Essays in Econometrics

by

Denis Kojevnikov

BA, Tel Aviv University, 2004

MA, Tel Aviv University, 2006

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Economics)

The University of British Columbia
(Vancouver)

August 2019

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the thesis entitled:

Essays in Econometrics

submitted by **Denis Kojevnikov** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy** in **Economics**.

Examining Committee:

Vadim Marmer, Economics

Co-Supervisor

Kyungchul (Kevin) Song, Economics

Co-Supervisor

Harry Sue Wah Joe, Statistics

University Examiner

Paul T. Schrimpf, Economics

University Examiner

Tim Conley, Economics, Western University

External Examiner

Abstract

Chapter 2, co-authored with Vadim Marmer and Kyungchul Song, considers a general form of network dependence, where dependence between two sets of random variables becomes weaker as their network distance increases. We show that such network dependence cannot be viewed 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 work applies Doukhan and Louhichi (1999)'s notion of weak dependence to networks by measuring the strength of dependence using the covariance between nonlinearly transformed random variables. While this approach covers examples such as strong mixing random fields on graphs and conditional dependency graphs, it is most useful when dependence arises through a functional-causal system of equations. The main results of this chapter include a law of large numbers and a central limit theorem for network dependent processes.

Chapter 3 focuses on the bootstrap for network dependent processes studied in Chapter 2. Such processes are distinct from other forms of random fields that are commonly used in the statistics and econometrics literature so that the existing bootstrap methods cannot be applied directly. I propose a block-based method and a modification of the dependent wild bootstrap for constructing confidence sets for the mean of a network dependent process. In addition, I establish the consistency of these methods for the smooth function model and provide the bootstrap alternatives to the network heteroskedasticity-autocorrelation consistent variance estimator obtained in Chapter 2.

Finally, Chapter 4, co-authored with Kyungchul Song, presents a large Bayesian game with multiple information groups and develops a bootstrap inference method that does not require a common prior assumption and allows each player to form beliefs differently from other players. By drawing on the intuition of Kalai (2004), this work introduces the notion of a hindsight regret, which measures a player's ex post value of other players' type information, and obtains its belief-free bound. Using this bound, we derive testable implications and propose a bootstrap inference procedure for the structural parameters of the game.

Lay Summary

This thesis is based on a series of papers dealing with econometric inference in the presence of dependence between observations. Chapter 2 focuses on asymptotic theory for observations lying on a network and provides tools needed for hypothesis testing when the sample size is large. Chapter 3 keeps on studying this framework and develops a number of non-parametric resampling methods, which are useful for statistical inference in small samples. Finally, Chapter 4 concentrates on strategic interactions between economic agents in a game-theoretic context and proposes a bootstrap inference procedure for the parameters of interest.

Preface

Chapter 2 of this thesis is an unpublished working paper that I co-authored with Vadim Marmer and Kyungchul Song. All co-authors contributed equally to all aspects of the project.

Chapter 3 of this thesis is my original work.

Chapter 4 of this thesis is an unpublished working paper that I co-authored with Kyungchul Song. All co-authors contributed equally to all aspects of the project.

Table of Contents

List of Tables

List of Figures

Acronyms

CLT central limit theorem
DGP data-generating process
DWB dependent wild bootstrap
GMM generalized method of moments
HAC heteroskedasticity-autocorrelation consistent
LLN law of large numbers
MDS multidimensional scaling
RMSE root mean square error

Chapter 1

Introduction

The main focus of this thesis is statistical inference in the presence of dependence which originates as a result of interactions of economics agents. Specifically, the first two chapters concentrate on network dependent observations and bring a novel view on modeling cross-sectional dependence patterns. The third chapter employs a game-theoretic setup to capture strategic interactions among many agents.

Network models have attracted an increasing attention in the economic literature. As pointed out by Matthew Jackson, networks play an important role in the transmission of information in various scenarios, e.g., information about vacancies in the labor market, advertising on online "network communities", trade of many differentiated products, etc. A substantial body of economic literature focuses on questions of network formation and the effect of network structures on the behavior of economic agents.

A large number of available data sets based on different networks allows to assess empirically many types of network-based models. However, network data introduces additional challenges in dealing with different statistical dependence structures as opposed to traditional cross-sectional or time series dependence. In these settings observations are often heteroskedastic, dependence between them is induced by the underlying network and may vary with location so that statistical inference in econometric models requires a special asymptotic theory.

The second chapter of this thesis focuses on developing such a theory, in joint work with Vadim Marmer and Kyungchul Song. First, we introduce the notion of network dependent processes and show that network dependence is quite different from the types of spatial dependence studied in the literature. We assume a particular form of stochastic dependence between observations that lie on a network, which is sufficiently general and, more importantly, useful in econometric modeling. We provide a number of broad classes of network dependent processes, including those where observations are generated from primitive random variables through nonlinear transformations. Finally, this chapter establishes a law of large numbers

(LLN) and a central limit theorem (CLT) for network dependent processes and provides non-parametric HAC estimators of the variance-covariance matrix for a vector of sample moments.

Network dependent processes are closely related to random fields indexed by elements of a lattice in a Euclidean space. However, the lack of regular structure in networks renders the use of existing resampling approaches for dependent observations impracticable. In Chapter 3 I develop resampling methods for the sample mean under network dependence introduced in Chapter 2. I extend this notion by allowing weighed networks, which can be useful in several respects: it allows modelling some additional random processes as network dependent; assuming varying intensity of connections enables one to handle denser networks in the sense of the total number of links. Then I propose a number of bootstrap methods that utilize the structure of an underlying network and show their consistency as the sample size grows to infinity. In addition, this chapter provides the bootstrap variance estimators of the sample mean, which are positive semidefinite by construction and can be used for asymptotic inference. All the results hold conditionally on a common shock of a general form, which may represent a random network formation process or may help to obtain a local form of stochastic dependence. In this connection, I present a number of general results regarding conditional bootstrap.

The last chapter of this thesis, in joint work with Kyungchul Song, extends the econometric literature on strategic interactions among many agents. The existing models often assume i.i.d. replications of a representative game and rely on a common prior assumption. We propose an alternative model of a single large Bayesian game in which players form beliefs differently from each other using different priors. Instead of attempting to recover these beliefs, this work introduces a hindsight regret approach. Based on the observed outcomes from a pure strategy Bayesian Nash equilibrium, we develop an inference procedure for the structural parameters of the underlying game. This procedure is robust to the equilibrium selection and the way players form their beliefs about other players' types. Specifically, in order to avoid overly conservative inference on the structural parameters we adapt a two-step bootstrap approach to get asymptotically uniformly valid confidence sets for the parameters of interest.

Chapter 2

Limit Theorems for Network Dependent Random Variables

2.1 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 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 therein.

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 between sample units, which can be viewed as a form of a network. To apply the random field modeling, one would need to transform these relations into a random field on a lattice in a Euclidean space using methods such as multidimensional scaling (MDS) (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 work shows that network dependence cannot be viewed 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 chapter, 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 sums 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 (2018) 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 CLT 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 chapter 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 work is that while 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 transformations applied to random variables in these sets.

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. 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. We provide various auxiliary results, which could be helpful in such situations. They are found in the appendix of this chapter.

The main results of this chapter, an LLN and a CLT for network dependent processes, 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 graphs. In addition, we establish the consistency of HAC variance estimators and investigate their performance through various simulation designs. As expected, these estimators perform better when the cross-sectional dependence is less extensive.

The chapter is organized as follows. In the next section, we define network dependence of stochastic processes and provide examples. In Section 2.3, we present the main results, i.e., the limit theorems. Section 2.4 focuses on HAC estimation and investigates its finite sample performance using Monte Carlo simulations. Some auxiliary results and the proofs of the main results are provided in the appendix.

2.2 Network dependence and examples

2.2.1 Network topology and a lattice in a Euclidean space

Let $N_n \equiv \{1, 2, \ldots, n\}$ 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 work 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, 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,

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

Define

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

When $k=1$, we simply write $\delta_n(s) \equiv \delta_n(s;1)$ and $\delta_n^\partial(s) \equiv \delta_n^\partial(s;1)$. 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 $\delta_n(s)$ as s increases.

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.

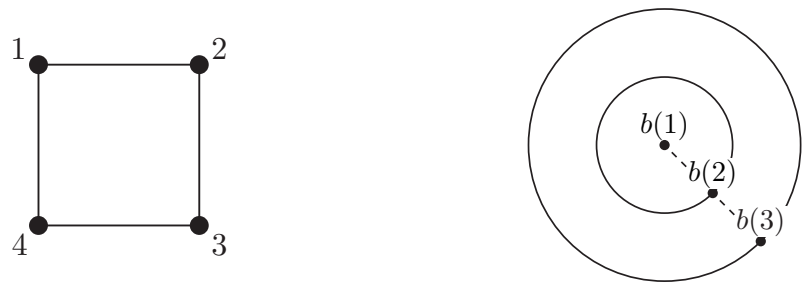
Definition 2.1. An isometric embedding of a network $G_n = (N_n, E_n)$ into a metric space $(\mathcal{X}, d_\mathcal{X})$ is an injective map $b : N_n \to \mathcal{X}$ such that for all $i, j \in N_n$

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

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.

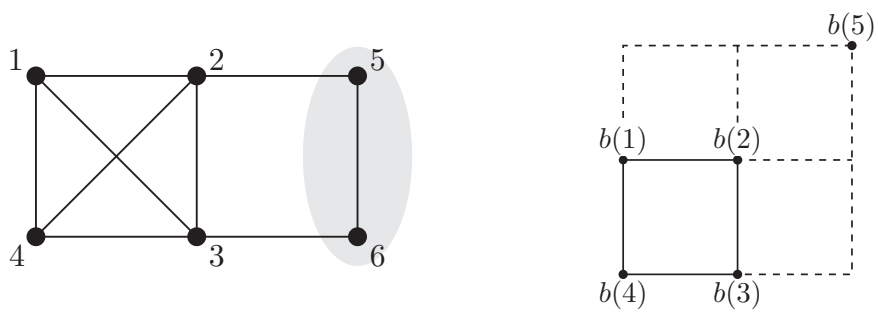
Proposition 2.1. *A connected network G_n is isometrically embeddable into a metric space $(\mathcal{X}, d_\mathcal{X})$ only if $\omega(G_n) \le e(\mathcal{X})$.*

[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.



(a) (b)

Figure 2.1: 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 $b(2)$.



(a) (b)

Figure 2.2: 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.

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 $|C| > e(\mathcal{X})$. ■

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 2.1 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 2.2 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 chapter 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 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.

2.2.2 Network dependent processes

Let us introduce the notion of network dependence that is the focus of this chapter. Suppose that we are given a sequence of networks $\{G_n\}$ and an associated triangular array of $\mathbf{R}^v$-valued random vectors $\{Y_{n,i}\}_{i \in N_n}$, $n \geq 1$, defined on a common probability space $(\Omega, \mathcal{H}, \mathsf{P})$. We adapt the weak dependence notion of Doukhan and Louhichi (1999) to our set up. Specifically, for any $v, a \in \mathbb{N}$ we define

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

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

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

with $\mathrm{Lip}(f)$ denoting the Lipschitz constant of f,[3] and $\|\cdot\|_\infty$ the sup-norm of f. For any $a, b \in \mathbb{N}$ and $s \geq 0$, consider two sets of nodes (of size a and b) with distance between each

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

other of at least s. Let $\mathcal{P}_n(a,b;s)$ denote the collection of all such pairs:

$$\mathcal{P}_n(a,b;s) := \{(A,B) \subset N_n^2 : |A| = a, |B| = b, d_n(A,B) \geq s\},$$

where

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

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 a subset $A \subset N_n$ we write $Y_{n,A} \equiv \{Y_{n,i}\}_{i \in A}$. We take $\{\mathcal{C}_n\}_{n \geq 1} \subset \mathcal{H}$ to be a given sequence of sub-σ-fields.

Definition 2.2. The triangular array $\{Y_{n,i}\}$ is *conditionally ψ-weakly dependent* given $\{\mathcal{C}_n\}$, if for each $n \in \mathbb{N}$, there exists a non-negative, $\mathcal{C}_n$-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 non-random 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 $g \in \mathcal{L}_{v,b}$,

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

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 $\{Y_{n,i}\}$ 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 being generated by some characteristics or actions of multiple central nodes which affect other nodes through their 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 $U_1, \ldots, U_n$ 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 $\mathsf{P}(\theta_{n,2} = 0 \mid \mathcal{C}_n) = 1$.

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

Assumption 2.1. (a) $\{Y_{n,i}\}$ is conditionally ψ-weakly dependent given $\{\mathcal{C}_n\}$ with the weak dependence coefficients $\{\theta_n\}$ and

$$\begin{aligned}\psi_{a,b}(f,g) = {} & c_1 \|f\|_\infty \|g\|_\infty + c_2 \,\mathrm{Lip}(f) \|g\|_\infty \\ & + c_3 \|f\|_\infty \,\mathrm{Lip}(g) + c_4 \,\mathrm{Lip}(f)\,\mathrm{Lip}(g),\end{aligned}$$

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 chapter. Assumption 2.1(a) is shown to be satisfied by all the examples we present in the next subsection. The restriction on the weak dependence coefficients can be relaxed at the cost of more complex notation. The restriction on the form of $\psi_{a,b}$ is convenient because the ψ-dependence of random vectors carries over to linear combinations of the elements of these vectors as shown in the following lemma.

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

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

2.2.3 Examples

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

2.2.3.1 Strong-mixing processes

Let $\mathcal{F}, \mathcal{G} \subset \mathcal{H}$ be given sub-σ-fields and let

$$\alpha(\mathcal{F}, \mathcal{G} \mid \mathcal{C}) := \sup_{F \in \mathcal{F}, G \in \mathcal{G}} |\mathrm{Cov}(\mathbf{1}_F, \mathbf{1}_G \mid \mathcal{C})|.$$

We define the strong mixing coefficients by

$$\alpha_{n,s} := \sup\{\alpha(\sigma(Y_{n,A}), \sigma(Y_{n,B}) \mid \mathcal{C}_n) : A, B \subset N_n, d_n(A, B) \geq s\}.^4$$

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

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

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

Therefore, if $\sup_{n \geq 1} \alpha_{n,s} \to 0$ a.s. as $s \to \infty$, 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 θ_n are given by the strong mixing coefficients $\alpha_n \equiv \{\alpha_{n,s}\}_{s=1}^\infty$.

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

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 $\alpha(\mathcal{F}, \mathcal{G} \mid \mathcal{C})$ with $\alpha(\mathcal{F} \vee \mathcal{C}, \mathcal{G} \vee \mathcal{C} \mid \mathcal{C})$.

2.2.3.2 Conditional dependency graphs

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). It is not hard to see that when $\{Y_{n,i}\}_{i \in N_n}$ has G_n as a conditional dependency graph given $\mathcal{C}_n$ for each $n \geq 1$, 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 θ_n are such that $\theta_{n,s} = 0$ for all $s \geq 1$.

2.2.3.3 Functional dependence on independent variables

Consider a collection of $\mathbf{R}^k$-valued random vectors, $\varepsilon_n \equiv \{\varepsilon_{n,i}\}_{i \in N_n}$, that are conditionally independent given $\mathcal{C}_n$ and a collection of $\mathbf{R}^v$-valued, measurable functions $\{\phi_{n,i}\}$. Suppose that each $Y_{n,i}$, $i \in N_n$ is generated as follows:

$$Y_{n,i} = \phi_{n,i}(\varepsilon_n). \tag{2.3}$$

Further, given $s \geq 0$ we define a modified version of $Y_{n,i}$:

$$Y_{n,i}^{(s)} := \phi_{n,i}\big(\varepsilon_{n,i}^{(s)}\big),$$

where $\varepsilon_{n,i}^{(s)} \equiv \{\varepsilon_j \mathbf{1}\{j \in N_n(i; s)\} : j \in N_n\}$.[5] Note 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 $\mathcal{C}_n$.

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

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

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

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

Proposition 2.3. *For any $(A, B) \in \mathcal{P}_n(a, b; 2s+1)$ and $f \in \mathcal{L}_{v,a}, g \in \mathcal{L}_{v,b}$ that are Lipschitz with respect to the distances $\boldsymbol{\delta}_a$ and $\boldsymbol{\delta}_b$, respectively,*

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

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

It follows that when $\{Y_{n,i}\}_{i\in N_n}$ is generated as in (2.3) for each $n \geq 1$, and $\sup_n \theta_{n,s} \to 0$ a.s. as $s \to \infty$, the array $\{Y_{n,i}\}$ is conditionally ψ-weakly dependent given $\{\mathcal{C}_n\}$ with

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

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

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

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

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

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

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

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 $s \to \infty$.

2.2.3.4 Functional dependence on associated or Gaussian variables

As in the previous example, consider a collection of random variables $\varepsilon_n \equiv \{\varepsilon_{n,i}\}_{i\in N_n}$ and a collection of real-valued, measurable functions $\{\varphi_{n,i}\}_{i\in N_n}$. We say that ε_n is a conditionally positively associated process given $\mathcal{C}_n$ if for all $A, B \subset N_n$ and all coordinatewise non-decreasing, real-valued, measurable functions f and g,

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

When the above inequality is reversed for all $A, B \subset N_n$, we say that ε_n is negatively associated.

Suppose further that each $Y_{n,i}$, $i \in N_n$ is generated as follows:

$$Y_{n,i} = \varphi_{n,i}(\varepsilon_n). \tag{2.4}$$

The following result is a consequence of a covariance inequality due to Theorem 3.1 of Birkel (1988) and Lemma 19 of Doukhan and Louhichi (1999).

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) ε_n is conditionally positively or negatively associated given $\mathcal{C}_n$ and functions $f, g \in \mathscr{C}_b^1$, or (ii) ε_n is conditionally Gaussian given $\mathcal{C}_n$ and functions $f, g \in \mathscr{C}_b^1 \cap \mathscr{C}_b$. Then for any $(A, B) \in \mathcal{P}_n(a, b; s)$,*

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

where

$$\theta_{n,s} = \max_{1 \le k_1, k_2 \le n} \sum_{i \in N_n} \sum_{j \in N_n} \left\| \frac{\partial \varphi_{n,k_1}}{\partial \varepsilon_{n,i}} \right\|_\infty \left\| \frac{\partial \varphi_{n,k_2}}{\partial \varepsilon_{n,j}} \right\|_\infty |\mathrm{Cov}(\varepsilon_{n,i}, \varepsilon_{n,j} \mid \mathcal{C}_n)|. \tag{2.5}$$

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 the elements of ε_n are conditionally independent given $\mathcal{C}_n$, the sequence $\theta_{n,s}$ reduces to the following:

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

Suppose further that $\partial \varphi_{n,k} / \partial \varepsilon_{n,i} = 0$ whenever i is at least m-edges away from k in G_n. Consider a graph G_n' in which i and j are adjacent if and only if i and j are within $2m$ edges away. Then G_n' is a conditional dependency graph for $Y_{n,i}$'s given $\mathcal{C}_n$. In this case $\theta_{n,s} = 0$ for all $s \ge 2m$.

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). Specifically, suppose that $X_1, \ldots, X_n$ are i.i.d., $\mathbf{R}^d$-valued random vectors. For a given parameter r we form an undirected graph $G_n = (N_n, E_n)$ such that $ij \in E_n$ if and only if $\|X_i - X_j\| \le r$. Let $\mathcal{C}_n' := \mathcal{C}_n \vee \sigma(X_1, \ldots, X_n)$. If $\{\varepsilon_{n,i}\}_{i \in N_n}$ is a conditionally associated process given $\mathcal{C}_n'$, Proposition 2.4 applies to this case with $\mathcal{C}_n$ replaced by $\mathcal{C}_n'$.

The following corollary shows that if $\{Y_{n,i}\}_{i \in N_n}$ is generated as in (2.4) for each $n \ge 1$, the array $\{Y_{n,i}\}$ 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 $n \ge 1$. Then $\{Y_{n,i}\}$ is conditionally ψ-weakly dependent given $\{\mathcal{C}_n\}$ with the weak dependence coefficients given by* (2.5)

and

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

provided that $\sup_{n\geq 1}\theta_{n,s} \to 0$ *a.s. 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). ■

2.3 Limit theorems

2.3.1 Law of large numbers

We first establish a law of large numbers. Since an LLN can be applied element-by-element in the vector case, without loss of generality we assume that $Y_{n,i} \in \mathbf{R}$, i.e., $v = 1$. We assume below that the process is uniformly integrable.

Assumption 2.2 (Uniform L_1-Integrability)**.**

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

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 $\sup_{i\in N_n}\mathsf{E}[|Y_{n,i}|^{1+\epsilon}] < \infty$ 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 $\|X\|_p \equiv (\mathsf{E}\|X\|^p)^{1/p}$ denote the L_p-norm of a random vector X.

Assumption 2.3. $n^{-1}\sum_{s=1}^{n-1}\delta_n^\partial(s)\|\theta_{n,s}\|_1 \to 0$ 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 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 2.3 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 2.4.2 for details of the graph generation). As shown in Figure 2.3, $\sup_{s\geq 1}\delta_n^\partial(s)$ increases much slower than n.

Assumption 2.3 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

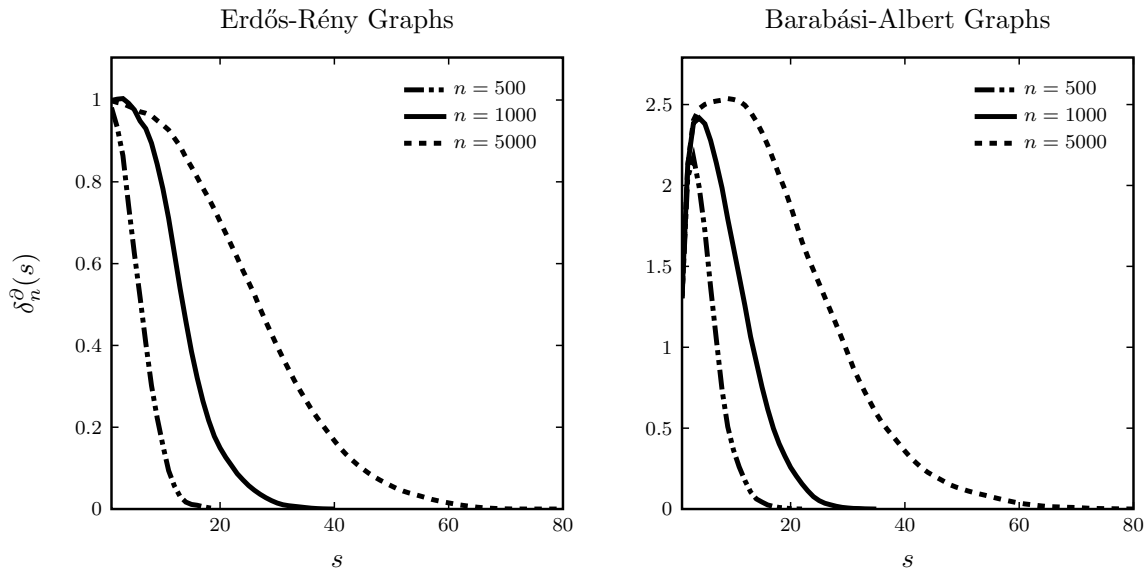


Figure 2.3: Plots of $\delta_n^\partial(s)$ for graphs generated using the Erdős-Rény (left panel) and Barabási-Albert (right panel) models (see Subsection 2.4.2 for details). The graphs show that regardless of the sample size, $\delta_n^\partial(s)$ vanishes as s becomes larger. It is not hard to see that in both cases $\sup_{s\geq 1}\delta_n^\partial(s)/n$ decreases as n becomes larger.

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 2.3 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 2.6(a) below. In that case $\delta_n^\partial(s) \leq 2$, and Assumption 2.3 holds when $n^{-1}\sum_{s=1}^{n-1}\|\theta_{n,s}\|_1 \to 0$. The latter condition holds, for example, if $\{\|\theta_{n,s}\|_1\}_{s=1}^{n-1}$ is summable as $n \to \infty$.

Theorem 2.1. *Suppose that Assumptions 2.2 and 2.3 hold. Then, as $n \to \infty$,*

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

When $f \in \mathcal{L}_{v,1}$, an 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,

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

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

$$\leq \frac{1}{n}\sup_{n,i}\mathsf{E}|f(Y_{n,i})|^2 + \psi_{1,1}(f,f)\frac{1}{n}\sum_{s=1}^{n-1}\delta_n^{\partial}(s;1)\|\theta_{n,s}\|_1.$$

We have the following result.

Corollary 2.2. *Suppose that Assumption 2.3 holds and* $\sup_{n,i}\mathsf{E}|f(Y_{n,i})|^2 < \infty$*, where* $f \in \mathcal{L}_{v,1}$*. Then,*

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

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. 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 $h(Y_{n,i}) - \mathsf{E}[h(Y_{n,i}) \mid \mathcal{C}_n]$ and $h(Y_{n,j}) - \mathsf{E}[h(Y_{n,j}) \mid \mathcal{C}_n]$ is bounded by

$$C\left\|\sup_{n,i}\|h(Y_{n,i})\|^2_{\mathcal{C}_n,p}\right\|_2 \times \left\|\theta^{1-\frac{2}{p}}_{n,d_n(i,j)}\right\|_2,^{7}$$

Therefore,

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

provided that $\mathsf{E}[\sup_{n,i}\|h(Y_{n,i})\|^2_{\mathcal{C}_n,p}] < \infty$, and a condition similar to that in Assumption 2.3 holds:

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

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 2.4.

2.3.2 Central limit theorem

In this section, we provide a central limit theorem for network dependent processes. Define $S_n := \sum_{i\in N_n} Y_{n,i}$,

$$\sigma_n^2 := \mathrm{Var}(S_n \mid \mathcal{C}_n), \quad \text{and} \quad \mu_{n,p} := \max_{i\in N_n}\|Y_{n,i}\|_{\mathcal{C}_n,p}.$$

We make a number of assumptions. The assumption below presents a moment condition.

[7] For a random vector X we write $\|X\|_{\mathcal{C}_n,p} \equiv (\mathsf{E}[\|X\|^p \mid \mathcal{C}_n])^{1/p}$.

Assumption 2.4. There exist $C > 0$ and $p \geq 4$ such that for all $n \geq 1$, $\mu_{n,p} \leq C$ a.s.

While requiring the moment condition in Assumption 2.4 is more restrictive than those conditions known for CLTs for special cases of ψ-weak dependence, such a moment condition is widely used in many models in practice. We define

$$H_n(s,m) := \left\{ (i,j,k,l) \in N_n^4 : \begin{smallmatrix} j \in N_n(i;m_n), l \in N_n(k;m_n), \\ d_n(\{i,j\},\{k,l\}) = s \end{smallmatrix} \right\}.$$

The following assumption limits the extent of the cross-sectional dependence between the random variables in $\{Y_{n,i}\}$ through restrictions on the network.

