcal{L}_{v,a} \times \mathcal{L}_{v,b} \to [0,\infty)$, such that for all $(A,B) \in \mathcal{P}_n(a,b;s)$ with $s > 0$ and all $f \in \mathcal{L}_{v,a}$ and $g \in \mathcal{L}_{v,b}$,

$$|\text{Cov}\,(f(Y_{n,A}), g(Y_{n,B}) \mid \mathcal{C}_n)| \leq \psi_{a,b}(f,g)\theta_{n,s} \quad \text{a.s.} \tag{2.2}$$

In this case, we call the sequence θ_n *the weak dependence coefficients* of $\{Y_{n,i}\}_{i \in N_n}$.

The σ-field $\mathcal{C}_n$ can be thought of as a "common shock" such that when we condition on it, the cross-sectional dependence of triangular array $\{Y_{n,i}\}$ becomes substantially weaker. However, we do not have to think of $\mathcal{C}_n$ as being originated from a variable that affects every node in the network. In many network set-ups, $\mathcal{C}_n$ can be thought of as having been generated by some characteristics or actions of multiple central nodes which affect many other nodes through their many links. For example, consider a star network, where node 1 is adjacent to the other $n-1$ nodes. Suppose that $Y_{n,1} = U_1$ corresponds to the central node, and for the remaining nodes ($i \geq 2$) $Y_{n,i} = U_1 + U_i$, where $\{U_i : i = 1, \ldots, n\}$ are independent. In that case, we can take $\mathcal{C}_n = \sigma(U_1)$. Then, conditionally on $\mathcal{C}_n$, $Y_{n,2}, \ldots, Y_{n,n}$ are i.i.d., and $P\{\theta_{n,2} = 0 \mid \mathcal{C}_n\} = 1$.

Unlike the unconditional version of ψ-weak dependence of Doukhan and Louhichi (1999), in our definition the weak dependence coefficients $\{\theta_n\}$ are random, due to our accommodation of the common shocks, $\mathcal{C}_n$. We make the following assumption.

Assumption 2.1.

(a) The triangular array $\{Y_{n,i}\}_{i\in N_n}$ is conditionally ψ-weakly dependent given $\{\mathcal{C}_n\}_{n\geq 1}$ with

$$\psi_{a,b}(f,g) = c_1 \|f\|_\infty \|g\|_\infty + c_2 \operatorname{Lip}(f) \|g\|_\infty + c_3 \|f\|_\infty \operatorname{Lip}(g) + c_4 \operatorname{Lip}(f) \operatorname{Lip}(g),$$

where $c_1, \ldots, c_4 \leq Cab$ for some constant $C > 0$.

(b) There exists a constant $M \in [1, \infty)$ such that $\theta_{n,s} \leq M$ a.s. for all $n, s \geq 1$.

The above assumption will be maintained throughout the paper. Assumption 2.1 is shown to be satisfied by all the examples we present in the next subsection. The restriction on the weak dependence coefficients in (b) can be relaxed at the cost of more complex notation. The restriction is convenient because the ψ-dependence of random vectors carries over to linear combinations of the elements in the random vectors as shown in the following lemma.

Lemma 2.1. *Suppose that a triangular array of random vectors $\{Y_{n,i}\}_{i\in N_n}$, $Y_{n,i} \in \mathbf{R}^v$, satisfies Assumption 2.1(a) with the weak dependence coefficients $\{\theta_n\}_{n\geq 1}$. Let $c_n \in \mathbf{R}^v$ be a $\mathcal{C}_n$-measurable sequence such that $\|c_n\| \leq 1$, a.s. Then the array $\{Z_{n,i}\}_{i\in N_n}$ defined by $Z_{n,i} = c_n^\top Y_{n,i}$ is conditionally ψ-weakly dependent given $\{\mathcal{C}_n\}_{n\geq 1}$ with the weak dependence coefficients $\{\theta_n\}_{n\geq 1}$.*

A result similar to Lemma 2.1 holds for nonlinear transforms of random variables, under certain conditions for the nonlinear transforms. See Appendix for details.

2.3. Examples

In this section, we consider four broad classes of examples of conditionally ψ-weakly dependent random vectors.

2.3.1. **Strong-Mixing Processes.** Let $(\Omega, \mathcal{F}, P)$ be an underlying probability space. For sub σ-fields, $\mathcal{G}$, $\mathcal{H}$, $\mathcal{C}$, of $\mathcal{F}$, let

$$\alpha(\mathcal{G}, \mathcal{H} \mid \mathcal{C}) = \sup_{G\in\mathcal{G}, H\in\mathcal{H}} |\operatorname{Cov}(\mathbf{1}_G, \mathbf{1}_H \mid \mathcal{C})| .$$

For a triangular array $\{Y_{n,i}\}$ and a sequence of σ-fields $\{\mathcal{C}_n\}$ we define the strong mixing coefficients by[4]

$$\alpha_{n,s} = \sup\left\{\alpha\left(\sigma(Y_{n,A}), \sigma(Y_{n,B}) \mid \mathcal{C}_n\right) : A, B \subset N_n, d_n(A,B) \geq s\right\}.$$

[4]These coefficients are different from those given in Jenish and Prucha (2009) because our weak θ_n coefficients do not depend on $|A|$ and $|B|$.

The proposition below provides a conditional covariance inequality that is due to Theorem 9 of Prakasa Rao (2013).

Proposition 2.2. *For $f \in \mathcal{L}_{v,a}$, $g \in \mathcal{L}_{v,b}$, and $(A, B) \in \mathcal{P}_n(a, b; s)$,*

$$|\mathrm{Cov}(f(Y_{n,A}), g(Y_{n,B}) \mid \mathcal{C}_n)| \leq 4 \|f\|_\infty \|g\|_\infty \alpha_{n,s} \quad \textit{a.s.}$$

Hence, if $\sup_{n \geq 1} \alpha_{n,s} \to_{a.s.} 0$ as $s \to \infty$, then the array $\{Y_{n,i}\}$ is conditionally ψ-weakly dependent given $\{\mathcal{C}_n\}$ with $\psi_{a,b}(f, g) = 4 \|f\|_\infty \|g\|_\infty$, and the weak dependence coefficients $\{\theta_n\}_{n \geq 1}$ are given by the strong mixing coefficients $\alpha_n \equiv \{\alpha_{n,s}\}_{s=1}^\infty$.

The proof of Proposition 2.2 follows by adapting the proof of Theorem A.5. of Hall and Heyde (1980) to the conditional settings and noticing that the strong mixing coefficients can be equivalently defined by replacing $\alpha(\mathcal{G}, \mathcal{H} \mid \mathcal{C})$ with $\alpha(\mathcal{G} \vee \mathcal{C}, \mathcal{H} \vee \mathcal{C} \mid \mathcal{C})$.

2.3.2. **Conditional Dependency Graphs.** Suppose that $\{Y_{n,i}\}_{i \in N_n}$ is a given array of random vectors and $G_n = (N_n, E_n)$ is a graph on the index set N_n. Let $\mathcal{C}_n$ be a given σ-field. We say that $\{Y_{n,i}\}_{i \in N_n}$ has G_n as a *conditional dependency graph given* $\mathcal{C}_n$, if for any set $A \subset N_n$, $Y_{n,A}$ and $\{Y_{n,i} : i \in N_n \setminus \overline{N}_n(A)\}$ are conditionally independent given $\mathcal{C}_n$, where

$$\overline{N}_n(A) = \bigcup_{i \in A} N_n(i; 1).$$

The notion of a conditional dependency graph is a conditional variant of a dependency graph introduced by Stein (1972). Now, it is not hard to see that when $\{Y_{n,i}\}_{i \in N_n}$ has G_n as a conditional dependency graph given $\mathcal{C}_n$, $\{Y_{n,i}\}_{i \in N_n}$ is conditionally ψ-weakly dependent given $\mathcal{C}_n$, with

$$\psi_{a,b}(f, g) = 4\|f\|_\infty \|g\|_\infty,$$

and $\theta_{n,s}$ is such that $\theta_{n,s} = 0$ for all $s \geq 1$.

2.3.3. **Functional Dependence on Independent Variables.** Consider a triangular array of $\mathbf{R}^k$-valued random vectors $\{\varepsilon_{n,i}\}_{i \in N_n}$ which is row-wise independent given $\mathcal{C}_n$. For $\mathbf{R}^v$-valued measurable functions $\{\boldsymbol{\phi}_{n,i}\}_{i \in N_n}$, let

$$Y_{n,i} = \boldsymbol{\phi}_{n,i}(\varepsilon_n), \quad i \in N_n,$$

where $\varepsilon_n = (\varepsilon_{n,j} : j \in N_n)$. Further, define a modified version of $Y_{n,i}$, which replaces too distant shocks $\varepsilon_{n,j}$ with zeros:

$$Y_{n,i}^{(s)} = \boldsymbol{\phi}_{n,i}\big(\varepsilon_n^{(s)}\big),$$

where $\varepsilon_n^{(s)} = (\varepsilon_{n,j} \mathbf{1}\{j \in N_n(i; s)\} : j \in N_n)$,[5] so that for any $A, B \subset N_n$ with $d_n(A, B) > 2s$, $Y_{n,A}^{(s)}$ and $Y_{n,B}^{(s)}$ are conditionally independent given $\mathcal{C}_n$.

[5] Zero can be replaced with another constant if the functions $\boldsymbol{\phi}_{n,i}$ are undefined at zero.

Let $\|\cdot\|$ be a norm on $\mathbf{R}^v$ and let $\boldsymbol{d}_u$ denote the distance on $\mathbf{R}^{v\times u}$ given by

$$\boldsymbol{d}_u(\mathbf{x},\mathbf{y}) = \sum_{l=1}^{u} \|x_l - y_l\|,$$

where $\mathbf{x} \equiv (x_1,\ldots,x_u)$ and $\mathbf{y} \equiv (y_1,\ldots,y_u)$ are points in $\mathbf{R}^{v\times u}$.

Proposition 2.3. *Let $\{Y_{n,i}\}$ be as described above. Then for any $(A,B) \in \mathcal{P}_n(a,b;2s+1)$ and $f \in \mathcal{L}_{v,a}$, $g \in \mathcal{L}_{v,b}$, which are Lipschitz with respect to the distances $\boldsymbol{d}_a$ and $\boldsymbol{d}_b$, respectively,*

$$\begin{aligned}&|\mathrm{Cov}(f(Y_{n,A}), g(Y_{n,B}) \mid \mathcal{C}_n)| \\ &\qquad \le \big(a\,\|g\|_\infty \,\mathrm{Lip}(f) + b\,\|f\|_\infty \,\mathrm{Lip}(g)\big)\theta_{n,s}, \quad \textit{a.s.,}\end{aligned}$$

where $\theta_{n,s} = 2\max_{i\in N_n}\mathbf{E}[\|Y_{n,i} - Y_{n,i}^{(s)}\| \mid \mathcal{C}_n]$.

It follows from Proposition 2.3 that $\{Y_{n,i}\}$ is conditionally ψ-weakly dependent given $\{\mathcal{C}_n\}$ provided that $\sup_n \theta_{n,s} \to_{a.s.} 0$ as $s \to \infty$, where the ψ function is given by

$$\psi_{a,b}(f,g) = \big(a\,\|g\|_\infty \,\mathrm{Lip}(f) + b\,\|f\|_\infty \,\mathrm{Lip}(g)\big).$$

As a concrete example, consider a simple linear case in which

$$Y_{n,i} = \sum_{m\ge 0} \gamma_m \sum_{j\in N_n^\partial(i;m)} \varepsilon_{n,j}.$$

This process is analogous to a linear process in time series. Since

$$\|Y_{n,i} - Y_{n,i}^{(s)}\| \le \sum_{m>s} |\gamma_m| \sum_{j\in N_n^\partial(i;m)} \|\varepsilon_{n,j}\|,$$

setting $\alpha_n = \max_{i\in N_n}\mathbf{E}[\|\varepsilon_{n,i}\| \mid \mathcal{C}_n]$, we find that

$$\theta_{n,s} \le 2\alpha_n \sum_{m>s} |\gamma_m|\, D_n^\partial(m) \quad \text{a.s.}$$

Consequently, $\{Y_{n,i}\}$ is conditionally ψ-weakly dependent when $\sup_n \alpha_n < \infty$ a.s. and $|\gamma_s| \sup_n D_n^\partial(s)$ converges to 0 fast enough, as $s \to \infty$.

2.3.4. **Functional Dependence on Associated or Gaussian Variables.** Let us consider the following process:

$$Y_{n,i} = \varphi_{n,i}(\varepsilon_{n,j}),$$

where $\varepsilon_n = (\varepsilon_{n,j} : j \in N_n)$ is a positively associated process, $\varepsilon_{n,j} \in \mathbf{R}$, conditional on certain σ-field $\mathcal{C}_n$, i.e., for all coordinatewise non-decreasing real-valued measurable functions f and g,

$$\mathrm{Cov}(f(\varepsilon_{n,A}), g(\varepsilon_{n,B}) \mid \mathcal{C}_n) \ge 0, \text{ a.s.},$$

for all finite subsets A and B of N_n. When the above inequality is reversed for all finite subsets A and B of N_n, we say that $\{\varepsilon_{n,i}\}_{i\in N_n}$ negatively associated. When a set of random variables is positively or negatively associated, independence between two random variables in the set is equivalent to their being uncorrelated. The following result follows as a consequence of a covariance inequality due to Theorem 3.1 of Birkel (1988) and Lemma 19 of Doukhan and Louhichi (1999). Let us endow N_n with a metric d_n. One example of such a metric is the distance in a graph G_n on N_n.

Proposition 2.4. *Suppose that for each $i \in N_n$, $\varphi_{n,i} \in \mathscr{C}_b^1$ (i.e. $\varphi_{n,i}$ is continuously differentiable with bounded derivatives). Suppose further that either (i) $\{\varepsilon_{n,i}\}_{i\in N_n}$ is conditionally positively or negatively associated given $\mathcal{C}_n$ and functions f, g are such that $f, g \in \mathscr{C}_b^1$, or (ii) $\{\varepsilon_{n,i}\}_{i\in N_n}$ is conditionally Gaussian given $\mathcal{C}_n$, and functions f, g are such that $f, g \in \mathscr{C}_b^1$ and f, g are bounded.*

Then for any $A, B \subset N_n$ such that $d_n(A, B) \geq s$,

$$|\mathrm{Cov}(f(Y_{n,A}), g(Y_{n,B}) \mid \mathcal{C}_n)| \leq ab\,\mathrm{Lip}(f)\,\mathrm{Lip}(g)\theta_{n,s} \quad a.s.,$$

where

$$\theta_{n,s} = \max_{1\leq k_1,k_2\leq n} \sum_{i\in N_n}\sum_{j\in N_n} \left\| \frac{\partial \varphi_{n,k_1}}{\partial \varepsilon_{n,i}} \right\|_\infty \left\| \frac{\partial \varphi_{n,k_2}}{\partial \varepsilon_{n,j}} \right\|_\infty |\mathrm{Cov}(\varepsilon_{n,i}, \varepsilon_{n,j} \mid \mathcal{C}_n)|\,. \tag{2.3}$$

The above proposition clearly shows that the dependence structure of $\{Y_{n,i}\}_{i\in N_n}$ is determined by the (conditional) local dependence structure of $\varepsilon_{n,i}$'s and $\varphi_{n,i}$'s. In the special case where $\varepsilon_{n,i}$'s are all conditionally independent given $\mathcal{C}_n$, the sequence $\theta_{n,s}$ is reduced to the following:

$$\theta_{n,s} = \max_{1\leq k_1,k_2\leq n} \left\| \frac{\partial \varphi_{n,k_1}}{\partial \varepsilon_{n,i}} \right\|_\infty \left\| \frac{\partial \varphi_{n,k_2}}{\partial \varepsilon_{n,i}} \right\|_\infty .$$

Suppose further that N_n is endowed with a graph G_n such that $\partial\varphi_{n,k}/\partial\varepsilon_{n,i} = 0$, whenever i is at least m-edges away from k in G_n. Then, $Y_{n,i}$'s have a graph G_n' as a conditional dependency graph given $\mathcal{C}_n$, where i and j are adjacent in G_n' if and only if i and j are within $2m$ edges away. Hence, it follows that $\theta_{n,s} = 0$, for all $s \geq 2m$.