Assumption 2.5. For each $n \geq 1$, there exist $c_n \to \infty$ such that $0 < c_n \leq \sigma_n$ a.s. and a positive sequence $m_n \to \infty$ such that for some $p \geq 4$ that appears in Assumption 2.4,

$$c_n^{-4} \sum_{s=0}^{n-2} |H_n(s,m_n)| \theta_{n,s}^{1-4/p} \to 0 \quad \text{a.s.}, \tag{2.6}$$

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

$$\frac{n\delta_n(m_n;2)}{c_n^3} \to 0, \quad \text{and} \quad \frac{n^2 \theta_{n,m_n}^{1-1/p}}{c_n} \to 0 \quad \text{a.s.} \tag{2.7}$$

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 the model of dependency graphs. Then, the conditions in (2.6) and (2.7) 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 they are not necessarily overly strong relative to the existing literature. For example, these conditions hold 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

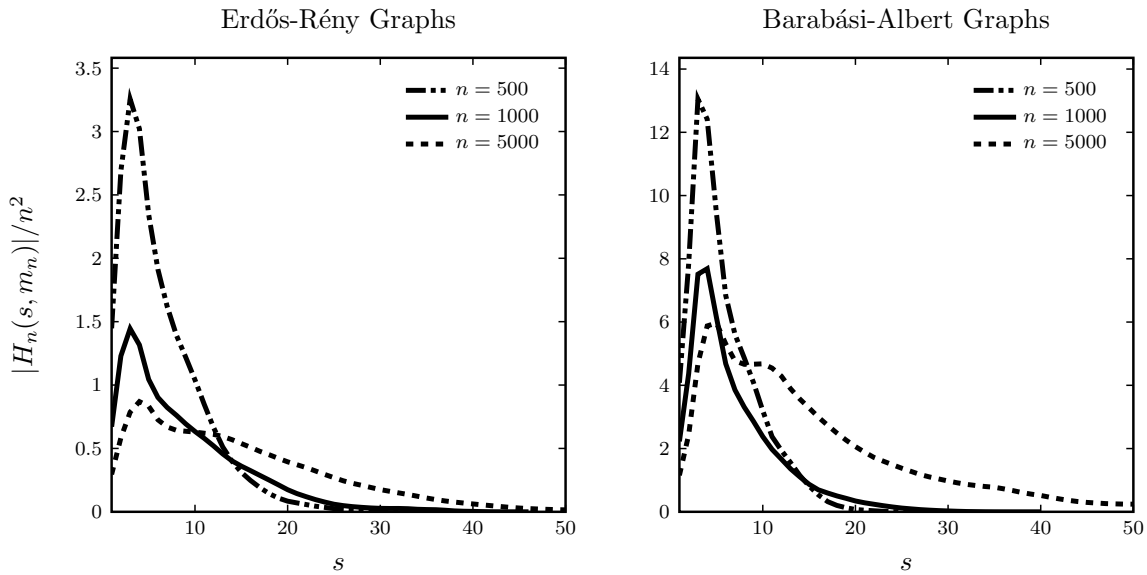


Figure 2.4: Plots of $|H_n(s; m_n)|/n^2$ for graphs generated using the Erdős-Rény (left panel) and Barabási-Albert (right panel) models (see Subsection 2.4.2 for details) with $m_n = \ln(1 + n)$. The graphs show that regardless of the sample size, $|H_n(s; m_n)|$ vanishes as s becomes larger. It is not hard to see that in both cases $\sup_{s\geq 1}|H_n(s; m_n)|/n^2$ decreases as n becomes larger.

that there exist $C > 0$ and $\tau \geq 1$ such that for all $s \geq 1$,

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

and

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

Then the conditions (2.6) and (2.7) hold 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 (2.7) is satisfied.

Figure 2.4 shows the plots of $|H_n(s; m_n)|/n^2$ with $m_n = \ln(1 + n)$ for two types of graphs used in Figure 2.3. The plots demonstrate that $|H_n(s; m_n)|/n^2$ decreases as n becomes larger and, therefore, the conditions in Assumption 2.5 seem plausible in these examples.

Theorem 2.2. *Suppose that Assumptions 2.1, 2.4-2.5 hold and* $\mathsf{E}[Y_{n,i} \mid \mathcal{C}_n] = 0$ *a.s. for all* $i \in N_n$ *and* $n \geq 1$. *Then*

$$\sup_{t\in\mathbf{R}} |\mathsf{P}(S_n/\sigma_n \leq t \mid \mathcal{C}_n) - \Phi(t)| \to 0 \quad \textit{a.s.}$$

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

2.4 HAC estimation

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 $\mathsf{E}[Y_{n,i} \mid \mathcal{C}_n] = 0$ a.s. for all $i \in N_n$ and $n \geq 1$. Let

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

Then the above-mentioned variance is given by

$$V_n := \mathrm{Var}(S_n/\sqrt{n} \mid \mathcal{C}_n) = \sum_{s\geq 0} \Omega_n(s) \quad \text{a.s.} \tag{2.8}$$

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

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

$$\hat{V}_n = \sum_{s=0}^{\lfloor b_n \rfloor} \omega_n(s) \hat{\Omega}_n(s), \tag{2.9}$$

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 to 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 $\mathsf{E}[Y_{n,i} \mid \mathcal{C}_n] = \Lambda_n$ a.s. for all $i \in N_n$, and the sequence of common conditional expectations $\{\Lambda_n\}$ is unknown. Under suitable assumptions Theorem 2.1 implies that $\bar{Y}_n := S_n/n$ is a consistent estimator of Λ_n. We redefine the kernel HAC estimator given in (2.9) as follows:

$$\tilde{V}_n = \sum_{s=0}^{\lfloor b_n \rfloor} \omega_n(s) \tilde{\Omega}_n(s), \tag{2.10}$$

where

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

When the elements of $\{Y_{n,i}\}_{i\in N_n}$ have different conditional expectations, it is hard to justify plugging the sample mean into $\tilde{\Omega}_n(\,\cdot\,)$ because $\bar{Y}_n$ is not a consistent estimator of $\mathsf{E}[Y_{n,i} \mid \mathcal{C}_n]$ in that case. 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.

2.4.1 Consistency

The consistency of the estimators defined in (2.9) and (2.10) 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 2.6. There exists $r > 2$ such that

(a) $\mu := \sup_{n\geq 1} \mu_{n,2r} < \infty$ a.s.,
(b) $\limsup_{n\to\infty} \sum_{s=1}^{n-1} \delta_n^{\partial}(s)\theta_{n,s}^{1-\frac{2}{r}} < \infty$ a.s.,
(c) $\lim_{n\to\infty} n^{-2} \sum_{s=0}^{n-1} |H_n(s,b_n)|\, \theta_{n,s}^{1-\frac{2}{r}} = 0$ a.s. with $\theta_{n,0} \equiv 1$,
(d) $\lim_{n\to\infty} |\omega_n(s) - 1| = 0$ for all $s \in \mathbb{N}$.

The first three conditions demonstrate the tradeoff between the conditional moments of $\{Y_{n,i}\}$ 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 2.6(c) 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 2.6(b) and

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

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

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

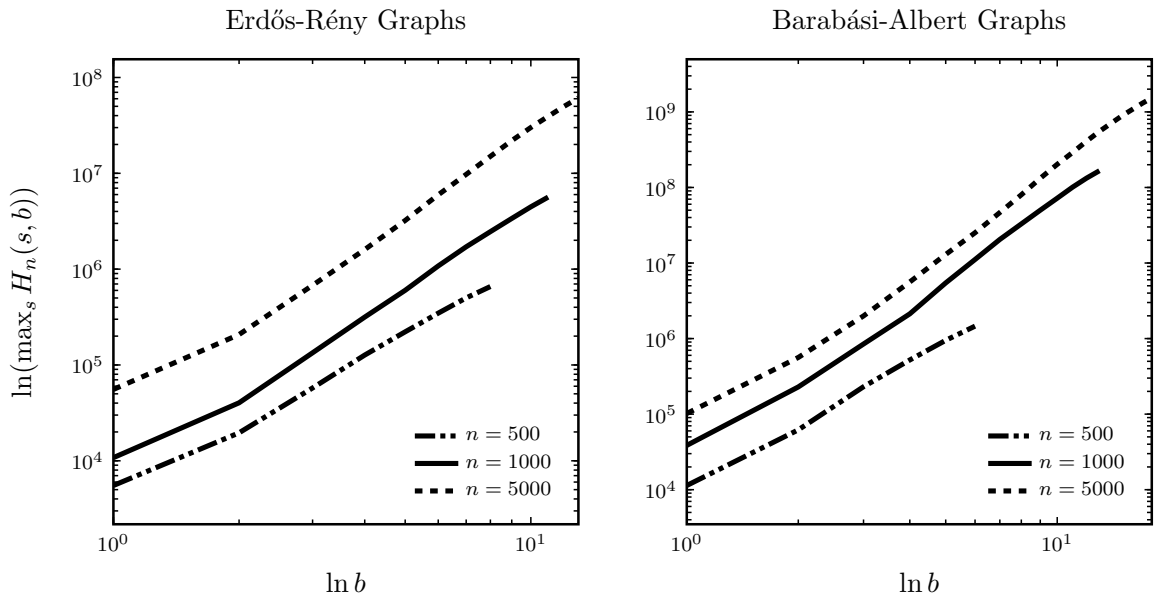


Figure 2.5: Log-log plots of $\max_s H_n(s,b)$ against b for graphs generated using the Erdős-Rény (left panel) and Barabási-Albert (right panel) models (see Subsection 2.4.2 for details).

Proposition 2.5. *Suppose that Assumption 2.6 holds. Then*

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

where $\|\cdot\|_F$ *denotes the Frobenius norm.*[8] *If, in addition,* $D_n(b_n)/n \to 0$*, then*

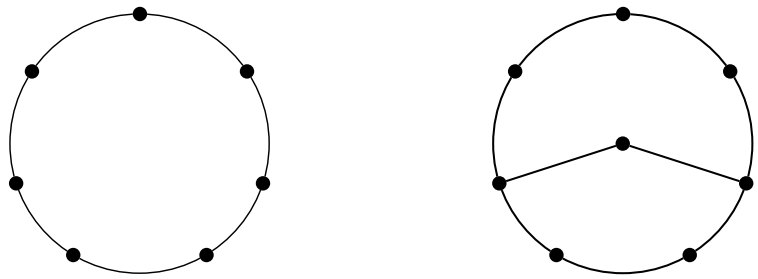
$$\mathsf{E}[\|\tilde{V}_n - V_n\|_F \mid \mathcal{C}_n] \to 0 \quad a.s.$$

Note that under Assumption 2.4.1(c') 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 $A_n(c) := c^\top\Big(\hat{V}_n - \tilde{V}_n\Big)c$, where $c \in \mathbf{R}^v$ is a fixed vector. Example 2.1 below illustrates such a case and shows that the positive definiteness of the kernel function does not imply automatically the 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 2.6(c). 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 $\lim_{n\to\infty} b_n = \infty$ and for some $\beta > 2$,

$$\sup_s H_n(s, b_n) = O\Big(b_n^\beta\Big). \tag{2.11}$$

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



(a) (b)

Figure 2.6: 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 $\omega(z) = \mathbf{1}\{|z| \le 1\}(1-|z|)$.

In such cases, a sufficient condition for Assumption 2.6(c) is that

$$b_n = o\Big(n^{2/\beta}\Big). \tag{2.12}$$

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

The parameter β in (2.11) 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 $\ln(\max_s H_n(s,b))$ against $\ln b$ 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 2.5 shows the plots of $\ln(\max_s H_n(s,b))$ against $\ln b$ for Erdős-Rényi and Barabási-Albert graphs. One can see that the condition in (2.11) 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 2.1. Consider a ring network (an example is shown in Figure 2.6(a)), where $N_n^{\partial}(i;s) = 2$ for $1 \le s \le \lfloor (n-1)/2 \rfloor$ and all $i \in N_n$. Suppose that $\Lambda_n = 0$ a.s. and let $\omega(z) = (1-|z|)\mathbf{1}\{|z| \le 1\}$ (Barlett kernel). For $b_n < (n-1)/2$,

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

Hence, $\hat{V}_n - \tilde{V}_n$ is positive semidefinite. In particular, $[\hat{V}_n - \tilde{V}_n]_{k,k} \ge 0$ for all $1 \le k \le v$ 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 2.6(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 $(Y_{n,i} - \mathsf{E}[Y_{n,i} \mid \mathcal{C}_n])$ 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 $x_i, x_j \in \mathbf{R}^p$, the positive definiteness of the kernel function ω_n 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 the 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 2.6(a) and the Barlett kernel with $b_n = 2$, it becomes indefinite after a slight modification shown in Figure 2.6(b).

2.4.2 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:

$$Y_{n,i} = \sum_{m \geq 0} \gamma^m \sum_{j \in N_n^\partial(i;m)} \varepsilon_{n,j}, \quad i \in N_n, \tag{2.13}$$

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 2.4.1.

We present the results of HAC estimation in Table 2.1. These results show that both the bias (in absolute value) and the RMSE of the estimators (2.9) and (2.10) decline as the sample size increases for all values of γ. In addition, as follows from Example 2.1, the second HAC estimator yields consistently lower estimates.

Table 2.1: 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 (2.13). Specifically, $\gamma = 0$ corresponds to the case of stochastic independence.

The next simulation results presented in Table 2.3 and Table 2.4 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 $\mu_0 = 0$. The consistency of the kernel HAC estimator and Theorem 2.2 guarantee that the limiting distribution of T_n is the standard normal. We use the 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) 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 2.2. In particular, $\bar{d}$ represents the average connected distance and is given by

$$\bar{d} := \frac{1}{k} \sum_{\substack{u,v \in N_n \\ 0 < d(u,v) < \infty}} d_n(u,v), \tag{2.14}$$

where k is the number of connected pairs $(u,v) \in N_n^2$. We choose the truncation parameter as $b_n = n^{2/\hat{\beta}}/\ln\ln n$, where $\hat{\beta}$ is the estimated slope coefficient from the log-log regression described in Section 2.4.1. It can be seen that the Barabási–Albert model generates denser networks, which results in slightly lower coverage probabilities.

Table 2.2: 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 (2.14).

Finally, we simulate the same t-test for a network of size $n = 5{,}000$ consisting of m equal-sized disconnected components (Block-ER graphs), where $m = 10, 25$, or 50 and each component is generated using the Erdős-Rényi model with parameter $p = 2m/n$. These graphs are denser than ones used previously (see Table 2.5). 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 2.6. Specifically, when the number of blocks is 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.

Table 2.3: Simulated power and size of the sample mean test at nominal size 5% for the ER graphs ($\mu_0 = 0$).

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 (2.13).

Table 2.4: Simulated power and size of the sample mean test at nominal size 5% for the BA graphs ($\mu_0 = 0$).

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 (2.13).

Table 2.5: Denseness measures of the Block-ER 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 (2.14).

Table 2.6: Simulated power of the sample mean test at nominal size 5% for the Block-ER graphs of size 5,000 ($\mu_0 = 0$).

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 (2.13).

2.5 Conclusion

Developing asymptotically valid inference methods for observations that are cross-sectionally dependent has long drawn a great deal of attention in the econometrics literature. This work contributes to this literature by establishing limit theorems for sums of random vectors

when the random vectors exhibit cross-sectional dependence along a given 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) that 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.

Chapter 3

The Bootstrap for Network Dependent Processes

3.1 Introduction

The aim of this chapter is developing bootstrap approaches for the sample mean of network dependent processes studied in Chapter 2. A network dependent process is a random field indexed by the set of nodes of a given undirected network. This network governs the stochastic dependence between the elements of the associated random field. Specifically, the latter is assumed to satisfy a conditional version of the ψ-weak dependence condition of Doukhan and Louhichi (1999) given a common shock of a general form. Chapter 2 provides an LLN and a CLT for a sequence of such processes, which hold under suitable assumptions on the networks' denseness and the strength of the stochastic dependence. In addition, it provides nonparametric HAC estimators of the variance-covariance matrix for a vector of sample moments, which are similar to the spatial HAC estimator developed in Kelejian and Prucha (2007).

These results provide an asymptotic approximation of the distribution of the sample mean, which can be used for inference on the true mean of a network dependent process. However, this approximation relies on the network HAC estimator, which has two major drawbacks. First, unlike its spatial or time-series counterparts it is not guaranteed to yield a positive semi-definite estimate. Second, these estimators are known to have poor finite sample properties (see, e.g., Matyas, 1999, Section 3.5). The aim of the current work is to provide an alternative nonparametric way to conduct inference in these settings.

The nonparametric bootstrap methods for the case of weakly dependent observations have been studies since the introduction of the non-overlapping block bootstrap in Carlstein (1986) and the moving block bootstrap in Künsch (1989) and Liu and Singh (1992) for stationary, mixing time-series. Since then, a number of block-based methods have been

considered in the statistics literature. They share the idea of resampling groups of consecutive observations to capture the stochastic dependence in the original series and include, among others, the circular block bootstrap (Politis and Romano, 1992), the stationary bootstrap (Politis and Romano, 1994) and the tapered block bootstrap (Paparoditis and Politis, 2001). A detailed exposition and comparison of some of these methods can be found in Lahiri (2003). Block-based bootstrap was also successfully applied to the case of weakly dependent random fields satisfying certain mixing conditions (see, e.g., Lahiri, 2003, Section 12 and references therein).

More recent developments in this area of research are discussed in Gonçalves and Politis (2011). In particular, the dependent wild bootstrap (DWB) proposed in Bühlmann (1993) and Shao (2010) departs from other methods. Instead of using blocks, it tries to mimic the autocovariance structure of the original data by introducing auxiliary random variables and can be applied to irregularly spaced data. A related method, the dependent random weighting, was recently introduced in Sengupta, Shao, and Wang (2015) and has wider applicability; specifically, it can be directly applied to irregularly spaced spatial data.

Another useful resampling technique developed for stationary and nonstationary times-series and homogenous random fields under mixing is subsampling. A comprehensive treatment of this method is given in Politis, Romano, and Wolf (1999). Interestingly, in the time-series case subsampling is similar to the moving block bootstrap where a single block is resampled. Finally, it is worth mentioning the spatial smoothed bootstrap suggested in Garcia-Soidan, Menezes, and Rubinos (2014). In this instance, assuming homogeneity of the underlying data generating process, bootstrap pseudo-samples are drawn from the estimated joint distribution of a given sample.

Network dependent processes are closely related to random fields indexed by elements of a lattice in $\mathbf{R}^d$ (see, e.g., Comets and Janžura, 1998; Conley, 1999). However, they are not a special case of the latter and so the existing bootstrap methods cannot be directly applied to our framework. The main reason for that is the irregularity of the structure of underlying networks. In particular, subsampling and all types of the block bootstrap for time-series and spatial data rely on the existence of ordered blocks of closely-located observations. The DWB uses a well-known property of kernel functions that guarantees the positive semi-definiteness of certain weighting matrices. However, as argued in Chapter 2, this relation does not necessarily hold when applied to networks. Finally, the homogeneity assumption of the spatial smoothed bootstrap and the spatial subsampling, that is the invariance of joint distributions under spatial shifts is not suitable for our case.

We propose two bootstrap approaches for constructing asymptotically valid confidence sets for the mean of a network dependent process and establish the first-order consistency of these methods for smooth functions of means conditionally on the common shock. The first approach is a block-based method in which blocks are constructed from certain neighborhoods

of each node in a network. The second is a modification of the DWB that employs the topology of a given graph to generate random weights instead of using a fixed kernel function. In addition, we provide the bootstrap variance estimators of the scaled sample mean, which yield positive semi-definite estimates and can be used as an alternative to the network HAC estimator. We find that the consistency of the modified DWB and the corresponding variance estimator holds under weaker conditions as compared with the block bootstrap. However, the bootstrap distribution corresponding to the former method may fail to match the higher-order cumulants of the underlying data generating process, thus preventing improvements over asymptotic approximations.

The rest of the chapter is organized as follows. The next section describes a modification of network dependent processes allowing for weighted networks. This modification can be useful for handling dense graphs once varying intensity of links is assumed. Section 3.3 provides some general result regarding conditional bootstrap. Specifically, we use the almost sure convergence of probability kernels to ensure that the bootstrap is valid for (almost) every realization of the common shock, which may also represent the stochastic network formation process. In Section 3.4 we present the above-mentioned bootstrap methods in detail and establish sufficient conditions for their conditional consistency. All the proofs and other technical details are presented in the appendix.

3.2 The setup

We consider a variation of network dependent processes characterized in Section 2.2 of Chapter 2. Namely, let $G \equiv (N, E)$ be an undirected graph (possibly infinite), where N is the set of nodes and E denotes the set of links (we identify N with integers $\{1, 2, \ldots\}$). Each edge $e \in E$ is associated with a weight $W(e) \in \bar{\mathbf{R}}$. Also let the function $d : N \times N \to \bar{\mathbf{R}}_{\geq 0}$ be a distance on G; for example, the shortest path distance for an unweighted graph. An $\mathcal{X}$-valued *network dependent process* $Y \equiv (Y, G)$ is a collection of $\mathcal{X}$-valued random elements defined on a common probability space indexed by N, i.e., $\{Y_i : i \in N\}$. The network G governs the stochastic dependence between random elements. Here we consider $\mathcal{X} = \mathbf{R}^v$ with $v \geq 1$.

Further, suppose that we are given a sequence of network dependent processes $\{(Y_n, G_n)\}$ defined on a common probability space $(\Omega, \mathcal{H}, \mathsf{P})$, where each $G_n \equiv (N_n, E_n)$ is a finite graph of size $m_n \to \infty$ as $n \to \infty$; w.l.o.g. we set $m_n = n$. Here, the sequence $\{G_n\}$ can be a sequence of subgraphs of an infinite network (N_∞, E_∞). In general, however, these graphs can be unrelated. In order to emphasize the dependence of the distance between two nodes on n, we denote it as $d_n(\,\cdot\,,\,\cdot\,)$. Additionally, since the distance function may implicitly depend on weights associated with the edges of a graph, we impose the following restriction in order to employ the results established in Chapter 2 with the least possible change.

Assumption 3.1. For all $n \geq 1$, $\min_{i,j \in N_n} d_n(i,j) \geq 1$ and $d_n(i,j) = \infty$ whenever $i, j \in N_n$ are disconnected (i.e., there is no path connecting i and j).

For example, if $W(e) \in [0,1]$ for all $e \in E$, which can be interpreted as the intensity of links, then the shortest weighed distance associated with $1/W(\,\cdot\,)$ satisfies this assumption, where implicitly we set $1/0 \equiv \infty$. In this case an unweighted network (N, E) is equivalent to a complete graph (N, E'), where for $e \in E'$, $W(e) = \mathbf{1}\{e \in E\}$. In a similar manner, the (at most countable) parameter space of a random field on a metric space $(\mathcal{Z}, \rho)$ can be modelled as a compete graph of suitable cardinality, where $W(x \leftrightarrow y)$ is a function of the distance ρ between two points $x, y \in \mathcal{Z}$. Then Assumption 3.1 corresponds to the case of increasing domain asymptotics (see, e.g., Conley, 1999; Jenish and Prucha, 2009).

Let $\mathcal{C} \subset \mathcal{H}$ be a given sub-σ-field. We assume that the sequence of network dependent processes is conditionally weakly dependent given $\mathcal{C}$. Specifically, for $a, b \in \mathbb{N}$ and $s \geq 0$ let

$$\mathcal{P}_n(a, b; s) \coloneqq \{(A, B) \subset N_n^2 : |A| = a, |B| = b, d_n(A, B) \geq s\},$$

with $d_n(A, B) \coloneqq \min_{i \in A, j \in B} d_n(i,j)$ and let $\mathcal{L}_v$ be the family of real-valued, bounded, Lipschitz functions, i.e., $\mathcal{L}_v \coloneqq \bigcup_{a \geq 1} \mathcal{L}_{v,a}$, where

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

The functions in $\mathcal{L}_{v,a}$ are Lipschitz with respect to the distance $\boldsymbol{\delta}_a$ on $\mathbf{R}^{v \times a}$ given by

$$\boldsymbol{\delta}_a(\mathbf{x}, \mathbf{y}) \coloneqq \sum_{l=1}^{a} \|x_l - y_l\|,$$

where $\|\cdot\|$ is a norm on $\mathbf{R}^v$ and $\mathbf{x} \equiv (x_1, \ldots, x_a)$ and $\mathbf{y} \equiv (y_1, \ldots, y_a)$ are points in $\mathbf{R}^{v \times a}$. In addition, for a set of nodes $A \subset N_n$ we write $Y_{n,A} \equiv \{Y_{n,i} : i \in A\}$.

Definition 3.1. The sequence $\{(Y_n, G_n)\}$ is $(\mathcal{L}_v, \psi, \mathcal{C})$-weakly dependent if for each $n \geq 1$ there exist a non-negative, $\mathcal{C}$-measurable sequence $\gamma_n \equiv \{\gamma_{n,s}\}_{s=1}^{\infty}$ such that $\sup_n \gamma_{n,s} \to 0$ a.s. as $s \to \infty$ and a collection of function $(\psi_{a,b})_{a,b \in \mathbb{N}}$, $\psi_{a,b} : \mathcal{L}_a \times \mathcal{L}_b \to \mathbf{R}_{\geq 0}$, such that for any $(A, B) \in \mathcal{P}_n(a, b; s)$ with $s \geq 1$, $f \in \mathcal{L}_{v,a}$, and $g \in \mathcal{L}_{v,b}$,

$$|\mathrm{Cov}(f(Y_{n,A}), g(Y_{n,B}) \mid \mathcal{C})| \leq \psi_{a,b}(f, g)\gamma_{n,\lfloor s \rfloor} \quad \text{a.s.} \tag{3.1}$$

Remark 3.1. (a) When it is clear from the context, we denote $\{(Y_n, G_n)\}$ as $\{Y_n\}$ omitting the reference to the associated networks. (b) $(\mathcal{L}_v, \psi) \equiv (\mathcal{L}_v, \psi, \{\emptyset, \Omega\})$. (c) The elements of $\{\gamma_{n,s}\}$ are called the weak-dependence coefficients associated with $\{Y_n\}$. (d) For convenience, we set $\gamma_{n,0} \equiv 1$.

A number of examples of network dependent processes that are $(\mathcal{L}_v, \psi, \mathcal{C})$-weakly dependent

are given in Section 2.2 of Chapter 2. For instance, strong mixing processes correspond to $\psi_{a,b}(f,g) = 4\|f\|_\infty \|g\|_\infty$. Also associated and Gaussian processes and their certain derivatives are $(\mathcal{L}_v, \psi, \mathcal{C})$-weakly dependent with $\psi_{a,b}(f,g) = ab \operatorname{Lip}(f) \operatorname{Lip}(g)$. It is worth mentioning that the corresponding weak dependence coefficients may depend on the topology of the underlying networks.

Conditioning on a σ-field $\mathcal{C}$ can be useful in various cases. First, if the underlying graphs are realizations of a stochastic network formation process, then one can potentially condition on the σ-field generated by that process and treat the observed graphs as fixed. Second, fixing nodes with high degree centrality may help to obtain local stochastic dependence.

Example 3.1. Consider a set independent random variables $\{\varepsilon_i : i \in N\}$ and let $C \subset N$ denote a set of nodes with "high" degree centrality (for clarity, we omitted the subscript n). Then $u_{N \setminus C}$, where $u_i := \varepsilon_i + \sum_{j \in C} \beta_{ij} \varepsilon_j$ and $\beta_{ij} \in \mathbf{R}$, are conditionally independent given ε_C. Moreover, for arbitrary measurable functions $\{\phi_i\}$ the process $\{Y_i := \phi_i(u_N)\}$ satisfies the covariance bound (3.1) with $\psi_{a,b}(f,g) = a\|g\|_\infty \operatorname{Lip}(f) + b\|f\|_\infty \operatorname{Lip}(g)$. In the context of social interaction models $\{u_i\}$ and $\{Y_i\}$ may represent idiosyncratic shocks and observable outcomes, respectively.

In order to facilitate the exposition, throughout the chapter we consider a sequence of network dependent processes $\{Y_n\}$ satisfying the covariance bound (3.1) with a specific form of the function $\psi_{a,b}$ and bounded weak dependence coefficients. The restricted $\psi_{a,b}$ function is fairly general and covers many useful examples of weakly dependent processes.

Assumption 3.2. $\{(Y_n, G_n)\}$ is $(\mathcal{L}_v, \psi, \mathcal{C})$-weakly dependent, there exist constants $M \geq 1$ and $C > 0$ such that $\gamma_{n,s} \leq M$ a.s. for all $n, s \geq 1$, and

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

where $c_1, \ldots, c_4 \leq Cab$.

It should be noted that processes satisfying Assumption 3.2 possess some hereditary properties. Specifically, if $\{Y_n\}$ is $(\mathcal{L}_v, \psi, \mathcal{C})$-weakly dependent with the weak dependence coefficients $\{\gamma_{n,s}\}$, then for any Lipschitz function $h : \mathbf{R}^v \to \mathbf{R}^w$ the sequence $\{h(Y_{n,i}) : i \in N_n\}$ is $(\mathcal{L}_w, \psi, \mathcal{C})$-weakly dependent with the same weak dependence coefficients. Moreover, this type of weak dependence is preserved under some locally Lipschitz functions as shown in Proposition 3.1 below, which is an extension of Proposition 2.1. in Dedecker, Doukhan, and Lang (2007) to our settings.