The same discussion carries over to the case where the graph G_n is generated through a model of Random Geometric Graphs (see Penrose, 2003). More specifically, suppose that we have i.i.d. random vectors $X_{n,i} \in \mathbf{R}^d$, $i \in N_n$. We form an undirected graph $G_n = (N_n, E_n)$ such that $ij \in E_n$ if and only if $\|X_i - X_j\| \leq r$, where r is a given parameter. If $\varepsilon_{n,i}$'s constitute an associated process conditional on $\mathcal{C}_n' \equiv \mathcal{C}_n \vee \sigma(\{X_{n,i}\}_{i\in N_n})$, Proposition 2.4 applies to this case, with the common shock $\mathcal{C}_n'$ taken to be a σ-field that contains $\{X_{n,i}\}_{i\in N_n}$.

The following corollary shows that $\{Y_{n,i}\}_{i\in N_n}$ is conditionally ψ-weakly dependent given $\mathcal{C}_n$.

Corollary 2.1. *Suppose that $\varphi_{n,i} \in \mathscr{C}_b^1$ for all $i \in N_n$ and $n \geq 1$. Then the triangular array $\{Y_{n,i}\}$ is conditionally ψ-weakly dependent given $\{\mathcal{C}_n\}$ with the weak dependence coefficients given by* (2.3) *and*

$$\psi_{a,b}(f,g) = ab\,\mathrm{Lip}(f)\,\mathrm{Lip}(g),$$

provided that $\sup_{n\geq 1}\theta_{n,s} \to_{a.s.} 0$ *as* $s \to \infty$.

Proof. The result follows from the fact that for any $\epsilon > 0$, a Lipschitz function f admits an approximation by a continuously differentiable function f_ϵ s.t. $\|f - f_\epsilon\|_\infty \leq \epsilon$ and $\mathrm{Lip}(f_\epsilon) \leq \mathrm{Lip}(f)$ (see, e.g., Jiménez-Sevilla and Sánchez-González, 2011, p. 174). ■

3. Limit Theorems for Network Dependent Processes

3.1. Law of Large Numbers

In order to specify what features of the network topology are relevant for limit theorems, we introduce some notation for network properties.

We first establish a Law of Large Numbers (LLN). Let $\{Y_{n,i}\}$ be conditionally ψ-weakly dependent given $\{\mathcal{C}_n\}$. Since a LLN can be applied element-by-element in the vector case, without loss of generality we can assume that $Y_{n,i} \in \mathbf{R}$ in this section, i.e., $v = 1$. We assume below that the process is uniformly integrable.

Assumption 3.1 (Uniform L_1-Integrability)**.**

$$\lim_{k\to\infty} \sup_{n\geq 1} \sup_{i\in N_n} \mathbf{E}\left[|Y_{n,i}|\, 1\{|Y_{n,i}| > k\}\right] = 0.$$

Uniform L_1-integrability has been used for establishing LLNs, for example, in Jenish and Prucha (2009). A sufficient condition for the assumption is that $\sup_{i\in N_n}\mathbf{E}\left[|Y_{n,i}|^{1+\epsilon}\right] < \infty$ for some $\epsilon > 0$, see Davidson (1994, Theorem 12.10).

Below we provide an additional condition that, in combination with uniform integrability, is sufficient for the LLN to hold. Let $\|X\|_p = (\mathbf{E}\,\|X\|^p)^{1/p}$ denote the L_p-norm of a random vector X.

Assumption 3.2. $n^{-1}\sum_{s=1}^{n-1}\delta_n^{\partial}(s)\,\|\theta_{n,s}\|_1 \to 0$ as $n \to \infty$.

To interpret the above condition, we can borrow the intuition from the time-series literature on strong or uniform mixing processes. In that literature, it is common to assume that mixing coefficients are summable. Suppose that such a summability condition holds

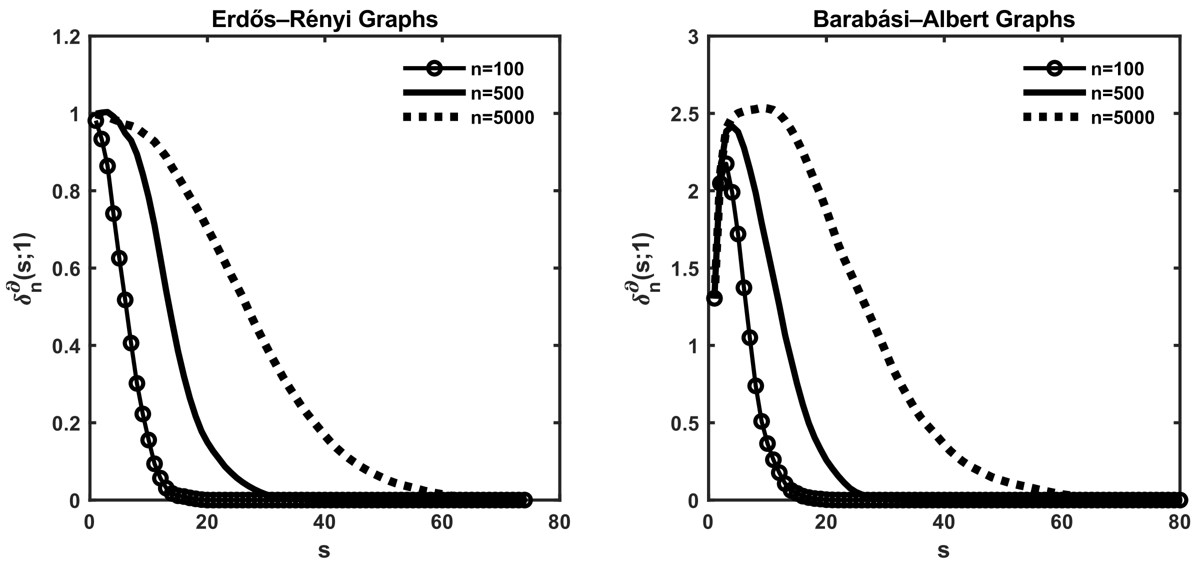

FIGURE 3. Plots of $\delta_n^\partial(s)$ for Graphs Generated as an Erdős-Rény Graphs (on the left panel) and Barabási-Albert Graphs (on the right panel): See the details of the graph generation in Section 4.2. The graph shows that regardless of the sample size, $\delta_n^\partial(s)$ vanishes as s becomes larger. It is not hard to see from the graph that $\sup_{s\geq 1}\delta_n^\partial(s)/n$ decreases as n becomes larger.

for $\|\theta_{n,s}\|_1$: $\sum_{s=1}^\infty \|\theta_{n,s}\|_1 = O(1)$, as $n\to\infty$. In such cases, a sufficient condition for Assumptions 3.2 is that $\sup_{s\geq 1}\delta_n^\partial(s) = o(n)$, i.e. the average number of neighbors at distance s across the network grows slower than the size of the network. This seems to be plausible in practice. For example, we experimented with two types of graphs generated from the random graph models of Erdős-Rény Graphs and Barabási-Albert Graphs. (See Section 4.2 for details of the graph generation.) As shown in Figure 3, $\sup_{s\geq 1}\delta_n^\partial(s)$ increases much slower than n.

Assumption 3.2 can fail, for example, if there is a node connected to almost every other node in the network as in this case. Consider a network with the *star* topology, which has a central node or hub connected to every other node. In this case, the distance between any two nodes does not exceed 2: $\delta_n^\partial(1) = 2(n-1)/n$, $\delta_n^\partial(2) = (n-2)(n-1)/n$, and $\delta_n^\partial(s) = 0$ for $s\geq 3$. Hence, for a star network, $n^{-1}\sum_{s=1}^{n-1}\delta_n^\partial(s)\,\|\theta_{n,s}\|_1 = O(\|\theta_{n,2}\|_1)$ as $n\to\infty$, and therefore Assumption 3.2 fails in this case.

Next, consider a network with the ring topology, where nodes are connected in a circular fashion to form a loop, see Figure 6(a) below. In that case, $\delta_n^\partial(s)\leq 2$, and Assumption 3.2 holds when $n^{-1}\sum_{s=1}^{n-1}\|\theta_{n,s}\|_1\to 0$. For example, if $\{\|\theta_{n,s}\|_1\}_{s=1}^{n-1}$ is summable as $n\to\infty$, Assumption 3.2 holds.

Theorem 3.1. *Suppose that $\{Y_{n,i}\}$ is conditionally ψ-weakly dependent given $\{\mathcal{C}_n\}$ and Assumptions 3.1 and 3.2 hold. Then, as $n \to \infty$,*

$$\mathbf{E}\left|\frac{1}{n}\sum_{i=1}^{n}\left(Y_{n,i} - \mathbf{E}[Y_{n,i} \mid \mathcal{C}_n]\right)\right| \to 0.$$

When $f \in \mathcal{L}_{v,1}$, a LLN for a nonlinear transformation $f(Y_{n,i})$ follows immediately from the definition of the ψ-weakly dependence in Definition 2.2.[6] In that case,

$$\mathbf{E}\left|\frac{1}{n}\sum_{i=1}^{n}\left(f(Y_{n,i}) - \mathbf{E}[f(Y_{n,i}) \mid \mathcal{C}_n]\right)\right|^2 \leq \frac{1}{n}\sup_{n,i}\mathbf{E}\left|f(Y_{n,i})\right|^2 + \psi_{1,1}(f,f)\frac{1}{n}\sum_{s=1}^{n-1}\delta_n^{\partial}(s;1)\left\|\theta_{n,s}\right\|_1.$$

We have the following result.

Corollary 3.1. *Suppose that $\{Y_{n,i}\}$ is conditionally ψ-weakly dependent given $\{\mathcal{C}_n\}$, Assumption 3.2 holds, and $\sup_{n,i}\mathbf{E}\left|f(Y_{n,i})\right|^2 < \infty$, where $f \in \mathcal{L}_{v,1}$. Then,*

$$\mathbf{E}\left|\frac{1}{n}\sum_{i=1}^{n}(f(Y_{n,i}) - \mathbf{E}[f(Y_{n,i}) \mid \mathcal{C}_n])\right|^2 \to 0.$$

In general, however, nonlinear transformations of ψ-weakly dependent processes are not necessarily ψ-weakly dependent. In such cases, LLNs for nonlinear transformations can be established using the covariance inequalities for transformation functions presented in Appendix A.1 in this paper. For example, suppose that the assumptions of Corollary A.1 in the appendix hold for some nonlinear function $h(\cdot)$ of a ψ-weakly-dependent process $\{Y_{n,i}\}$. In that case for some constants $C > 0$ and $p > 2$, the covariance between $h(Y_{n,i}) - \mathbf{E}[h(Y_{n,i}) \mid \mathcal{C}_n]$ and $h(Y_{n,j}) - \mathbf{E}[h(Y_{n,j}) \mid \mathcal{C}_n]$ is bounded by

$$C \cdot \left\|\sup_{n,i}\|h(Y_{n,i})\|^2_{\mathcal{C}_n,p}\right\|_2 \cdot \left\|\theta^{1-\frac{2}{p}}_{n,d_n(i,j)}\right\|_2,$$

where $\|X\|_{\mathcal{C}_n,p} = (\mathbf{E}[|X|^p \mid \mathcal{C}_n])^{1/p}$. Therefore,

$$\mathbf{E}\left|\frac{1}{n}\sum_{i=1}^{n}(h(Y_{n,i}) - \mathbf{E}[h(Y_{n,i}) \mid \mathcal{C}_n])\right|^2 \to 0,$$

[6] Note that compositions of bounded Lipschitz functions are also bounded and Lipschitz; hence $f(Y_{n,i})$ is also ψ-weakly dependent when $f \in \mathcal{L}_v$.

provided that $\mathbf{E}[\sup_{n,i}\|h(Y_{n,i})\|^2_{\mathcal{C}_n,p}] < \infty$, and a condition similar to that in Assumption 3.2 holds:

$$\frac{1}{n}\sum_{s=1}^{n-1}\delta_n^{\partial}(s)\left\|\theta_{n,s}^{1-\frac{2}{p}}\right\|_2 = o(1).$$

Cases not covered by Corollary A.1 can be handled in a similar manner using the covariance inequality of Theorem A.2 in Appendix A.1. We use such a strategy to show the consistency of the HAC estimator in Section 4.

3.2. Central Limit Theorem

In this section, we study the central limit theorem for a sum of random variables that are conditionally ψ-weakly dependent. Define

$$\sigma_n^2 = \mathrm{Var}(S_n \mid \mathcal{C}_n),$$

where $S_n = \sum_{i\in N_n} Y_{n,i}$. We make the following assumptions. The assumption below presents a moment condition.

Assumption 3.3. There exist $C > 0$ and $p \geq 4$ such that for all $n \geq 1$, $\max_{i\in N_n}\|Y_{n,i}\|_{\mathcal{C}_n,p} \leq C$ a.s.

While requiring the moment condition in Assumption 3.3 is more restrictive than those conditions known for the central limit theorem for special cases of ψ-dependence, such a moment condition is widely used in many models in practice. The following assumption limits the extent of the cross-sectional dependence of the random variables through restrictions on the network.

Assumption 3.4. For each $n \geq 1$, there exists $c_n \to \infty$ such that $0 < c_n \leq \sigma_n$ a.s., and there exists a positive sequence $m_n \to \infty$ such that for some $p \geq 4$ that appears in Assumption 3.3,

$$\frac{1}{c_n^4}\sum_{s=0}^{n-2}|H_n(s,m_n)|\theta_{n,s}^{1-4/p} \to_{a.s.} 0, \tag{3.1}$$

where we take $\theta_{n,0} = 1$, and define

$$H_n(s,m_n) = \{(i,j,k,l)\in N_n^4 : j\in N_n(i;m_n), l\in N_n(k;m_n), d_n(\{i,j\},\{k,l\}) = s\},$$

and

$$\frac{n\delta_n(m_n;2)}{c_n^3} \to 0, \text{ and } \frac{n^2\theta_{n,m_n}^{1-1/p}}{c_n} \to_{a.s.} 0, \tag{3.2}$$

as $n \to \infty$.

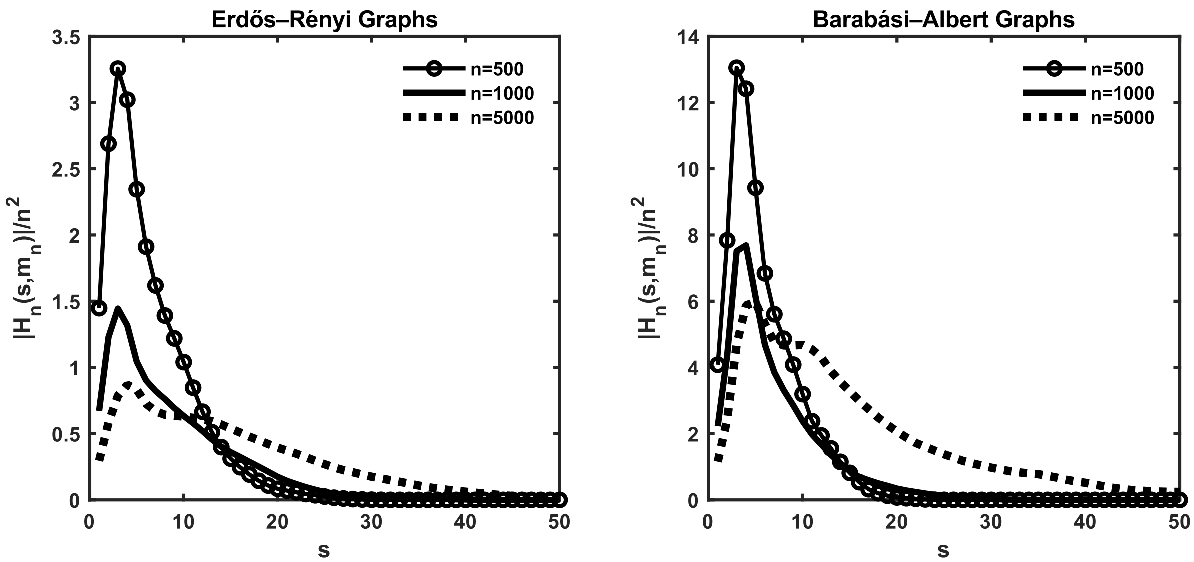

FIGURE 4. Plots of $|H_n(s;m_n)|/n^2$ for Graphs Generated as an Erdős-Rény Graph (on the left panel) and Barabási-Albert Graph (on the right panel) With $m_n = \log(1+n)$: See the details of the graph generation in Section 4.2. The graph shows that regardless of the sample size, $|H_n(s;m_n)|$ vanishes as s becomes larger. It is not hard to see from the graph that $\sup_{s \geq 1}|H_n(s;m_n)|/n^2$ decreases as n becomes larger.