Proposition 3.1. *Suppose that $\{Y_n\}$ satisfies Assumption 3.2 and there exist $L < \infty$ and $p > 1$ s.t. $\sup_{n,i\in N_n} \mathsf{E}[\|Y_{n,i}\|_\infty^p \mid \mathcal{C}] \le L$ a.s. Let $h : \mathbf{R}^v \to \mathbf{R}^w$ be s.t.*

$$\|h(x) - h(y)\| \le \eta\|x - y\|\Big(\|x\|^{\tau-1} + \|y\|^{\tau-1}\Big) \tag{3.2}$$

for some $\eta > 0$ and $\tau \in [1, p)$. Then $\{h(Y_{n,i}) : i \in N_n\}$ is $(\mathcal{L}_w, \psi, \mathcal{C})$-weakly dependent with the weak dependence coefficients

$$\gamma'_{n,s} = KM\gamma^r_{n,s},$$

where K is a constants depending on η, v, and L and

$$r = \begin{cases} (p-\tau)/(p-1), & \text{if } c_4 = 0, \\ (p-\tau)/(p+\tau-2), & \text{otherwise.} \end{cases}$$

Remark 3.2. The boundedness of the conditional moments of $\|Y_{n,i}\|_\infty$ is required in order to maintain Assumption 3.2. Once this condition is relaxed, it suffices to assume that these moments are a.s. finite.

3.2.1 Asymptotic results

Introducing weighted networks is useful in several scenarios. First, as we have already mentioned it allows incorporating some additional random processes into the current framework. Second, assuming varying intensity of connections enables one to handle denser networks in the sense of the total number of links. Finally, some commonly used statistical models explicitly use weights and can be adapted to our framework, e.g., the spatial Cliff-Ord-type linear model in Kelejian and Prucha (2010).

Example 3.2. For each $n \ge 1$, let u_n be a $n \times 1$ vector of independent random variables and let $\widetilde{W}_n$ be an $n \times n$ matrix which is a function of weights associated with a given network. Consider a linear model with disturbances following the next autoregressive process:

$$\varepsilon_n = \lambda\widetilde{W}_n\varepsilon_n + u_n, \quad |\lambda| < 1,$$

Typically the original weighting matrix is modified to ensure that the spectral radius of $\widetilde{W}_n$ is bounded by 1. Under certain restrictions on the denseness of underlying networks, the process $\{\varepsilon_n\}$ is weakly dependent with $\psi_{a,b}(f, g) = a\|g\|_\infty \operatorname{Lip}(f) + b\|f\|_\infty \operatorname{Lip}(g)$ so that the model can be accommodated within the current framework.

Assume that $C_n := (I - \lambda\widetilde{W}_n)^{-1}$ exists for each $n \ge 1$ and $\mu := \sup_{n,i\in N_n} \mathsf{E}|u_{n,i}| < \infty$. Then $\varepsilon_n = C_n u_n$ and, letting $\varepsilon^{(s)}_{n,i} := \sum_{j\in N_n: d_n(i,j)<s+1} [C_n]_{ij} u_{n,j}$,

$$\mathsf{E}[|\varepsilon_{n,i} - \varepsilon^{(s)}_{n,i}|] \le \mu \max_{i\in N_n} \sum_{j\in N_n: d_n(i,j)\ge s+1} |[C_n]_{ij}| \equiv \gamma_{n,s}.$$

Thus, $\{\varepsilon_n\}$ is $(\mathcal{L}_1, \psi)$-weakly dependent provided that $\sup_n \gamma_{n,s} \to 0$ as $s \to \infty$. In order to appreciate the magnitude of the elements of C_n consider a simple case when $\widetilde{W}_n = A_n/\rho(A_n)$ and A_n is the adjacency matrix of the underlying graph. Then $C_n = \sum_{k\geq 0}(\lambda/\rho(A_n))^k A_n^k$, where $[A_n^k]_{ij}$ measures the number of paths of length k between nodes i and j, which is directly related to the denseness of the network.

For a given network G_n let $N_n(i;s)$ denote the open neighborhood of radius $s > 0$ around $i \in N_n$, i.e,

$$N_n(i;s) := \{j \in N_n : d_n(i,j) < s\},^1$$

and let $N_n^\partial(i;s) := N_n(i;s+1) \setminus N_n(i;s)$. In addition, we define the following aggregate measures of the network denseness:

$$\begin{aligned} \delta_n(s;k) &:= n^{-1}\sum_{i\in N_n}|N_n(i;s+1)|^k, \quad \delta_n^\partial(s;k) := n^{-1}\sum_{i\in N_n}|N_n^\partial(i;s)|^k, \\ D_n(s) &:= \max_{i\in N_n}|N_n(i;s+1)|, \quad \text{and} \quad D_n^\partial(s) := \max_{i\in N_n}|N_n^\partial(i;s)|. \end{aligned} \tag{3.3}$$

It is straightforward to see that under Assumption 3.1, which restricts the minimum distance between any two nodes of a network, the asymptotic results derived in Section 2.3 of Chapter 2 remain valid once we replace their measures of network denseness with those given in (3.3) and redefine $H_n(s,m)$ as follows:

$$H_n(s,m) := \left\{(i,j,k,l) \in N_n^4 : \begin{smallmatrix} j \in N_n(i;m+1), l \in N_n(k;m+1), \\ \lfloor d_n(\{i,j\},\{k,l\})\rfloor = s \end{smallmatrix}\right\}. \tag{3.4}$$

In the case of random networks, however, the measures of network denseness are also random. Therefore, one needs a conditional version of the LLN in order to be able to condition on the common shock $\mathcal{C}$. Note that the other result are stated in the conditional form and can be directly applied to this case if we assume certain measurability conditions. Let $\mathcal{D}(G_n)$ denote the distance matrix associated with G_n, i.e., $[\mathcal{D}(G_n)]_{ij} = d_n(i,j)$. If $\mathcal{D}(G_n)$ is $\mathcal{C}$-measurable, then $N_n(i;s) = \sum_{j\in N_n}\mathbf{1}\{[\mathcal{D}(G_n)]_{ij} < s\}$ is also $\mathcal{C}$-measurable as well as the quantities given in (3.3) and (3.4). We make the following assumption:

Assumption 3.3. The distance matrix $\mathcal{D}(G_n)$ is $\mathcal{C}$-measurable for all $n \geq 1$.

In addition, we introduce the notion of asymptotically negligible random functions, which is useful for defining the conditional versions of the asymptotic tightness and uniform integrability.

Definition 3.2. Let $\mathcal{F} \subset \mathcal{H}$ and let $f : \mathcal{Y} \times \Omega \to \mathbf{R}_{\geq 0}$ be such that $f(y, \cdot)$ is $\mathcal{F}$-measurable for all $y \in \mathcal{Y}$. A sequence of such functions $\{f_n\}$ is *asymptotically negligible* (a.n.) if for

[1] Note that this definition of the open neighborhood of a node differs from one commonly used in graph theory.

almost all $\omega \in \Omega$,

$$\operatorname*{ess\,inf}_{s\in S} \limsup_{n\to\infty} f_n(s,\omega) = 0.^2$$

In particular, an array of random vectors $\{X_{n,i}\}$ is

– *$\mathcal{F}$-asymptotically tight* if $\max_i \mathsf{P}(\|X_{n,i}\| > y \mid \mathcal{F})$ is a.n. and
– *$\mathcal{F}$-asymptotically uniformly integrable* (u.i.) if $\max_i \mathsf{E}[\|X_{n,i}\|\mathbf{1}\{\|X_{n,i}\| > y\} \mid \mathcal{F}]$ is a.n.

Theorem 3.1 (Conditional Weak Law of Large Numbers)**.** *Let $\{(Y_n, G_n)\}$ be $(\mathcal{L}_v, \psi, \mathcal{C})$-weakly dependent satisfying Assumption 3.1, 3.2, and 3.3. Suppose that $\{Y_n\}$ is $\mathcal{C}$-asymptotically u.i. and*

$$\frac{1}{n}\sum_{s\geq 1} \delta_n^{\partial}(s;1)\gamma_{n,s} \to 0 \quad a.s.$$

Then

$$\left\| \frac{1}{n}\sum_{i\in N_n} (Y_{n,i} - \mathsf{E}[Y_{n,i} \mid \mathcal{C}]) \right\|_{\mathcal{C},1} \to 0 \quad a.s.^3$$

Remark 3.3. Similarly to the unconditional case a sufficient condition for the $\mathcal{C}$-asymptotic uniform integrability of $\{Y_n\}$ is the a.s. finiteness of $\sup_{n,i\in N_n} \mathsf{E}[\|Y_{n,i}\|^p \mid \mathcal{C}]$ for some $p > 1$.

Finally, let $\bar{Y}_n := n^{-1}\sum_{i\in N_n} Y_{n,i}$ and $\Sigma_n := \mathrm{Var}(\sqrt{n}\bar{Y}_n \mid \mathcal{C})$. Then the network HAC estimator of Σ_n,

$$\hat{\Sigma}_n = \frac{1}{n}\sum_{i,j\in N_n} \kappa\left(\frac{d_n(i,j)}{b_n+1}\right)(Y_{n,i} - \bar{Y}_n)(Y_{n,j} - \bar{Y}_n)^\top, \tag{3.5}$$

where $\kappa : \bar{\mathbf{R}} \to [-1,1]$ is a kernel function satisfying: $\kappa(0) = 1$, $\kappa(z) = \kappa(-z)$, and $\kappa(z) = 0$ for $|z| > 1$ and b_n is the lag truncation parameter, is consistent under the same set of assumptions. Unfortunately, due to the irregularity of a network's structure, this estimator is not guaranteed to be positive semi-definite. However, once the minimal eigenvalue of Σ_n is a.s. bounded from below or it converges to an a.s. positive definite matrix, a simple way to fix this issue is available. The details are given in Appendix B.2.

3.3 Conditional bootstrap

In this section we present some general result regarding conditional bootstrap. The latter is useful for an inference which is asymptotically valid for almost all $\omega \in \Omega$ (or almost all

[2]The essential infimum of an arbitrary family of random variables $\{Z_\alpha : \alpha \in \mathcal{A}\}$ a random variable Z s.t. (a) $Z \leq Z_\alpha$ a.s. for all $\alpha \in \mathcal{A}$ and (b) $Z \geq Z'$ a.s. for any random variable Z' satisfying $Z' \leq Z_\alpha$ a.s. for all $\alpha \in \mathcal{A}$. In particular, there exists a sequence $\{\alpha_n\}$ s.t. $Z = \inf_n Z_{\alpha_n}$ a.s. (see, e.g., Cohen and Elliott, 2015, Theorem 1.3.40). The essential supremum is defined similarly and the following identity holds: $\operatorname{ess\,sup}_\alpha\{Z_\alpha\} = -\operatorname{ess\,inf}_\alpha\{-Z_\alpha\}$.

[3]For a random vector X and $\mathcal{F} \subset \mathcal{H}$ we write $\|X\|_{\mathcal{F},p} \equiv \mathsf{E}[\|X\|^p \mid \mathcal{F}]^{1/p}$.

realizations of the common shock). These results do not depend on the underlying data generating process. However, we use the present framework for convenience.

Suppose that $\{(Y_n, G_n)\}$ is a sequence of network dependent processes. For a given $n \geq 1$ let θ_n be a $\mathcal{C}$-measurable parameter taking values in $\Theta \subseteq \mathbf{R}^w$ with $w \geq 1$ and let

$$T_n(\theta_n) \coloneqq \mathcal{T}_n(Y_n, \theta_n; \vartheta_n),$$

where $\mathcal{T}_n$ is a measurable, real-valued function and ϑ_n is a $\mathcal{C}$-measurable nuisance parameter, denote a statistic used to conduct inference on θ_n based on a realization of (Y_n, G_n) conditionally on $\mathcal{C}$.

Let F_n denote the conditional cdf of T_n given $\mathcal{C}$.[4] The goal of this section is to provide sufficient conditions for the conditional first-order consistency of resampling estimators of F_n. Specifically, let $\mathcal{G}_n \coloneqq \mathcal{C} \vee \sigma(Y_n)$ and let Y_n^* be a pseudo-sample drawn using a realization of Y_n. Then the bootstrap counterpart of T_n is $T_n^* \coloneqq \mathcal{T}_m(Y_n^*, \theta_n^*)$, where m is the size of Y_n^* and $\theta_n^* \equiv \theta_n^*(Y_n)$ is an estimator of θ_n. The conditional cdf F_n^* of T_n^* is used as an approximation of F_n. If the latter explicitly depends on the nuisance parameter ϑ_n, then one needs to provide its consistent estimator based on both Y_n and Y_n^*.

A typical way of showing the consistency of the bootstrap estimators is bounding the Kolmogorov distance between the cdfs of T_n and T_n^* (see, e.g., Shao and Tu, 1995, Chapter 3). For random variables X and Y and sub-σ-fields $\mathcal{F} \subset \mathcal{G} \subset \mathcal{H}$ the conditional version of the latter is defined by

$$d_K(X, Y \mid \mathcal{G}, \mathcal{F}) \coloneqq \sup_{x \in \mathbf{R}} \left| F_X^{\mathcal{G}}(\,\cdot\,, x) - F_Y^{\mathcal{F}}(\,\cdot\,, x) \right|,^{5}$$

where $F_X^{\mathcal{G}}$ and $F_Y^{\mathcal{F}}$ are the conditional cdfs of X and Y, respectively (when $\mathcal{F} = \mathcal{G}$ we denote this measure by $d_K(X, Y \mid \mathcal{F})$). In addition, we define the conditional convergence in probability and the almost sure convergence of conditional distributions.[6]

Definition 3.3. Let $\mathcal{F} \subset \mathcal{H}$ be a sub-σ-field and let Z be a $\mathcal{F}$-measurable random vector in $\mathbf{R}^v$ with $v \geq 1$. A sequence of $\mathbf{R}^v$-valued random vectors $Z_n \xrightarrow{\mathcal{F}-p} Z$ a.s. if for any $\epsilon > 0$, $\mathsf{P}(\|Z_n - Z\| > \epsilon \mid \mathcal{F}) \to 0$ a.s.[7]

[4] A (regular) conditional cdf $F_X^{\mathcal{F}}$ of $X \in \mathbf{R}$ given $\mathcal{F} \subset \mathcal{H}$ satisfies: (i) $\forall x \in \mathbf{R}$, $F_X^{\mathcal{F}}(\,\cdot\,, x)$ is a version of $\mathsf{P}(X \leq x \mid \mathcal{F})$, and (ii) $\forall \omega \in \Omega$, $F_X^{\mathcal{F}}(\omega, \cdot\,)$ is a distribution function. We omit the subscript X or the superscript $\mathcal{F}$ whenever clear from the context.

[5] Note that $d_K(\,\cdot\,, \cdot \mid \mathcal{G}, \mathcal{F})$ is $\mathcal{G}$-measurable because $\{Z_x\}$, where $Z_x \coloneqq \left| F_X^{\mathcal{G}}(\,\cdot\,, x) - F_Y^{\mathcal{F}}(\,\cdot\,, x) \right|$, is a cádlág stochastic process).

[6] A (regular) conditional distribution $\mathsf{Q}_X^{\mathcal{F}}$ of $X \in \mathbf{R}^v$ given $\mathcal{F} \subset \mathcal{H}$ satisfies: (i) $\forall B \in \mathcal{B}(\mathbf{R}^v)$, $\mathsf{Q}_X^{\mathcal{F}}(\,\cdot\,, B)$ is a version of $\mathsf{P}(X \in B \mid \mathcal{F})$ and (ii) $\forall \omega \in \Omega$, $\mathsf{Q}_X^{\mathcal{F}}(\omega, \cdot\,)$ is a probability measure on $(\mathbf{R}, \mathcal{B}(\mathbf{R}^v))$. We omit the subscript X or the superscript $\mathcal{F}$ whenever clear from the context.

[7] Note that this implies convergence in probability due to the dominated convergence theorem. In addition, for an a.s. positive $\mathcal{F}$-measurable random variable ν, $\mathsf{P}(\|Z_n - Z\| > \nu \mid \mathcal{F}) \to 0$ a.s.

Definition 3.4. Suppose that $\{X_n\}$ is a sequence of random vectors on $(\Omega, \mathcal{H}, \mathsf{P})$ and $\mathcal{F} \subset \mathcal{H}$. Let Q_n be the conditional distribution of X_n given $\mathcal{F}$. We say that X_n converges $\mathcal{F}$-weakly to X having the conditional distribution Q if for almost all $\omega \in \Omega$ the sequence $\{\mathsf{Q}_n(\omega, \cdot)\}$ converges weakly to $\mathsf{Q}(\omega, \cdot)$.

Remark 3.4. (a) Equivalently, the $\mathcal{F}$-weak convergence can be defined using the notion of probability kernels. So the limiting random vector X is an artificial construct which is used to describe the limiting kernel. (b) A more general notion of the almost sure convergence of conditional probability measures and some of its properties are presented in Berti, Pratelli, and Rigo (2006). (c) The notion of the $\mathcal{F}$-weak convergence is stronger than the $\mathcal{F}$-stable convergence and the usual weak convergence. In particular, if $X_n \to X$ $\mathcal{F}$-weakly, then for any real-valued, bounded, continuous function f, $\mathsf{E}[f(X_n) \mid \mathcal{F}] \to \mathsf{E}[f(X) \mid \mathcal{F}]$ a.s. which implies that X_n converges to X $\mathcal{F}$-stably and in distribution.

Assume for a moment that $\mathcal{C} = \{\emptyset, \Omega\}$. Then if there exists a sequence of random variables $\{S_n\}$ such that $d_K(T_n, S_n \mid \mathcal{C})$ converges to 0 as $n \to \infty$ and $d_K(T_n^*, S_n \mid \mathcal{G}_n, \mathcal{C})$ converges to 0 a.s. (in probability), then the bootstrap estimator is first-order strongly (weakly) consistent. Moreover, if S_n converges weakly to a continuous limit, then the conditional quantiles of F_n^* are a good approximation to those of F_n. This typically happens when the statistic T_n is pivotal. However, in the case of a non-pivotal statistic, which is useful when a consistent estimator of ϑ_n is hard to obtain or the available estimators have poor finite sample properties, the cdfs of $\{T_n\}$ need not converge.[8] In this case, the convergence of the Kolmogorov distance between T_n^* and T_n to zero does not necessarily imply that $F_n(c_n^*(\alpha)) \to \alpha$ as $n \to \infty$, where $c_n^*(\alpha)$ is the conditional α-quantile of F_n^*. Nevertheless, as shown in the next result, a sufficient condition for the latter to happen is the continuity of the cdfs of $\{S_n\}$.

Theorem 3.2. *Suppose that for all $n \geq 1$, S_n is conditionally independent of Y_n given $\mathcal{C}$ and the conditional cdf of S_n given $\mathcal{C}$ is (a.s.) continuous. Then if*

(a) $d_K(T_n, S_n \mid \mathcal{C}) \to 0$ *a.s. and*
(b) $d_K(T_n^*, S_n \mid \mathcal{G}_n, \mathcal{C}) \xrightarrow{\mathcal{C}-p} 0$ *a.s.,*

$d_K(T_n^*, T_n \mid \mathcal{G}_n, \mathcal{C}) \xrightarrow{\mathcal{C}-p} 0$ *a.s. and*

$$\operatorname*{ess\,sup}_{\alpha \in (0,1)} |\mathsf{P}(T_n \leq c_n^*(\alpha) \mid \mathcal{C}) - \alpha| \to 0 \quad a.s.$$

Remark 3.5. (a) Usually when $\mathcal{C} = \{\emptyset, \Omega\}$ and the statistic T_n is pivotal, we have $S_n = S_\infty$, which is the weak limit of T_n. (b) A variant of this result in the context of a Gaussian

[8] Consider, for example, the case of the linearized statistic T_n' given in (3.6). It does not have a nondegenerate weak limit when the sequence of parameters $\{\theta_n\}$ is not convergent.

multiplier bootstrap can be found in Chernozhukov et al. (2013). (c) Theorem 3.2 also implies that the conditional quantiles $\{c_n^*(\alpha) : \alpha \in (0,1)\}$ approximate the unconditional quantiles of T_n because, by the dominated convergence theorem,

$$\sup_{\alpha\in(0,1)} |\mathsf{P}(T_n \le c_n^*(\alpha)) - \alpha| \le \mathsf{E} \operatorname*{ess\,sup}_{\alpha\in(0,1)} |\mathsf{P}(T_n \le c_n^*(\alpha) \mid \mathcal{C}) - \alpha| \to 0.$$

Definition 3.5. We say that F_n^* is conditionally d_K-consistent given $\mathcal{C}$ if the conclusion of Theorem 3.2 holds.

Typically it is not hard to show that condition (a) of Theorem 3.2 holds (for example, when the elements of Y_n^* are conditionally i.i.d. given $\mathcal{G}_n$). On contrary, establishing (b) may be a difficult task, especially when T_n is a nonlinear transformation of Y_n in the presence of stochastic dependence between its elements as in the current framework. However, in the case when the statistic T_n converges $\mathcal{C}$-weakly to S and the limiting kernel (i.e., the regular conditional cdf of S given $\mathcal{C}$) is continuous, Lemma B.4 implies that this convergence is equivalent to one with respect to the conditional Kolmogorov distance. In addition, by Lemma B.3 the almost sure convergence of conditional distributions enjoys a number of useful properties associated with the usual weak convergence such as the continuous mapping theorem, converging together lemma, and the Cramér–Wold device. In this situation we have the following simple corollary.

Corollary 3.1. *Suppose that S is conditionally independent of $\{Y_n\}$ given $\mathcal{C}$ and the conditional cdf of S given $\mathcal{C}$ is (a.s.) continuous. Then if*

(a) $T_n \to S$ *$\mathcal{C}$-weakly and*
(b) $d_K(T_n^*, S \mid \mathcal{G}_n, \mathcal{C}) \xrightarrow{\mathcal{C}-p} 0$ *a.s.,*

F_n^* is conditionally d_K-consistent given $\mathcal{C}$.

Next, we consider the case in which the statistic T_n takes the following form:

$$T_n(\theta_n) = \tau_n\Big(\phi(\hat\theta_n) - \phi(\theta_n)\Big),$$

where $\phi : \Theta \to \mathbf{R}$ is a continuously differentiable function, $\hat\theta_n$ is a consistent estimator of θ_n (in the sense of Definition 3.3) and τ_n is a normalizing coefficient. In particular, the *smooth function model* (see, e.g., Lahiri, 2003, Section 4.2 and Hall, 1992, Section 2.4) falls into this case. The resampling version of the statistic T_n is

$$T_n^* = \tau_n^*\Big(\phi(\hat\theta_n^*) - \phi(\theta_n^*)\Big),$$

where θ_n^* is a consistent estimator of θ_n, which may differ from $\hat\theta_n$, and τ_n^* is the bootstrap counterpart of τ_n. Let $\xi_n := \tau_n(\hat\theta_n - \theta_n)$ and $\xi_n^* := \tau_n^*(\hat\theta_n^* - \theta_n^*)$. Consider the linearized

statistics

$$T'_n := \nabla\phi(\theta_n)^\top \zeta_n \quad \text{and} \quad T'^*_n := \nabla\phi(\theta^*_n)^\top \xi^*_n. \tag{3.6}$$

The following result shows that it suffices to find a "smooth" approximation S'_n of the linearized statistics in order to apply Theorem 3.2 to this setup. In particular, the result largely depends on the asymptotic behavior of the conditional Lévy concentration function of S'_n. For a random variable X, $\epsilon > 0$, and a sub-σ-field $\mathcal{F} \subset \mathcal{H}$ the latter is given by

$$Q(\epsilon, X \mid \mathcal{F}) := \sup_{x\in\mathbf{R}} \big(F^{\mathcal{F}}_X(\,\cdot\,, x+\epsilon) - F^{\mathcal{F}}_X(\,\cdot\,, x-)\big).$$

Lemma 3.1. *Suppose that $\hat\theta^*_n - \theta^*_n \xrightarrow{\mathcal{C}-p} 0$ a.s., ξ^*_n and ξ_n are $\mathcal{C}$-asymptotically tight and $\sup_n\|\theta_n\| < \infty$ a.s. Furthermore, assume that*

(a) $d_K(T'_n, S'_n \mid \mathcal{C}) \to 0$ *a.s.,*
(b) $d_K(T'^*_n, S'_n \mid \mathcal{G}_n, \mathcal{C}) \xrightarrow{\mathcal{C}-p} 0$ *a.s., and*
(c) $Q(\epsilon, S'_n \mid \mathcal{C})$ *is a.n.*

Then w.p.1,

$$d_K\big(T^*_n, S'_n \mid \mathcal{G}_n, \mathcal{C}\big) \xrightarrow{\mathcal{C}-p} 0 \quad \textit{and} \quad d_K\big(T_n, S'_n \mid \mathcal{C}\big) \to 0.$$

Consequently, the continuity of the conditional cdfs of $\{S'_n\}$ ensures the bootstrap consistency in the sense of Definition 3.5.

Theorem 3.3. *Suppose that the conditions of Lemma 3.1 hold and, in addition, $\{S_n\}$ satisfy the independence and continuity conditions of Theorem 3.2. Then F^*_n is conditionally d_K-consistent given $\mathcal{C}$.*

Similarly to the general case, when ξ_n converges $\mathcal{C}$-weakly to some random vector ξ and the sequence of parameters $\{\theta_n\}$ converges a.s. to a $\mathcal{C}$-measurable random variable θ, Lemma B.3 implies that T'_n converges $\mathcal{C}$-weakly to $\nabla\phi(\theta)^\top\xi$. In addition, if the conditional cdf of the latter is (a.s.) continuous, it satisfies assumption (b) of Lemma 3.1.

Corollary 3.2. *Suppose that $\hat\theta^*_n - \theta^*_n \xrightarrow{\mathcal{C}-p} 0$ a.s., ξ^*_n is $\mathcal{C}$-asymptotically tight and $S' := \nabla\phi(\theta)^\top\xi$ satisfies the independence and continuity conditions of Corollary 3.1. Then if*

(a) $\xi_n \to \xi$ *$\mathcal{C}$-weakly,*
(b) $\theta_n \to \theta$ *a.s., and*
(c) $d_K(T'^*_n, S' \mid \mathcal{G}_n, \mathcal{C}) \xrightarrow{\mathcal{C}-p} 0$ *a.s.,*

F^*_n is conditionally d_K-consistent given $\mathcal{C}$.

Remark 3.6. The assumption regarding convergence of the sequence of parameters $\{\theta_n\}$ can be relaxed. In the unconditional case it suffices to assume that $\sup_n \|\theta_n\| < \infty$. Then one needs to provide a uniform bound on $|\mathsf{P}(\xi_n \in A) - \mathsf{P}(\xi \in A)|$, where A ranges over the class of half-spaces for a network dependent process similar to that established in Bentkus (2003). The conditioning on $\mathcal{C}$ complicates the problem even more so it falls out of the scope of this work.

3.4 Bootstrap of the mean

Consider a sequence of network dependent processes $\{(Y_n, G_n)\}$ satisfying Assumptions 3.1, 3.2, and 3.3. As an application of the results given in the preceding section, we consider the mean of a Y_n, $\mu_n \equiv \mathsf{E}[Y_{n,i} \mid \mathcal{C}]$, which may vary with n but not across $i \in N_n$. The parameter of interest μ_n is estimated using the sample mean $\bar{Y}_n$, which is a consistent estimator of μ_n under the assumptions of Theorem 3.1. In this section we provide a number of resampling based methods for constructing the asymptotically valid confidence sets for μ_n. In addition, we establish consistency of a restricted version of the *smooth function model*, in which we are interested in $\phi(\mu_n)$ for a continuously differential function $\phi : \mathbf{R}^v \to \mathbf{R}$. When the elements of Y_n have the same marginal conditional distributions given $\mathcal{C}$, we may consider $\phi(\mathsf{E}[f(Y_{n,1}) \mid \mathcal{C}])$, where $f : \mathbf{R}^v \to \mathbf{R}^w$ is a locally Lipschitz function satisfying (3.2) and the domain of ϕ is $\mathbf{R}^w$ in this case. Since the process $\{f(Y_{n,i}) : i \in N_n\}$ is $(\mathcal{L}_w, \psi, \mathcal{C})$-weakly dependent by Proposition 3.1, without loss of generality we examine the first version. In addition, we provide consistent positive semi-definite estimators of Σ_n.

The corresponding test statistics are given by

$$\begin{aligned} T_{1,n}(\mu_n) &= \sqrt{n}\|\bar{Y}_n - \mu_n\|, \quad \text{and} \\ T_{2,n}(\mu_n) &= \sqrt{n}\big(\phi(\bar{Y}_n) - \phi(\mu_n)\big), \end{aligned} \tag{3.7}$$

where $\|\cdot\|$ is the Euclidean norm on $\mathbf{R}^v$. Their conditional distributions given $\mathcal{C}$ are denoted by $F_{1,n}$ and $F_{2,n}$, respectively, and the bootstrap approximations of these distributions are denoted by $F^*_{1,n}$ and $F^*_{2,n}$. The confidence sets for μ_n are obtained by test inversion, i.e

$$CS_{n,1-\alpha} := \big\{\mu \in \mathbf{R}^v : T_{1,n}(\mu) \le c^*_{n,1-\alpha}\big\},$$

where $c^*_{n,\alpha}(\omega) := \inf\{x : F^*_{1,n}(\omega, x) \ge \alpha\}$ is the conditional α-quantile of $F^*_{1,n}$. In practice, if the exact distribution of $T^*_{j,n}$, $j = 1, 2$ is not available, it can be replaced with a suitable Monte-Carlo estimator.

3.4.1 Block bootstrap

First, we suggest a variant of the block bootstrap, which is extensively studied in the time-series and spatial literature. Specifically, we choose the maximal block radius $s_n > 0$ and define n overlapping blocks $\{B_{n,1}, \ldots, B_{n,n}\}$ with $B_{n,k} := N_n(k; s_n + 1)$. That is $B_{n,k}$ is an $(s_n + 1)$ open neighborhood of the node k. Then we randomly select $K_n := \lfloor n/\delta_n(s_n) \rfloor$ blocks $\{B^*_{n,1}, \ldots, B^*_{n,K_n}\}$ with *replacement* (note that $\delta_n(s_n)$ is the average block size), which yields a bootstrap sample

$$Y^*_n = \big\{Y_{n,B^*_{n,k}} : 1 \le k \le K_n\big\}.$$

Formally, let $\{u_1, \ldots, u_{K_n}\}$ be i.i.d. $U\{1, n\}$ random variables defined on $(\Omega, \mathcal{H}, \mathsf{P})$ and independent of $\mathcal{G}_n$. Then the k-th resampled block is defined as $B^*_{n,k} = B_{n,u_k}$ and, therefore, for $1 \le k \le K_n$ and $1 \le l \le n$, $\mathsf{P}\Big(B^*_{n,k} = B_{n,l} \mid \mathcal{G}_n\Big) = n^{-1}$ a.s. For the ease of exposition we assume that $n/\delta_n(s_n)$ is an integer.

The size of the bootstrap sample $L_n := \sum_{k=1}^{K_n} |B^*_{n,k}|$ is random conditional on the data and depends on the distribution of $|N_n(\,\cdot\,; s_n)|$ given the network G_n. However, on average it is expected to be close to n, (in fact, the conditional expectation of L_n given $\mathcal{C}$ is exactly n). Also in the time series case this approach reduces to a variant of the moving blocks bootstrap with unequally sized blocks such that blocks located near the endpoints have smaller size.

Let $Z^*_{n,k} := \sum_{j \in B^*_{n,k}} Y_{n,j}$ and let $\tilde{Y}^*_n := n^{-1} \sum_{k=1}^{K_n} Z^*_{n,k}$ be the quasi-average of the bootstrap sample Y^*_n, which replaces the sample average in the bootstrap versions of $T_{1,n}$ and $T_{2,n}$. We could also consider the true average of a pseudo-sample, i.e., $\bar{Y}^*_n := L_n^{-1} \sum_{k=1}^{K_n} Z^*_{n,k}$. However, L_n is not independent of the blocks sums and, as mentioned before, its distribution depends on the underlying network structure. As a result, it is relatively difficult to find a "smooth" approximation of the distribution of $\sqrt{L_n}\bar{Y}^*_n$ which guarantees the first-order consistency of the bootstrap (in particular, the suggested resampling scheme may not be appropriate in this case). In addition, since the conditional expectation of $\tilde{Y}_n$ given $\mathcal{G}_n$ differs from the sample average, we replace the true parameter μ_n with $\mu^*_n := \mathsf{E}[\tilde{Y}^*_n \mid \mathcal{G}_n]$. As indicated in Lahiri (1992) in the time-series context, replacing μ_n with $\bar{Y}_n$ introduces an additional bias which does not allow for second-order improvements over the normal approximation (see also Lahiri, 2003, Section 2.7.1). The bootstrap counterparts of the test statistics in (3.7) are given by

$$T^*_{1,n} = \sqrt{n}\|\tilde{Y}^*_n - \mu^*_n\|, \quad \text{and}$$
$$T^*_{2,n} = \sqrt{n}\big(\phi(\tilde{Y}^*_n) - \phi(\mu^*_n)\big).$$

The conditional variance of the scaled sample mean Σ_n can be estimated using the bootstrap version $\Sigma^*_n \equiv \mathrm{Var}(\sqrt{n}\tilde{Y}^*_n \mid \mathcal{G}_n)$. Since $\{Z^*_{n,1}, \ldots, Z^*_{n,K_n}\}$ are conditionally independent

given $\mathcal{G}_n$,

$$\Sigma_n^* = \frac{1}{\delta_n(s_n)}\left(\frac{1}{n}\sum_{i\in N_n} Z_{n,i}Z_{n,i}^\top - \bar{Z}_n\bar{Z}_n^\top\right) \quad \text{a.s.},$$

where $Z_{n,i} := \sum_{j\in B_{n,j}} Y_{n,j}$ and $\bar{Z}_n := n^{-1}\sum_{i\in N_n} Z_{n,i}$. By construction the matrix Σ_n^* is positive semidefinite and its form is similar to the network HAC estimator (3.5). To see this let

$$\omega_n(i,j) := \frac{|N_n(i;s_n+1)\cap N_n(j;s_n+1)|}{\delta_n(s_n)} \tag{3.8}$$

(when $i=j$ we denote this quantity by $\omega_n(i)$). Then

$$\Sigma_n^* = \frac{1}{n}\sum_{i,j\in N_n} \omega_n(i,j)(Y_{n,i}-\mu_n)(Y_{n,j}-\mu_n)^\top + R_n \quad \text{a.s.},$$

where $\mathsf{E}[\|R_n\|_F \mid \mathcal{C}] \to 0$ a.s. under some conditions, given later. It is worth mentioning that when $\mu_n = 0$ a.s., the remainder term $R_n = 0$ a.s. Unlike a typical kernel, the weighting functions $\omega_n(\,\cdot\,,\,\cdot\,)$ depends on the network topology and it is not bounded by 1. However, for fixed $i \in N_n$ it is decreasing in the distance between i and j. Let $\tilde{\omega} := \sup_n \max_{i\neq j}\omega_n(i,j)$, $\tilde{\mu}_p := \sup_{n,i\in N_n}\|Y_{n,i}\|_{\mathcal{C},p}$ for $p>0$, and

$$\Delta_n(s;k) := \frac{1}{n}\sum_{i\in N_n} \big||N_n(i;s+1)| - \delta_n(s)\big|^k, \tag{3.9}$$

which is the k-th absolute central moment of the sizes of the $(s+1)$-neighborhoods. The following assumptions provide sufficient conditions for the consistency of Σ_n^*.

Assumption 3.4. The sequence $\{(G_n, s_n)\}$ is such that w.p.1 $\tilde{\omega} < \infty$ and

(a) $\Delta_n(s_n;2)/\delta_n(s_n) + D_n(s_n)/\sqrt{\delta_n(s_n)n} \to 0$,
(b) $\max_{i\in N_n}\Big|\sum_{j\in B_{n,i}}(\omega_n(j)-1)\Big|/\sqrt{n} \to 0$,
(c) $n^{-1}\sum_{i\in N_n}\sum_{j\in N_n^\partial(i;s)}|\omega_n(i,j)-1|\gamma_{n,s} \to 0$ for all $s \geq 1$.

Assumption 3.4 imposes restrictions on the admissible networks topologies. Specifically, the consistency of Σ_n^* requires a certain degree of homogeneity of the resampled blocks, which is characterized by various moments of the weights $\{\omega_n(i,j) : i,j \in N_n\}$. For example, condition (a) requires that the sample variance of $\{|B_{n,i}|\}$ increases at a lower rate than the average block size. It also guarantees that μ_n^* is a consistent estimator of the mean μ_n and that for large samples the size of a pseudo-sample, L_n is close to n. In fact,

$\mathsf{E}[|L_n/n - 1| \mid \mathcal{C}] \to 0$ a.s. because

$$\begin{aligned}\mathsf{E}[|L_n/n - 1| \mid \mathcal{C}] &= \frac{1}{n}\mathsf{E}\left[\left|\sum_{k=1}^{K_n}\left(\left|B^*_{n,k}\right| - \delta_n(s_n)\right)\right| \mid \mathcal{C}\right] \\ &\leq \frac{1}{n}\sum_{k=1}^{K_n}\Delta_n(s_n;1) \leq \frac{\sqrt{\Delta_n(s_n;2)}}{\delta_n(s_n)} \quad \text{a.s.}\end{aligned}$$

This condition is clearly satisfied in the time series context when $s_n = o(\sqrt{n})$ (although, it has been shown that the consistency of the moving block bootstrap in this case holds for $s_n = o(n)$ (see, e.g., Calhoun, 2018)). However, it does not hold for unweighted "star" networks and $s_n \equiv 1$ because $\Delta_n(1;2) \geq [\Delta_n(1;1)]^2 \to 4$ and $\delta_n(1) \to 3$ as $n \to \infty$. In practice, one can compute $\Delta_n(s_n;2)$ for a given graph to see whether this quantity is small relative to the average block size.

Condition (c) ensures that all the non-zero autocovariances are estimated consistently. It is similar to an assumption on kernel functions used in HAC estimation, that is in the limit the value of a kernel at each s must converge to 1. In addition, if $\tilde{\gamma}_s > 0$ with positive probability for all $s \geq 1$, then the parameter s_n must go to infinity for this condition to hold.

Assumption 3.5. There exists $r > 2$ such that w.p.1 $\tilde{\mu}_{2r} < \infty$ and

(a) $\limsup_{n\to\infty}\sum_{s\geq 1}\delta_n^{\partial}(s)\gamma_{n,s}^{1-\frac{2}{r}} < \infty$,

(b) $n^{-2}\sum_{s\geq 0}|H_n(s, 2s_n+1)|\gamma_{n,s}^{1-\frac{2}{r}} \to 0$.

The conditions of Assumption 3.5 are similar to those needed for the consistency of the network HAC estimator (3.5). In particular, condition (b) gives a rule of thumb for the choice of the truncation parameter s_n (see Section 2.4 of Chapter 2). Also condition (a) implies that the elements of the true variance Σ_n do not diverge to $\pm\infty$. To see this note that for $1 \leq k, l \leq v$ and some constant $C > 0$,

$$|[\Sigma_n]_{kl}| \leq C(\tilde{\mu}_{2r}^2 \vee 1)\left(1 + \sum_{s\geq 1}\delta_n^{\partial}(s)\gamma_{n,s}^{1-\frac{2}{r}}\right) \quad \text{a.s.}$$

Therefore, $\limsup_{n\geq 1}|[\Sigma_n]_{kl}| < \infty$ a.s.

Proposition 3.2. *Suppose that Assumptions 3.4 and 3.5 hold. Then*

$$\mathsf{E}[\|\Sigma_n^* - \Sigma_n\|_F \mid \mathcal{C}] \to 0 \quad \textit{a.s.}$$

The result of Proposition 3.2 implies that Σ_n^* is a consistent estimator of Σ_n. Therefore, assuming that $\Sigma_n \to \Sigma$ a.s. and $\sqrt{n}(\bar{Y}_n - \mu_n)$ converges $\mathcal{C}$-weakly to a conditionally normal random vector with variance Σ, we may use Corollaries 3.1 and 3.2 to establish the consistency

of the bootstrap distributions. For example, one may employ Theorem 2.2 in Chapter 2 together with the Cramér–Wold device and Lemma B.4.

Assumption 3.6. Σ_n converges a.s. a $\mathcal{C}$-measurable, a.s. positive definite matrix Σ, and

$$\sqrt{n}(\bar{Y}_n - \mu_n) \to \Sigma^{1/2}\eta \quad \mathcal{C}\text{-weakly},$$

where $\eta \sim \mathcal{N}(0, I_v)$ independent of $\mathcal{C}$.[9]

In addition, we introduce the local versions of some measures of the network denseness. Specifically, for $s, m \geq 0$ we define

$$\delta^{\partial}_{loc,n}(s,m) := \max_{i \in N_n} \sum_{j \in N_n(i;m)} \frac{\left|N^{\partial}_n(j;s) \cap N_n(i;m)\right|}{|N_n(i;m)|}$$

and

$$h_{loc,n}(s,m) := \max_{i \in N_n} \frac{\left|H_n(s,\infty) \cap N^4_n(i;m)\right|}{|N_n(i;m)|^3}.$$

These measures are constructed in a way such that for any $m \geq 0$, $\delta^{\partial}_{loc,n}(0,m) = h_{loc,n}(0,m) = 1$. Also note that $h_{loc,n}(s,m) \leq \delta^{\partial}_{loc,n}(s,m)$.

Assumption 3.7. There exists $p > 2$ such that w.p.1 $\tilde{\mu}_{2p} < \infty$ and

$$\begin{aligned}(\delta_n(s_n)/n)^{1/3} &\sum_{s \geq 0} \delta^{\partial}_{loc,n}(s, s_n)\gamma_{n,s}^{1-\frac{2}{p}} \\ &+ \left(\delta_n^{5/2}(s_n)/n\right)^{2/3} \sum_{s \geq 0} h_{loc,n}(s, s_n)\gamma_{n,s}^{1-\frac{2}{p}} \to 0.\end{aligned}$$

When the following summability condition holds:

$$\limsup_{n \to \infty} \sum_{s \geq 1} \delta^{\partial}_{loc,n}(s, s_n)\gamma_{n,s}^{1-\frac{2}{p}} < \infty \quad \text{a.s.},$$

Assumption 3.6 reduces to $\delta_n^{5/2}(s_n)/n \to 0$ a.s. In particular, if for each $n \geq 1$ the blocks $\{B_{n,k}\}$ have the same size $l_n < n$, it suffices to assume that the weak dependence coefficients raised to the power $1 - 2/p$ are a.s. summable and $l_n = o(n^{2/5})$. Note that this assumption also explicitly requires $K_n \to \infty$ as $n \to \infty$.

Proposition 3.3. *Suppose that Assumptions 3.4-3.7 hold. Then $F^*_{1,n}$ is conditionally d_K-consistent given $\mathcal{C}$. If, in addition, μ_n converges a.s. to a $\mathcal{C}$-measurable random vector μ and $\nabla\phi(\mu) \neq 0$ a.s., then $F^*_{2,n}$ is conditionally d_K-consistent given $\mathcal{C}$.*

[9] Assumption 3.6 is made merely for ease of exposition. In view of Theorem 3.2 it can be omitted at the expense of establishing additional Berry-Essen type bounds.

3.4.2 Dependent wild bootstrap

The dependent wild bootstrap for time-series was introduced in Shao (2010). This method approximates the finite-sample distribution of T_n by mimicking the autocovariance structure of the underlying sample. In particular, adapting to our framework, assume that $\mathcal{C} = \{\emptyset, \Omega\}$ and let G_n be an unweighted "line" network. Consider an n-dimensional, zero mean random vector W_n defined on $(\Omega, \mathcal{H}, \mathsf{P})$ and independent of Y_n such that $\mathrm{Var}(W_{n,i}) = 1$ and $\mathrm{Cov}(W_{n,i}, W_{n,j}) = \kappa(d_n(i,j)/(s_n+1))$, where $\kappa(\,\cdot\,)$ is a positive definite kernel function and s_n is a bandwidth parameter. The DWB pseudo-sample Y_n^* is defined as follows:

$$Y_{n,i}^* = \bar{Y}_n + (Y_{n,i} - \bar{Y}_n)W_{n,i}, \quad i \in N_n.$$

Let $\bar{Y}_n^* := n^{-1}\sum_{i\in N_n} Y_{n,i}^*$. By construction, $\mathsf{E}[\bar{Y}_n^* \mid \mathcal{G}_n] = \bar{Y}_n$ so that in contrast to the block bootstrap, the statistic $\sqrt{n}(\bar{Y}_n^* - \bar{Y}_n)$ is unbiased given $\mathcal{G}_n$. In addition, noticing that $\kappa(0) = 1$, the conditional variance of the scaled bootstrap mean given $\mathcal{G}_n$ is

$$\begin{aligned}\Sigma_n^* &= \frac{1}{n}\sum_{i,j\in N_n} \mathrm{Cov}(W_{n,i}, W_{n,j})(Y_{n,i} - \bar{Y}_n)(Y_{n,j} - \bar{Y}_n)^\top \\ &= \frac{1}{n}\sum_{i,j\in N_n} \kappa\left(\frac{d_n(i,j)}{s_n+1}\right)(Y_{n,i} - \bar{Y}_n)(Y_{n,j} - \bar{Y}_n)^\top,\end{aligned}$$

which is a version of the network HAC estimator (3.5). Then under certain regularity conditions the DWB is first-order consistent for smooth functions of the mean.

For general graphs, however, positive definiteness of the kernel function κ does not imply that the matrix $[\kappa(d_n(i,j)/(s_n+1))]_{i,j\in N_n}$ is positive semi-definite (see Section 2.4 of Chapter 2). Therefore, in general, we cannot guarantee the existence of a random vector with the required covariance structure. A simple way to overcome this issue is to rely on the topology of a given network. Consider the matrix $\Omega_n = [\omega_n(i,j)]_{i,j\in N_n}$, where ω_n is defined in (3.8).

Claim 3.1. *Ω_n is positive semi-definite.*

Proof. Let $c \in \mathbf{R}^n$ and $\xi_i := \sum_{j\in N_n(i;s_n+1)} c_j$. Then, since $(j,k) \in N_n^2(i; s_n+1)$ if and only if $i \in N_n(j; s_n+1) \cap N_n(k; s_n+1)$,

$$\sum_{i\in N_n} \xi_i^2 = \sum_{i\in N_n} \sum_{j,k\in N_n(i;s_n+1)} c_j c_k = \sum_{i,j\in N_n} c_i c_j \omega_n(i,j)\delta_n(s_n),$$

Therefore,

$$c^\top \Omega_n c = \sum_{i,j\in N_n} c_i c_j \omega_n(i,j) = \sum_{i\in N_n} \xi_i^2/\delta_n(s_n) \geq 0. \qquad \blacksquare$$

Consequently, we consider a random vector W_n satisfying the following assumption.

Assumption 3.8. W_n is conditionally independent of Y_n given $\mathcal{C}$ with $\mathsf{E}[W_n \mid \mathcal{C}] = 0$ a.s and $\mathsf{E}[W_n W_n^\top \mid \mathcal{C}] = \Omega_n$ a.s.

Under Assumption 3.8 the bootstrap variance estimator given by

$$\Sigma_n^* = \frac{1}{n} \sum_{i,j \in N_n} \omega_n(i,j)(Y_{n,i} - \bar{Y}_n)(Y_{n,j} - \bar{Y}_n)^\top \quad \text{a.s.} \tag{3.10}$$

is positive semi-definite. We impose the next conditions on the sequence of networks, which in combination with Assumption 3.5, ensure the consistency of Σ_n^*.

Assumption 3.9. The sequence $\{(G_n, s_n)\}$ is such that w.p.1 $\tilde{\omega} < \infty$ and

(a) $\Delta_n(s_n; 1)/\delta_n(s_n) + D_n(s_n)/n \to 0$,

(b) $n^{-1} \sum_{i \in N_n} \sum_{j \in N_n^\partial(i;s)} |\omega_n(i,j) - 1| \gamma_{n,s} \to 0$ for all $s \geq 1$.

The conditions given in Assumption 3.8 are clearly weaker than those needed for the consistency of the BB variance estimator, which follows from the fact that $\Delta_n(s_n; 1) \leq \sqrt{\Delta_n(s_n; 2)}$ and $\delta_n(s_n) \leq n$. Therefore, the DWB estimator (3.10) is likely to be consistent for a wider class of networks. As in the case of the block bootstrap we assume that the true variance Σ_n converges a.s. to a $\mathcal{C}$-measurable matrix Σ.

Proposition 3.4. *Suppose that Assumptions 3.9 and 3.5 hold. Then*

$$\mathsf{E}[\|\Sigma_n^* - \Sigma_n\|_F \mid \mathcal{C}] \to 0 \quad \textit{a.s.}$$

First, we consider the Gaussian case. That is, we take $W_n = \Omega_n^{1/2} \zeta_n$, where ζ_n is the standard normal random vector in $\mathbf{R}^n$ independent of $\mathcal{G}_n$. From the practical perspective it is a convenient choice, especially when n is large because a sample from a multivariate normal distribution can be easily generated. Moreover, efficient algorithms for finding the square root of positive semidefinite matrices are available. We refer to Higham (2008) for details. As noted in Shao (2010), although the DWB sample with Gaussian weights may not match non-zero higher-order cumulants of the original process, it is difficult to choose the joint distribution of W_n that fits those cumulants, and performance of the DWB primarily depends on the choice of the truncation parameter s_n.

In this case, conditionally on $\mathcal{G}_n$, the statistic

$$\sqrt{n}(\bar{Y}_n^* - \bar{Y}_n) = \frac{1}{\sqrt{n}} \sum_{i \in N_n} W_{n,i}(Y_{n,i} - \bar{Y}_n)$$

is also normal with zero mean and variance given in (3.10). Therefore, the conditional distribution of the DWB counterpart of the test statistic $T_{1,n}$,

$$T_{1,n}^* = \sqrt{n} \left\| \bar{Y}_n^* - \bar{Y}_n \right\|$$

given $\mathcal{G}_n$ is known and is the same as the conditional distribution of the asymptotic Gaussian approximation $\|\Sigma_n^{*1/2}\eta\|$, where η is as v-dimensional standard normal random vector independent of $\mathcal{G}_n$, and the latter converges $\mathcal{C}$-weakly to $\|\Sigma^{1/2}\eta\|$ by Lemma B.3. A more interesting case, however, arises when considering the second test statistic $T_{2,n}$ because for nonlinear transformations the conditional distribution of its bootstrap analog,

$$T_{2,n}^* = \sqrt{n}\big(\phi(\bar{Y}_n^*) - \phi(\bar{Y}_n)\big),$$

is typically unavailable. Then in the Gaussian case the DWB is consistent without any further restriction on the topology of the sequence of networks $\{G_n\}$. We only need to assume that $\sqrt{n}(\bar{Y}_n - \mu_n)$ converges $\mathcal{C}$-weakly to a conditionally normal random vector and the asymptotic variance of $T_{2,n}$ is a.s. positive.

Proposition 3.5. *Suppose that W_n is Gaussian, Assumptions 3.8, 3.9, 3.5, and 3.6 hold. Then $F_{1,n}^*$ is conditionally d_K-consistent given $\mathcal{C}$. If, in addition, μ_n converges a.s. to a $\mathcal{C}$-measurable random vector μ and $\nabla\phi(\mu) \neq 0$ a.s., then $F_{2,n}^*$ is conditionally d_K-consistent given $\mathcal{C}$.*

Given another choice of W_n, the process $\{\xi_{n,i} := W_{n,i}(Y_{n,i} - \bar{Y}_n) : i \in N_n\}$ is s_n-dependent conditionally on $\mathcal{G}_n$, i.e., $\xi_{n,i}$ and $\xi_{n,j}$ are conditionally independent given $\mathcal{G}_n$ whenever $j \notin B_{n,i} := N_n(i; s_n + 1)$. Consequently, in addition to the assumptions of Proposition 3.5, we need to control the behavior of the third conditional moments of W_n and the neighborhoods $\{B_{n,i}\}$ such that the bootstrap distributions $F_{1,n}^*$ and $F_{2,n}^*$ in this case approach ones under the Gaussian weights as $n \to \infty$.

Proposition 3.6. *Suppose that Assumptions 3.8, 3.9, 3.5, and 3.6 hold, and*

$$\frac{1}{n^{3/2}} \sum_{i \in N_n} \sum_{j \in B_{n,i}} \sum_{k \in B_{n,i} \cup B_{n,j}} \prod_{l \in \{i,j,k\}} \|W_{n,l}\|_{\mathcal{C},3} \to 0 \quad a.s. \tag{3.11}$$

Then $F_{1,n}^*$ is conditionally d_K-consistent given $\mathcal{C}$. If, in addition, μ_n converges a.s. to a $\mathcal{C}$-measurable random vector μ and $\nabla\phi(\mu) \neq 0$ a.s., then $F_{2,n}^*$ is conditionally d_K-consistent given $\mathcal{C}$.

Remark 3.7. The convergence condition in (3.11) is quite strong. In particular, in a simple case when the neighborhoods $\{B_{n,i}\}$ have the same size $l_n < n$ for all $n \geq 1$ and $\sup_{n,i \in N_n} \|W_{n,i}\|_{\mathcal{C},3} < \infty$ a.s., it requires $l_n = o(n^{1/4})$. Therefore, it is of high interest to find a better way to handle network dependent processes under m-dependence.

3.5 Conclusion

Nonparametric bootstrapping for time series and spatial processes has been extensively studied in the past decades. Thus, various resampling methods are now available for

statistical analysis of dependent data in these cases. However, the lack of regular structure in networks renders the use of these techniques for bootstrap-based inference in the case of network dependent processes impracticable. In this work we proposed a block-based method and a variant of the DWB suitable for the latter processes satisfying the conditional version of Doukhan and Louhichi (1999)'s ψ-weak dependence condition. We established the first-order validity of these methods to construct confidence sets for the mean of a network dependent process. In addition, we showed their consistency under the smooth function model conditionally on a common shock of a general form. Finally, the corresponding bootstrap variance estimators can be used for asymptotic inference instead of the network HAC estimator, which is not necessarily positive semi-definite.

As for the future directions, having a continuity theorem and other related results similar to ones established in Belyaev and Sjöstedt-de Luna (2000) but under convergence in conditional probability would significantly weaken the bootstrap consistency conditions derived in this chapter. In addition, an extension of these methods for bootstrapping M-estimators and empirical processes is of great importance for applied research.

Chapter 4

Econometric Inference on a Large Bayesian Game with Heterogeneous Beliefs

4.1 Introduction

Many economic outcomes arise as a consequence of agents' decisions under the influence of others' choices. Endogeneity and simultaneity of such influence pose challenge for an empirical researcher. In response to this challenge, a strand of empirical methods employ game-theoretic models to capture strategic interactions among agents. These models often share two main features. First, they assume observation of many i.i.d. replications from a single representative game, so that statistical independence is imposed across the replications, whereas strategic interdependence is kept within each replication. Second, they rely on a common prior assumption, where the types distribution is assumed to be common knowledge among players.[1] However, such a framework does not fit well in a situation where there are interactions between many agents.

We propose an alternative model, in which heterogeneity across the games and heterogeneity across the players are given characteristics of a single large Bayesian game.[2] Here an econometrician observes outcomes from a pure strategy Bayesian Nash equilibrium (possibly among multiple equilibria), and attempts to make inference about the structural parameters of the game. While the equilibrium is driven by the subjective beliefs of the players, the

[1] See Bresnahan and Reiss (1991), Tamer (2003), Ciliberto and Tamer (2009), Aradillas-Lopez (2010), Beresteanu et al. (2011), Aradillas-Lopez and Tamer (2008), and de Paula and Tang (2012), among many others, for methodological contributions.