The requirement for the graph is naturally tied to the strength and the extensiveness of the cross-sectional dependence of the random variables. If the cross-sectional dependence is substantially local, i.e., for each random variable, there is only a small set of other random variables that it is allowed to stochastically dependent with, then the requirement for the graph can be weak. For example, let us assume that $\mathcal{C}_n$ is a trivial σ-field, and the variance of S_n increases at the rate of n, so that we can take $c_n = cn^{1/2}$ for all $n \geq 1$ for some $c > 0$. Furthermore, assume that $\theta_{n,s} = 0$ for all $s > 1$. An example of such a model is a model of dependency graphs. Then, Conditions in (3.1) and (3.2) are satisfied if

$$\frac{\delta_n^\partial(1)D_n^\partial(1)}{\sqrt{n}} + \frac{\delta_n^\partial(1)D_n^\partial(1)^2}{n} \to 0,$$

as $n \to \infty$. These conditions are easily satisfied by many sparse graphs, and are not necessarily overly strong given what exists in the literature. For example, these conditions are satisfied under the condition that the Berry-Esseen bound for normal approximation in Theorem 2.3 of Penrose (2003) converges to zero.

More generally, suppose that m_n is a slowly increasing sequence such that θ_{n,m_n} decays to zero faster than any polynomial rate in n and $\delta_n(m_n;2) = o(\sqrt{n})$. Furthermore, assume that there exist $C > 0$ and $\tau \geq 1$ such that

$$|H_n(s,m_n)| \leq Cs^\tau o(n^2)$$

for all $s \geq 1$, and

$$\sum_{s=1}^{n-2} s^{\tau} \theta_{n,s}^{1-4/p} < \infty.$$

Then Conditions in (3.1) and (3.2) follow with $c_n = cn^{1/2}$ for some constant $c > 0$, if the variance of S_n increases at the rate n. When $\theta_{n,s}$ does not decay to zero fast enough as s becomes large (i.e., the cross-sectional dependence is extensive), it is difficult to find a slowly increasing sequence m_n such that the second condition in (3.2) is satisfied.

We have performed experiments to check the plausibility of the conditions using some examples of two types of graphs used in Figure 3. The results are shown in Figure 4. We have chosen $m_n = \log(1 + n)$. The graph shows that $|H_n(s; m_n)|/n^2$ decreases as n becomes larger. Hence, the conditions in Assumption 3.4 seem plausible in these examples.

Theorem 3.2. *Suppose that Assumptions 2.1, 3.3-3.4 hold, and that* $\mathbf{E}[Y_{n,i} \mid \mathcal{C}_n] = 0$ *a.s. Then,*

$$\sup_{t \in \mathbf{R}} \left| P\left\{ \frac{1}{\sigma_n} \sum_{i \in N_n} Y_{n,i} \leq t \mid \mathcal{C}_n \right\} - \Phi(t) \right| \to_{a.s.} 0, \text{ as } n \to \infty,$$

where Φ *denotes the distribution function of* $N(0, 1)$.

The proof of the central limit theorem uses Stein's Lemma (Stein, 1986). The central limit theorem immediately gives a stable convergence of a normalized sum of random variables under appropriate conditions. More specifically, suppose that

$$\sigma_n^2/n \to_{a.s.} v^2,$$

where v^2 is a random variable that is $\mathcal{C}$-measurable and $\mathcal{C}$ is a sub σ-field of $\mathcal{C}_n$. Then it follows that $\frac{1}{\sqrt{n}} \sum_{i \in N_n} Y_{n,i}$ converges stably to a mixture normal random variable.

4. HAC Estimation

Suppose that we observe a network dependent process $\{Y_{n,i}\}$ satisfying Assumption 2.1. In this section we provide two kernel HAC estimators for the conditional variance of $S_n/\sqrt{n}$ given $\mathcal{C}_n$, where $S_n = \sum_{i \in N_n} Y_{n,i}$. First, we assume that $\mathbf{E}[Y_{n,i} \mid \mathcal{C}_n] = 0$ a.s. for all $i \in N_n$. Let

$$\Omega_n(s) = n^{-1} \sum_{i \in N_n} \sum_{j \in N_n^{\partial}(i;s)} \mathbf{E}[Y_{n,i} Y_{n,j}^{\top} \mid \mathcal{C}_n].$$

Then the above-mentioned variance is given by

$$V_n = \mathrm{Var}(S_n/\sqrt{n} \mid \mathcal{C}_n) = \sum_{s \geq 0} \Omega_n(s) \quad \text{a.s.} \tag{4.1}$$

Similarly to the time-series case the asymptotic consistency of an estimator of V_n requires a restriction on weights given to the estimated "autocovariance" terms $\Omega_n(\,\cdot\,)$. Consider a kernel function $\omega : \bar{\mathbf{R}} \to [-1,1]$ such that

(1) $\omega(0) = 1$;
(2) $\omega(z) = \omega(-z)$, for $z \in \bar{\mathbf{R}}$;
(3) $\omega(z) = 0$ for $|z| > 1$.

Let $b_n \le \max_{i,j \in N_n}\{d_n(i,j)\} < n$ denote the bandwidth or the lag truncation parameter. Then the kernel HAC estimator of V_n is given by

$$\hat{V}_n = \sum_{s=0}^{\lfloor b_n \rfloor} \omega_n(s)\hat{\Omega}_n(s), \tag{4.2}$$

where $\omega_n(s) = \omega(s/b_n)$,

$$\hat{\Omega}_n(s) = n^{-1} \sum_{i \in N_n} \sum_{j \in N_n^{\partial}(i;s)} Y_{n,i} Y_{n,j}^{\top},$$

and $\lfloor b_n \rfloor$ is the greatest integer less than or equal to b_n. The weight given for each sample covariance term $\hat{\Omega}_n(s)$ is a function of distance s implied by the structure of a network. Also notice that if nodes i and j are disconnected then $d_n(i,j) = \infty$ so that $\omega_n(d_n(i,j)) = 0$.

Next, assume that $\mathbf{E}[Y_{n,i} \mid \mathcal{C}_n] = \Lambda_n$ a.s. for all $i \in N_n$ and the sequence of common conditional expectations $\{\Lambda_n\}$ is unknown. Under suitable assumptions Theorem 3.1 implies that $\bar{Y}_n = S_n/n$ is a consistent estimator of Λ_n in the sense that their difference converges to zero (in probability). We redefine the kernel HAC estimator given in (4.2) as follows:

$$\tilde{V}_n = \sum_{s=0}^{\lfloor b_n \rfloor} \omega_n(s)\tilde{\Omega}_n(s), \tag{4.3}$$

where

$$\tilde{\Omega}_n(s) = n^{-1} \sum_{i \in N_n} \sum_{j \in N_n^{\partial}(i;s)} \left(Y_{n,i} - \bar{Y}_n\right)\left(Y_{n,j} - \bar{Y}_n\right)^{\top}.$$

If random vectors $\{Y_{n,i}\}_{i \in N_n}$ do not share a common expectation, it is hard to justify plugging the sample mean into $\tilde{\Omega}_n(\,\cdot\,)$ because $\bar{Y}_n$ is not a consistent estimator of $\mathbf{E}[Y_{n,i} \mid \mathcal{C}_n]$. The role of truncation is even more important in this case; for example, for the rectangular kernel and the maximal possible bandwidth $b_n = \max_{i,j \in N_n}\{d(i,j)\}$ the estimator $\tilde{V}_n$ is identically zero.

4.1. Consistency

The consistency of the estimators defined in (4.2) and (4.3) is established by imposing suitable conditions on the moments of the array $\{Y_{n,i}\}$, the denseness of a sequence of networks, and the rate of growth of the bandwidth parameter.

Assumption 4.1. There exists $r > 2$ such that

(i) $\mu = \sup_{n\geq 1} \sup_{i\in N_n} \|Y_{n,i}\|_{\mathcal{C}_n,2r} < \infty$ a.s.;

(ii) $\limsup_{n\to\infty} \sum_{s=1}^{n-1} \delta_n^{\partial}(s)\theta_{n,s}^{1-\frac{2}{r}} < \infty$ a.s.;

(iii) $\lim_{n\to\infty} n^{-2} \sum_{s=0}^{n-1} |H_n(s,b_n)|\,\theta_{n,s}^{1-\frac{2}{r}} = 0$ a.s. with $\theta_{n,0} \equiv 1$;

(iv) $\lim_{n\to\infty} |\omega_n(s) - 1| = 0$ for all $s \in \mathbb{N}$.

The first three conditions demonstrate the tradeoff between the conditional moments of $\{Y_{n,i}\}$ given $\{\mathcal{C}_n\}$ and the magnitude of the network dependence. For a given sequence of networks, a stronger network dependence requires the finiteness of higher conditional moments, i.e. a larger value of r. On the other hand, sparse networks allow for either weaker moments conditions or a stronger dependence along the network.

Assumption 4.1(iii) determines the admissible rate of growth of the sequence of bandwidths $\{b_n\}$. In particular, it strongly depends on the network topology. For example, if the number of neighbors at any distance is uniformly bounded over a sequence of networks, i.e. $\sup_{s,n} \sup_{i\in N_n} N_n^{\partial}(i;s) < \infty$, then the bandwidth is allowed to grow with a rate slightly lower than $n^{1/2}$. To see this, it worth noting that for sufficiently sparse networks this condition can be replaced with the combination of Assumption 4.1(ii) and

(iii') $D_n(b_n)/\sqrt{n} \to 0$.

Indeed, since

$$\{(i,j,k,l) \in N_n^4 : d_n(\{i,j\},\{k,l\}) = s\} \subseteq \bigcup_{\tau_1,\tau_2\in N_n^2} \{d_n(\tau_1,\tau_2) = s\},$$

we have

$$\frac{1}{n^2}\sum_{s=0}^{n-1} |H_n(s,b_n)|\,\theta_{n,s}^{1-\frac{2}{r}} \leq \frac{4D_n(b_n)^2}{n}\left(1 + \sum_{s=1}^{n-1} \delta^{\partial}(s;1)\theta_{n,s}^{1-\frac{2}{r}}\right).$$

Proposition 4.1. *Suppose that Assumption 4.1 holds. Then*

$$\mathbf{E}[\|\hat{V}_n - V_n\|_F \mid \mathcal{C}_n] \to 0 \quad a.s.,$$

where $\|\cdot\|_F$ denotes the Frobenius norm.[7] *If, in addition, $D_n(b_n)/n \to 0$, then*

$$\mathbf{E}[\|\tilde{V}_n - V_n\|_F \mid \mathcal{C}_n] \to 0 \quad a.s.$$

[7] For a real matrix A, $\|A\|_F = \sqrt{\mathrm{tr}(A^\top A)}$.

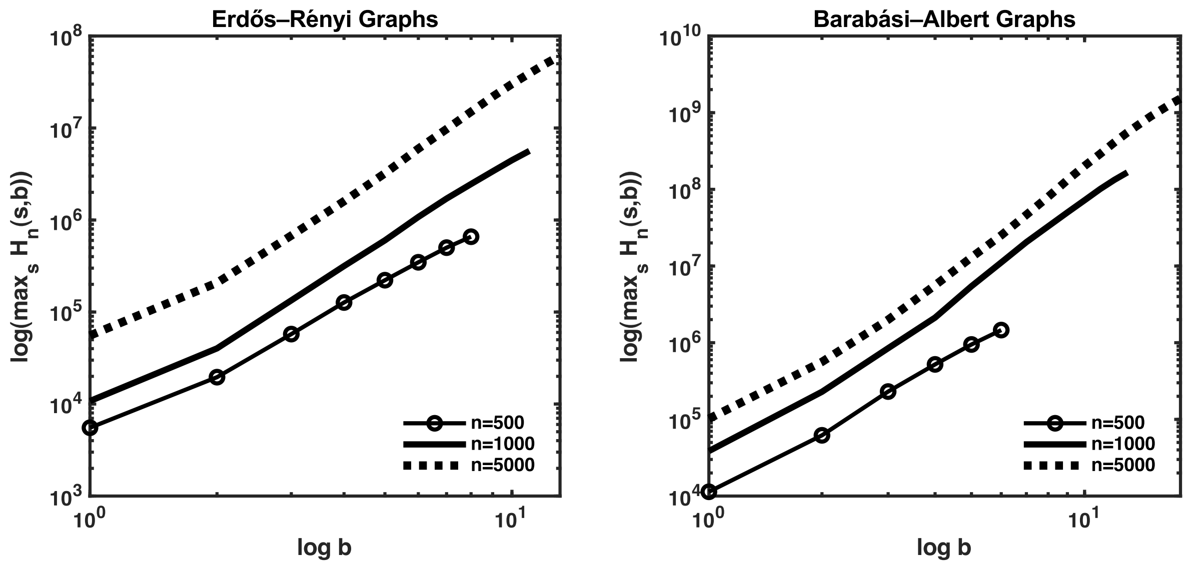

FIGURE 5. The log-log plots of $\max_s H_n(s,b)$ against b for graphs generated using the Erdős-Rény (on the left panel) and Barabási-Albert (on the right panel) models: See the details of the graph generation in Section 4.2.

Note that under Assumption 4.1(iii') the additional condition for consistency of the second HAC estimator $\tilde{V}_n$ becomes redundant. Also for simple network topologies it is possible to derive an explicit expression for the difference between two estimators $A_n(c) = c^\top \left(\hat{V}_n - \tilde{V}_n\right) c$, where $c \in \mathbf{R}^v$ is a fixed vector. Example 4.1 below illustrates such a case and shows that positive-definiteness of the kernel function does not imply automatically positive-semidefiniteness of the estimated variance-covariance matrix.

Construction of the HAC estimator requires selection of the truncation parameter b_n in a way that satisfies Assumption 4.1(iii). Suppose that the sequence $\{\theta_{n,s}^{1-2/r}\}$ is summable, i.e. $\sum_{s=0}^{n-1} \theta_{n,s}^{1-2/r} = O_{a.s.}(1)$. Suppose further that for some $\beta > 2$, and $\lim_{n\to\infty} b_n = \infty$,

$$\sup_s H_n(s, b_n) = O(b_n^\beta). \tag{4.4}$$

In such cases, a sufficient condition for Assumption 4.1(iii) is that

$$b_n = o(n^{2/\beta}). \tag{4.5}$$

For example, in practice one can use $b_n = n^{2/\beta}/\log\log n$.

The parameter β in (4.4) depends on the asymptotic behavior of a sequence of networks. Since the network is observed, $\max_s H_n(s,b)$ can be computed from data for a range of values of b. The coefficient β can be estimated by regressing $\log(\max_s H_n(s,b))$ against $\log b$ and a constant. Note that since we only observe a finite network, one should exclude large values of b to avoid biasing the estimates of β. For example, Figure 5 shows

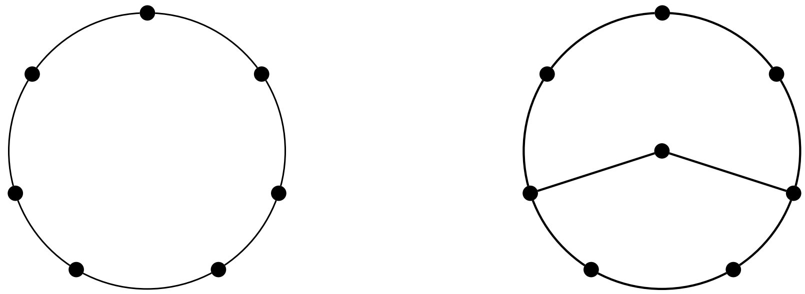

FIGURE 6. An example of networks for which the corresponding weighting matrices $W = [\omega(d_n(i,j)/2)]_{i,j\in N_n}$ are either positive semidefinite (A) or indefinite (B) for the same positive-definite kernel function $\omega(z) = \mathbf{1}\{|z| \leq 1\}(1 - |z|)$.

the plots of $\log(\max_s H_n(s,b))$ against $\log b$ for Erdős-Rényi and Barabási-Albert graphs. One can see that the condition in (4.4) matches the behavior of $H_n(s,b)$ for these graphs with the estimates of β in the range 2.4–3.1 for Erdős-Rényi, and 2.8–3.5 for Barabási-Albert (for different sample sizes n).

Example 4.1. Consider a ring network (an example is shown in Figure 6(A)) where $N_n^{\partial}(i;s) = 2$ for $1 \leq s \leq \lfloor (n-1)/2 \rfloor$ and all $i \in N_n$. Suppose that $\Lambda_n = 0$ a.s. and let $\omega(z) = (1 - |z|)\mathbf{1}\{|z| \leq 1\}$ (Barlett kernel). For $b_n < (n-1)/2$,

$$A_n(c) = 2\bar{y}_n^2 \sum_{s=0}^{b_n} \left(1 - \frac{s}{b_n + 1}\right) = \bar{y}_n^2(2 + b_n) \geq 0.$$

Hence, $\hat{V}_n - \tilde{V}_n$ is positive-semidefinite. In particular, $[\hat{V}_n - \tilde{V}_n]_{k,k} \geq 0$ for all $1 \leq k \leq v$ so that the HAC estimator $\tilde{V}_n$ yields lower variances in finite samples.