[2] This work focuses on simultaneous-move games with an unordered finite action space. Hence, auction models with continuous bids are excluded. Global network games with endogenous network formation or matching games are also excluded because the action space increases as the number of players increases.

validity of the econometrician's inference is measured in terms of the Nature's objective probability. Thus, the inference procedure in described in Section 4.5 is valid regardless of how an equilibrium is selected, or whether the sets of equilibria across different games are identical or not.[3]

The main departure of this model from the existing literature is that the players are allowed to form beliefs differently from each other. Regarding the subjective beliefs of the players, it is assumed that they commonly (and correctly) believe that the types are conditionally independent given their publicly available information, and that any event a player believes highly likely to occur does occur with high probability. Within the boundary of these two conditions, each player is allowed to form beliefs differently using potentially different priors. Thus, our inference procedure is robust to the way individual players form their beliefs about other players' types or beliefs. If it is desirable that the predictions from game models be robust to particular assumptions about players' belief formation process, as emphasized by Wilson (1987) and Bergemann and Morris (2005), the same may apply all the more to econometric inference on such models.[4]

Subjective beliefs are hard to recover in this environment. Manski (2004) proposes using data on subjective probabilities in choice studies. See Dominitz and Manski (1997) for a study on subjective income expectations, and Li and Lee (2009) for an investigation of rational expectations assumptions in social interactions using subjective expectations data. However, in strategic environments with many players, it is often not easy to procure data on players' expectations on other players' types prior to the play.

Instead of attempting to recover subjective beliefs, we develop a *hindsight regret approach*, based on the insights of Kalai (2004) and Deb and Kalai (2015). The *hindsight regret* of a player measures the *ex post* payoff loss due to his inability to (fully) observe the other players' types. More specifically, the hindsight regret quantifies the amount of additional compensation needed to preserve each player's incentive compatibility constraint in equilibrium even after all the players' types are revealed. We provide a general form of belief-free hindsight regret, which is used to form an *ex post* version of incentive compatibility constraints for each player in equilibrium.

Using the *ex post* incentive compatibility constraints, we derive testable implications,

[3]Note that the notion of an equilibrium selection rule in the literature already presumes replications of a single representative game. When two games are different with different sets of equilibria, there cannot be an equilibrium selection rule that applies to both the games.

[4]Note the unique contribution by Aradillas-Lopez and Tamer (2008) for various implications of assumptions of higher order beliefs in econometric game models. The main difference between Aradillas-Lopez and Tamer (2008) and this work is twofold. First, the former work focused on level k rationalizability where uncertainty faced by a player is about other players' actions, whereas uncertainty faced by a player in our game is confined to other players' payoff types. Second, the main purpose of the study by Aradillas-Lopez and Tamer (2008) was to explore implications of various higher order belief configurations for econometric inference, whereas we pursue an inference method that is robust to various order belief configurations within the Nash equilibrium framework.

which can be used for inference without knowledge of a particular way beliefs are formed. The implications are formulated as moment inequalities in a spirit similar to Ciliberto and Tamer (2009). The tightness of the moment inequalities (thus, the nontriviality of subsequent inference) depends on how strongly any two players are strategically interdependent. When the reference group is large and each player's payoff is affected by the action of another player in inverse proportion to the group sizes, the inequalities can be fairly tight, opening the possibility of nontrivial inference. On the other hand, this condition excludes the situation where the econometrician observes many small private information games as in Aradillas-Lopez (2010) and de Paula and Tang (2012).

For inference we propose a bootstrap based approach and establish its asymptotic validity, as the number of the players increases to infinity. The asymptotic validity is uniform over the probabilities that the Nature adopts for drawing the players' types. Our approach for inference is inspired by the work of Andrews (2005) who investigated the inference problem in the presence of common shocks in short panel data (see Kuersteiner and Prucha, 2013 for a related research on dynamic panel models). However, we cannot use the random norming as he did to pivotize the test statistic, because the restrictions here are inequalities rather than equalities. The method of bootstrap proposed in this work obviates the need to pivotize the test statistic.

The existing econometrics literature of games often assumes observation of many independent replications from a single representative game, so that statistical independence is imposed across the replications, whereas strategic interdependence is kept within each replication. See Bresnahan and Reiss (1991), Tamer (2003), Ciliberto and Tamer (2009), Aradillas-Lopez (2010), Bajari et al. (2010), Beresteanu et al. (2011), Aradillas-Lopez and Tamer (2008), and de Paula and Tang (2012) for examples. See de Paula (2013) for references. See Krauth (2006) and de Paula (2013) for using the framework of many replications of a representative game in analysis of social interactions among people. In contrast to such a framework, our approach does not require that one observe many games for asymptotically valid inference. The asymptotic validity holds as long as the number of the players is large, regardless of how they are partitioned into different subgames. This feature is convenient in particular, for it accommodates a single large game such as a social interactions model with multiple large overlapping reference groups.

This chapter's framework is most relevant to various models of social interactions. As a seminal paper in the structural modeling and estimation of social interactions, Brock and Durlauf (2001) developed discrete choice-based models of social interactions. Their discrete-choice based approach influenced many researches such as Krauth (2006), Ioannides and Zabel (2008), and Li and Lee (2009) to name a few. See Blume et al. (2011) and the monograph by Ioannides (2013) for recent methodological progresses in the literature of social interactions. In contrast with this approach, we relax the symmetry of equilibrium

strategies or rational expectations. Furthermore, this work fully develops a bootstrap inference procedure which is asymptotically valid regardless of whether there are multiple equilibria, or how the equilibrium is selected across different games. On the other hand, the approach of Brock and Durlauf (2001), within the boundary of their set-up, is simpler to use than ours.

It is also worthwhile to compare our approach with recent contributions by Xu (2018), Bisin et al. (2011) and Menzel (2012). These papers are more explicit about the inferential issues in a large game model. Xu (2018) studied a single large Bayesian game framework similar to ours, but his inference procedure requires various conditions that yield uniqueness of the equilibrium and point-identification of the parameters. On the other hand, Bisin et al. (2011) admit multiple equilibria, but their equilibrium concept requires asymptotic stability of the aggregate quantities (as the number of the players increases). Menzel (2012) recently developed asymptotic theory for inference based on large complete information games where type-action profiles are (conditionally) exchangeable sequences.[5]

This chapter is organized as follows. The first section formally introduces a large Bayesian game, and discusses examples. Section 4.3 introduces a belief-free version of hindsight regrets. Section 4.4 turns to the assumptions for econometric inference and derives testable implications. Section 4.5 proposes general inference methods and establishes their uniform asymptotic validity. For simplicity of exposition, most of the results in the chapter are obtained assuming a binary action space. Their extension to the case of a general finite action set is provided in the appendix.

4.2 A large Bayesian game with iInformation groups

4.2.1 The setup

In this section we formally introduce a Bayesian game that defines the scope of the chapter. Assumptions relevant to econometric inference appear in Section 4.4. The game is played by a finite set N of players. Each player belongs to at least one information group from a finite set of groups G and chooses an action form a finite action set A.

Consider a probability space $(\Omega, \mathcal{H}, \mathsf{P})$. At the beginning of the game, the Nature draws an outcome $\omega \in \Omega$ and each player $i \in N$ observes an event from $\tau_i \equiv \tau_i^o \vee \tau_i^u$, where $\tau_i^o \subset \mathcal{H}$ and $\tau_i^u \subset \mathcal{H}$ constitute the observable and unobservable components of i's type, respectively. Players belonging to group $g \in G$ are directly affected by an event from $\mathcal{C}_g \subset \mathcal{H}$, which we

[5]The fundamental difference between Menzel (2012) and this work lies in modeling the probability of observations. Menzel employs a complete information game model where the randomness of the observed outcomes is mainly due to the sampling variations. Thus, random sampling schemes and variants justify his exhangeability conditions. On the other hand, we considers an incomplete information game, where the randomness of observed outcomes stems from the inherent heterogeneity across players due to Nature's drawing of types.

refer to as the group-specific public signal. We denote the set of such players by $N_g \subseteq N$ and let $G_i := \{g \in G : \{i\} \cap N_g \neq \emptyset\}$, i.e., the subset of groups to which player i belongs. In many empirical applications, different groups can be thought of either as separate games, or as large overlapping reference groups of players in a single game.[6] We assume that $\mathcal{F} := \bigvee_{i \in N} \tau_i^o \vee \bigvee_{g \in G} \mathcal{C}_g$ represents the information available to all the players. No player, however, shares his unobservable type information with others.

Each player $i \in N$ is endowed with a subjective belief Q_i, which is a probability measure on $(\Omega, \mathcal{H})$. Thus, P is the *objective probability* that the econometrician uses to express the validity of his inference method, whereas Q_i is a *subjective probability* formed by player i possibly through her higher order beliefs about other players' beliefs. As pointed out in Aumann (1976), when P belongs to common knowledge, $\mathsf{Q}_i = \mathsf{P}$ for all $i \in N$ so that the distinction between the objective and subjective probabilities is not necessary. We introduce a conditional independence assumption for these probability measures.

Assumption 4.1 (*Conditional Independence Under Objective and Subjective Probabilities*)**.** The σ-fields $\{\tau_i^u : i \in N\}$ are conditionally independent given $\mathcal{F}$ under P and Q_i for all $i \in N$.

In addition, we assume that any event that a player believes strongly to occur after observing his own type and receiving the publicly available information is highly likely to occur according to the objective probability. We denote the information observed by player $i \in N$ as $\mathcal{I}_i$, which is defined as $\mathcal{I}_i := \tau_i^u \vee \mathcal{F}$.

Assumption 4.2 (*One-Sided Rational Expectations about High-Probability Events*)**.** There exists a small number $\rho \in (0, 1)$ such that for each $i \in N$, $\mathsf{P}(H \mid \mathcal{I}_i) \geq 1 - \rho$ (P-a.s.) for any $H \in \mathcal{H}$ with $\mathsf{Q}_i(H \mid \mathcal{I}_i) \geq 1 - \rho$ (Q_i-a.s.).

This assumption imposes a limited version of rational expectations on the players' beliefs about events that are believed to be highly likely by the players. The version is one-sided in the sense that a high-probability event (according to the Nature's experiment) is not necessarily viewed as a high-probability event by each player.

Consider player $i \in N$ and let ξ_i be an τ_i-measurable random element taking values in $(E_i, \mathcal{E}_i)$. Once the Nature draws $\omega \in \Omega$, this player, facing the other players choosing $b \in A^{|N|-1}$, receives a payoff $u_i(a, b; t)$ by choosing $a \in A$ and observing $t = \xi_i(\omega)$, i.e.,

[6] The information group structure belongs to common knowledge among the players, and is exogenously formed prior to the current game. The assumption of exogenous group formation is plausible when the players are randomly assigned to groups only based on some public signals, or the group formation has almost no relevance to the current game. For example, consider a study on the presidential election among reference groups with similar demographic characteristics. In this case, the formation of the demographic groups has little relevance to the subsequent decisions in the election. However, there are also many other situations where the group formation is directly relevant to the current game. Extending the framework to endogenous group formation requires further research.

$u_i : A \times A^{|N|-1} \times E_i \to \mathbf{R}$ is a measurable function that may depend on the player's type.[7] A *pure strategy* $Y_i : \Omega \to A$ of player $i \in N$ is an $\mathcal{I}_i$-measurable function from the state space to the action set, and a *pure strategy profile* $\mathcal{Y} \equiv (Y_i : i \in N)$ is the vector of individual pure strategies. For $i \in N$ we write $Y_{-i} \equiv (Y_j : j \in N \setminus \{i\})$.[8]

Definition 4.1. A strategy profile $\mathcal{Y}$ is a *pure strategy Bayesian Nash equilibrium* if for each player $i \in N$ and any pure strategy Y',

$$\mathsf{E}_{\mathsf{Q}_i}[u_i(Y_i, Y_{-i}) \mid \mathcal{I}_i] \geq \mathsf{E}_{\mathsf{Q}_i}[u_i(Y', Y_{-i}) \mid \mathcal{I}_i] \quad (\mathsf{Q}_i\text{-a.s.}).^9 \tag{4.1}$$

Note that $\mathcal{Y}$ can be viewed as a random vector on $(\Omega, \mathcal{H})$, and by Assumption 4.1, the elements of $\mathcal{Y}$ are conditionally independent given $\mathcal{F}$ under P and Q_i for all $i \in N$. In addition, since the subjective beliefs are heterogeneous, the distributions of the elements of $\mathcal{Y}$ are not necessarily identical, even if we focus on symmetric equilibria.

4.2.2 Examples

Example 4.1 (*Large Games with Social Interactions*)**.** Suppose that the sets $\{N_g : g \in G\}$ are disjoint. Each group of players corresponds to a game with private information. For player $i \in N$ belonging to group $g \in G$ we follow Brock and Durlauf (2001) (see (4) and (5) there) and consider either of the following two specifications of payoff functions:

$$u_i(a, b; t) = v_1(a, t) + v_2(a, t) \sum_{j \in N_g \setminus \{i\}} w_{j,g} b_j, \tag{4.2}$$

or

$$u_i(a, b; t) = v_1(a, t) - \frac{v_2(t)}{2} \left(a_i - \sum_{j \in N_g \setminus \{i\}} w_{j,g} b_j \right)^2, \tag{4.3}$$

where v_1 and v_2 are some measurable functions, and $\{w_{j,g} : j \in N_g\}$ are non-negative weights with $\sum_{j \in N_g \setminus \{i\}} w_{j,g} = 1$. The first specification expresses interaction between player's action and the average actions of the other players. The second one captures a preference for conformity to the average actions of other players.

Example 4.2 (*Large Game with Multiple Overlapping Reference Groups*)**.** Suppose that the game is a large private information game with multiple overlapping information groups. The information groups are reference groups such that the average of the actions by players

[7] For ease of exposition we omit the last argument of u_i when clear from the context.

[8] For a collection of σ-fields $\{\mathcal{F}_i : i \in N\}$ we write $\mathcal{F}_{-i} \equiv \bigvee_{i \in N \setminus \{i\}} \mathcal{F}_i$.

[9] Existence of a pure strategy equilibrium can be established by invoking a more special structure of the game in application. For example, see Milgrom and Weber (1985), Athey (2001), McAdams (2003), and Reny (2011) and references therein for general results.

in each group affects the payoff of the players in the group. More specifically, the playoff function takes the following form:

$$u_i(a,b;t) = v_i(a,t) + \frac{a\theta}{|G_i|}\sum_{g\in G_i}\left(\frac{1}{|N_g|-1}\sum_{j\in N_g\setminus\{i\}} b_j\right), \tag{4.4}$$

where v_i is a measurable function. The within-group correlation among the types $\{\xi_i : i \in N_g\}$ is permitted through the public signal $\mathcal{C}_g$. More importantly the reference groups are allowed to be overlapping, so that each player may belong to multiple reference groups simultaneously and yet differently from other players.

For ease of exposition in the rest of the chapter we focus on the binary action space case, i.e., $A = \{0,1\}$. The general case of a multinomial action set is presented in Appendix C.1.

4.3 Belief-free hindsight regrets

In this section we introduce the notion of hindsight regret and establish its belief-free version. This version is used later to derive testable implications from the large game model. First, we rewrite the equilibrium constraints in (4.1) as follows. For a given equilibrium $\mathcal{Y}$ and each $i \in N$,

$$\mathsf{E}_{\mathsf{Q}_i}[u_i^{\Delta}(Y_i, Y_{-i}) \mid \mathcal{I}_i] \geq 0 \quad (\mathsf{Q}_i\text{-a.s.}), \tag{4.5}$$

where

$$u_i^{\Delta}(a,b;t) \coloneqq u_i(a,b;t) - u_i(1-a,b;t).^{10}$$

Such constraints are generally useful for deriving moment inequalities for inference. However, they cannot be directly used here due to the heterogeneous subjective beliefs of the players.

The hindsight regret approach of this work replaces the inequality (4.5) by the following *ex post* version:

$$u_i^{\Delta}(Y_i, Y_{-i}) \geq -\lambda, \tag{4.6}$$

which is ensured to hold with large probability according to player i's belief Q_i conditionally on $\mathcal{I}_i$ by choosing a compensation scheme $\lambda \geq 0$ appropriately. The compensation λ is anticipated (with high probability) to prevent player i from switching from the chosen action Y_i in equilibrium to $1 - Y_i$ after the types of all the players are revealed.

For use in econometric inference we seek to find a minimal compensation scheme that does not rely on the players' beliefs (except through already given equilibrium $\mathcal{Y}$). As we

[10] Similarly to u_i, we omit the last argument of u_i^{Δ} when clear from the context. Also note that $|u_i^{\Delta}(1,\cdot)| = |u_i^{\Delta}(0,\cdot)|$.

will see later, the quality of prediction and the econometrician's inference improves with the use of a tighter compensation scheme.

4.3.1 Strategic interdependence among players

A player's hindsight regret measures the *ex post* loss of payoff due to the inability to fully observe the types of other players. The notion of hindsight regret is directly related to strategic interdependence among the players. In order to formally introduce a measure of strategic interdependence, we first define the maximal variation of a real function. Consider a function $f : \mathcal{X}^d \to \mathbf{R}$ with $d \geq 1$ and let

$$V_j(f) := \sup_{x \in \mathcal{X}^d, x' \in \mathcal{X}} \big| f(x) - f(x_1, \ldots, x_{j-1}, x', x_{j+1}, \ldots, x_d) \big|. \tag{4.7}$$

We call $V_j(f)$ a *maximal variation of f at the j-th coordinate.* For $i, j \in N$ such that $i \neq j$ we define

$$\Delta_{ij}(t) := V_j(b \mapsto u_i^{\Delta}(1, b; t)). \tag{4.8}$$

The function Δ_{ij} measures the largest variation in the player i's payoff differential u_i^{Δ} between actions 1 and 0 which can be caused by player j's arbitrary choice of action. Hence, Δ_{ij} summarizes the strategic relevance of player j to player i, and is used to formulate the belief-free hindsight regrets later. We make the following assumption which imposes a restriction on the admissible payoff functions and ensures that Δ_{ij} is well-defined.

Assumption 4.3. For each $i, j \in N$ with $i \neq j$, Δ_{ij} is $(\mathcal{E}_i, \mathcal{B}_{\mathbf{R}})$-measurable.

4.3.2 Belief-free hindsight regrets for large games

Definition 4.2. Given an equilibrium $\mathcal{Y}$ and $\rho \in (0, 1)$, an $\mathcal{I}_i$-measurable non-negative random variable λ is a *ρ-hindsight regret* for player $i \in N$ if

$$\mathsf{Q}_i\big(u_i^{\Delta}(Y_i, Y_{-i}) \geq -\lambda \mid \mathcal{I}_i\big) \geq 1 - \rho \quad (\mathsf{Q}_i\text{-a.s.}). \tag{4.9}$$

A ρ-hindsight regret λ for player $i \in N$ represents the amount of compensations that induces him to maintain the chosen strategy with probability of at least $1 - \rho$. By definition, if λ is a ρ-hindsight regret, any map dominating it also satisfies (4.9). Therefore, we need to find a version of λ that is tight enough for use by the econometrician.

In order to characterize a belief-free hindsight regret, we let

$$\lambda_{i,\rho}(t) := \sqrt{-\frac{\ln \rho}{2} \cdot \Lambda_i(t)}, \quad \text{where} \quad \Lambda_i(t) := \sum_{j \in N \setminus \{i\}} (\Delta_{ij}(t))^2. \tag{4.10}$$

The hindsight regret $\lambda_{i,\rho} \equiv \lambda_{i,\rho}(\xi_i)$ is *belief-free* in the sense that it does not depend on the subjective beliefs Q_i, other than through the given pure strategy equilibrium. Furthermore, $\lambda_{i,\rho}$ is a ρ-hindsight regret as formalized in the following theorem.

Theorem 4.1. *Suppose that Assumptions 4.1-4.3 hold. Then for any pure strategy equilibrium $\mathcal{Y}$ and any $\rho \in (0,1)$, $\lambda_{i,\rho}$ is a ρ-hindsight regret for player $i \in N$. Moreover, there exists $\rho \in (0,1)$ such that*

$$\mathsf{P}\big(u_i^{\Delta}(Y_i, Y_{-i}) \geq -\lambda_{i,\rho}(\xi_i) \mid \mathcal{I}_i\big) \geq 1 - \rho_i(\xi_i) \quad (\mathsf{P}\text{-}a.s.), \tag{4.11}$$

where $\rho_i(t) \coloneqq \rho\, \mathbf{1}\{\lambda_{i,\rho}(t) > 0\}$.

The function Λ_i in (4.10) measures the overall strategic relevance of other players to player $i \in N$. Therefore, the hindsight regret increases with the strategic interdependence among the players. This is intuitive because player i's *ex post* payoff loss due to the inability to fully observe the types of other players is large when actions by those players can have a large impact on player i's payoff.

Let us see how $\lambda_{i,\rho}$ defined in (4.10) becomes a ρ-hindsight regret. For any $\mathcal{I}_i$-measurable, positive λ,

$$\begin{aligned} &\mathsf{Q}_i\big(u_i^{\Delta}(Y_i, Y_{-i}) \leq -\lambda \mid \mathcal{I}_i\big) \\ &\qquad \leq \mathsf{Q}_i\big(u_i^{\Delta}(Y_i, Y_{-i}) - \mathsf{E}_{\mathsf{Q}_i}[u_i^{\Delta}(Y_i, Y_{-i}) \mid \mathcal{I}_i] \leq -\lambda \mid \mathcal{I}_i\big) \\ &\qquad \leq \exp\big(-2\lambda^2/\Lambda_i(\xi_i)\big) \quad (\mathsf{Q}_i\text{-a.s.}). \end{aligned} \tag{4.12}$$

The first inequality follows by the Nash equilibrium constraint, and the second inequality follows from the conditional McDiarmid's inequality. The result follows by setting $\lambda = \lambda_{i,\rho}(\xi_i)$ on $\{\lambda_{i,\rho} > 0\}$ and noticing that $u_i^{\Delta}(Y_i, Y_{-i}) = 0$ on $\{\lambda_{i,\rho} = 0\}$. The inequality (4.11) is an immediate consequence of (4.12) combined with Assumption 4.2. Later we use the inequality (4.12) to obtain testable implications.

4.3.3 Examples revisited

Example 4.1 cont'd. As for the belief-free hindsight regrets, we first consider that in both cases

$$u_i^{\Delta}(1, b; t) = v_1^{\Delta}(t) + v_2^{\Delta}(t) \sum_{j \in N_g \setminus \{i\}} w_{j,g} b_j,$$

where in the specification (4.2),

$$v_1^{\Delta}(t) \coloneqq v_1(1,t) - v_1(0,t) \quad \text{and} \quad v_2^{\Delta}(t) \coloneqq v_2(1,t) - v_2(0,t),$$

and in the specification (4.3),

$$v_1^{\Delta}(t) \coloneqq v_1(1,t) - v_1(0,t) - v_2(t)/2 \quad \text{and}$$
$$v_2^{\Delta}(t) \coloneqq v_2(t).$$

Therefore, in both cases $\Delta_{ij}(t) = \left|v_2^{\Delta}(t)\right| w_{j,s}$ for all $j \neq i$. Using this, we define $\lambda_{i,\rho}$ as in (4.10). In particular, when $w_{j,g} = (|N_g| - 1)^{-1}$ for all $j \in N_g \setminus \{i\}$, we get

$$\lambda_{i,\rho} = \left|v_2^{\Delta}(\xi_i)\right| \sqrt{-\frac{\ln \rho}{2(|N_g| - 1)}}.$$

The hindsight regret is heterogeneous across information groups, depending on the number of the players $|N_g|$ in each information group $g \in G$. The more players in a group, the smaller the hindsight regret for that group.

Example 4.2 cont'd. From the payoff specification (4.4), we observe that for $i, j \in N$, such that $i \neq j$,

$$\Delta_{ij}(t) = \frac{|\theta|}{|G_i|} \sum_{g \in G_i} \frac{\mathbf{1}\{j \in N_g\}}{|N_g| - 1}.$$

Only those players who belong to at least one of player i's reference groups are strategically relevant. Consequently,

$$\lambda_{i,\rho} = |\theta| \sqrt{-\frac{\ln \rho}{2} \sum_{j \in N \setminus \{i\}} \left(\frac{1}{|G_i|} \sum_{g \in G_i} \frac{\mathbf{1}\{j \in N_g\}}{|N_g| - 1} \right)^2}.$$

It follows that those players with large reference groups tend to have negligible hindsight regrets.

4.4 The econometrician's observations and testable iImplications

In the rest of the chapter we set $|N| = n \in \mathbb{N}$ and identify N with integers $\{1, \ldots, n\}$. We assume that the econometrician observes a realization of $(\{(Y_i, X_i)\}_{i \in N}, \{C_g\}_{g \in G})$, where Y_i is an action taken by player $i \in N$, X_i is the vector of observable covariates of that player, and C_g is group g's public signal.

Assumption 4.4 (*The econometrician's observations*)**.** For each $i \in N$ and $g \in G$:

(a) Y_i corresponds to a pure strategy equilibrium $\mathcal{Y}$;
(b) $\tau_i^o = \sigma(X_i)$, $\tau_i^u = \sigma(\eta_i)$, and $\mathcal{C}_g = \sigma(C_g)$, where $X_i \in \mathbf{R}^v$, $\eta_i \in \mathbf{R}^w$, and $C_g \in \mathbf{R}^c$;
(c) η_i is conditionally independent of τ_{-i} given X_i and $C \equiv (C_1, \ldots, C_G)$ (under P).

The distribution of the observable quantities that the econometrician focuses on stems from the Nature's objective probability P and a pure strategy equilibrium $\mathcal{Y}$. When there are multiple equilibria, the researcher does not know which equilibrium the observed outcomes are associated with. The players' subjective beliefs affect the distribution of $\{Y_i\}_{i\in N}$ through their impact on the associated equilibrium.

Assumption 4.5 (*Model Parameterization*)**.** For each $i \in N$,

$$\mathsf{P}(\eta_i \leq \cdot \mid X_i = x, C = c) = F_{i,\theta}(\cdot \mid x, c) \quad \text{and} \quad u_i(\,\cdot\,,\,\cdot\,) \equiv u_{i,\theta}(\,\cdot\,,\,\cdot\,),$$

where $\theta \in \Theta \subseteq \mathbf{R}^d$ and $F_{i,\theta}$ is a parametric distribution function.

Assumption 4.5 states that the conditional CDF of η_i given X_i and the payoff function are parameterized by a finite dimensional vector $\theta \in \Theta$. It is worth noting that Assumptions 4.4 and 4.5 are concerned only with the primitives of the game. They do not impose restrictions on the equilibrium $\mathcal{Y}$ or the way the agents' beliefs are formed. These assumptions are only concerned with the objective probability P.

4.4.1 Testable implications implied by belief-free hindsight regrets

We derive testable implications using Theorem 4.1. Consider a pure strategy equilibrium $\mathcal{Y}$ and player $i \in N$. Using the equilibrium condition 4.5, for almost all $\omega \in \Omega$ such that $Y_i(\omega) = 1$ we have $\mathsf{E}_{\mathsf{Q}_i}[u_i^{\Delta}(1, Y_{-i}) \mid \mathcal{I}_i](\omega) \geq 0$. Therefore,

$$\mathbf{1}\{Y_i = 1\} \leq \mathbf{1}\{\mathsf{E}_{\mathsf{Q}_i}[u_i^{\Delta}(1, Y_{-i}) \mid \mathcal{I}_i] \geq 0\} \quad (\mathsf{Q}_i\text{-a.s.}). \tag{4.13}$$

Similarly,

$$\mathbf{1}\{Y_i = 0\} \leq \mathbf{1}\{\mathsf{E}_{\mathsf{Q}_i}[u_i^{\Delta}(0, Y_{-i}) \mid \mathcal{I}_i] \geq 0\} \quad (\mathsf{Q}_i\text{-a.s.}). \tag{4.14}$$

Taking conditional expectations of both sides of (4.13) and (4.14) and assuming that P is absolutely continuous w.r.t. each Q_i, we deduce that for each $i \in N$,

$$1 - \pi_{i,L}^* \leq \mathsf{P}(Y_i = 1 \mid \mathcal{F}) \leq \pi_{i,U}^* \quad (\mathsf{P}\text{-a.s.}), \tag{4.15}$$

where $\pi_{i,L}^* := \mathsf{P}\big(\mathsf{E}_{\mathsf{Q}_i}[u_i^{\Delta}(0, Y_{-i}) \mid \mathcal{I}_i] \geq 0 \mid \mathcal{F}\big)$ and $\pi_{i,U}^* := \mathsf{P}\big(\mathsf{E}_{\mathsf{Q}_i}[u_i^{\Delta}(1, Y_{-i}) \mid \mathcal{I}_i] \geq 0 \mid \mathcal{F}\big)$.

Unfortunately, the inequalities in (4.15) cannot be directly used in our setup for inference for two reasons. First, the bounds involve heterogeneous subjective beliefs which the econometrician has difficulty recovering from the observations. Second, the probabilities in both bounds of (4.15) cannot be simulated because the bounds depend on the unknown distribution of Y_{-i}, which is a nonprimitive quantity. We use Theorem 4.1 to address both issues.

First, we construct the following conditional probabilities:

$$\begin{aligned} \pi_{i,L} &\coloneqq \mathsf{P}\big(u_i^{\Delta}(0, Y_{-i}) \geq -\lambda_{i,\rho} \mid \mathcal{G}_i\big) \quad \text{and} \\ \pi_{i,U} &\coloneqq \mathsf{P}\big(u_i^{\Delta}(1, Y_{-i}) \geq -\lambda_{i,\rho} \mid \mathcal{G}_i\big), \end{aligned} \tag{4.16}$$

where $\mathcal{G}_i \coloneqq \mathcal{F} \vee \sigma(Y_{-i})$. Unlike $\pi^*_{i,L}$ and $\pi^*_{i,U}$, these probabilities can be simulated as explained at the end of this subsection. To construct testable implications, we define

$$\begin{aligned} e_{i,L} &\coloneqq \mathsf{P}(Y_i = 1 \mid \mathcal{F}) - \left(1 - \frac{\mathsf{E}_{\mathsf{P}}[\pi_{i,L} \mid \mathcal{F}]}{1 - \rho_i}\right) \quad \text{and} \\ e_{i,U} &\coloneqq \mathsf{P}(Y_i = 1 \mid \mathcal{F}) - \frac{\mathsf{E}_{\mathsf{P}}[\pi_{i,U} \mid \mathcal{F}]}{1 - \rho_i}, \end{aligned} \tag{4.17}$$

where $\rho_i \coloneqq \rho\, \mathbf{1}\{\sup_{t \in E_i} \lambda_{i,\rho}(t) > 0\}$. Choosing a vector of non-negative measurable functions $\mathbf{g}_i \equiv [g_{i,1}, \ldots, g_{i,m}]^\top : \mathbf{R}^v \to \mathbf{R}^m_{\geq 0}$, we consider the following moment inequalities in a spirit similar to Andrews and Shi (2013):

$$\begin{aligned} \boldsymbol{\mu}_L &\coloneqq n^{-1} \sum_{i=1}^n e_{i,L}\, \mathbf{g}_i(X_i) \geq 0 \quad (\mathsf{P}\text{-a.s.}), \\ \boldsymbol{\mu}_U &\coloneqq n^{-1} \sum_{i=1}^n e_{i,U}\, \mathbf{g}_i(X_i) \leq 0 \quad (\mathsf{P}\text{-a.s.}). \end{aligned} \tag{4.18}$$

The following result confirms that the inequality restrictions in (4.18) hold.

Proposition 4.1. *Suppose that Assumptions 4.1-4.5 are satisfied. Then the inequality restrictions in* (4.18) *hold.*

Consider the simulation of $\pi_{i,L}$ and $\pi_{i,U}$. Assuming that $\xi_i = (X_i, \eta_i)$ and using Assumptions 4.4 and 4.5 we notice that these quantities are functions of (X_i, C, Y_{-i}) and can be written as

$$\pi_{i,L}(x, c, b) = \int_{H_i(0,b,x)} F_{i,\theta}(d\eta \mid x, c), \quad \pi_{i,U}(x, c, b) = \int_{H_i(1,b,x)} F_{i,\theta}(d\eta \mid x, c), \tag{4.19}$$

where $H_i(a, b, x) \coloneqq \{\eta \in \mathbf{R}^w : u_i^{\Delta}(a, b; (x, \eta)) \geq -\lambda_{i,\rho}(x, \eta)\}$.

The integrals in (4.19) can be either evaluated explicitly or simulated by drawing η_i from $F_{i,\theta}(\cdot \mid x, c)$. For example, consider the payoff differential for player $i \in N_g$ with $G_i = \{g\}$:

$$u_i^{\Delta}(1, b; (x, \eta)) = v_{i,1}(x) + \frac{v_{i,2}(x)}{|N|_g - 1} \sum_{j \in N_g \setminus \{i\}} b_j - \eta$$

for some functions $v_{i,1}$, and $v_{i,2}$. Then

$$\pi_{i,L}(x,c,b) = \bar{F}_{i,\theta}\left(v_{i,1}(x) + \frac{v_{i,2}(x)}{|N_g|-1} \sum_{j \in N_g \setminus \{i\}} b_j - \lambda_{i,\rho}(x) \mid x, c \right) \quad \text{and}$$

$$\pi_{i,U}(x,c,b) = F_{i,\theta}\left(v_{i,1}(x) + \frac{v_{i,2}(x)}{|N_g|-1} \sum_{j \in N_g \setminus \{i\}} b_j + \lambda_{i,\rho}(x) \mid x, c \right).$$

where $\bar{F}_{i,\theta}$ is the complementary (conditional) cdf and

$$\lambda_{i,\rho}(x) = |v_{2,i}(x)| \sqrt{-\frac{\ln \rho}{2(|N_g|-1)}}.$$

Hence, in this case there is no need to resort to simulations.

In general, the inequality restrictions in (4.18) become weaker when $\lambda_{i,\rho}$ increases, i.e., the strategic relevance of the players among each other is stronger. This is a cost to the econometrician for not being able to recover fully the beliefs of individual players despite strong strategic interactions among them.

4.5 Bootstrap inference and asymptotic validity

4.5.1 Test statistics

For inference we compare the actual actions of the players and their predicted actions conditional on $\mathcal{F}$. Let $\mathcal{P}_0$ be a family of probability measures on $(\Omega, \mathcal{H})$ satisfying Assumptions 4.1-4.2 and 4.4-4.5. We seek for inference about the true model's parameter $\theta_0 \in \Theta$ that is robust to any choice of $\mathsf{P} \in \mathcal{P}_0$, a configurations of subjective beliefs (within the boundary set by Assumptions 4.1-4.2), and any pure strategy equilibrium among multiple equilibria implied by these beliefs.

First, we define sample analogues of $e_{i,L}$ and $e_{i,U}$ given in (4.17):

$$\hat{e}_{i,L} := \mathbf{1}\{Y_i = 1\} - \left(1 - \frac{\pi_{i,L}}{1-\rho_i}\right) \quad \text{and} \quad \hat{e}_{i,U} := \mathbf{1}\{Y_i = 1\} - \frac{\pi_{i,U}}{1-\rho_i}.$$

Further, let

$$\hat{\boldsymbol{\mu}}_L := n^{-1} \sum_{i=1}^{n} \hat{e}_{i,L}\, \mathbf{g}_i(X_i) \quad \text{and} \quad \hat{\boldsymbol{\mu}}_U := n^{-1} \sum_{i=1}^{n} \hat{e}_{i,U}\, \mathbf{g}_i(X_i). \tag{4.20}$$

Although the sample moments in (4.20) are similar to those employed in many researches in the literature of moment inequalities (see, e.g., Andrews and Shi, 2013; Andrews and Soares, 2010; Rosen, 2008), $\hat{\boldsymbol{\mu}}_L$ and $\hat{\boldsymbol{\mu}}_U$ here are not necessarily the sums of independent

or conditionally independent random variables. The summands $\hat{e}_{i,L}\,\mathbf{g}_i(X_i)$ and $\hat{e}_{i,U}\,\mathbf{g}_i(X_i)$ involve Y_{-i} so that they are dependent across i's in a complicated manner. In order to handle this issue, we introduce non-negative random vectors $\mathbf{w}_L, \mathbf{w}_U \in \mathbf{R}^m_{\geq 0}$ and take the following as our test statistic:

$$T \coloneqq \mathcal{T}\big(\sqrt{n}(\hat{\boldsymbol{\mu}}_L + \mathbf{w}_L), \sqrt{n}(\hat{\boldsymbol{\mu}}_U - \mathbf{w}_U)\big),$$

where $\mathcal{T} : \mathbf{R}^m \times \mathbf{R}^m \to \mathbf{R}$ is a function defined by $\mathcal{T}(x, y) \coloneqq \|[x]_- + [y]_+\|_1$.[11]

The random vectors $\mathbf{w}_L$ and $\mathbf{w}_L$ are defined as follows. First, recall that $\pi_{i,L}$ and $\pi_{i,U}$ are functions of (X_i, C, Y_{-i}). For $j \neq i$ and $1 \leq l \leq m$ let

$$\begin{aligned} d_{l,L} &\coloneqq n^{-1} \sum_{j \in N} \sum_{i \in N \setminus \{j\}} \frac{V_{\sigma(i,j)}(b \mapsto \pi_{i,L}(X_i, C, b)) g_{i,l}(X_i)}{1 - \rho_i} \quad \text{and} \\ d_{l,U} &\coloneqq n^{-1} \sum_{j \in N} \sum_{i \in N \setminus \{j\}} \frac{V_{\sigma(i,j)}(b \mapsto \pi_{i,U}(X_i, C, b)) g_{i,l}(X_i)}{1 - \rho_i}, \end{aligned} \tag{4.21}$$

where $\sigma(i, j)$ is the index of the j-th element of $N = \{1, \ldots, n\}$ in N_{-i}. The corresponding quantities in (4.21) determine the maximal approximation error due to the use of $\pi_{i,L}$ and $\pi_{i,U}$ instead of $\mathsf{E}[\pi_{i,L} \mid \mathcal{F}]$ and $\mathsf{E}[\pi_{i,U} \mid \mathcal{F}]$. For a given $\varrho \in (0, 1)$ we define the l-th element of $\mathbf{w}_L$ to be

$$\sqrt{-\frac{1}{2} \ln\Big(\frac{\varrho}{4m}\Big) \Big(d^*_{l,L}\Big)^2}.$$

where $d^*_{l,L}$ is the minimal measurable majorant of $d_{l,L}$. The elements of $\mathbf{w}_U$ are defined similarly.

Let $\tilde{\boldsymbol{\mu}}_L$ and $\tilde{\boldsymbol{\mu}}_U$ be defined as $\hat{\boldsymbol{\mu}}_L$ and $\hat{\boldsymbol{\mu}}_U$ in (4.20), where we replace $\pi_{i,L}$ and $\pi_{i,U}$ with their conditional expectations given $\mathcal{F}$. The following proposition shows that $\mathbf{w}_L$ and $\mathbf{w}_U$ control the difference between these moments with predetermined probability.

Proposition 4.2. *Suppose that Assumptions 4.1-4.5 hold. Then for a given $\varrho \in (0, 1)$,*

$$\mathsf{P}(|\hat{\boldsymbol{\mu}}_L - \tilde{\boldsymbol{\mu}}_L| \leq \mathbf{w}_L, |\hat{\boldsymbol{\mu}}_U - \tilde{\boldsymbol{\mu}}_U| \leq \mathbf{w}_U) \geq 1 - \varrho.$$

The result of Proposition 4.2 implies that

$$T \leq \mathcal{T}\big(\sqrt{n}\tilde{\boldsymbol{\mu}}_L, \sqrt{n}\tilde{\boldsymbol{\mu}}_U\big)$$

with probability of at least $1 - \varrho$. Consequently, since $\tilde{\boldsymbol{\mu}}_L$ and $\tilde{\boldsymbol{\mu}}_U$ are the sums of conditionally independent independent random vectors given $\mathcal{F}$, we employ these infeasible moments to

[11] For a vector $x \in \mathbf{R}^d$, $[x]_+ \equiv [x_j \vee 0]_{j=1}^d$ and $[x]_- \equiv -[x_j \wedge 0]_{j=1}^d$.

establish the validity of the bootstrap procedure described in the next subsection.

4.5.2 Bootstrap critical values

We adapt the idea of Romano et al. (2014) and propose the following bootstrap procedure. Consider i.i.d. standard normal random variables $\{\varepsilon_1, \ldots, \varepsilon_n\}$ and let

$$\mu_i^* := \frac{1}{2}\left(1 - \frac{\pi_{i,L} - \pi_{i,U}}{1 - \rho_i}\right) \vee 0 \wedge 1.$$

Then we define

$$\boldsymbol{\zeta}^* := n^{-1} \sum_{i=1}^{n} (\mathbf{1}\{Y_i = 1\} - \mu_i^*) \mathbf{g}_i(X_i)\, \varepsilon_i$$

Since we are unable to estimate the conditional expectation of $\mathbf{1}\{Y_i = 1\}$ given $\mathcal{F}$, the random variable μ_i^* serves as its proxy. In addition, given fixed $\kappa \in (0, 1)$ we let

$$\begin{aligned} \hat{\boldsymbol{\varphi}}_L &:= \left[\hat{\boldsymbol{\mu}}_L - \mathbf{w}_{\tau,L} - \mathbf{1}_K \cdot q^*(1 - \kappa/2)/\sqrt{n}\right]_+ \quad \text{and} \\ \hat{\boldsymbol{\varphi}}_U &:= \left[\hat{\boldsymbol{\mu}}_U + \mathbf{w}_{\tau,U} + \mathbf{1}_K \cdot q^*(1 - \kappa/2)/\sqrt{n}\right]_-, \end{aligned}$$

where $q^*(z)$ is a certain random variable that dominates the z-quantile of the conditional distribution of the maximum of $\tilde{\boldsymbol{\mu}}_L - \boldsymbol{\mu}_L$ (or $\tilde{\boldsymbol{\mu}}_U - \boldsymbol{\mu}_U$) given $\mathcal{F}$. The computation of $q^*(z)$ is described in the end of this section.

We consider the following bootstrap test statistic:

$$T^* := \mathcal{T}\big(\sqrt{n}(\boldsymbol{\zeta}^* + \hat{\boldsymbol{\varphi}}_L \wedge \hat{\boldsymbol{\varphi}}_U), \sqrt{n}(\boldsymbol{\zeta}^* - \hat{\boldsymbol{\varphi}}_L \wedge \hat{\boldsymbol{\varphi}}_U)\big),$$

where the minimum between $\hat{\boldsymbol{\varphi}}_L$ and $\hat{\boldsymbol{\varphi}}_U$ is taken element-wise. Consequently, the confidence set for $\theta_0 \in \Theta$ at nominal level $1 - \alpha$ is given by

$$CS_\epsilon := \{\theta \in \Theta : T \leq c^*(\gamma) \vee \epsilon\}, \tag{4.22}$$

where $\epsilon > 0$ is a fixed small number and $c^*(\gamma)$ is the $\gamma \equiv (1 - \alpha + \varrho + \kappa)$-quantile of the bootstrap distribution of T^*. The tuning parameters ϱ and κ should obviously satisfy $\varrho + \kappa < \alpha$ and can be chosen according to a Monte Carlo study.

Theorem 4.2. *Suppose that Assumptions 1-4 hold and $\exists C_{\mathbf{g}} > 0$ s.t. $\|g_{il}\|_\infty \leq C_{\mathbf{g}}$ for all $i \geq 1$ and $1 \leq l \leq m$. Furthermore, assume that there exists a sequence $\{\lambda_n\}$ s.t. $\lambda^{-1} = O(n^{1/(3+\delta)})$ for some $\delta > 0$ and*

$$\lim_{n\to\infty} \sup_{\mathsf{P} \in \mathcal{P}_0} \mathsf{P}(\underline{\lambda}(V) < \lambda) = 0.$$

Then for any $\epsilon > 0$,

$$\limsup_{n\to\infty} \sup_{\mathsf{P}\in\mathcal{P}_0} \mathsf{P}(\theta_0 \notin CS_\epsilon) \leq \alpha.$$

The observations are cross-sectionally dependent due to the public signals $\{\mathcal{C}_g : g \in G\}$. Cameron et al. (2008) proposed a wild bootstrap procedure for regression models with clustered errors. It is worth comparing our wild bootstrap procedure with theirs. Their wild bootstrap procedure requires that the simulated multipliers ε_i be group-specific. We cannot apply their method here, because we do not require the number of the groups to grow to infinity as the sample size increases. In contrast, our bootstrap procedure remains valid regardless of whether the number of the groups is small or large.

Consider the computation of $q^*(z)$. Let $\mathscr{S}_m \subset \mathbf{R}^{m\times m}$ denote the space of semi-positive definite matrices, and for $x \equiv (x_1, \ldots, x_n)$ let

$$B(x) \coloneqq n^{-1} \sum \mathbf{g}_i(x_i)\mathbf{g}_i(x_i)^\top.$$

We denote the smallest eigenvalue of $A \in \mathscr{S}_m$ as $\underline{\lambda}(A)$. Then $q^*(z)$ is the minimal measurable majorant of $q^*(z, (X_1, \ldots, X_n))$, where

$$q^*(z, x) \coloneqq \sup\{H^{-1}(z; A) : A \in \mathscr{S}_m,\ \underline{\lambda}(A) \geq \lambda, |A| \leq [B(x)]\},$$

$H(\,\cdot\,;\Sigma)$ is the cdf of the maximum of $Z \sim \mathcal{N}(0, \Sigma)$, $|A|$ is the element-wise absolute value of A, and λ is the parameter appearing in Theorem 4.2.

4.6 Conclusion

This chapter focuses on a large Bayesian game perspective for social interactions models, and develops an inference method that is robust to heterogeneous formation of beliefs among the players. Utilizing the strategic interdependence among the players and the assumption of conditionally independent types, this work derives testable implications from the equilibrium constraints.

The framework proposed in this chapter may have limitations in some applications for several reasons. First, the framework assumes that the information groups are exogenously given in the beginning of the game. This does not cause any problem, if the current game's types satisfy the conditional independence assumption given any information used by the agents in the endogenous group formation that occurs prior to the game. However, this conditional independence assumption is violated when the agents observe the groups formed, before entering the current game. Second, the framework assumes that the idiosyncratic component of the types is not shared between two different players. This assumption excludes a large network model where the information flow among the agents can be highly complex. Researches on both fronts require further research.

Bibliography

Andrews, D. W. K., 2005. Cross-section regression with common shocks. Econometrica 73 (5), 1551–1585. 52

Andrews, D. W. K., Shi, X., 2013. Inference based on conditional moment inequalities. Econometrica 81 (2), 609–666. 61, 62

Andrews, D. W. K., Soares, G., 2010. Inference for parameters defined by moment inequalities using generalized moment selection. Econometrica 78 (1), 119–157. 62

Aradillas-Lopez, A., 2010. Semiparametric estimation of a simultaneous game with incomplete information. Journal of Econometrics 157 (2), 409–431. 50, 52

Aradillas-Lopez, A., Tamer, E., 2008. The identification power of equilibrium in simple games. Journal of Business & Economic Statistics 26 (3), 261–283. 50, 51, 52

Athey, S., 2001. Single crossing properties and the existence of pure strategy equilibria in games of incomplete information. Econometrica 69 (4), 861–889. 55

Athreya, K. B., Lahiri, S. N., 2006. Measure Theory and Probability Theory (Springer Texts in Statistics). Springer-Verlag, Berlin, Heidelberg. 101, 106

Aumann, R. J., 11 1976. Agreeing to disagree. Ann. Statist. 4 (6), 1236–1239. 54

Bajari, P., Hong, H., Krainer, J., Nekipelov, D., 2010. Estimating static models of strategic interactions. Journal of Business & Economic Statistics 28 (4), 469–482. 52

Baldi, P., Rinott, Y., 1989. On normal approximations of distributions in terms of dependency graphs. Annals of Probability 17, 1646–1650. 4

Belyaev, Y., Sjöstedt-de Luna, S., 2000. Weakly approaching sequences of random distributions. Journal of Applied Probability 37 (3), 807–822. 49

Bentkus, V., 2003. On the dependence of the Berry-Esseen bound on dimension. Journal of Statistical Planning and Inference 113, 385–402. 41

Beresteanu, A., Molchanov, I., Molinari, F., 2011. Sharp identification regions in models with convex moment predictions. Econometrica 79 (6), 1785–1821. 50, 52

Bergemann, D., Morris, S., 2005. Robust mechanism design. Econometrica 73 (6), 1771–1813. 51

Berti, P., Pratelli, L., Rigo, P., 2006. Almost sure weak convergence of random probability measures. Stochastics 78 (2), 91–97. 38

Bühlmann, P. L., 1993. The blockwise bootstrap in time series and empirical processes. Ph.D. thesis, Swiss Federal Institute of Technology Zürich. 30

Birkel, T., 1988. On the convergence rate in the central limit theorem for associated processes. Annals of Probability 16, 1685–1698. 13

Bisin, A., Moro, A., Topa, G., 2011. The empirical content of models with multiple equilibria in economies with social interactions. NBER Working Papers 17196, National Bureau of Economic Research, Inc. 53

Bläsius, T., Friedrich, T., Krohmer, A., 2017. Cliques in hyperbolic random graphs. Algorithmica, 2324–2344. 8

Blume, L. E., Brock, W. A., Durlauf, S. N., Jayaraman, R., 2011. Linear social network models, working Paper. 52

Borg, I., Groenen, P. J. F., 2005. Modern Multidimensional Scaling. Springer Verlag, New York. 3, 8

Boucher, V., Mourifie, I., 2017. My friend far, far away: A random field approach to exponential random graph models. Econometrics Journal 20, S14–S46. 3

Bresnahan, T. F., Reiss, P. C., 1991. Empirical models of discrete games. Journal of Econometrics 48 (1), 57–81. 50, 52

Brock, W. A., Durlauf, S. N., 2001. Discrete choice with social interactions. The Review of Economic Studies 68 (2), 235–260. 52, 53, 55

Calhoun, G., 2018. Block bootstrap consistency under weak assumptions. Econometric Theory, 1–24. 44

Cameron, A. C., Gelbach, J. B., Miller, D. L., 2008. Bootstrap-based improvements for inference with clustered errors. The Review of Economics and Statistics 90 (3), 414–427. 65

Carlstein, E., 1986. The use of subseries values for estimating the variance of a general statistic from a stationary sequence. The Annals of Statistics 14 (3), 1171–1179. 29

Chen, L. H. Y., Goldstein, L., Shao, Q.-M., 2011. Normal Approximation by Stein's Method. Springer-Verlag, New York, USA. 85

Chen, L. H. Y., Shao, Q. M., 2004. Normal approximation under local dependence. Annals of Probability 32, 1985–2028. 4

Chernozhukov, V., Chetverikov, D., Kato, K., 2013. Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. Annals of Statistics 41 (6), 2786–2819. 39, 128

Chernozhukov, V., Chetverikov, D., Kato, K., 2015. Comparison and anti-concentration bounds for maxima of Gaussian random vectors. Probability Theory and Related Fields 162, 47–70. 130

Ciliberto, F., Tamer, E., 2009. Market structure and multiple equilibria in airline markets. Econometrica 77 (6), 1791–1828. 50, 52

Cohen, S., Elliott, R. J., 2015. Stochastic Calculus and Applications, 2nd Edition. Probability and Its Applications. Birkhäuser Basel. 36

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. 3, 30

Conley, T. G., 1999. GMM estimation with cross-sectional dependence. Journal of Econometrics 92, 1–45. 3, 8, 30, 32

Crimaldi, I., 2009. An almost sure conditional convergence result and an application to a generalized Pólya urn. International Mathematical Forum 4 (23), 1139–1156. 100

Da Prato, G., Zabczyk, J., 2014. Stochastic Equations in Infinite Dimensions, 2nd Edition. Encyclopedia of Mathematics and Its Applications. Cambridge University Press. 74

Davidson, J., 1994. Stochastic Limit Theory: An Introduction for Econometricians. Oxford University Press, New York. 14

de Paula, u., 2013. Econometric analysis of games with multiple equilibria. Annual Review of Economics 5 (1), 107–131. 52

de Paula, u., Tang, X., 2012. Inference of signs of interaction effects in simultaneous games with incomplete information. Econometrica 80 (1), 143–172. 50, 52

Deb, J., Kalai, E., 2015. Stability in large Bayesian games with heterogeneous players. Journal of Economic Theory 157, 1041–1055. 51

Dedecker, J., Doukhan, P., Lang, G., 2007. Weak Dependence: With Examples and Applications. Lecture Notes in Statistics. Springer, New York. 33

Dominitz, J., Manski, C. F., 1997. Using expectations data to study subjective income expectations. Journal of the American Statistical Association 92 (439), 855–867. 51

Doukhan, P., Louhichi, S., 1999. A new weak dependence condition and applications to moment inequalities. Stochastic Processes and their Applications 84 (2), 313–342. iii, 4, 8, 9, 13, 28, 29, 49

Durrett, R., 2010. Probability: Theory and Examples, 4th Edition. Cambridge University Press. 107

Embrechts, P., Hofert, M., 2013. A note on generalized inverses. Mathematical Methods of Operations Research 77 (3), 423–432. 91

Gaetan, C., Guyon, X., 2010. Spatial Statistics and Modeling. Springer, New York. 4

Garcia-Soidan, P., Menezes, R., Rubinos, O., 2014. Bootstrap approaches for spatial data. Stochastic Environmental Research and Risk Assessment 28 (5), 1207–1219. 30

Gonçalves, S., Politis, D., 2011. Discussion: Bootstrap methods for dependent data: A review. Journal of the Korean Statistical Society 40 (4), 383–386. 30

Hall, P., 1992. The Bootstrap and Edgeworth Expansion. Springer Series in Statistics. Springer-Verlag, New York, Berlin, Paris. 39

Hall, P. G., Heyde, C. C., 1980. Martingale Limit Theory and Its Applications. Academic Press, New York; London. 11

Higham, N., 1988. Computing a nearest symmetric positive semidefinite matrix. Linear Algebra and its Applications 103, 103–118. 98

Higham, N., 2002. Computing the nearest correlation matrix - a problem from finance. IMA Journal of Numerical Analysis 22 (3), 329–343. 98

Higham, N. J., 2008. Functions of Matrices: Theory and Computation. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA. 47

Ioannides, Y. M., 2013. From Neighborhoods to Nations: The Economics of Social Interactions. Princeton University Press. 52

Ioannides, Y. M., Zabel, J. E., 2008. Interactions, neighborhood selection and housing demand. Journal of Urban Economics 63 (1), 229–252. 52

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. 4

Jenish, N., Prucha, I. R., 2009. Central limit theorems and uniform laws of large numbers for arrays of random fields. J. Econometrics 150 (1), 86–98. 3, 8, 10, 14, 32, 80

Jenish, N., Prucha, I. R., 2012. On spatial processes and asymptotic inference under near-epoch dependence. Journal of Econometrics 170, 178–190. 3

Jensen, D. R., 1984. Ordering ellipsoidal measures: Scale and peakness ordering. SIAM Journal on Applied Mathematics 44 (6), 1226–1231. 117

Jia, P., 2008. What happens when wal-mart comes to town: An empirical analysis of the discount retailing industry. Econometrica 170, 1263–1316. 3

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. 14

Kalai, E., 2004. Large robust games. Econometrica 72 (6), 1631–1665. iii, 51

Kallenberg, O., 2002. Foundations of Modern Probability, 2nd Edition. Probability and its Applications (New York). Springer-Verlag, New York. 91, 109, 135

Kelejian, H. H., Prucha, I. R., 2007. HAC estimation in aa spatial framework. Journal of Econometrics, 131–154. 23, 29

Kelejian, H. H., Prucha, I. R., 2010. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. Journal of Econometrics 157 (1), 53–67, nonlinear and Nonparametric Methods in Econometrics. 34

Künsch, H. R., 1989. The jackknife and the bootstrap for general stationary observations. The Annals of Statistics 17, 1217–1241. 29

Krauth, B. V., 2006. Simulation-based estimation of peer effects. Journal of Econometrics 133 (1), 243–271. 52

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. 52

Lahiri, S. N., 1992. Edgeworth correction by 'moving block' bootstrap for stationary and nonstationary data. In: LePage, R., Billard, L. (Eds.), Exploring the limits of bootstrap. John Wiley & Sons, New York; Chichester, pp. 183–214. 42

Lahiri, S. N., 2003. Resampling Methods for Dependent Data. Springer Series in Statistics. 30, 39, 42

Lauritzen, S. L., 1996. Graphical Models. Clarendon Press, Oxford. 4

Lee, J. H., Song, K., 2018. Stable limit theorems for empirical processes under conditional neighborhood dependence. Bernoulli, Forthcoming. 4