In addition, it is easy to verify that given the network in Figure 6(A) and the Barlett kernel each estimator yields a positive-semidefinite covariance matrix. Generally, if the weighting matrix $W = [\omega_n(d_n(i,j))]_{i,j\in N_n}$ is positive-semidefinite, there exists a matrix L with $W = LL^\top$ so that

$$\hat{V}_n = n^{-1}(\hat{Y}L)(\hat{Y}L)^\top \text{ and } \tilde{V}_n = n^{-1}(\tilde{Y}L)(\tilde{Y}L)^\top,$$

where $\hat{Y}$ and $\tilde{Y}$ are $d \times n$ matrices whose columns are given by $(Y_{n,i} - \mathbf{E}[Y_{n,i} \mid \mathcal{C}_n])$ and $(Y_{n,i} - \bar{Y}_n)$, respectively. Hence, both $\hat{V}_n$ and $\tilde{V}_n$ are positive semidefinite. Consequently, in a context, in which the distance measure corresponds to the Euclidean norm on $\mathbf{R}^p$, $p \geq 1$ i.e. $d(i,j) = \|x_i - x_j\|_2$ for some vectors of characteristics $x_i, x_j \in \mathbf{R}^p$, positive-definiteness of the kernel function implies that W is positive semidefinite (see, e.g., Kelejian and Prucha, 2007 and Wendland, 2004, Chapter 6).

This result, however, is not applicable to our case and positive-semidefiniteness of the weighting matrix strongly depends on the network topology. For example, while W is positive semidefinite for the ring network in Figure 6(A) and the Barlett kernel with $b_n = 2$, it becomes indefinite after a slight modification shown in Figure 6(B).

4.2. Monte Carlo Study

For our simulation study we use a simple ring network topology with different number of nodes $n = 500, 1,000$, and $5,000$ and the Parzen kernel given by

$$\omega(x) = \begin{cases} 1 - 6x^2 + 6|x|^3 & \text{for } 0 \leq |x| \leq 1/2, \\ 2(1-|x|)^3 & \text{for } 1/2 < |x| \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

These calculations are based on $10,000$ simulated samples of n random variables generated by the following linear model:

$$Y_{n,i} = \sum_{m \geq 0} \gamma^m \sum_{j \in N_n^{\partial}(i;m)} \varepsilon_{n,j}, \quad i \in N_n, \tag{4.6}$$

where $\{\varepsilon_{n,i}\}$ are independent $\mathcal{N}(0,1)$ r.v.s. and $\gamma \in [0,1)$. The triangular array $\{Y_{n,i}\}$ is ψ-weakly dependent because $\sup_n \theta_{n,s} \leq 2\gamma^s/(1-\gamma) \to 0$ as $s \to \infty$. Note that when $\gamma = 0$, $Y_{n,i} = \varepsilon_{n,i}$ so that all the nodes of a network are independent. In this case the true variance takes a simple form and for large n it approximately equals $[(1+\gamma)/(1-\gamma)]^2$. Since the relevant graphs are sparse enough, we set the bandwidth parameter $b_n = \lfloor n^{2/5} \rfloor + 1$, which satisfies Assumption 4.1.

We present results Table 1 on bias and MSEs of HAC estimators in Table 1. The results show small sample biases and RMSEs for the HAC estimators (4.2) and (4.3). It is clear that both the bias (in absolute value) and the RMSE of the estimators decline as the sample size increases for all values of γ. In addition, as follows from Example 4.1, the second HAC estimator yields consistently lower estimates.

The next simulation results presented in Table 3 and Table 4 correspond to a test for the sample mean. The following t-type test statistic is used for testing $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$:

$$T_n = \frac{\bar{Y}_n - \mu_0}{\sqrt{\tilde{V}_n/n}}.$$

In the simulations, we set $\mu_0 = 0$. Consistency of the kernel HAC estimator and Theorem 3.2 guarantee that the limiting distribution of T_n is standard normal. We use the

TABLE 1. Bias and RMSE of the HAC estimators $\hat{V}_n$ and $\tilde{V}_n$ relative to the true variance V_n for ring networks.

n	V_n	$\hat{V}_n$		$\tilde{V}_n$	
		Bias	RMSE	Bias	RMSE
$\gamma = 0$[a]					
500	1.0	-0.002	0.170	-0.021	0.167
1,000	1.0	-0.000	0.133	-0.013	0.132
5,000	1.0	-0.000	0.080	-0.006	0.080
$\gamma = 1/3$					
500	4.0	-0.322	0.725	-0.401	0.755
1,000	4.0	-0.226	0.562	-0.274	0.580
5,000	4.0	-0.070	0.326	-0.089	0.329
$\gamma = 1/2$					
500	9.0	-1.542	2.069	-1.719	2.191
1,000	9.0	-1.136	1.588	-1.244	1.660
5,000	9.0	-0.379	0.802	-0.421	0.820

[a] γ controls the strength of the stochastic dependence in the DGP given by Equation (4.6). Specifically, $\gamma = 0$ corresponds to the case of stochastic independence.

TABLE 2. Denseness measures of the *ER* and *BA* graphs used in the simulations.

Graph	n	Avg. degree	Max. degree	Diam.	$\bar{d}$[a]
ER	500	1.02	5	19	6.27
	1,000	0.98	5	30	9.24
	5,000	1.01	7	56	15.77
BA	500	1.31	7	15	5.56
	1,000	1.33	9	29	10.27
	5,000	1.27	10	43	13.16

[a] The average connected distance; see Equation (4.7).

same simulation setup except for the underlying network structures which are generated using (a) the $G(n,p)$ Erdős-Rényi model (*ER* graphs) with parameter $p = 1/n$, (b)

TABLE 3. Simulated power and size of the sample mean test of nominal size 5% for the *ER* graphs; the true mean $\mu_0 = 0$.

n	μ				
	-0.2	-0.1	**0.0**	0.1	0.2
$\gamma = 0$[(a)]					
500	0.993	0.611	**0.056**	0.627	0.994
1,000	1.000	0.889	**0.056**	0.886	1.000
5,000	1.000	1.000	**0.052**	1.000	1.000
$\gamma = 1/3$					
500	0.815	0.326	**0.066**	0.337	0.817
1,000	0.976	0.542	**0.063**	0.554	0.978
5,000	1.000	0.993	**0.056**	0.991	1.000
$\gamma = 1/2$					
500	0.575	0.215	**0.070**	0.220	0.578
1,000	0.811	0.340	**0.065**	0.346	0.814
5,000	1.000	0.864	**0.059**	0.857	1.000

[(a)] γ controls the strength of the stochastic dependence in the DGP given by Equation (4.6). Specifically, $\gamma = 0$ corresponds to the case of stochastic independence.

the Barabási–Albert model (*BA* graphs) with the connectivity parameter 1 and the seed graph $G(m, 1/m)$, where $m = 0.7n$. These parameters are chosen to obtain sufficiently sparse networks. Some useful statistics describing the denseness of the generated graphs are shown in Table 2. In particular, $\bar{d}$ represents the average connected distance and is given by

$$\bar{d} = \frac{1}{k} \sum_{\substack{u,v \in N_n \\ 0 < d(u,v) < \infty}} d_n(u,v), \tag{4.7}$$

where k is the number of connected pairs $(u,v) \in N_n^2$. We choose the truncation parameter as $b_n = n^{2/\hat{\beta}}/\log\log n$, where $\hat{\beta}$ is the estimated slope coefficient from the log-log regression described in Section 4.1. It can be seen that the Barabási–Albert model generates denser networks which results in slightly lower coverage probabilities.

Finally, we simulate the same t-test for a network of size $n = 5,000$ consisting of m equal-sized disconnected components (*Block* graphs), where $m = 10, 25$, or 50 and each

TABLE 4. Simulated power and size of the sample mean test of nominal size 5% for the *BA* graphs; the true mean $\mu_0 = 0$.

n	μ				
	-0.2	-0.1	**0.0**	0.1	0.2
$\gamma = 0$[(a)]					
500	0.993	0.619	**0.063**	0.629	0.994
1,000	1.000	0.889	**0.059**	0.891	1.000
5,000	1.000	1.000	**0.052**	1.000	1.000
$\gamma = 1/3$					
500	0.665	0.262	**0.077**	0.267	0.674
1,000	0.896	0.413	**0.075**	0.407	0.897
5,000	1.000	0.962	**0.057**	0.964	1.000
$\gamma = 1/2$					
500	0.370	0.165	**0.084**	0.162	0.373
1,000	0.550	0.232	**0.087**	0.221	0.545
5,000	0.993	0.656	**0.059**	0.654	0.993

(a) γ controls the strength of the stochastic dependence in the DGP given by Equation (4.6). Specifically, $\gamma = 0$ corresponds to the case of stochastic independence.

TABLE 5. Denseness measures of the *Block* graphs used in the simulations; $n = 5,000$.

Num. of blocks	Avg. degree	Max. degree	Diam.	$\bar{d}$[(a)]
10	1.99	9	22	8.04
25	1.98	8	23	6.61
50	1.99	8	21	5.60

(a) The average connected distance; see Equation (4.7).

component is generated using the Erdös-Rényi model with parameter $p = 2m/n$. These graphs are denser than ones used previously (see Table 5). However, since the blocks are disconnected there is no long-range dependence which is common to the previous setup. The simulation results are shawn in Table 6. Specifically, when the number of blocks is

TABLE 6. Simulated power and size of the sample mean test of nominal size 5% for the *Block* graphs of size $5,000$; the true mean $\mu_0 = 0$.

Num. of blocks	μ				
	-0.2	-0.1	**0.0**	0.1	0.2
$\gamma = 0$[(a)]					
10	1.000	1.000	**0.062**	1.000	1.000
25	1.000	1.000	**0.057**	1.000	1.000
50	1.000	1.000	**0.056**	1.000	1.000
$\gamma = 1/3$					
10	0.998	0.708	**0.084**	0.716	0.998
25	0.999	0.715	**0.064**	0.720	0.999
50	0.999	0.748	**0.060**	0.746	0.999
$\gamma = 1/2$					
10	0.563	0.253	**0.122**	0.256	0.570
25	0.613	0.231	**0.071**	0.232	0.620
50	0.725	0.261	**0.062**	0.263	0.723

[(a)] γ controls the strength of the stochastic dependence in the DGP given by Equation (4.6). Specifically, $\gamma = 0$ corresponds to the case of stochastic independence.

large enough, the overall dependence in a network is low so that the simulated size of the test is close to the nominal one even for high values of γ.

5. Conclusion

Developing an asymptotically valid inference method for observations that are cross-sectionally dependent has long drawn a great deal of attention in the literature of econometrics. This paper contributes to this literature by establishing limit theorems for sums of random vectors when the random vectors exhibit cross-sectional dependence along a large network. For the notion of dependence, we adopt the approach of Doukhan and Louhichi (1999). The normal approximation of the distribution of the sum of such random vectors is limited by the extensiveness of the cross-sectional dependence which can be summarized in terms of graph statistics. An interesting question is whether there would be an alternative inference method (such as based on certain resampling methods) which exhibits stable performance over a wide range of the extent of the cross-sectional dependence. Standard nonparametric bootstrap methods do not directly apply for these random vectors because of heterogeneous distributions and heterogeneity in local dependence across observations. Thus, developing a new inference method that overcomes this challenge seems both promising and challenging.

References

Andrews, D., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica 59 (3), 817–58.

Andrews, D., Monahan, J. C., 1992. An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. Econometrica 60 (4), 953–66.

Baldi, P., Rinott, Y., 1989. On normal approximations of distributions in terms of dependency graphs. Annals of Probability 17, 1646–1650.

Birkel, T., 1988. On the convergence rate in the central limit theorem for associated processes. Annals of Probability 16, 1685–1698.

Bläsius, T., Friedrich, T., Krohmer, A., 2017. Cliques in hyperbolic random graphs. Algorithmica.

Borg, I., Groenen, P. J. F., 2005. Modern Multidimensional Scaling. Springer Verlag, New York.

Boucher, V., Mourifie, I., 2017. My friend far, far away: A random field approach to exponential random graph models. Econometrics Journal 20, S14–S46.

Chen, L. H. Y., Goldstein, L., Shao, Q.-M., 2011. Normal approximation by Stein's method. Springer-Verlag, New York, USA.

Chen, L. H. Y., Shao, Q.-M., 2004. Normal approximation under local dependence. Annals of Probability 32, 1985–2028.

Comets, F., Janžura, M., 1998. A central limit theorem for conditionally centered random fields with an application to markov fields. Journal of Applied Probability 35, 608–621.

Conley, T. G., 1999. Gmm estimation with cross-sectional dependence. Journal of Econometrics 92, 1–45.

Da Prato, G., Zabczyk, J., 2014. Stochastic Equations in Infinite Dimensions, 2nd Edition. Encyclopedia of Mathematics and Its Applications. Cambridge University Press.

Davidson, J., 1994. Stochastic Limit Theory: An Introduction for Econometricians. Oxford University Press, New York.

Dedecker, J., Doukhan, P., Lang, G., 2007. Weak dependence: with examples and applications. Lecture notes in statistics. Springer, New York.

Dedecker, J., Louhichi, S., 2002. Maximal inequalities and empirical central limit theorems. In: Empirical Process Techniques for Dependent Data. pp. 137–159.

Doukhan, P., Louhichi, S., 1999. A new weak dependence condition and applications to moment inequalities. Stochastic Processes and their Applications 84 (2), 313–42.

Gaetan, C., Guyon, X., 2010. Spatial Statistics and Modeling. Springer, New York.

Hall, P. G., Heyde, C. C., 1980. Martingale limit theory and its applications. Academic Press, New York; London.

Hwang, E., Shin, D. W., 2013. A study on moment inequalities under a weak dependence. Journal of the Korean Statistical Society 42 (1), 133–41.

Janson, S., 1988. Normal convergence by higher semiinvariants with applications to sums of dependent random variables and random graphs. Annals of Probability 16, 305–312.

Jenish, N., Prucha, I. R., 2009. Central limit theorems and uniform laws of large numbers for arrays of random fields. Journal of Econometrics 150 (1), 86–98.

Jenish, N., Prucha, I. R., 2012. On spatial processes and asymptotic inference under near-epoch dependence. Journal of Econometrics 170, 178–190.

Jia, P., 2008. What happens when wal-mart comes to town: An empirical analysis of the discount retailing industry. Econometrica 170, 1263–11316.

Jiménez-Sevilla, M., Sánchez-González, L., 2011. Smooth extension of functions on a certain class of non-separable banach spaces. Journal of Mathematical Analysis and Applications 378 (1), 173–183.

Kelejian, H. H., Prucha, I. R., 2007. HAC estimation in a spatial framework. Journal of Econometrics, 131–54.

Kuersteiner, G. M., Prucha, I. R., 2013. Limit theory for panel data models with cross sectional dependence and sequential exogeneity. Journal of Econometrics 174 (2), 107–126.

Kuersteiner, G. M., Prucha, I. R., 2015. Dynamic spatial panel models: Networks, common shocks, and sequential exogeneity. Working Paper.

Lauritzen, S. L., 1996. Graphical Models. Clarendon Press, Oxford.

Lee, J. H., Song, K., 2018. Stable limit theorems for empirical processes under conditional neighborhood dependence. Bernoulli, Forthcoming.

Lee, L.-F., 2004. Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. Econometrica 72, 1899–1925.

Lee, L.-F., Liu, X., Lin, X., 2010. Specification and estimation of social interaction models with network structures. Econometrics Journal 13, 145–176.

Leung, M. P., 2016. Treatment and spillover effects under network interference. Working Paper.

Murphy, K. P., 2012. Machine Learning: A Probabilistic Perspective. The MIT Press, New York, USA.

Newey, W., West, K., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. Econometrica 55 (3), 703–08.

Pearl, J., 2009. Causality. Cambridge University Press, New York.

Penrose, M., 2003. Random Geometric Graphs. Oxford University Press, New York, USA.

Petty, C. M., 1971. Equilateral sets in minkowski spaces. Proc. Amer. Math. Soc. 29, 369–374.

Prakasa Rao, B. L. S., 2013. Conditional independence, conditional mixing and conditional association. Annals of the Institute of Statistical Mathematics 61, 441–460.