Lee, L.-F., 2004. Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. Econometrica 72, 1899–1925. 4

Lee, L.-F., Liu, X., Lin, X., 2010. Specification and estimation of social interaction models with network structures. Econometrics Journal 13, 145–176. 4

Leung, M. P., 2016. Treatment and spillover effects under network interference, working Paper. 4

Li, J., Lee, L.-f., 2009. Binary choice under social interactions: An empirical study with and without subjective data on expectations. Journal of Applied Econometrics 24 (2), 257–281. 51, 52

Liu, R. Y., Singh, K., 1992. Moving blocks jackknife and bootstrap capture weak dependence. In: LePage, R., Billard, L. (Eds.), Exploring the limits of bootstrap. John Wiley & Sons, New York; Chichester, pp. 225–248. 29

Manski, C. F., 2004. Measuring expectations. Econometrica 72 (5), 1329–1376. 51

Matyas, L., 1999. Generalized Method of Moments Estimation. Cambridge University Press, Cambridge. 29

McAdams, D., 2003. Isotone equilibrium in games of incomplete information. Econometrica 71 (4), 1191–1214. 55

McDiarmid, C., 1989. On the method of bounded differences. In: Siemons, J. (Ed.), Surveys in Combinatorics. Vol. 141 of London Mathematical Society Lecture Notes. Cambridge University Press, pp. 148–188. 120

Menzel, K., 2012. Inference for large games with exchangeable players, working paper. 53

Milgrom, P. R., Weber, R. J., 1985. Distributional strategies for games with incomplete information. Mathematics of Operations Research 10 (4), 619–632. 55

Murphy, K. P., 2012. Machine Learning: A Probabilistic Perspective. The MIT Press, New York, USA. 4

Paparoditis, E., Politis, D. N., 2001. Tapered block bootstrap. Biometrika 88 (4), 1105–1119. 30

Pearl, J., 2009. Causality. Cambridge University Press, New York. 4

Penrose, M., 2003. Random Geometric Graphs. Oxford University Press, New York, USA. 13, 17, 84

Petty, C. M., 1971. Equilateral sets in minkowski spaces. Proc. Amer. Math. Soc. 29, 369–374. 8

Politis, D. N., Romano, J. P., 1992. A circular block-resampling procedure for stationary data. In: LePage, R., Billard, L. (Eds.), Exploring the limits of bootstrap. John Wiley & Sons, New York; Chichester, pp. 263–270. 30

Politis, D. N., Romano, J. P., 1994. The stationary bootstrap. Journal of the American Statistical Association 89 (428), 1303–1313. 30

Politis, D. N., Romano, J. P., Wolf, M., 1999. Subsampling. Springer. 30

Prakasa Rao, B. L. S., 2013. Conditional independence, conditional mixing and conditional association. Annals of the Institute of Statistical Mathematics 61, 441–460. 10

Reny, P. J., 2011. On the existence of monotone pure-strategy equilibria in bayesian games. Econometrica 79 (2), 499–553. 55

Rhee, W., Talagrand, M., 1986. Uniform bound in the central limit theorem for Banach space valued dependent random variables. Journal of Multivariate Analysis 20 (2), 303–320. 105, 131, 137

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. 4

Röllin, A., 2013. Stein's method in high dimensions with applications. Ann. Inst. H. Poincaré Probab. Statist. 49 (2), 529–549. 109

Romano, J. P., Shaikh, A. M., Wolf, M., 2014. A practical two-step method for testing moment inequalities. Econometrica 82 (5), 1979–2002. 64

Rosen, A. M., 2008. Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities. Journal of Econometrics 146 (1), 107–117. 62

Sengupta, S., Shao, X., Wang, Y., 2015. The dependent random weighting. Journal of Time Series Analysis 36 (3), 315–326. 30

Shao, J., Tu, D., 1995. The Jackknife and Bootstrap. Springer-Verlag, Berlin; New York. 37

Shao, X., 2010. The dependent wild bootstrap. Journal of the American Statistical Association 105 (489), 218–235. 30, 46, 47

Shiryaev, A. N., 2016. Probability-1, 3rd Edition. Graduate Texts in Mathematics. Springer-Verlag New York. 101

Song, K., 2018. Measuring the graph concordance of locally dependent observations. Review of Economics and Statistics 100, 535–549. 4

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. 4, 11

Talagrand, M., 2011. Mean Field Models for Spin Glasses, Volume I: Basic Examples. Vol. 54 of A Series of Modern Surveys in Mathematics. Springer-Verlag. 107, 126

Tamer, E., 01 2003. Incomplete simultaneous discrete response model with multiple equilibria. The Review of Economic Studies 70 (1), 147–165. 50, 52

Wendland, H., 2004. Scattered Data Approximation. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press. 23

Wilson, R., 1987. Game-Theoretic Analyses of Trading Processes. Econometric Society Monographs. Cambridge University Press, pp. 33–70. 51

Xu, H., 2018. Social interactions in large networks: A game theoretic approach. International Economic Review 59 (1), 257–284. 53

Zhang, F., 2011. Matrix Theory: Basic Results and Techniques, 2nd Edition. Universitext. Springer-Verlag New York. 97

Appendix A

Appendix to Chapter 2

A.1 Auxiliary results for ψ-weakly dependent processes

In this section we present covariance inequalities for functions of general ψ-weak dependent processes. Consider a triangular array of $\mathbf{R}^v$-valued random vectors $\{Y_{n,i}\}_{i\in N_n}$, $n \geq 1$, associated with a given sequence of networks $\{G_n\}$. Let $\mathcal{F}_a$ and $\mathcal{G}_a$ be some classes of functions on $\mathbf{R}^{v\times a}$, 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 $\{Y_{n,i}\}$ is conditionally $(\mathcal{F}, \mathcal{G}, \psi)$-weakly dependent given $\{\mathcal{C}_n\}$, if for each $n \in \mathbb{N}$, there exists a non-negative, $\mathcal{C}_n$-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 non-random 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 $\{\theta_{n,s}\}$. 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 $g \notin \mathcal{G}_b$. For given $(A, B) \in \mathcal{P}_n(a, b; s)$ with $s > 0$ we define

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

We proceed assuming that the weak dependence coefficients are a.s. bounded by a constant $M \geq 1$. This assumption can be easily relaxed by truncating θ-s and noticing that $\theta = (\theta \vee 1)(\theta \wedge 1)$. However, we replace Assumption 2.1(a) with the following restriction on the classes of functions $\mathcal{F}$ and $\mathcal{G}$:

Assumption A.1. (i) $\mathcal{F}$ and $\mathcal{G}$ are stable under multiplication by constants, i.e., if $f \in \mathcal{F}$, $g \in \mathcal{G}$, and $c \in \mathbf{R}$, then $cf \in \mathcal{F}$ and $cg \in \mathcal{G}$. (ii) 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 covariance inequality that permits the nonlinear transforms to be random functions. Suppose that $\mathcal{Z}_j$, $j = 1,2$ is a random element taking values in a separable metric space $(\mathcal{Z}_j, \rho_j)$ equipped with the Borel σ-field $\mathcal{B}(\mathcal{Z}_j)$ 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). 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 $\mathcal{Z}_1 \times \mathcal{Z}_2$.*[1] *If Z_1 and Z_2 are $\mathcal{C}_n$-measurable and $\bar{f}, \bar{g} \in L^2$, then*

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

Proof. Suppose w.l.o.g. that $\mathsf{E}[f(Y_{n,A}, Z_1) \mid \mathcal{C}_n] = 0$ and $\mathsf{E}[g(Y_{n,B}, Z_2) \mid \mathcal{C}_n] = 0$ 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 $m \ge 1$, $Z_{j,m} = \sum_{k=1}^m z_{j,k}\mathbf{1}_{A_{j,k}}$, 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 $k \ne l$. Then, letting $B_{k,l} := A_{1,k} \cap A_{2,l}$,

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

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 $(z_1,z_2) \in \mathcal{Z}_1 \times \mathcal{Z}_2$, $F(z_1,z_2) \le \tilde{F}(z_1,z_2)$.

[1] Note that the continuity of F implies that it is Borel measurable. Moreover, if $\mathcal{Z}_j = \mathcal{Z}_{j,1} \times \mathbb{Z}_{j,2}$, $j = 1,2$, 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.

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 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*

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

Then

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

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

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

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

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

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,

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

Consider each term in the last inequality separately. By Lemma A.1 and Assumption A.1 we find that[2]

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

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

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

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

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

Similarly,

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

Finally,

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

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

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

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

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

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

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

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

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 (b) 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.[3] 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*

(a) *The functions f and g are continuous,*
(b) *$\mu_{\xi,p} < \infty$ 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

(c) *$\gamma_1 \coloneqq \max_{i\in A}\max_{1\le k\le v}\|h_1^{-1}(|[Y_{n,i}]_k|)\|_{\mathcal{C}_n,p} < \infty$ a.s. and*
$\gamma_2 \coloneqq \max_{i\in B}\max_{1\le k\le v}\|h_2^{-1}(|[Y_{n,i}]_k|)\|_{\mathcal{C}_n,q} < \infty$ a.s.,
(d) *$f_K \coloneqq \varphi_{K_1} \circ f \circ \varphi_{h_1(K_2)} \in \mathcal{F}_a$ and $g_K \coloneqq \varphi_{K_1} \circ g \circ \varphi_{h_2(K_2)} \in \mathcal{G}_b$ for all $K \in (0,\infty)^2$,*
(e) *$(K,L) \mapsto \psi_{a,b}(f_K, g_L)$ is continuous on $(0,\infty)^4$.*

Then

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

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

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

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

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

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 (c) 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

$$\pi_1 = \prod_{i\in A} \|Y_{n,i}\|_{\mathcal{C}_n,p_i}, \quad \tilde{\gamma}_1 = \max_{i\in A} \|Y_{n,i}^{a-1}\|_{\mathcal{C}_n,p}, \quad \textit{and}$$
$$\pi_2 = \prod_{i\in B} \|Y_{n,i}\|_{\mathcal{C}_n,q_i}, \quad \tilde{\gamma}_2 = \max_{i\in B} \|Y_{n,i}^{b-1}\|_{\mathcal{C}_n,q},$$

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

$$\begin{aligned}
\xi_{\kappa\kappa} &\coloneqq \mu_{\xi,p}^{-1} f_{(\kappa\mu_{\xi,p},\kappa\gamma_1)}(Y_{A,n})\mathbf{1}_{\Xi}, & \xi^*_{\kappa\kappa} &\coloneqq \xi_{\kappa\kappa} - \mathsf{E}[\xi_{\kappa\kappa} \mid \mathcal{C}_n],\\
\hat{\xi}_{\kappa\kappa} &\coloneqq \xi_{\kappa} - \xi_{\kappa\kappa}, & \hat{\xi}^*_{\kappa\kappa} &\coloneqq \hat{\xi}_{\kappa\kappa} - \mathsf{E}[\hat{\xi}_{\kappa\kappa} \mid \mathcal{C}_n],
\end{aligned}$$

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

$$\begin{aligned}
|\mathsf{E}[\xi^*_{\kappa}\zeta^*_{\lambda} \mid \mathcal{C}_n]| \le{}& |\mathsf{E}[\xi^*_{\kappa\kappa}\zeta^*_{\lambda\lambda} \mid \mathcal{C}_n]| + |\mathsf{E}[\xi^*_{\kappa\kappa}\hat{\zeta}^*_{\lambda\lambda} \mid \mathcal{C}_n]|\\
&+ |\mathsf{E}[\hat{\xi}^*_{\kappa\kappa}\zeta^*_{\lambda\lambda} \mid \mathcal{C}_n]| + |\mathsf{E}[\hat{\xi}^*_{\kappa\kappa}\hat{\zeta}^*_{\lambda\lambda} \mid \mathcal{C}_n]| \quad \text{a.s.}
\end{aligned}$$

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

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

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

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

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

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

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

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

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 the inequality (A.4) established in the proof of Theorem A.1. ■

A.2 Proofs of the main results

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

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

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

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

■

Proof of Theorem 2.1. 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.[4] Consider a censored version of $Y_{n,i}$. For some $k > 0$ let

$$Y_{n,i} = Y_{n,i}^{(k)} + \tilde{Y}_{n,i}^{(k)},$$

where

$$Y_{n,i}^{(k)} := \varphi_k(Y_{n,i}) \quad \text{and} \quad \tilde{Y}_{n,i}^{(k)} := Y_{n,i} - \varphi_k(Y_{n,i}).$$

We have:

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

Since $\tilde{Y}_{n,i}^{(k)} = 0$ on $\{|Y_{i,k}| \le k\}$, $\mathsf{E}|\tilde{Y}_{n,i}^{(k)}| = \mathsf{E}|\tilde{Y}_{n,i}^{(k)}|\mathbf{1}\{|Y_{n,i}| > k\} \le 2\mathsf{E}|Y_{n,i}|\mathbf{1}\{|Y_{n,i}| > k\}$. Hence, using the triangle inequality, the second term on the right-hand side of (A.6) can be bounded by $2\sup_{i \in N_n} \mathbf{E}|\tilde{Y}_{n,i}^{(k)}| = 4\sup_{i \in N_n} \mathbf{E}|Y_{n,i}|\mathbf{1}\{|Y_{n,i}| > k\}$. By Assumption 2.2, for each $\varepsilon > 0$ one now can find k such that the second term on the right-hand side of (A.6) is

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

smaller than $\varepsilon/2$ for all n large.

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

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

By the norm inequality,

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

where

$$\begin{aligned}
\sigma_{n,k}^2 &\coloneqq \mathsf{E}\left[\sum_{i=1}^{n}\Big(Y_{n,i}^{(k)} - \mathsf{E}[Y_{n,i}^{(k)} \mid \mathcal{C}_n]\Big)\right]^2 \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n}\mathsf{E}\Big[(Y_{n,i}^{(k)} - \mathsf{E}[Y_{n,i}^{(k)} \mid \mathcal{C}_n])(Y_{j,n}^{(k)} - \mathsf{E}[Y_{j,n}^{(k)} \mid \mathcal{C}_n])\Big] \\
&\leq 4nk^2 + \sum_{i=1}^{n}\sum_{s=1}^{n-1}\sum_{j:d_n(i,j)=s}\mathbf{E}\Big|\mathrm{Cov}(Y_{n,i}^{(k)}, Y_{j,n}^{(k)} \mid \mathcal{C}_n)\Big|.
\end{aligned}$$

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

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

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

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

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

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

The result in (A.7) now follows by Assumption 2.3 for all n sufficiently large. ■

Lemma A.2. *Suppose that the conditions of Theorem 2.2 hold and let $g : \mathbf{R} \to \mathbf{R}$ be a twice continuously differentiable bounded function with bounded derivatives. Then*

$$\left|\mathsf{E}[g'(S_n) - S_n g(S_n) \mid \mathcal{C}_n]\right| \lesssim \bar{\Delta}_n(g) + \frac{n\mu_{n,3}^3 \|g''\|_\infty \delta_n(m_n;2)}{\sigma_n^3} \quad a.s.,$$

where

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

and $H_n(\,\cdot\,,\,\cdot\,)$ is as defined in Assumption 2.5.

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

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

where

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

First, applying Taylor expansion and the arithmetic-geometric mean inequality,

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

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

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

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

$$A_{n,1} = g'(\tilde{S}_n) \left(1 - \sum_{i \in N_n} \sum_{j \in N_n(i; m_n)} \tilde{Y}_{n,i} \tilde{Y}_{n,j} \right).$$

Since, by the definition of σ_n,

$$1 = \mathsf{E}\left[\sum_{i \in N_n} \sum_{j \in N_n} \tilde{Y}_{n,i} \tilde{Y}_{n,j} \mid \mathcal{C}_n \right] \quad \text{a.s.},$$

we find that

$$\begin{aligned}\mathsf{E}[A_{n,1} \mid \mathcal{C}_n] &= -\mathsf{E}\left[g'(\tilde{S}_n) \left(\sum_{i \in N_n} \sum_{j \in N_n(i; m_n)} (\tilde{Y}_{n,i} \tilde{Y}_{n,j} - \mathsf{E}[\tilde{Y}_{n,i} \tilde{Y}_{n,j} \mid \mathcal{C}_n]) \right) \mid \mathcal{C}_n \right] \\ &\quad + \mathsf{E}\Big[g'(\tilde{S}_n) \mid \mathcal{C}_n \Big] \left(\sum_{i \in N_n} \sum_{j \in N_n \setminus N_n(i; m_n)} \mathsf{E}[\tilde{Y}_{n,i} \tilde{Y}_{n,j} \mid \mathcal{C}_n] \right) \\ &\equiv -R_{n,1} + R_{n,2}.\end{aligned}$$

Using the Cauchy-Schwarz inequality and letting $Z_{n,ij} \coloneqq c_n^2 (\tilde{Y}_{n,i} \tilde{Y}_{n,j} - \mathsf{E}[\tilde{Y}_{n,i} \tilde{Y}_{n,j} \mid \mathcal{C}_n])$,

$$|R_{n,1}| \leq \frac{1}{c_n^2} \sqrt{\mathsf{E}[(g'(\tilde{S}_n))^2 \mid \mathcal{C}_n]} \cdot \sqrt{\mathsf{E}\left[\left(\sum_{i \in N_n} \sum_{j \in N_n(i; m_n)} Z_{n,ij} \right)^2 \mid \mathcal{C}_n \right]} \quad \text{a.s.}$$

Corollary A.2 implies that

$$\begin{aligned}\mathsf{E}\left[\left(\sum_{i \in N_n} \sum_{j \in N_n(i; m_n)} Z_{n,ij} \right)^2 \mid \mathcal{C}_n \right] &= \sum_{s=0}^{n-2} \sum_{(i,j,k,l) \in H_n(s; m_n)} \mathsf{E}[Z_{n,ij} Z_{n,kl} \mid \mathcal{C}_n] \\ &\leq M(\mu_{n,p}^2 + \mu_{n,p}^4) \sum_{s=0}^{n-2} |H_n(s; m_n)| \theta_{n,m}^{1-4/p} \quad \text{a.s.},\end{aligned}$$

Thus,

$$|R_{n,1}| \le \frac{C_1\|g'\|_\infty}{c_n^2}\sqrt{(\mu_{n,p}^2+\mu_{n,p}^4)\sum_{s=0}^{n-2}|H_n(s;m_n)|\theta_{n,s}^{1-4/p}} \quad \text{a.s.},$$

where $C_1 > 0$ is a constant. Using Corollary A.1,

$$|R_{n,2}| \le \frac{C_2 n(n-\delta_n(m_n))\mu_{n,p}^2\|g'\|_\infty\theta_{n,m_n}^{1-4/p}}{\sigma_n^2} \quad \text{a.s.}, \tag{A.11}$$

where $C_2 > 0$ is a constant.

Finally, consider $A_{n,3}$. Let $S_{n,i} := \tilde{S}_{n,i}\sigma_n$ and $h_n(x) := g(x/c_n)$. Note that

$$\mathsf{E}[A_{n,3} \mid \mathcal{C}_n] \le c_n^{-1}\sum_{i\in N_n}\left|\mathrm{Cov}\left(\frac{Y_{n,i}c_n}{\sigma_n}, h_n\left(\frac{S_{n,i}c_n}{\sigma_n}\right) \mid \mathcal{C}_n\right)\right| \quad \text{a.s.}$$

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

$$\begin{aligned}&\left(M\sup_{K,L\ge 1}\frac{1}{KL}\psi_{a,b}(\varphi_{K\mu_{n,p}c_n/\sigma_n}\circ f_n, \varphi_{L\|g\|_\infty}\circ h_n) + \frac{16\mu_{n,p}c_n\|g\|_\infty}{\sigma_n}\right)\theta_{n,m_n}^{1-1/p-1/q}\\ &\quad\le C_3(n-\delta_n(m_n))\left(\frac{\mu_{n,p}c_n}{\sigma_n}+1\right)\left(\|g\|_\infty + \frac{\|g'\|_\infty}{c_n}\right)\theta_{n,m_n}^{1-1/p-1/q} \quad \text{a.s.},\end{aligned}$$

where q is such that $p^{-1}+q^{-1} < 1$, $f_n(x) := xc_n/\sigma_n$, and $C_3 > 0$ is a constant. Thus, taking $q \to \infty$, we conclude that

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

Since $c_n \le \sigma_n$, 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* $\mathsf{E}[Y_{n,i} \mid \mathcal{C}_n] = 0$ *a.s. Then*

$$\begin{aligned}&\sup_{t\in\mathbf{R}}|\mathsf{P}(S_n/\sigma_n \le t \mid \mathcal{C}_n) - \Phi(t)|\\ &\quad\lesssim \left(\frac{\mu_{n,3}}{\sigma_n}\right)^{3/2}\sqrt{n\delta_n(m_n;2)} + \frac{1}{c_n^2}\sqrt{(\mu_{n,p}^2+\mu_{n,p}^4)\sum_{s=0}^{n-2}|H_n(s;m_n)|\,\theta_{n,s}^{1-4/p}}\\ &\qquad + \frac{n(n-\delta_n(m_n))(\mu_{n,p}+1)}{c_n}\left(1+\frac{1}{c_n}\right)\theta_{n,m_n}^{1-1/p} \quad \textit{a.s.},\end{aligned}$$

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

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 $x \le t$, $h_+(x) = 0$ for $x \ge t + \varepsilon$, and h_+ is continuous and linear on $[x, x+\varepsilon]$. Similarly, we also take $h_-(x) = 1$ for $x \le t - \varepsilon$, $h_-(x) = 0$ for $x \ge t$, and h_- is continuous and linear on $[x - \varepsilon, x]$. Define for any real function g,

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

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),

$$\left|\mathsf{E}[g'(S_n) - S_n g(S_n) \mid \mathcal{C}_n]\right| = \Delta_n(h) \quad \text{a.s.} \tag{A.12}$$

where

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

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

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

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

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

for some constant $C > 0$. Next,

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

where ϕ is the density of $\mathcal{N}(0,1)$. Similarly,

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

Hence, we find that

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

Finally, choosing

$$\varepsilon = \sqrt{\frac{n\delta_n(m_n;2)\mu_{n,3}^3}{\phi(0)\sigma_n^3}}$$

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

Proof of Theorem 2.2. The desired result follows from Lemma A.3 in combination with the conditions given in the theorem. Details are omitted. ■

Proof of Proposition 2.5. For the first implication it suffices to show that for any vector $c \in \mathbf{R}^v$ with $\|c\| = 1$, $\mathsf{E}[|A_n(c)| \mid \mathcal{C}_n] \to 0$ 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, $\mathsf{E}[y_{n,i} \mid \mathcal{C}_n] = 0$ a.s. and by Assumption 2.6(a)

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

Then

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

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

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

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

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

and it follows from Assumption 2.6(b & d) and the dominated convergence theorem that $|R_{n,2}| \to 0$ a.s.

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

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

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

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

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

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

we find that

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

Hence, it follows from Assumption 2.6(c) that $\mathsf{E}[R_{n,1}^2 \mid \mathcal{C}_n] \to 0$ a.s.

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

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

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

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

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

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

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

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 $|B_{n,i}| \le B_n$ for all $i \in N_n$, Assumption 2.6 implies that

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

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

Appendix B

Appendix to Chapter 3

B.1 Proofs of the main results

In the following, let φ_K with $K \in \mathbf{R}_+$ denote the element-wise censoring function, i.e., for an indexed family of real numbers $\mathbf{x} \equiv (x_i)_{i \in I}$,

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

Proof of Proporsition 3.1. Fix $\kappa \geq 1$ and let $\xi \coloneqq (f \circ h)(Z_{n,A})$ and $\zeta \coloneqq (g \circ h)(Z_{n,B})$, where $f, g \in \mathcal{L}_w$ and $(A, B) \in \mathcal{P}_n(a, b; s)$. Define the censored versions

$$\xi_\kappa \coloneqq (f \circ h \circ \varphi_\kappa)(Z_{n,A}) \quad \text{and} \quad \zeta_\kappa \coloneqq (g \circ h \circ \varphi_\kappa)(Z_{n,B}).$$

Then

$$\begin{aligned} |\mathrm{Cov}(\xi, \zeta \mid \mathcal{C})| &\leq |\mathrm{Cov}(\xi - \xi_\kappa, \zeta - \zeta_\kappa \mid \mathcal{C})| + |\mathrm{Cov}(\xi_\kappa, \zeta_\kappa \mid \mathcal{C})| \\ &\leq 2\|f\|_\infty \mathsf{E}[|\zeta - \zeta_\kappa| \mid \mathcal{C}] + 2\|g\|_\infty \mathsf{E}[|\xi - \xi_\kappa| \mid \mathcal{C}] \\ &\quad + |\mathrm{Cov}(\xi_\kappa, \zeta_\kappa \mid \mathcal{C})| \quad \text{a.s.} \end{aligned}$$

First, $\mathrm{Lip}(f \circ h \circ \varphi_\kappa) \leq 2\eta\kappa^{\tau-1} \mathrm{Lip}(f)$ and $\mathrm{Lip}(g \circ h \circ \varphi_\kappa) \leq 2\eta\kappa^{\tau-1} \mathrm{Lip}(g)$. Therefore,

$$\begin{aligned} |\mathrm{Cov}(\xi_\kappa, \zeta_\kappa \mid \mathcal{C})| \leq \big(& c_1 \|f\|_\infty \|g\|_\infty \\ & + 2\eta\kappa^{\tau-1} \{c_2 \mathrm{Lip}(f) \|g\|_\infty + c_3 \|f\|_\infty \mathrm{Lip}(g)\} \\ & + (2\eta\kappa^{\tau-1})^2 c_4 \mathrm{Lip}(f) \mathrm{Lip}(g) \big) \gamma_{n,s} \quad \text{a.s.} \end{aligned} \tag{B.1}$$

Second,

$$\begin{aligned}\mathsf{E}[|\xi-\xi_\kappa| \mid \mathcal{C}] &\le \operatorname{Lip}(f)\sum_{i\in A}\mathsf{E}[\|h(Z_{n,i})-(h\circ\varphi_\kappa)(Z_{n,i})\| \mid \mathcal{C}] \\ &\le C_v \operatorname{Lip}(f)\sum_{i\in A}\mathsf{E}[\|Z_{n,i}\|_\infty^\tau \mathbf{1}\{\|Z_{n,i}\|_\infty>\kappa\} \mid \mathcal{C}] \\ &\le C_v \operatorname{Lip}(f) aL\kappa^{\tau-p} \quad \text{a.s.,}\end{aligned} \tag{B.2}$$

where $C_v > 0$ is a constant depending on v and η. Similarly,

$$\mathsf{E}[|\zeta-\zeta_\kappa| \mid \mathcal{C}] \le C_v \operatorname{Lip}(g) bL\kappa^{\tau-p} \quad \text{a.s.} \tag{B.3}$$

Since inequalities (B.1)-(B.3) hold for all $\kappa \ge 1$ a.s., they also hold for random κ on $\{\kappa \in [1,\infty)\}$. The result follows by setting $\kappa = (\gamma_{n,s}\wedge 1)^{1/(1-p)}$, if $c_4 = 0$ and $\kappa = (\gamma_{n,s}\wedge 1)^{1/(2-p-\tau)}$, otherwise, and, noticing that $\operatorname{Cov}(\xi,\zeta \mid \mathcal{C}) = 0$ a.s. on $\{\gamma_{n,s}=0\}$. ■

Proof of Theorem 3.1. First, it suffices to show that

$$\left\|c^\top(\bar{Y}_n - \mathsf{E}[\bar{Y}_n \mid \mathcal{C}])\right\|_{\mathcal{C},1} \to 0 \quad \text{a.s.}$$

for any $c \in \mathbf{R}^v$ with $\|c\| = 1$. Then the proof is similar to one given in the unconditional case. Specifically, for $k > 0$, let $\xi_{n,i}^{(k)} := \varphi_k(c^\top Y_{n,i})$ and $\zeta_{n,i}^{(k)} := c^\top Y_{n,i} - \xi_{n,i}^{(k)}$ so that

$$\begin{aligned}\left\|c^\top(\bar{Y}_n - \mathsf{E}[\bar{Y}_n \mid \mathcal{C}])\right\|_{\mathcal{C},1} &\le 2\max_{i\in N_n}\left\|\zeta_{n,i}^{(k)}\right\|_{\mathcal{C},1} \\ &\quad + \left\|n^{-1}\sum_{i\in N_n}\left(\xi_{n,i}^{(k)} - \mathsf{E}[\xi_{n,i}^{(k)} \mid \mathcal{C}]\right)\right\|_{\mathcal{C},2} \quad \text{a.s.}\end{aligned}$$

The result then follows from the definition of the essential infimum and the following inequalities:

$$\left\|\zeta_{n,i}^{(k)}\right\|_{\mathcal{C},1} \le \mathsf{E}[\|Y_{n,i}\|\mathbf{1}\{\|Y_{n,i}\|>k\} \mid \mathcal{C}] \quad \text{a.s.}$$

and, since $\psi_{1,1}(\varphi_k,\varphi_k) \le Ck^2$,

$$\left\|\sum_{i\in N_n}\left(\xi_{n,i}^{(k)} - \mathsf{E}[\xi_{n,i}^{(k)} \mid \mathcal{C}]\right)\right\|_{\mathcal{C},2} \le \sqrt{n}k\left(4 + C\sum_{s\ge 1}\delta_n^\partial(s;1)\theta_{n,s}\right)^{1/2} \quad \text{a.s.}$$ ■

Proof of Theorem 3.2. The first assertion follows trivially from the triangle inequality. Consider the second assertion. First, note that for a sub-σ-field $\mathcal{F} \subset \mathcal{H}$, random variables X and Y, and an $\mathcal{F}$-measurable random variable Z, $\mathsf{P}(X \le Z \mid \mathcal{F}) = F_X^{\mathcal{F}}(\,\cdot\,, Z)$ a.s. and

$\mathsf{P}(Y \le Z \mid \mathcal{F}) = F_Y^{\mathcal{F}}(\,\cdot\,, Z)$ a.s. (see, e.g., Kallenberg, 2002, Theorem 5.4). Therefore,

$$|\mathsf{P}(X \le Z \mid \mathcal{F}) - \mathsf{P}(Y \le Z \mid \mathcal{F})| \le d_K(X, Y \mid \mathcal{F}) \quad \text{a.s.}$$

In addition, if $\mathcal{F} = \sigma(\mathcal{A} \cup \mathcal{B})$, where $\mathcal{A}$ and $\mathcal{B}$ are sub-σ-fields of $\mathcal{H}$, and Y is conditionally independent of $\mathcal{A}$ given $\mathcal{B}$, then $d_K(X, Y \mid \mathcal{F}) = d_K(X, Y \mid \mathcal{F}, \mathcal{B})$ a.s.