Rinott, Y., Rotar, V., 1996. A multivariate CLT for local dependence with $n^{1/2} \log n$ rate and applications to multivariate graph related statistics. Journal of Multivariate Analysis 56, 333–350.

Song, K., 2017. Measuring the graph concordance of locally dependent observations. Review of Economics and Statistics, Forthcoming.

Stein, C., 1972. A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. Proceedings in the Sixth Berkeley Symposium on Mathematical Statistics 2, 583–602.

Stein, C., 1986. Approximate computation of expectations. Lecture Notes-Monograph Series 7, i–164.

Wendland, H., 2004. Scattered Data Approximation. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press.

Appendix A.

A.1. Auxiliary Results for ψ-weak Dependent Processes

In this section, we present covariance inequalities for functions of general ψ-weak dependent processes. Let $\mathcal{F}_a$ and $\mathcal{G}_a$ be some classes of functions on $\mathbf{R}^{v\times a}$ with $v, a \geq 1$ and let $\mathcal{F} = \bigcup_{a\geq 1} \mathcal{F}_a$ and $\mathcal{G} = \bigcup_{a\geq 1} \mathcal{G}_a$.

Definition A.1. The triangular array $\{Y_{n,i}\}_{i\in N_n, n\geq 1}$, $Y_{n,i} \in \mathbf{R}^v$, is conditionally $(\mathcal{F}, \mathcal{G}, \psi)$-weakly dependent given $\{\mathcal{C}_n\}_{n\geq 1}$, if for each $n \in \mathbb{N}$, there exists a $\mathcal{C}_n$-measurable sequence $\theta_n \equiv \{\theta_{n,s}\}_{s=1}^{\infty}$ such that $\sup_{n\geq 1} \theta_{n,s} \to 0$ a.s. as $s \to \infty$, and a collection of nonrandom functions $(\psi_{a,b})_{a,b\in\mathbb{N}}$, $\psi_{a,b} : \mathcal{F}_a \times \mathcal{G}_b \to [0, \infty)$, such that for all $(A, B) \in \mathcal{P}_n(a, b; s)$ with $s > 0$ and all $f \in \mathcal{F}_a$ and $g \in \mathcal{G}_b$,

$$|\mathrm{Cov}\,(f(Y_{n,A}), g(Y_{n,B}) \mid \mathcal{C}_n)| \leq \psi_{a,b}(f, g)\theta_{n,s} \quad \text{a.s.} \tag{A.1}$$

Let $\{Y_{n,i}\}$ be conditionally $(\mathcal{F}, \mathcal{G}, \psi)$-weakly dependent given $\{\mathcal{C}_n\}$ with the weak dependence coefficients $\{\theta_{n,s}\}$ and $Y_{n,i} \in \mathbf{R}^v$. Also let $(A, B) \in \mathcal{P}_n(a, b; s)$ with $s > 0$. Consider measurable functions $f : \mathbf{R}^{v\times a} \to \mathbf{R}$ and $g : \mathbf{R}^{v\times b} \to \mathbf{R}$ such that $f \notin \mathcal{F}_a$ and $g \notin \mathcal{G}_b$, and define

$$\xi = f(Y_{n,A}) \quad \text{and} \quad \zeta = g(Y_{n,B}).$$

We proceed under Assumption 2.1(b) assuming the weak dependence coefficients are a.s. bounded by some constant $M \geq 1$. This assumption can be easily relaxed by truncating θ-s and noticing that $\theta = (\theta \vee 1)(\theta \wedge 1)$. In addition, we replace Assumption 2.1(a) with the following restriction:

Assumption A.1.
(a) $\mathcal{F}$ and $\mathcal{G}$ are stable under multiplication by constants, that is, if $f \in \mathcal{F}$, $g \in \mathcal{G}$ and $c \in \mathbf{R}$, then $cf \in \mathcal{F}$ and $cg \in \mathcal{G}$;
(b) If $f \in \mathcal{F}_a, g \in \mathcal{G}_b$ and $c_1, c_2 \in \mathbf{R}$, then $\psi_{a,b}(c_1 f, c_2 g) = |c_1 c_2| \cdot \psi_{a,b}(f, g)$.

In the following let $\mu_{\xi,p} = \|\xi\|_{\mathcal{C}_n,p}$ and $\mu_{\zeta,p} = \|\zeta\|_{\mathcal{C}_n,p}$, $p > 0$, and let φ_K with $K \in [0, \infty)$ denote the element-wise censoring function, i.e., for an indexed family of real numbers $\mathbf{x} \equiv (x_i)_{i\in I}$,

$$[\varphi_K(\mathbf{x})]_i = (-K) \vee (K \wedge x_i), \quad i \in I, \tag{A.2}$$

where $[A]_i$ denotes the i-th element of an indexed family A.

We first provide a result of a covariance inequality that permits the nonlinear transforms to be random functions. Suppose that Z_j, $j = 1, 2$ is a random element taking values in a separable metric space $(\mathcal{Z}_j, \rho_j)$ equipped with the Borel σ-algebra $\mathcal{B}(\mathcal{Z}_j)$ and

f and g are real-valued, measurable functions defined on $\mathbf{R}^{v\times a} \times \mathcal{Z}_1$ and $\mathbf{R}^{v\times b} \times \mathcal{Z}_2$, respectively. Let f^z be the z-section of f, i.e. $f^z(y) = f(y,z)$ (the z-section g^z of g is defined similarly) and note that if $f^{z_1} \in \mathcal{F}_a$ and $g^{z_2} \in \mathcal{G}_b$, then $\psi_{a,b}(f^{z_1}, g^{z_2})$ is well defined. In addition, let $\bar{f}(y) = \sup_{z\in\mathcal{Z}_1} |f(y,z)|$ and $\bar{g}(y) = \sup_{z\in\mathcal{Z}_2} |g(y,z)|$.

Lemma A.1. *Suppose that $f^{z_1} \in \mathcal{F}_a$ and $g^{z_2} \in \mathcal{G}_b$ for all $z_j \in \mathcal{Z}_j$, f and g are continuous in the second arguments, and the function $F(z_1,z_2) = \psi_{a,b}(f^{z_1}, g^{z_2})$ is continuous on $\mathcal{Z}_1 \times \mathcal{Z}_2$.*[8] *If Z_1 and Z_2 are $\mathcal{C}_n$-measurable and $\bar{f}, \bar{g} \in L^2$, then*

$$|\mathrm{Cov}\,(f(Y_{A,n}, Z_1), g(Y_{B,n}, Z_2) \mid \mathcal{C}_n)| \leq F(Z_1, Z_2)\theta_{n,s} \quad \textit{a.s.}$$

Proof. Suppose w.l.o.g. that $\mathbf{E}[f(Y_{A,n}, Z_1) \mid \mathcal{C}_n] = 0$ and $\mathbf{E}[g(Y_{B,n}, Z_2) \mid \mathcal{C}_n] = 0$ a.s. By Lemma 1.3 in Da Prato and Zabczyk (2014) we can approximate Z_j by a sequence of simple functions $\{Z_{j,m}\}$ s.t. $\rho_j(Z_{j,m}, Z_j) \searrow 0$ pointwise and for each $m \geq 1$, $Z_{j,m} = \sum_{k=1}^m z_{j,k}\mathbf{1}_{A_{j,k}}$, where $z_{j,k} \in \mathcal{Z}_j$, $A_{j,k} \in \mathcal{C}_n$ and $A_{j,k} \cap A_{j,l} = \emptyset$ for $k \neq l$. Then, letting $B_{k,l} = A_{1,k} \cap A_{2,l}$,

$$\begin{aligned}
|\mathbf{E}[f(Y_{A,n}, Z_{1,m})g(Y_{B,n}, Z_{2,m}) \mid \mathcal{C}_n]| &\leq \sum_{k,l=1}^m |\mathbf{E}[f(Y_{A,n}, z_{1,k})g(Y_{B,n}, z_{2,l}) \mid \mathcal{C}_n]|\, \mathbf{1}_{B_{k,l}} \\
&\leq \sum_{k,l=1}^m \psi_{a,b}(f^{z_{1,k}}, g^{z_{2,l}})\mathbf{1}_{B_{k,l}}\theta_{n,s} \qquad \because (2.2) \\
&= F(Z_{1,m}, Z_{2,m})\theta_{n,s} \quad \text{a.s.}
\end{aligned}$$

The second inequality above is due to (2.2). Consequently, the result follows by the conditional dominated convergence theorem. ■

The continuity requirement of the function $F(z_1,z_2)$ in Lemma A.1 can be relaxed by considering a continuous function $\tilde{F}$ such that for all $(z_1,z_2) \in \mathcal{Z}_1 \times \mathcal{Z}_2$, $F(z_1,z_2) \leq \tilde{F}(z_1,z_2)$. Consider, for example, the case when $h : \mathbf{R} \to \mathbf{R}$ is piece-wise linear and $f(x,z) = \varphi_z(h(x))$. If the ψ function depends on the Lipschitz coefficient of f^{z_1} as in Assumption 2.1, then the corresponding $F(z_1,z_2)$ is not continuous in z_1. It is clear, however, that the result of Lemma A.1 holds if we replace F with a smooth dominating function.

The following result establishes a bound for the conditional covariance between ξ and ζ given $\mathcal{C}_n$ in the case in which the censored functions $\varphi_K \circ f$ and $\varphi_L \circ g$, $K, L > 0$, belong to the classes $\mathcal{F}$ and $\mathcal{G}$, respectively. The result therefore does not require truncation of the

[8] Note that the continuity of F implies that it is Borel measurable. Moreover, if $\mathcal{Z}_j = \mathcal{Z}_{j,1} \times \mathcal{Z}_{j,2}$, $j = 1,2$, where each $\mathcal{Z}_{j,k}$ is a separable metric space, the supremum of F taken over $\mathcal{Z}_{1,1}$ and $\mathcal{Z}_{2,1}$ is also Borel measurable. The last observation is essential for other result presented in this section.

domains of the transformation functions. We apply the definition of ψ-weak dependence to the censored counterparts of f and g.

Theorem A.1. *Let $\{Y_{n,i}\}$, ξ, and ζ be as described above. Suppose that*

(i) $\mu_{\xi,p} < \infty$ *and* $\mu_{\zeta,q} < \infty$ *a.s. for some* $p, q > 1$ *with* $p^{-1} + q^{-1} < 1$*;*
(ii) $\varphi_K \circ f \in \mathcal{F}_a$ *and* $\varphi_K \circ g \in \mathcal{G}_b$ *for all* $K \in (0, \infty)$*;*
(iii) $(K, L) \mapsto \psi_{a,b}(\varphi_K \circ f, \varphi_L \circ g)$ *is continuous on* $(0, \infty)^2$.

Then

$$|\mathrm{Cov}(\xi, \zeta \mid \mathcal{C}_n)| \leq \left(M\bar{\psi}_{a,b}(\mu_{\xi,p}, \mu_{\zeta,q}) + 16\mu_{\xi,p}\mu_{\zeta,q}\right)\theta_{n,s}^{1-\frac{1}{p}-\frac{1}{q}} \quad \textit{a.s.}, \tag{A.3}$$

where for $z_1, z_2 \in (0, \infty)$*,*

$$\bar{\psi}_{a,b}(z_1, z_2; f, g) = \sup_{K,L \geq 1} (KL)^{-1}\psi_{a,b}(\varphi_{Kz_1} \circ f, \varphi_{Lz_2} \circ g).$$

It is not hard to check that under Assumption A.1 the bound in (A.3) preserves the scale-equivariance property because for any $c_1, c_2 \in \mathbf{R}$,

$$\bar{\psi}_{a,b}(c_1 z_1, c_2 z_2; c_1 f, c_2 g) = |c_1 c_2|\, \bar{\psi}_{a,b}(z_1, z_2; f, g).$$

Proof of Theorem A.1. Fix $\kappa, \lambda \geq 1$ and let $\Xi = \{(\mu_{\xi,p}, \mu_{\zeta,q}) \in (0, \infty)^2\}$. Next we define $\xi' = \mu_{\xi,p}^{-1}\xi\mathbf{1}_\Xi$,

$$\begin{aligned} \xi_\kappa &= (\varphi_\kappa \circ \mu_{\xi,p}^{-1} f)(Y_{A,n})\mathbf{1}_\Xi, & \xi_\kappa^* &= \xi_\kappa - \mathbf{E}[\xi_\kappa \mid \mathcal{C}_n], \\ \hat{\xi}_\kappa &= \xi' - \xi_\kappa, & \hat{\xi}_\kappa^* &= \hat{\xi}_\kappa - \mathbf{E}[\hat{\xi}_\kappa \mid \mathcal{C}_n], \end{aligned}$$

and, similarly, $\zeta', \zeta_\lambda, \zeta_\lambda^*, \hat{\zeta}_\lambda$, and $\hat{\zeta}_\lambda^*$, where we use g, $\mu_{\zeta,q}$, and λ instead of f, $\mu_{\xi,p}$, and κ. First,

$$\begin{aligned} |\mathrm{Cov}(\xi', \zeta' \mid \mathcal{C}_n)| &= |\mathbf{E}[(\xi_\kappa^* + \hat{\xi}_\kappa^*)(\zeta_\lambda^* + \hat{\zeta}_\lambda^*) \mid \mathcal{C}_n]| \\ &\leq |\mathbf{E}[\xi_\kappa^*\zeta_\lambda^* \mid \mathcal{C}_n]| + |\mathbf{E}([\xi_\kappa^*\hat{\zeta}_\lambda^* \mid \mathcal{C}_n]| \\ &\quad + |\mathbf{E}[\hat{\xi}_\kappa^*\zeta_\lambda^* \mid \mathcal{C}_n]| + |\mathbf{E}[\hat{\xi}_\kappa^*\hat{\zeta}_\lambda^* \mid \mathcal{C}_n]| \quad \text{a.s.} \end{aligned}$$

Consider each term in the last inequality separately. By Lemma A.1 and Assumption A.1 we find that[9]

$$\begin{aligned} |\mathbf{E}[\xi_\kappa^*\zeta_\lambda^* \mid \mathcal{C}_n]| &\leq \psi_{a,b}(\varphi_\kappa \circ \mu_{\xi,p}^{-1} f, \varphi_\lambda \circ \mu_{\zeta,q}^{-1} g)\theta_{n,s} \\ &\leq \frac{\kappa\lambda}{\mu_{\xi,p}\mu_{\zeta,q}}\bar{\psi}_{a,b}(\mu_{\xi,p}, \mu_{\zeta,q})\theta_{n,s} \quad \text{a.s. on } \Xi. \end{aligned}$$

[9] Note that for $x \geq 0$ and $z \neq 0$, $\varphi_x \circ z^{-1} f = z^{-1}(\varphi_{xz} \circ f)$.

As for the other terms, noticing that $|\xi^*_\kappa| \le 2\kappa$ a.s., we have

$$\begin{aligned}|\mathbf{E}[\xi^*_\kappa \hat{\zeta}^*_\lambda \mid \mathcal{C}_n]| &= |\mathrm{Cov}(\xi^*_\kappa, \hat{\zeta}^*_\lambda \mid \mathcal{C}_n)| = |\mathrm{Cov}(\xi^*_\kappa, \hat{\zeta}_\lambda \mid \mathcal{C}_n)| \\ &\le \mathbf{E}[|\xi^*_\kappa||\hat{\zeta}_\lambda| \mid \mathcal{C}_n] \le 2\kappa \mathbf{E}[|\hat{\zeta}_\lambda| \mid \mathcal{C}_n] \\ &\le 4\kappa\lambda^{1-q} \quad \text{a.s. on } \Xi\end{aligned}$$

because $\|\zeta'\|_{C_n,q} = \mathbf{1}_\Xi$ a.s. and

$$\begin{aligned}\mathbf{E}[|\hat{\zeta}_\lambda| \mid \mathcal{C}_n] &= \mathbf{E}[|\zeta' - \zeta_\lambda| \mathbf{1}\{\zeta' > \lambda\} \mid \mathcal{C}_n] \\ &\le \left(\mathbf{E}[|\zeta' - \zeta_\lambda|^q \mid \mathcal{C}_n]\right)^{1/q} \left(\mathbf{P}\left(\zeta' > \lambda \mid \mathcal{C}_n\right)\right)^{1-1/q} \\ &\le 2\,\|\zeta'\|_{\mathcal{C}_n,q} \left(\lambda^{-q}\mathbf{E}[|\zeta'|^q \mid \mathcal{C}_n]\right)^{1-1/q} \\ &= 2\lambda^{1-q} \quad \text{a.s. on } \Xi\end{aligned}$$

Similarly,

$$|\mathbf{E}[\hat{\xi}^*_\kappa \zeta^*_\lambda \mid \mathcal{C}_n]| \le 4\kappa^{1-p}\lambda \quad \text{a.s. on } \Xi$$

Finally,

$$\begin{aligned}|\mathbf{E}[\hat{\xi}^*_\kappa \hat{\zeta}^*_\lambda \mid \mathcal{C}_n]| &= |\mathrm{Cov}(\hat{\xi}^*_\kappa, \hat{\zeta}^*_\lambda \mid \mathcal{C}_n)| = |\mathrm{Cov}(\hat{\xi}_\kappa, \hat{\zeta}_\lambda \mid \mathcal{C}_n)| \\ &\le |\mathbf{E}[\hat{\xi}_\kappa \hat{\zeta}_\lambda \mid \mathcal{C}_n]| + \mathbf{E}[|\hat{\xi}_\kappa| \mid \mathcal{C}_n]\mathbf{E}[|\hat{\zeta}_\lambda| \mid \mathcal{C}_n] \\ &\le |\mathbf{E}[\hat{\xi}_\kappa \hat{\zeta}_\lambda \mid \mathcal{C}_n]| + 4\kappa^{1-p}\lambda^{1-q} \quad \text{a.s. on } \Xi,\end{aligned}$$

and for p', q' s.t. $1/p' + 1/q' = 1 - 1/p - 1/q$ we find that

$$\begin{aligned}|\mathbf{E}[\hat{\xi}_\kappa \hat{\zeta}_\lambda \mid \mathcal{C}_n]| &\le \mathbf{E}[|\hat{\xi}_\kappa \hat{\zeta}_\lambda| \mid \mathcal{C}_n] \\ &\le \left(\mathbf{E}[|\xi' - \xi_\kappa|^p \mid \mathcal{C}_n]\right)^{1/p} \left(\mathbf{P}\left(\xi' > \kappa \mid \mathcal{C}_n\right)\right)^{1/p'} \\ &\quad \times \left(\mathbf{E}[|\zeta' - \zeta_\lambda|^q \mid \mathcal{C}_n]\right)^{1/q} \left(\mathbf{P}\left(\zeta' > \lambda \mid \mathcal{C}_n\right)\right)^{1/q'} \\ &\le 4\kappa^{-p/p'}\lambda^{-q/q'} \quad \text{a.s. on } \Xi.\end{aligned}$$

Combining these inequalities and multiplying by $\mu_{\xi,p}\mu_{\zeta,q}$ we get a.s. on Ξ,

$$\begin{aligned}|\mathrm{Cov}(\xi, \zeta \mid \mathcal{C}_n)| &\le \bar{\psi}_{a,b}(\mu_{\xi,p}\mu_{\zeta,q})\kappa\lambda\theta_{n,s} + 4\mu_{\xi,p}\mu_{\zeta,q} \\ &\quad \times \left(\kappa\lambda^{1-q} + \kappa^{1-p}\lambda + \kappa^{-p/p'}\lambda^{-q/q'} + \kappa^{1-p}\lambda^{1-q}\right)\end{aligned} \tag{A.4}$$

Since (A.4) holds for all $\kappa, \lambda \ge 1$ a.s. on Ξ, it also holds for random κ and λ a.s. on $\Xi' = \Xi \cap \{(\kappa, \lambda) \in [1, \infty)^2\}$. Thus, setting $\kappa = (\theta_{n,s} \wedge 1)^{-1/p}$ and $\lambda = (\theta_{n,s} \wedge 1)^{-1/q}$ we get (A.3) on Ξ'. As for the set $\Xi \cap \Xi'^c$, note that $\mathrm{Cov}(\xi, \zeta \mid \mathcal{C}_n) = 0$ a.s. on $\{\theta_{n,s} = 0\}$. Similarly, $\mathrm{Cov}(\xi, \zeta \mid \mathcal{C}_n) = 0$ a.s. on $\{\mu_{\xi,p} = 0\} \cup \{\mu_{\zeta,q} = 0\}$, and $\{\mu_{\xi,p} = \infty\}$ and $\{\mu_{\zeta,q} = \infty\}$ are null sets. ■

Corollary A.1. *Suppose that the assumptions of Theorem A.1 hold. If*

$$\sup_{K,L\in(0,\infty)} (KL)^{-1}\psi_{a,b}(\varphi_K \circ f, \varphi_L \circ g) < \infty,$$

then

$$|\mathrm{Cov}(\xi,\zeta \mid \mathcal{C}_n)| \le C\mu_{\xi,p}\mu_{\zeta,q}\theta_{n,s}^{1-\frac{1}{p}-\frac{1}{q}} \quad \textit{a.s.},$$

where $C > 0$ is a constant.

The latter result applies trivially to the strong mixing processes and any measurable functions f and g satisfying relevant moment conditions because $\psi_{a,b}(f,g) = 4\|f\|_\infty \|g\|_\infty$. However, for some types of ψ-weak dependence Condition (ii) of Theorem A.1 may not be satisfied. Consider, for example, the case in which $\mathcal{F} = \mathcal{L}_v$ and $f(x,y) = xy$ with $x, y \in \mathbf{R}$. For any $K > 0$, the set $\{|f| \le K\}$ is unbounded so that $\varphi_K \circ f$ is not Lipschitz.[10] To handle such cases we use truncated domains in addition to censoring of transformation functions.

Theorem A.2. *Let $\{Y_{n,i}\}$, ξ, and ζ be as described above. Suppose that*

(i) *the functions f and g are continuous, and*

(ii) *$\mu_{\xi,p} < \infty$ and $\mu_{\zeta,q} < \infty$ a.s. for some $p, q > 1$ s.t. $p^{-1} + q^{-1} < 1$.*

Furthermore, there exist increasing continuous functions $h_1, h_2 : [0,\infty] \to [0,\infty]$ such that

(iii) *$\gamma_1 = \max_{i\in A}\max_{1\le k\le v}\|h_1^{-1}(|[Y_{n,i}]_k|)\|_{\mathcal{C}_n,p} < \infty$ a.s. and*
$\gamma_2 = \max_{i\in B}\max_{1\le k\le v}\|h_2^{-1}(|[Y_{n,i}]_k|)\|_{\mathcal{C}_n,q} < \infty$ a.s.

(iv) *$f_K = \varphi_{K_1} \circ f \circ \varphi_{h_1(K_2)} \in \mathcal{F}_a$ and $g_K = \varphi_{K_1} \circ g \circ \varphi_{h_2(K_2)} \in \mathcal{G}_b$ for all $K \in (0,\infty)^2$.*

(v) *$(K,L) \mapsto \psi_{a,b}(f_K, g_L)$ is continuous on $(0,\infty)^4$.*

Then

$$|\mathrm{Cov}(\xi,\zeta \mid \mathcal{C}_n)| \le \left(M\tilde{\psi}_{a,b}(\mu_{\xi,p},\mu_{\zeta,q},\gamma_1,\gamma_2) + 16(abv^2+1)\mu_{\xi,p}\mu_{\zeta,q}\right)\theta_{n,s}^{1-\frac{1}{p}-\frac{1}{q}} \quad \textit{a.s.}, \tag{A.5}$$

where for $(z_j, w_j) \in (0,\infty)^2$, $j = 1, 2$,

$$\tilde{\psi}_{a,b}(z_1,z_2,w_1,w_2; f,g) = \sup_{K,L\ge 1}(KL)^{-1}\psi_{a,b}(f_{(Kz_1,Kw_1)}, g_{(Lz_2,Lw_2)}).$$

It can be seen from the proof that when $\varphi_K \circ f \in \mathcal{F}_a$ for all $K > 0$ and g satisfies the conditions of Theorem A.2, there is no need to truncate the domain of f. In such a case we do not require the continuity of f and the covariance inequality becomes

$$|\mathrm{Cov}(\xi,\zeta \mid \mathcal{C}_n)| \le \left(M\tilde{\psi}_{a,b}(\mu_{\xi,p},\mu_{\zeta,q},0,\gamma_2) + 4(bv+4)\mu_{\xi,p}\mu_{\zeta,q}\right)\theta_{n,s}^{1-\frac{1}{p}-\frac{1}{q}} \quad \text{a.s.},$$

[10] Since $\partial(xy)/\partial x = y$, one can choose $x = 0$ so that the function is bounded by any $K > 0$, but the partial derivative is unbounded.

where $h_1 \equiv \infty$. Similarly, if both $\varphi_K \circ f \in \mathcal{F}_a$ and $\varphi_K \circ g \in \mathcal{G}_b$ for all $K > 0$, we are back to the result of Theorem A.1.

Condition (iii) is a moment condition on the original process, where the required moments are defined through the functions h_1 and h_2. In the special case in which Assumption 2.1 holds (i.e. $\mathcal{F} = \mathcal{G} = \mathcal{L}_v$ and the ψ functions are of a certain form) and f and g are the product functions on $\mathbf{R}^{1\times a}$ and $\mathbf{R}^{1\times b}$, respectively, i.e.

$$f(Y_{A,n}) = \prod_{i\in A} Y_{n,i} \quad \text{and} \quad g(Y_{B,n}) = \prod_{i\in B} Y_{n,i},$$

it suffices to choose $h_1(x) = x^{\frac{1}{a-1}}$ and $h_2(x) = x^{\frac{1}{b-1}}$ in order to guarantee that $\tilde{\psi}_{a,b}$ is finite valued. Indeed, with this choice of functions h_1 and h_2 it is not hard to see that $\mathrm{Lip}(f_{(K_1,K_2)})$ and $\mathrm{Lip}(g_{(K_1,K_2)})$ are bounded by K_2.

Corollary A.2. *Let $\{Y_{n,i}\}$ be an array of random variables satisfying Assumption 2.1, $\xi = \prod_{i\in A} Y_{n,i}$, and $\zeta = \prod_{i\in B} Y_{n,i}$. Let $\{p_i : i \in A\}$ and $\{q_i : i \in B\}$ be collections of positive reals such that $p^{-1} + q^{-1} < 1$, where $p = \left(\sum_{i\in A} 1/p_i\right)^{-1}$ and $q = \left(\sum_{i\in B} 1/q_i\right)^{-1}$. Suppose that $\|Y_{n,i}\|_{\mathcal{C}_n,p^*}, \|Y_{n,j}\|_{\mathcal{C}_n,q^*} < \infty$ a.s. for $p^* = \max_{i\in A} p_i$, $q^* = \max_{i\in B} q_i$ and all $i \in A$, $j \in B$. Then*

$$|\mathrm{Cov}(\xi, \zeta \mid \mathcal{C}_n)| \leq M\Pi_{a,b}(\pi_1, \pi_2, \tilde{\gamma}_1, \tilde{\gamma}_2)\theta_{n,s}^{1-\frac{1}{p}-\frac{1}{q}} \quad \textit{a.s.},$$

where

$$\pi_1 = \prod_{i\in A} \|Y_{n,i}\|_{\mathcal{C}_n,p_i}, \quad \tilde{\gamma}_1 = \max_{i\in A} \left\|Y_{n,i}^{a-1}\right\|_{\mathcal{C}_n,p}, \quad \textit{and}$$

$$\pi_2 = \prod_{i\in B} \|Y_{n,i}\|_{\mathcal{C}_n,q_i}, \quad \tilde{\gamma}_2 = \max_{i\in B} \left\|Y_{n,i}^{b-1}\right\|_{\mathcal{C}_n,q},$$

and for $c_1' \equiv c_1 + 16(ab+1)$,

$$\Pi_{a,b}(z_1, z_2, w_1, w_2) = c_1' z_1 z_2 + c_2 z_1 w_2 + c_3 w_1 z_2 + c_4 w_1 w_2.$$

Proof of Theorem A.2. We reuse the notation and bounds established in the proof of Theorem A.1. In addition, let

$$\xi_{\kappa\kappa} = \mu_{\xi,p}^{-1} f_{(\kappa\mu_{\xi,p},\kappa\gamma_1)}(Y_{A,n})\mathbf{1}_{\Xi}, \quad \xi_{\kappa\kappa}^* = \xi_{\kappa\kappa} - \mathbf{E}[\xi_{\kappa\kappa} \mid \mathcal{C}_n],$$
$$\hat{\xi}_{\kappa\kappa} = \xi_\kappa - \xi_{\kappa\kappa}, \qquad \hat{\xi}_{\kappa\kappa}^* = \hat{\xi}_{\kappa\kappa} - \mathbf{E}[\hat{\xi}_{\kappa\kappa} \mid \mathcal{C}_n],$$

and, similarly, $\zeta_{\lambda\lambda}$, $\zeta_{\lambda\lambda}^*$, $\hat{\zeta}_{\lambda\lambda}$, and $\hat{\zeta}_{\lambda\lambda}^*$, where f, $\mu_{\xi,p}$, and γ_1 are replaced by g, $\mu_{\zeta,q}$, and γ_2, respectively. Then

$$|\mathbf{E}[\xi_\kappa^*\zeta_\lambda^* \mid \mathcal{C}_n]| \leq |\mathbf{E}[\xi_{\kappa\kappa}^*\zeta_{\lambda\lambda}^* \mid \mathcal{C}_n]| + |\mathbf{E}[\xi_{\kappa\kappa}^*\hat{\zeta}_{\lambda\lambda}^* \mid \mathcal{C}_n]|$$
$$+ |\mathbf{E}[\hat{\xi}_{\kappa\kappa}^*\zeta_{\lambda\lambda}^* \mid \mathcal{C}_n]| + |\mathbf{E}[\hat{\xi}_{\kappa\kappa}^*\hat{\zeta}_{\lambda\lambda}^* \mid \mathcal{C}_n]| \quad \text{a.s.}$$

Let Ξ be as in the proof of Theorem A.1. Letting $\tilde{\Xi} = \Xi \cap \{(\gamma_1, \gamma_2) \in (0, \infty)^2\}$, by Lemma A.1 and Assumption A.1 we find that

$$|\mathbf{E}[\xi^*_{\kappa\kappa}\zeta^*_{\lambda\lambda} \mid \mathcal{C}_n]| \leq \frac{\kappa\lambda}{\mu_{\xi,p}\mu_{\zeta,q}}\tilde{\psi}_{a,b}(\mu_{\xi,p}, \mu_{\zeta,q}, \gamma_1, \gamma_2)\theta_{n,s} \quad \text{a.s. on } \tilde{\Xi}.$$

Second, noticing that $\{|\hat{\zeta}_{\lambda\lambda}| > 0\} \subseteq \bigcup_{i \in B} \bigcup_{1 \leq k \leq v} \{|[Y_{n,i}]_k| > h_2(\lambda\gamma_2)\}$ ($\because \zeta_\lambda \neq \zeta_{\lambda\lambda}$ only if $Y_{n,B} \neq \varphi_{h_2(\lambda\gamma_2)}(Y_{n,B})$),

$$\begin{aligned}
|\mathbf{E}[\xi^*_{\kappa\kappa}\hat{\zeta}^*_{\lambda\lambda} \mid \mathcal{C}_n]| &= |\mathbf{E}[\xi^*_{\kappa\kappa}\hat{\zeta}_{\lambda\lambda} \mid \mathcal{C}_n]| \leq 2\kappa\mathbf{E}[|\hat{\zeta}_{\lambda\lambda}| \mid \mathcal{C}_n] \\
&\leq 4\kappa\lambda \sum_{i \in B} \sum_{1 \leq k \leq v} \mathbf{P}\left(|[Y_{n,i}]_k| > h_2(\lambda\gamma_2) \mid \mathcal{C}_n\right) \\
&\leq 4bv \cdot \kappa\lambda^{1-q} \quad \text{a.s. on } \tilde{\Xi},
\end{aligned}$$

where the third line follows by the conditional Markov inequality on $\{\gamma_2 \in (0, \infty)\}$ and $|[Y_{n,i}]_k| = 0$ on $\{\gamma_2 = 0\}$ for all $i \in B$ and $1 \leq k \leq v$. Similarly,

$$|\mathbf{E}[\hat{\xi}^*_{\kappa\kappa}\zeta^*_{\lambda\lambda} \mid \mathcal{C}_n]| \leq 4av \cdot \kappa^{1-p}\lambda \quad \text{a.s. on } \tilde{\Xi},$$

and for p', q' s.t. $1/p' + 1/q' = 1 - 1/p - 1/q$,

$$|\mathbf{E}[\hat{\xi}^*_{\kappa\kappa}\hat{\zeta}^*_{\lambda\lambda} \mid \mathcal{C}_n]| \leq 4abv^2 \left(\kappa^{-p/p'}\lambda^{-q/q'} + \kappa^{1-p}\lambda^{1-q}\right) \quad \text{a.s. on } \tilde{\Xi}.$$

Finally, the result follows by choosing $\kappa = (\theta_{n,s} \wedge 1)^{-1/p}$ and $\lambda = (\theta_{n,s} \wedge 1)^{-1/q}$ and using inequality (A.4) established in the proof of Theorem A.1. ∎

A.2. Limit Theorems for ψ-weak Dependent Processes

Proof of Proposition 2.3. Let $\xi = f(Y_{n,A})$ and $\zeta = g(Y_{n,B})$ and

$$\xi^{(s)} = f(Y^{(s)}_{n,i} : i \in A), \quad \text{and} \quad \zeta^{(s)} = g(Y^{(s)}_{n,i} : i \in B).$$

Then, since $\xi^{(s)}$ and $\zeta^{(s)}$ are conditionally independent given $\mathcal{C}_n$, we find that

$$\begin{aligned}
|\mathrm{Cov}(\xi, \zeta \mid \mathcal{C}_n)| &\leq |\mathrm{Cov}((\xi - \xi^{(s)}), \zeta \mid \mathcal{C}_n)| + |\mathrm{Cov}(\xi^{(s)}, (\zeta - \zeta^{(s)}) \mid \mathcal{C}_n)| \\
&\leq 2\left\|g\right\|_\infty \mathbf{E}[|\xi - \xi^{(s)}| \mid \mathcal{C}_n] + 2\left\|f\right\|_\infty \mathbf{E}[|\zeta - \zeta^{(s)}| \mid \mathcal{C}_n] \\
&\leq 2\left\|g\right\|_\infty \mathrm{Lip}(f) \sum_{i \in A} \mathbf{E}[\|Y_{n,i} - Y^{(s)}_{n,i}\| \mid \mathcal{C}_n] \\
&\quad + 2\left\|f\right\|_\infty \mathrm{Lip}(g) \sum_{i \in B} \mathbf{E}[\|Y_{n,i} - Y^{(s)}_{n,i}\| \mid \mathcal{C}_n] \\
&\leq \left(a\left\|g\right\|_\infty \mathrm{Lip}(f) + b\left\|f\right\|_\infty \mathrm{Lip}(g)\right) \times \theta_{n,s} \quad \text{a.s.}
\end{aligned}$$

∎

Proof of Theorem 3.1. We follow the approach of Jenish and Prucha (2009), see the proof of Theorem 3 therein. However, instead of truncation used in Jenish and Prucha, we rely on censoring functions $\varphi_k(x)$ defined in (A.2) in order to be able to use the notion of ψ-weak dependence.[11] Consider a censored version of $Y_{n,i}$: for some $k > 0$, let

$$\begin{aligned} Y_{n,i} &= Y_{n,i}^{(k)} + \tilde{Y}_{n,i}^{(k)}, \text{ where} \\ Y_{n,i}^{(k)} &= \varphi_k(Y_{n,i}), \\ \tilde{Y}_{n,i}^{(k)} &= Y_{n,i} - \varphi_k(Y_{n,i}) = (Y_{n,i} - \operatorname{sgn}(Y_{n,i})k)1\left(|Y_{n,i}| > k\right). \end{aligned}$$

We have:

$$\begin{aligned} \mathbf{E}\left|\frac{1}{n}\sum_{i=1}^{n}(Y_{n,i} - \mathbf{E}[Y_{n,i} \mid \mathcal{C}_n])\right| &\leq \mathbf{E}\left|\frac{1}{n}\sum_{i=1}^{n}\left(Y_{n,i}^{(k)} - \mathbf{E}[Y_{n,i}^{(k)} \mid \mathcal{C}_n]\right)\right| \\ &\quad + \mathbf{E}\left|\frac{1}{n}\sum_{i=1}^{n}\left(\tilde{Y}_{n,i}^{(k)} - \mathbf{E}[\tilde{Y}_{n,i}^{(k)} \mid \mathcal{C}_n]\right)\right|. \end{aligned} \tag{A.6}$$

Since $\tilde{Y}_{n,i}^{(k)} = 0$ when $|Y_{i,k}| \leq k$, $\mathbf{E}|\tilde{Y}_{n,i}^{(k)}| = \mathbf{E}[|\tilde{Y}_{n,i}^{(k)}|1\left(|Y_{n,i}| > k\right)] \leq 2\mathbf{E}[|Y_{n,i}|1\{|Y_{n,i}| > k\}|]$. Hence, using the triangle inequality, the second term on the right-hand side of (A.6) can be bounded by $2\sup_{i\in N_n}\mathbf{E}|\tilde{Y}_{n,i}^{(k)}| = 4\sup_{i\in N_n}\mathbf{E}[|Y_{n,i}|1\{|Y_{n,i}| > k\}|]$. By Assumption 3.1, for each $\varepsilon > 0$ one now can find k such that the second term on the right-hand side of (A.6) is smaller than $\varepsilon/2$ for all n large.

It remains to show that for the same k and all n large,

$$\mathbf{E}\left|\frac{1}{n}\sum_{i=1}^{n}\left(Y_{n,i}^{(k)} - \mathbf{E}[Y_{n,i}^{(k)} \mid \mathcal{C}_n]\right)\right| < \epsilon/2. \tag{A.7}$$

By the norm inequality,

$$\mathbf{E}\left|\frac{1}{n}\sum_{i=1}^{n}\left(Y_{n,i}^{(k)} - \mathbf{E}[Y_{n,i}^{(k)} \mid \mathcal{C}_n]\right)\right| \leq \frac{\sigma_{n,k}}{n}, \tag{A.8}$$

where

$$\begin{aligned} \sigma_{n,k}^2 &= \mathbf{E}\left[\sum_{i=1}^{n}\left(Y_{n,i}^{(k)} - \mathbf{E}[Y_{n,i}^{(k)} \mid \mathcal{C}_n]\right)\right]^2 \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{E}\left[(Y_{n,i}^{(k)} - \mathbf{E}[Y_{n,i}^{(k)} \mid \mathcal{C}_n])(Y_{j,n}^{(k)} - \mathbf{E}[Y_{j,n}^{(k)} \mid \mathcal{C}_n])\right] \end{aligned}$$

[11]Unlike discontinuous truncation functions $x \cdot 1(|x| \leq k)$, censoring functions $\varphi_k(x)$ are continuous and have a finite Lipschitz constant: $\operatorname{Lip}(\varphi_k) = 1$.

$$\leq 4nk^2 + \sum_{i=1}^{n} \sum_{s=1}^{n-1} \sum_{j:d_n(i,j)=s} \mathbf{E} \left| \mathrm{Cov}(Y_{n,i}^{(k)}, Y_{j,n}^{(k)} \mid \mathcal{C}_n) \right|.$$

In view of Definition 2.2, we have for $d_n(i,j) = s$,

$$\left| \mathrm{Cov}(Y_{n,i}^{(k)}, Y_{j,n}^{(k)} \mid \mathcal{C}_n) \right| \leq \psi_{1,1}(\varphi_k, \varphi_k) \cdot \theta_{n,m} \quad \text{a.s.} \tag{A.9}$$

Using the definitions of $N_n^{\partial}(i; s)$ and $\delta_n^{\partial}(s, 1)$, we obtain:

$$\begin{aligned} \sigma_{n,k}^2 &\leq 4nk^2 + \psi_{1,1}(\varphi_k, \varphi_k) \sum_{s=1}^{n-1} \mathbf{E}[\theta_{n,s}] \sum_{i=1}^{n} |N_n^{\partial}(i; s)| \\ &= n \left(4k^2 + \psi_{1,1}(\varphi_k, \varphi_k) \sum_{s=1}^{n-1} \delta_n^{\partial}(s, 1) \left\| \theta_{n,s} \right\|_1 \right). \end{aligned} \tag{A.10}$$

By (A.8) and (A.10),

$$\mathbf{E} \left| \frac{1}{n} \sum_{i=1}^{n} \left(Y_{n,i}^{(k)} - \mathbf{E}[Y_{n,i}^{(k)} \mid C_n] \right) \right| \leq \left(\frac{4k^2}{n} + \psi_{1,1}(\varphi_k, \varphi_k) \frac{1}{n} \sum_{s=1}^{n-1} \delta_n^{\partial}(s; 1) \left\| \theta_{n,s} \right\|_1 \right)^{1/2}.$$

The result in (A.7) now follows by Assumption 3.2 for all n sufficiently large. ■

For each $p \geq 1$,

$$\mu_{n,p} = \max_{i \in N_n} \left(\mathbf{E}[|Y_{n,i}|^p \mid \mathcal{C}_n] \right)^{1/p}.$$

The following lemma is used for the central limit theorem.

Lemma A.2. *Suppose that the conditions of Theorem 3.2 are satisfied for $\{Y_{n,i}\}$. Let $g : \mathbf{R} \to \mathbf{R}$ be a twice continuously differentiable bounded function with bounded derivatives.*

Then there exists a constant $C > 0$ such that for any positive integer m_n and any $n \geq 1$,

$$|\mathbf{E}[g'(S_n) - S_n g(S_n) \mid \mathcal{C}_n]| \leq \bar{\Delta}_n(g) + \frac{Cn\mu_{n,3}^3 \|g''\|_\infty \delta_n(m_n; 2)}{\sigma_n^3},$$

where

$$\begin{aligned} \bar{\Delta}_n(g) = &\frac{C\|g'\|_\infty}{c_n^2} \sqrt{(\mu_{n,p}^2 + \mu_{n,p}^4) \sum_{m=0}^{n-2} |H_n(m; m_n)| \theta_{n,m}^{1-4/p}} \\ &+ \frac{Cn(n - \delta_n(m_n))(\mu_{n,p} + 1)}{c_n} \left\{ \|g\|_\infty + \frac{\|g'\|_\infty}{c_n} \right\} \theta_{n,m_n}^{1-1/p}, \end{aligned}$$

and $H_n(\cdot, \cdot)$ is as defined in Assumption 3.4.

Proof. We set an increasing sequence of positive integers m_n, and define for each $i \in N_n$,

$$\tilde{S}_{n,i} = \sum_{j \in N_n \setminus N_n(i;m_n)} \tilde{Y}_{n,j},$$

where $\tilde{Y}_{n,i} = Y_{n,i}/\sigma_n$. Also, let

$$\tilde{S}_n = \sum_{j \in N_n} \tilde{Y}_{n,j}.$$

We write

$$g'(\tilde{S}_n) - \tilde{S}_n g(\tilde{S}_n) = A_{n,1} + A_{n,2} + A_{n,3},$$

where

$$\begin{aligned}
A_{n,1} &= g'(\tilde{S}_n)\left(1 - \sum_{i \in N_n} \tilde{Y}_{n,i}(\tilde{S}_n - \tilde{S}_{n,i})\right) \\
A_{n,2} &= \sum_{i \in N_n} \tilde{Y}_{n,i}\left(g'(\tilde{S}_n)(\tilde{S}_n - \tilde{S}_{n,i}) - (g(\tilde{S}_n) - g(\tilde{S}_{n,i}))\right), \text{ and} \\
A_{n,3} &= \sum_{i \in N_n} \tilde{Y}_{n,i} g(\tilde{S}_{n,i}).
\end{aligned}$$

Let us turn to $A_{n,2}$. Applying Taylor expansion,

$$|\mathbf{E}[A_{n,2} \mid \mathcal{C}_n]| \le \frac{\|g''\|_\infty}{2} \sum_{i \in N_n} \mathbf{E}[|\tilde{Y}_{n,i}|(\tilde{S}_{n,i} - \tilde{S}_n)^2 \mid \mathcal{C}_n].$$

The last bound is bounded by

$$\begin{aligned}
&\frac{\|g''\|_\infty}{2} \sum_{i \in N_n} \sum_{j \in N_n(i;m_n)} \sum_{k \in N_n(i;m_n)} \mathbf{E}[|\tilde{Y}_{n,i}\tilde{Y}_{n,j}\tilde{Y}_{n,k}| \mid \mathcal{C}_n] \\
&\quad \le \frac{\|g''\|_\infty}{2} \sum_{i \in N_n} \sum_{j \in N_n(i;m_n)} \sum_{k \in N_n(i;m_n)} \max_{i \in N_n} \mathbf{E}[|\tilde{Y}_{n,i}|^3 \mid \mathcal{C}_n],
\end{aligned}$$

by the arithmetic-geometric mean inequality. Thus, it follows that

$$\begin{aligned}
|\mathbf{E}[A_{n,2} \mid \mathcal{C}_n]| &\le \frac{\|g''\|_\infty}{2} \sum_{i \in N_n} |N_n(i;m_n)|^2 \frac{\max_{i \in N_n} \mathbf{E}[|Y_{n,i}|^3 \mid \mathcal{C}_n]}{\sigma_n^3} \\
&\le \frac{n\|g''\|_\infty \delta_n(m_n;2)\mu_{n,3}^3}{2\sigma_n^3}.
\end{aligned}$$

Let us now turn to $A_{n,1}$. Write

$$|\mathbf{E}[A_{n,1} \mid \mathcal{C}_n]| = \mathbf{E}\left[g'(\tilde{S}_n)\left(1 - \sum_{i \in N_n} \sum_{j \in N_n(i;m_n)} \tilde{Y}_{n,i}\tilde{Y}_{n,j}\right) \mid \mathcal{C}_n\right].$$

Since, by the definition of σ_n,

$$1 = \mathbf{E}\left[\sum_{i\in N_n}\sum_{j\in N_n}\tilde{Y}_{n,i}\tilde{Y}_{n,j} \mid \mathcal{C}_n\right],$$

we rewrite

$$\mathbf{E}[A_{n,1} \mid \mathcal{C}_n] = -\mathbf{E}\left[g'(\tilde{S}_n)\left(\sum_{i\in N_n}\sum_{j\in N_n(i;m_n)}(\tilde{Y}_{n,i}\tilde{Y}_{n,j} - \mathbf{E}[\tilde{Y}_{n,i}\tilde{Y}_{n,j} \mid \mathcal{C}_n])\right) \mid \mathcal{C}_n\right]$$
$$+ \mathbf{E}\left[g'(\tilde{S}_n) \mid \mathcal{C}_n\right]\left(\sum_{i\in N_n}\sum_{j\in N_n\setminus N_n(i;m_n)}\mathbf{E}[\tilde{Y}_{n,i}\tilde{Y}_{n,j} \mid \mathcal{C}_n]\right) = -R_{n,1} + R_{n,2}, \text{ say.}$$

Using Cauchy-Schwarz inequality and letting $Z_{n,ij} = c_n^2(\tilde{Y}_{n,i}\tilde{Y}_{n,j} - \mathbf{E}[\tilde{Y}_{n,i}\tilde{Y}_{n,j} \mid \mathcal{C}_n])$, we bound $|R_{n,1}|$ by

$$\frac{1}{c_n^2}\sqrt{\mathbf{E}[(g'(\tilde{S}_n))^2 \mid \mathcal{C}_n]}\sqrt{\mathbf{E}\left[\left(\sum_{i\in N_n}\sum_{j\in N_n(i;m_n)}Z_{n,ij}\right)^2 \mid \mathcal{C}_n\right]}.$$

Let us write the last conditional expectation in the preceding display as

$$\sum_{i\in N_n}\sum_{j\in N_n(i;m_n)}\sum_{k\in N_n}\sum_{l\in N_n(k;m_n)}\mathbf{E}[Z_{n,ij}Z_{n,kl} \mid \mathcal{C}_n]$$
$$= \sum_{s=0}^{n-2}\sum_{(i,j,k,l)\in H_n(s;m_n)}\mathbf{E}[Z_{n,ij}Z_{n,kl} \mid \mathcal{C}_n]$$
$$\le M(\mu_{n,p}^2 + \mu_{n,p}^4)\sum_{s=0}^{n-2}|H_n(s;m_n)|\theta_{n,m}^{1-4/p},$$

by Corollary A.2. This gives the following bound for $|R_{n,1}|$:

$$\frac{C\|g'\|_\infty}{c_n^2}\sqrt{(\mu_{n,p}^2 + \mu_{n,p}^4)\sum_{s=0}^{n-2}|H_n(s;m_n)|\theta_{n,s}^{1-4/p}}.$$

Let us turn to $R_{n,2}$. We bound $|R_{n,2}|$ by

$$\frac{Cn(n-\delta_n(m_n))\mu_{n,p}^2\|g'\|_\infty\theta_{n,m_n}^{1-4/p}}{\sigma_n^2}, \tag{A.11}$$

using Corollary A.1. Finally, let us consider $A_{n,3}$. Note that

$$\mathbf{E}[A_{n,3}\mathcal{C}_n] \le \sum_{i\in N_n}\frac{1}{c_n}\left|\mathrm{Cov}\left(\frac{Y_{n,i}c_n}{\sigma_n}, h_n\left(\frac{S_{n,i}c_n}{\sigma_n}\right) \mid \mathcal{C}_n\right)\right|,$$

where $S_{n,i} = \tilde{S}_{n,i}\sigma_n$, $h_n(x) = g(x/c_n)$ and c_n is a sequence of constants in Assumption 3.4. By Lemma 2.1, $\{Y_{n,i}c_n/\sigma_n\}$ is conditionally ψ-weakly dependent with the coefficients $\{\theta_{n,s}\}$. Thus, we use Theorem A.1 to bound the last term by

$$\left(M \sup_{K,L \geq 1} \frac{1}{KL} \psi_{a,b}(\varphi_{K\mu_{n,p}c_n/\sigma_n} \circ f_n, \varphi_{L\|g\|_\infty} \circ h_n) + \frac{16\mu_{n,p}c_n\|g\|_\infty}{\sigma_n} \right) \theta_{n,m_n}^{1-1/p-1/q}$$
$$\leq C(n - \delta_n(m_n)) \left(\frac{\mu_{n,p}c_n}{\sigma_n} + 1 \right) \left\{ \|g\|_\infty + \frac{\|g'\|_\infty}{c_n} \right\} \theta_{n,m_n}^{1-1/p-1/q},$$

where q is such that $p^{-1} + q^{-1} < 1$ and $f_n(x) = xc_n/\sigma_n$. Thus by taking $q \to \infty$, we conclude that

$$|\mathbf{E}[A_{n,3} \mid \mathcal{C}_n]| \leq \frac{Cn(n - \delta_n(m_n))(\mu_{n,p} + 1)}{c_n} \left\{ \|g\|_\infty + \frac{\|g'\|_\infty}{c_n} \right\} \theta_{n,m_n}^{1-1/p} \quad \text{a.s.}$$

Since $c_n \leq \sigma_n$, we subsuming the bound in (A.11) into this bound, and obtain the desired result of the lemma. ∎

Lemma A.3. *Suppose that Assumption 2.1 holds, and that* $\mathbf{E}[Y_{n,i} \mid \mathcal{C}_n] = 0$ *a.s.*
Then, there exists $C > 0$ *such that for all* $n \geq 1$,

$$\sup_{t \in \mathbf{R}} \left| P\left\{ \frac{1}{\sigma_n} \sum_{i \in N_n} Y_{n,i} \leq t \mid \mathcal{C}_n \right\} - \Phi(t) \right|$$
$$\leq C \left(\frac{\mu_{n,3}}{\sigma_n} \right)^{3/2} \sqrt{n\delta_n(m_n; 2)} + \frac{C}{c_n^2} \sqrt{(\mu_{n,p}^2 + \mu_{n,p}^4) \sum_{s=0}^{n-2} |H_n(s; m_n)| \theta_{n,s}^{1-4/p}}$$
$$+ \frac{Cn(n - \delta_n(m_n))(\mu_{n,p} + 1)}{c_n} \left\{ 1 + \frac{1}{c_n} \right\} \theta_{n,m_n}^{1-1/p} \quad a.s.,$$

where Φ *denotes the distribution function of* $N(0, 1)$.

Proof. The proof is an adaptation of the proof of Theorem 2.4 of Penrose (2003) to our set-up. Let $\bar{\Delta}_n(g)$ be as defined in Lemma A.2. Let us define $h_+(x) = 1$ for $x \leq t$ and $h_+(x) = 0$ and for $x \geq t + \varepsilon$ and h_+ is continuous and linear on $[x, x + \varepsilon]$. Similarly, we also take $h_-(x) = 1$ for $x \leq t - \varepsilon$ and $h_-(x) = 0$ and for $x \geq t$ and h_- is continuous and linear on $[x - \varepsilon, x]$. Define for any real function g,

$$\Delta_n(g) = |\mathbf{E}[g(S_n) \mid \mathcal{C}_n] - \mathbf{E}[g(Z)]|.$$

Let us find a bound for $\Delta_n(h_+)$ and $\Delta_n(h_-)$. First, note that by Stein's Lemma (e.g. Chen et al., 2011, p. 15)

$$|\mathbf{E}[g'(S_n) - S_n g(S_n) \mid \mathcal{C}_n]| = \Delta_n(h), \tag{A.12}$$

where

$$g(x) = e^{x^2/2} \int_{-\infty}^{x} (h(w) - \mathbf{E}[h(Z)])e^{-w^2/2} dw.$$

Since for $h = h_+$ or $h = h_-$, (see Lemma 2.4 of Chen et al., 2011)

$$\begin{aligned} &\|g\|_\infty \le \sqrt{\pi/2}\, \|h - \mathbf{E}[h(Z)]\|_\infty \le \sqrt{\pi/2}, \\ &\|g'\|_\infty \le 2\, \|h - \mathbf{E}[h(Z)]\|_\infty \le 2, \quad \text{and} \\ &\|g''\|_\infty \le 2\, \|h'\|_\infty \le 2/\varepsilon, \end{aligned} \tag{A.13}$$

we apply Lemma A.2 to (A.12) to deduce that for $h = h_+$ or $h = h_-$,

$$\Delta_n(h) \le \bar{\Delta}_n(g) + \frac{C}{\varepsilon} \frac{n\delta_n(m_n; 2)\mu_{n,3}^3}{\sigma_n^3}.$$

Let us now bound

$$\begin{aligned} P\{S_n \le t \mid \mathcal{C}_n\} \le \mathbf{E}[h_+(S_n) \mid \mathcal{C}_n] &\le \mathbf{E}[h_+(Z)] + \Delta_n(h_+) \\ &\le P\{Z \le t + \varepsilon\} + \Delta_n(h_+) \\ &\le P\{Z \le t\} + \phi(0)\varepsilon + \Delta_n(h_+), \end{aligned}$$

where ϕ is the density of $N(0, 1)$. Similarly, we also bound

$$P\{S_n \le t \mid \mathcal{C}_n\} \ge P\{Z \le t\} - \phi(0)\varepsilon - \Delta_n(h_-).$$

Hence, we have

$$|P\{S_n \le t \mid \mathcal{C}_n\} - P\{Z \le t\}| \le 2\phi(0)\varepsilon + \frac{C}{\varepsilon} \frac{n\delta_n(m_n; 2)\mu_{n,3}^3}{\sigma_n^3} + \bar{\Delta}_n(g).$$

Choose

$$\varepsilon = \left(\frac{Cn\delta_n(m_n; 2)\mu_{n,3}^3}{2\phi(0)\sigma_n^3} \right)^{1/2}.$$

Applying the bounds in (A.13) to $\bar{\Delta}_n(g)$, we obtain the desired result. ■

Proof of Theorem 3.2. The desired result follows from Lemma A.3 in combination with the conditions given in the theorem. Details are omitted. ■

Proof of Proposition 4.1. For the first implication it suffices to show that for any vector $c \in \mathbf{R}^v$ with $\|c\| = 1$, $\mathbf{E}[|A_n(c)| \mid \mathcal{C}_n] \to 0$ a.s., where $A_n(c) = c^\top(\hat{V}_n - V_n)c$. Let $y_{n,i} = c^\top Y_{n,i}$ and notice that $\{y_{n,i}\}$ is $(\mathcal{L}_1, \phi)$-weakly dependent with the weak dependence coefficients $\{\theta_{n,s}\}$. In addition, $\mathbf{E}[y_{n,i} \mid \mathcal{C}_n] = 0$ a.s. and by Assumption 4.1(i)

$$\sup_n \sup_{i \in N_n} \|y_{n,i}\|_{\mathcal{C}_n, 2r} \le \mu < \infty \quad \text{a.s.}$$

Then

$$
\begin{aligned}
A_n(c) &= \frac{1}{n}\sum_{i\in N_n}\left(y_{n,i}^2 - \mathbf{E}[y_{n,i}^2 \mid \mathcal{C}_n]\right) \\
&\quad + \sum_{s=1}^{n-1}\omega_n(s)\times\frac{1}{n}\sum_{i\in N_n}\sum_{j\in N_n^\partial(i;s)}\left(y_{n,i}y_{n,j} - \mathbf{E}[y_{n,i}y_{n,j} \mid \mathcal{C}_n]\right) \\
&\quad + \sum_{s=1}^{n-1}[\omega_n(s)-1]\times\frac{1}{n}\sum_{i\in N_n}\sum_{j\in N_n^\partial(i;s)}\mathbf{E}[y_{n,i}y_{n,j} \mid \mathcal{C}_n] \\
&\equiv R_{n,0} + R_{n,1} + R_{n,2}.
\end{aligned}
\tag{A.14}
$$

Consider each term in the last line of (A.14) separately. Using Theorem A.1 for $y_{n,i}$ and $y_{n,j}$ with $d_n(i,j) = s \geq 1$ we get

$$
|\mathbf{E}[y_{n,i}y_{n,j} \mid \mathcal{C}_n)]| \leq C_2\theta_{n,s}^{1-\frac{2}{r}} \quad \text{a.s.},
$$

where $C_2 = C(\mu^2 \vee 1)$ for some constant $C \geq 1$. Therefore,

$$
\begin{aligned}
|R_{n,2}| &\leq \sum_{s=1}^{n-1}|\omega_n(s)-1|\times\frac{1}{n}\sum_{i\in N_n}\sum_{j\in N_n^\partial(i;s)}|\mathbf{E}[y_{n,i}y_{n,j} \mid \mathcal{C}_n]| \\
&\leq C_2\sum_{s=1}^{n-1}|\omega_n(s)-1|\,\theta_{n,s}^{1-\frac{2}{r}}\times\frac{1}{n}\sum_{i\in N_n}\left|N_n^\partial(i;s)\right| \\
&= C_2\sum_{s=1}^{n-1}|\omega_n(s)-1|\,\delta_n^\partial(s)\theta_{n,s}^{1-\frac{2}{r}} \quad \text{a.s.},
\end{aligned}
$$

and it follows from Assumption 4.1(ii & iv) and the dominated convergence theorem that $|R_{n,2}| \to 0$ a.s.

Let $z_{n,i,j} = y_{n,i}y_{n,j} - \mathbf{E}[y_{n,i}y_{n,j} \mid \mathcal{C}_n]$ so that $\mathbf{E}[z_{n,i,j} \mid \mathcal{C}_n] = 0$ a.s. Then, using Corollary A.2 for $z_{n,i,j}$ and $z_{n,k,l}$ with $d_n(\{i,j\},\{k,l\}) = s \geq 1$,

$$
|\mathbf{E}[z_{n,i,j}z_{n,k,l} \mid \mathcal{C}_n]| \leq C_1\theta_{n,s}^{1-\frac{2}{r}} \quad \text{a.s.},
$$

where $C_1 = C(\mu^4 \vee 1)$ for some constant $C \geq 1$. To deal with the case in which $d_n(\{i,j\},\{k,l\}) = 0$ note that $r > 2$ so that

$$
|\mathbf{E}[z_{n,i,j}z_{n,k,l} \mid \mathcal{C}_n]| \leq [\operatorname{Var}(y_{n,i}y_{n,j} \mid \mathcal{C}_n)\operatorname{Var}(y_{n,k}y_{n,l} \mid \mathcal{C}_n)]^{1/2} \leq \mu^4 \quad \text{a.s.}
$$

Noticing that $|\omega(\,\cdot\,)| \leq 1$ and letting

$$
G(s) = \mathbf{1}\{s=0\} + \theta_{n,s}^{1-\frac{2}{r}}\mathbf{1}\{s>0\},
$$

we find that

$$\begin{aligned}
\mathbf{E}[R_{n,1}^2 \mid \mathcal{C}_n] &\leq \frac{1}{n^2} \sum_{\substack{i,j \in N_n: \\ 1 \leq d_n(i,j) \leq b_n}} \sum_{\substack{k,l \in N_n: \\ 1 \leq d_n(k,l) \leq b_n}} |\mathbf{E}[z_{n,i,j} z_{n,k,l} \mid \mathcal{C}_n]| \\
&\leq \frac{C_1}{n^2} \sum_{s=0}^{n-1} \sum_{\substack{i,j \in N_n: \\ 1 \leq d_n(i,j) \leq b_n}} \sum_{\substack{k,l \in N_n: 1 \leq d_n(k,l) \leq b_n, \\ d_n(\{i,j\},\{k,l\}) = s}} G(s) \\
&\leq \frac{C_1}{n^2} \sum_{s=0}^{n-1} |H_n(s,b_n)| \, \theta_{n,s}^{1-\frac{2}{r}}, \quad \text{a.s.}
\end{aligned}$$

Hence, it follows from Assumption 4.1(iii) that $\mathbf{E}[R_{n,1}^2 \mid \mathcal{C}_n] \to 0$ a.s.

Finally, using similar arguments it is not hard to show that

$$\begin{aligned}
\mathbf{E}[R_{n,0}^2 \mid \mathcal{C}_n] &\leq \frac{1}{n^2} \sum_{s=0}^{n-1} \sum_{i \in N_n} \sum_{j \in N_n^{\partial}(i;s)} \left| \mathrm{Cov}(y_{n,i}^2, y_{n,j}^2 \mid \mathcal{C}_n) \right| \\
&\leq \frac{C_0}{n} \left(1 + \sum_{s=1}^{n-1} \delta_n^{\partial}(s) \theta_{n,s}^{1-\frac{2}{r}} \right) \to 0, \quad \text{a.s.},
\end{aligned}$$

where $C_0 = C(\mu^4 \vee 1)$ for some constant $C \geq 1$.

As for the second implication define $\bar{y}_n = c^\top \bar{Y}_n$, $\lambda_n = c^\top \Lambda_n$ and consider the difference between two estimators, $A_n'(c) = c^\top \left(\tilde{V}_n - \hat{V}_n \right) c$ which can be written as follows:

$$\begin{aligned}
A_n'(c) &= \sum_{s=0}^{b_n} \omega_n(s) c^\top \left(\tilde{\Omega}_n(s) - \hat{\Omega}_n(s) \right) c \\
&= (\bar{y}_n - \lambda_n)^2 \sum_{s=0}^{b_n} \omega_n(s) \times \frac{1}{n} \sum_{i \in N_n} |N_n^{\partial}(i;s)| \\
&\quad - (\bar{y}_n - \lambda_n) \sum_{s=0}^{b_n} \omega_n(s) \times \frac{2}{n} \sum_{i \in N_n} |N_n^{\partial}(i;s)| (y_{n,i} - \lambda_n).
\end{aligned}$$

Let $B_{n,i} = \sum_{s=0}^{b_n} \omega_n(s) |N_n^{\partial}(i;s)|$ and $B_n = D_n(b_n)$. Then, noticing that $|\omega_n(\cdot)| \leq 1$, we get

$$|A_n'(c)| \leq \left(\sqrt{B_n} \, |\bar{y}_n - \lambda_n| \right)^2 + \left(\sqrt{B_n} \, |\bar{y}_n - \lambda_n| \right) \times 2R_{3,n},$$

where

$$R_{3,n} \equiv \frac{1}{n\sqrt{B_n}} \left| \sum_{i \in N_n} B_{n,i}(y_{n,i} - \lambda_n) \right|.$$

Since $|B_{n,i}| \leq B_n$ for all $i \in N_n$, Assumption 4.1 implies that

$$\begin{aligned}\mathbf{E}[R_{3,n}^2 \mid \mathcal{C}_n] &\leq \frac{B_n}{n^2} \sum_{i \in N_n} \mathrm{Var}(y_{n,i} \mid \mathcal{C}_n) \\ &\quad + \frac{B_n}{n^2} \sum_{s=1}^{n-1} \sum_{i \in N_n} \sum_{j \in N_n^{\partial}(i;s)} |\mathrm{Cov}(y_{n,i} y_{n,j} \mid \mathcal{C}_n)| \\ &\leq \frac{C_2 B_n}{n} \left(1 + \sum_{s=1}^{n-1} \delta_n^{\partial}(s) \theta_{n,s}^{1-\frac{2}{r}} \right) \to 0 \quad \text{a.s..}\end{aligned}$$

Finally, it is not hard to show that $\mathbf{E}[|\sqrt{B_n}(\bar{y}_n - \lambda_n)|^2 \mid \mathcal{C}_n]$ is bounded by the same quantity. Hence, $\mathbf{E}[|A_n'(c)| \mid \mathcal{C}_n] \to 0$ a.s. ■