Let $c_n(\alpha)$ denote the conditional α-quantile of S_n given $\mathcal{C}$. Fix $\eta > 0$ such that $\alpha \pm 2\eta \in (0, 1)$ and let $\Delta_n \equiv d_K(T_n^*, S_n \mid \mathcal{G}_n, \mathcal{C})$. Then, using the properties of generalized inverses (see, e.g., Embrechts and Hofert, 2013, Proposition 1) and the conditional independence of S_n and Y_n given $\mathcal{C}$, we get

$$\begin{aligned}
\mathsf{P}(S_n \le c_n^*(\alpha + \eta) \mid \mathcal{G}_n) &\ge \mathsf{P}(T_n^* \le c_n^*(\alpha + \eta) \mid \mathcal{G}_n) - \eta \\
&\ge \alpha \\
&= \mathsf{P}(S_n \le c_n(\alpha) \mid \mathcal{G}_n) \quad \text{a.s. on } \{\Delta_n \le \eta\}
\end{aligned}$$

and

$$\begin{aligned}
\mathsf{P}(T_n^* \le c_n(\alpha + 2\eta) \mid \mathcal{G}_n) &\ge \mathsf{P}(S_n \le c_n(\alpha + 2\eta) \mid \mathcal{G}_n) - \eta \\
&= \alpha + \eta \\
&\ge \mathsf{P}(T_n^* \le c_n^*(\alpha) \mid \mathcal{G}_n) \quad \text{a.s. on } \{\Delta_n \le \eta\}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathsf{P}(c_n^*(\alpha) \ge c_n(\alpha - \eta) \mid \mathcal{C}) &\ge \mathsf{P}(c_n^*(\alpha) \ge c_n(\alpha - \eta), \Delta_n \le \eta \mid \mathcal{C}) \\
&= \mathsf{P}(\Delta_n \le \eta \mid \mathcal{C}) \quad \text{a.s.}
\end{aligned}$$

and

$$\begin{aligned}
\mathsf{P}(c_n^*(\alpha) \le c_n(\alpha + 2\eta) \mid \mathcal{C}) &\ge \mathsf{P}(c_n^*(\alpha) \le c_n(\alpha + 2\eta), \Delta_n \le \eta \mid \mathcal{C}) \\
&= \mathsf{P}(\Delta_n \le \eta \mid \mathcal{C}) \quad \text{a.s.}
\end{aligned}$$

Using the last two inequalities we find that

$$\begin{aligned}
&\mathsf{P}(c_n(\alpha) \wedge c_n^*(\alpha) < T_n \le c_n(\alpha) \vee c_n^*(\alpha) \mid \mathcal{C}) \\
&\quad \le \mathsf{P}(c_n(\alpha - \eta) < T_n \le c_n(\alpha + 2\eta) \mid \mathcal{C}) \\
&\qquad + \mathsf{P}(c_n(\alpha - \eta) > c_n^*(\alpha) \mid \mathcal{C}) + \mathsf{P}(c_n(\alpha + 2\eta) < c_n^*(\alpha) \mid \mathcal{C}) \\
&\quad \le \mathsf{P}(c_n(\alpha - \eta) < S_n \le c_n(\alpha + 2\eta) \mid \mathcal{C}) \\
&\qquad + 2\mathsf{P}(\Delta_n > \eta \mid \mathcal{C}) + 2d_K(T_n, S_n \mid \mathcal{C})
\end{aligned}$$

$$= 3\eta + 2\mathsf{P}(\Delta_n > \eta \mid \mathcal{C}) + 2d_K(T_n, S_n \mid \mathcal{C}) \quad \text{a.s.}$$

and

$$\begin{aligned} A_{n,\alpha} &\coloneqq |\mathsf{P}(T_n \le c_n^*(\alpha) \mid \mathcal{C}) - \alpha| \\ &\le 3\eta + 2\mathsf{P}(\Delta_n > \eta \mid \mathcal{C}) + 3d_K(T_n, S_n \mid \mathcal{C}) \quad \text{a.s.} \end{aligned} \tag{B.4}$$

Finally, there exists a sequence $\{\alpha_k\}$ s.t. $\operatorname{ess\,sup}_{\alpha\in(0,1)} A_{n,\alpha} = \sup_k A_{n,\alpha_k}$ a.s., and the latter is a.s. bounded by the RHS of (B.4). Therefore,

$$\limsup_{n\to\infty} \Big(\operatorname*{ess\,sup}_{\alpha\in(0,1)} A_{n,\alpha} \Big) \le 3\eta \quad \text{a.s.},$$

and the result follows by considering a sequence $\eta_m \searrow 0$. ■

Proof of Lemma 3.1. By the mean value theorem, we may write

$$\begin{aligned} T_n &= \nabla\phi(\tilde{\theta}_n)^\top \tau_n(\hat{\theta}_n - \theta_n) \quad \text{and} \\ T_n^* &= \nabla\phi(\tilde{\theta}_n^*)^\top \tau_n^*(\hat{\theta}_n^* - \theta_n^*), \end{aligned} \tag{B.5}$$

where $\tilde{\theta}_n$ and $\tilde{\theta}_n^*$ are such that $\|\tilde{\theta}_n - \theta_n\| \le \|\hat{\theta}_n - \theta_n\|$ and $\|\tilde{\theta}_n^* - \theta_n^*\| \le \|\hat{\theta}_n^* - \theta_n^*\|$. Then for any $r \in \mathbf{R}$ and $\epsilon > 0$,

$$\begin{aligned} &\big|\mathsf{P}(T_n^* \le r \mid \mathcal{G}_n) - \mathsf{P}\big(T_n'^* \le r \mid \mathcal{G}_n\big)\big| \\ &\quad \le \mathsf{P}\big(T_n'^* \le r + R_n^* \mid \mathcal{G}_n\big) - \mathsf{P}\big(T_n'^* \le r - R_n^* \mid \mathcal{G}_n\big) \\ &\quad \le 2d_K\big(T_n'^*, S_n' \mid \mathcal{G}_n, \mathcal{C}\big) + Q\big(S_n', 2\epsilon \mid \mathcal{C}\big) + \mathsf{P}(R_n^* > \epsilon \mid \mathcal{G}_n) \quad \text{a.s.}, \end{aligned}$$

where

$$R_n^* \equiv \Big|(\nabla\phi(\tilde{\theta}_n^*) - \nabla\phi(\theta_n^*))^\top \tau_n^*(\hat{\theta}_n^* - \theta_n^*)\Big|.$$

Similarly, for any $r \in \mathbf{R}$ and $\epsilon > 0$,

$$\begin{aligned} &\big|\mathsf{P}(T_n \le r \mid \mathcal{C}) - \mathsf{P}\big(T_n' \le r \mid \mathcal{C}\big)\big| \\ &\quad \le 2d_K\big(T_n', S_n' \mid \mathcal{C}\big) + Q\big(S_n', 2\epsilon \mid \mathcal{C}\big) + \mathsf{P}(R_n > \epsilon \mid \mathcal{C}) \quad \text{a.s.}, \end{aligned}$$

where

$$R_n \equiv \Big|(\nabla\phi(\tilde{\theta}_n) - \nabla\phi(\theta_n))^\top \tau_n(\hat{\theta}_n - \theta_n)\Big|.$$

By Lemma B.6 the sequence $\{\theta_n^*\}$ is $\mathcal{C}$-asymptotically tight. Therefore, using Lemma B.5 together with the $\mathcal{C}$-asymptotic tightness of $\tau_n^*(\hat{\theta}_n^* - \theta_n^*)$ and $\tau_n(\hat{\theta}_n - \theta_n)$ it follows that

$\mathsf{P}(R_n^* > \epsilon \mid \mathcal{C}) \to 0$ a.s. and $\mathsf{P}(R_n > \epsilon \mid \mathcal{C}) \to 0$ a.s. Consequently, for any $\nu > 0$,

$$\begin{aligned}
&\limsup_{n\to\infty} \mathsf{P}\big(d_K\big(T_n^*, T_n'^* \mid \mathcal{G}_n\big) > \nu \mid \mathcal{C}\big) \\
&\qquad \le 3\nu^{-1} \operatorname*{ess\,inf}_{\epsilon>0} \limsup_{n\to\infty} Q\big(S_n', \epsilon \mid \mathcal{C}\big) = 0 \quad \text{a.s.}
\end{aligned}$$

and

$$\begin{aligned}
&\limsup_{n\to\infty} d_K\big(T_n, T_n' \mid \mathcal{C}\big) \\
&\qquad \le \operatorname*{ess\,inf}_{\epsilon>0} \limsup_{n\to\infty} Q\big(S_n', \epsilon \mid \mathcal{C}\big) = 0 \quad \text{a.s.}
\end{aligned}$$

The result then follows from the triangle inequality. ■

Proof of Theorem 3.3. Follows immediately from Lemma 3.1 and Theorem 3.2. ■

Proof of Corollary 3.2. Consider Equation (B.5) in the proof of Lemma 3.1. By Lemma B.3 T_n converges $\mathcal{C}$-weakly to S' ($\because \tilde{\theta}_n \xrightarrow{\mathcal{C}-p} \theta$ a.s. and $x \mapsto \nabla\phi(x)$ is continuous). Hence, $d_K(T_n, S' \mid \mathcal{C}) \to 0$ a.s. by Lemma B.4. Convergence of $d_K(T_n^*, S' \mid \mathcal{G}_n, \mathcal{C})$ follows from arguments similar to those given in the proof of Lemma 3.1. Finally, the result holds by Theorem 3.2. ■

Proof of Proposition 3.2. Let $\zeta_{n,i} := \sum_{j\in B_{n,i}} Y_{n,i}'$, where $Y_{n,i}' := Y_{n,i} - \mu_{n,i}$, and let $\zeta_{n,i}^*$ be its resampling version. Then, using the conditional independence of the elements of $\{\zeta_{n,i}^*\}$ given $\mathcal{G}_n$, we find that

$$\begin{aligned}
\tilde{\Sigma}_n &:= \operatorname{Var}\left(\frac{1}{\sqrt{n}} \sum_{k=1}^{K_n} \zeta_{n,k}^* \mid \mathcal{G}_n\right) = \frac{1}{n} \sum_{k=1}^{K_n} \operatorname{Var}(\zeta_{n,k}^* \mid \mathcal{G}_n) \\
&= \frac{1}{\delta_n(s_n)} \left(\frac{1}{n} \sum_{i\in N_n} \zeta_{n,i}\zeta_{n,i}^\top - \bar{\zeta}_n \bar{\zeta}_n^\top\right) \\
&= \Sigma_n^* - \frac{1}{n} \sum_{i\in N_n} (\omega_n(i) - 1)(\zeta_{n,i}\mu_n^\top + \mu_n\zeta_{n,i}^\top) - \frac{\Delta_n(s_n; 2)}{\delta_n(s_n)} \mu_n\mu_n^\top \\
&\equiv \Sigma_n^* - A_{n,1} + A_{n,2},
\end{aligned} \tag{B.6}$$

where $\bar{\zeta}_n := n^{-1} \sum_{i\in N_n} \zeta_{n,i}$. On the other hand, using the second line of (B.6),

$$\begin{aligned}
\tilde{\Sigma}_n &= \frac{1}{n} \sum_{i,j\in N_n} \omega_n(i,j) Y_{n,i}' Y_{n,j}'^\top \\
&\qquad - \delta_n(s_n) \times \frac{1}{n} \sum_{i\in N_n} \omega_n(i) Y_{n,i}' \times \frac{1}{n} \sum_{i\in N_n} \omega_n(i) Y_{n,i}'^\top \\
&\equiv B_{n,1} + B_{n,2}.
\end{aligned} \tag{B.7}$$

Let $y'_{n,i} \coloneqq c^\top Y'_{n,i}$ and $\mu'_n \coloneqq c^\top \mu_n$, where $c \in \mathbf{R}^v$ with $\|c\| = 1$ and note that by Lemma B.1 it suffices to show that $\mathsf{E}[|c^\top(\Sigma^*_n - \Sigma_n)c| \mid \mathcal{C}] \to 0$ a.s. Also by Lemma 2.1 in Chapter 2 the process $\{y'_n\}$ is $(\mathcal{L}_1, \psi, \mathcal{C})$-weakly dependent with the weak dependence coefficients $\{\gamma_n\}$. In the following, let

$$\Xi_n \coloneqq \sum_{s \geq 1} \delta^\partial_n(s) \gamma_{n,s}^{1-\frac{2}{r}}.$$

Claim B.1. $\mathsf{E}[|c^\top(\tilde{\Sigma}_n - \Sigma_n)c| \mid \mathcal{C}] \to 0$ *a.s.*

Proof. Consider the first term on the last line of (B.7). Write

$$\begin{aligned} c^\top(B_{n,1} - \Sigma_n)c &= \frac{1}{n} \sum_{i \in N_n} (y'^2_{n,i} - \mathsf{E}[y'^2_{n,i} \mid \mathcal{C}]) + \frac{1}{n} \sum_{i \in N_n} (\omega_n(i) - 1) y'^2_{n,i} \\ &\quad + \frac{1}{n} \sum_{i \in N_n} \sum_{j \in N_n \setminus \{i\}} \omega_n(i,j)(y'_{n,i} y'_{n,j} - \mathsf{E}[y'_{n,i} y'_{n,j} \mid \mathcal{C}]) \\ &\quad + \frac{1}{n} \sum_{i \in N_n} \sum_{j \in N_n \setminus \{i\}} (\omega_n(i,j) - 1) \mathsf{E}[y'_{n,i} y'_{n,j} \mid \mathcal{C}] \\ &\equiv R_{n,0} + R_{n,1} + R_{n,2} + R_{n,3}. \end{aligned}$$

Using the covariance inequalities established in Section A.1,

$$|R_{n,3}| \leq C_3 \sum_{s \geq 1} \gamma_{n,s}^{1-\frac{2}{r}} \times \frac{1}{n} \sum_{i \in N_n} \sum_{j \in N^\partial_n(i;s)} |\omega_n(i,j) - 1| \quad \text{a.s.,} \tag{B.8}$$

where $C_3 = C(\mu^2_{2r} \vee 1)$ for some constant $C \geq 1$. Since $\tilde{\omega} < \infty$ a.s., the RHS of (B.8) is bounded by $C_3(\tilde{\omega} + 1)\Xi_n < \infty$ a.s. Therefore, by the dominated convergence theorem $|R_{n,3}| \to 0$ a.s. Also letting $w_{n,i,j} \coloneqq y'_{n,i} y'_{n,j} - \mathsf{E}[y'_{n,i} y'_{n,j} \mid \mathcal{C}]$, we find that

$$\begin{aligned} \mathsf{E}[R^2_{n,2} \mid \mathcal{C}] &\leq \frac{\bar{\omega}^2}{n^2} \sum_{\substack{i,j \in N_n \\ 1 \leq d_n(i,j) < 2(s_n+1)}} \sum_{\substack{k,l \in N_n \\ 1 \leq d_n(k,l) < 2(s_n+1)}} |\mathsf{E}[w_{n,i,j} w_{n,k,l} \mid \mathcal{C}]| \\ &\leq \frac{C_2 \bar{\omega}^2}{n^2} \sum_{s \geq 0} |H_n(s, 2s_n + 1)| \gamma_{n,s}^{1-\frac{2}{r}} \to 0 \quad \text{a.s.,} \end{aligned}$$

where $C_2 = C(\mu^4_{2r} \vee 1)$ for some constant $C \geq 1$. Finally,

$$\mathsf{E}[|R_{n,1}| \mid \mathcal{C}] \leq \frac{\mu^2_{2r}}{n} \sum_{i \in N_n} |\omega_n(i) - 1| \to 0 \quad \text{a.s.}$$

and

$$\mathsf{E}[R^2_{n,0} \mid \mathcal{C}] \leq \frac{1}{n^2} \sum_{i,j \in N_n} \left| \mathrm{Cov}(y'^2_{n,i}, y'^2_{n,j} \mid \mathcal{C}) \right|$$

$$\leq \frac{C_0}{n}(1+\Xi_n) \to 0 \quad \text{a.s.},$$

where $C_0 = C(\mu_{2r}^4 \vee 1)$ for some constant $C \geq 1$.

As for the second term, note that $c^\top B_{n,2} c \geq 0$ a.s. and

$$\begin{aligned}\mathsf{E}[c^\top B_{n,2} c \mid \mathcal{C}] &\leq \frac{(D_n(s_n))^2}{\delta_n(s_n) n^2} \sum_{i,j \in N_n} \left|\mathsf{E}[y'_{n,i} y'_{n,j} \mid \mathcal{C}]\right| \\ &\leq \frac{(D_n(s_n))^2 C_3}{\delta_n(s_n) n}(1+\Xi_n) \to 0 \quad \text{a.s.}\end{aligned}$$

□

Consider the last two terms on the last line of Equation (B.6). Let

$$\alpha_{n,i} := \sum_{j \in B_{n,i}} (\omega_n(j) - 1) \quad \text{and} \quad \alpha_n := \max_{i \in N_n} |\alpha_{n,i}|.$$

Then, since

$$\sum_{i \in N_n} (\omega_n(i) - 1)\zeta_{n,i} = \sum_{i \in N_n} \alpha_{n,i} Y'_{n,i},$$

we have $c^\top A_{n,1} c = 2n^{-1} \sum_{i \in N_n} \alpha_{n,i} y'_{n,i} \mu'_n$ and, therefore,

$$\mathsf{E}\Big[(c^\top A_{n,1} c)^2 \mid \mathcal{C}\Big] \leq \frac{4C_2 \alpha_n^2}{n}(1+\Xi_n) \to 0 \quad \text{a.s.}$$

Finally,

$$\|A_{n,2}\|_F \leq \mu_{2r}^2 \times \frac{\Delta_n(s_n;2)}{\delta_n(s_n)} \to 0 \quad a.s.$$

■

Proof of Proposition 3.4. We use the notation from the proof of Proposition 3.1. For a vector $c \in \mathbf{R}^v$ with $\|c\| = 1$ we have

$$\begin{aligned}c^\top(\Sigma_n^* - B_{n,1})c &= \bar{y}_n'^2 \times \frac{\delta_n(s_n;2)}{\delta_n(s_n)} - \frac{2\bar{y}'_n}{n} \sum_{i \in N_n} y'_{n,i} \sum_{j \in N_n} \omega_n(i,j) \\ &\equiv Q_{n,1} + Q_{n,2},\end{aligned} \tag{B.9}$$

where $\bar{y}_n := n^{-1} \sum_{i \in N_n} y'_{n,i}$. Consider the second term in the last line of (B.9). Letting $\tau_{n,i} := \sum_{j \in N_n} \omega_n(i,j)$ and noticing that

$$\begin{aligned}\max_{i \in N_n} \tau_{n,i} &\leq \max_{i \in N_n} (\omega_n(i) + \tilde{\omega}|N_n(i; s_n+1)|) \\ &\leq (\tilde{\omega}+1) D_n(s_n) \equiv \tau_n,\end{aligned}$$

it follows that

$$|Q_{n,2}| \le 2\big|\sqrt{\tau_n}\bar{y}'_n\big| \times \frac{1}{n\sqrt{\tau_n}}\left|\sum_{i\in N_n}\tau_{n,i}y'_{n,i}\right| \equiv 2\big|\sqrt{\tau_n}\bar{y}'_n\big| \times Q_{n,3}.$$

Similarly to the proof of Proposition 3.2,

$$\mathsf{E}[Q_{n,3}^2 \mid \mathcal{C}] \le \frac{C_3\tau_n}{n}(1+\Xi_n) \to 0 \quad \text{a.s.}$$

Finally, $\mathsf{E}[\tau_n\bar{y}_n'^2 \mid \mathcal{C}]$ is bounded the same quantity and since $\delta_n(s_n;2)/\delta_n(s_n) \le \tau_n$,

$$\mathsf{E}[Q_{n,1} \mid \mathcal{C}] \le \mathsf{E}[\tau_n\bar{y}_n'^2 \mid \mathcal{C}] \to 0 \quad \text{a.s.} \qquad \blacksquare$$

Proof of Proposition 3.3. Consider $T^*_{1,n}$ first. Let $\tilde{Y}_{n,i} \coloneqq Y_{n,i} - \mu_n$, $\tilde{Z}_{n,i} \coloneqq \sum_{k\in B_{n,i}}\tilde{Y}_{n,k}$ and let $\tilde{Z}^*_{n,i}$ be the bootstrap version of the letter. Also define $W^*_{n,i} \coloneqq \tilde{Z}^*_{n,i} - \mathsf{E}[\tilde{Z}^*_{n,i} \mid \mathcal{G}_n]$. Conditionally on $\mathcal{G}_n$, $\{\xi^*_{n,i}\}$ are row-wise i.i.d. random vectors with $\mathsf{E}[W^*_{n,1} \mid \mathcal{G}_n] = 0$ and $\mathrm{Var}(W^*_{n,1} \mid \mathcal{G}_n) = \delta_n(s_n)\Sigma^*_n$ a.s. Write

$$\sqrt{n}(\tilde{Y}^*_n - \mu^*_n) = \frac{1}{\sqrt{K_n}}\sum_{k=1}^{K_n}(\delta_n(s_n))^{-1/2}W^*_{n,k}.$$

Then, letting $\underline{\lambda}(A)$ denote the minimal eigenvalue of a square matrix A, by Corollary B.1,

$$d_K\big(T^*_{1,n}, S^*_{1,n} \mid \mathcal{G}_n\big) \le \frac{C_v}{(\underline{\lambda}(\Sigma)/2)^{3/8}}\left(\frac{\mathsf{E}[\|W^*_{n,1}\|^3 \mid \mathcal{G}_n]}{\sqrt{n}\delta_n(s_n)}\right)^{1/4}$$

a.s. on $\{\underline{\lambda}(\Sigma^*_n) \ge \underline{\lambda}(\Sigma)/2\}$, where $S^*_{1,n} = \|Q_n\|$ and Q_n is conditionally normal given $\mathcal{G}_n$ with zero mean and variance Σ^*_n.

Claim B.2. $\mathsf{E}[\|W^*_{n,i}\|^3 \mid \mathcal{C}]/(\sqrt{n}\delta_n(s_n)) \to 0$ *a.s.*

Proof. It suffices to show that $\mathsf{E}[|c^\top W_{n,i}|^3 \mid \mathcal{C}]/(\sqrt{n}\delta_n(s_n)) \to 0$ a.s. for any $c \in \mathbf{R}^v$ s.t. $\|c\| = 1$. By the c_r-inequality,[1]

$$\mathsf{E}[|c^\top W^*_{n,1}|^3 \mid \mathcal{G}_n] \le 8\mathsf{E}[|c^\top \tilde{Z}^*_{n,1}|^3 \mid \mathcal{G}_n] = \frac{8}{n}\sum_{i\in N_n}|c^\top \tilde{Z}_{n,i}|^3 \quad \text{a.s.}$$

Let $\tilde{y}_{n,i} \coloneqq c^\top\tilde{Y}_{n,i}$. Then

$$\mathsf{E}[|c^\top \tilde{Z}_{n,i}|^4 \mid \mathcal{C}] \le \sum_{j_1,j_2,j_3,j_4\in B_{n,i}}|\mathrm{Cov}(\tilde{y}_{n,j_1}\tilde{y}_{n,j_2}, \tilde{y}_{n,j_3}\tilde{y}_{n,j_4} \mid \mathcal{C})|$$

[1] We consider the conditional versions of all inequalities used in this proof.

$$+\left(\sum_{j_1,j_2\in B_{n,i}}|\mathrm{Cov}(\tilde{y}_{n,j_1},\tilde{y}_{n,j_2}\mid\mathcal{C})|\right)^2\equiv A_{n,i}+C_{n,i}^2\quad\text{a.s.}$$

Similarly to the proof of Proposition 3.2, we find that w.p.1,

$$A_{n,i}\le C_1(\tilde{\mu}_{2p}^4\vee 1)|B_{n,i}|^3\sum_{s\ge 0}h_{loc,n}(s,s_n)\gamma_{n,s}^{1-\frac{2}{p}}\quad\text{and}$$
$$C_{n,i}\le C_2(\tilde{\mu}_{2p}^2\vee 1)|B_{n,i}|\sum_{s\ge 0}\delta_{loc,n}^{\partial}(s,s_n)\gamma_{n,s}^{1-\frac{2}{p}},$$

where C_1 and C_2 are some positive constants. The result then follows by noticing that $\mathsf{E}[|c^\top\tilde{Z}_{n,i}|^3\mid\mathcal{C}]\le(\mathsf{E}[|c^\top\tilde{Z}_{n,i}|^4\mid\mathcal{C}])^{3/4}$ a.s. and the fact that $(A_{n,i}+C_{n,i}^2)^{3/4}\le A_{n,i}^{3/4}+C_{n,i}^{3/2}$. □

Using Jensen's inequality, we find that for any $\epsilon>0$,

$$\begin{aligned}&\mathsf{P}\big(d_K\big(T_{1,n}^*,S_{1,n}^*\mid\mathcal{G}_n\big)>\epsilon\mid\mathcal{C}\big)\\&\quad\le\frac{C_v}{(\underline{\lambda}(\Sigma)/2)^{3/8}}\left(\frac{\mathsf{E}[\|W_{n,1}^*\|^3\mid\mathcal{C}]}{\sqrt{n}\delta_n(s_n)}\right)^{1/4}+\mathsf{P}(\underline{\lambda}(\Sigma_n^*)<\underline{\lambda}(\Sigma)/2\mid\mathcal{C})\\&\quad\to 0\quad\text{a.s.},\end{aligned}$$

where the convergence of $\mathsf{P}(\underline{\lambda}(\Sigma_n^*)<\underline{\lambda}(\Sigma)/2\mid\mathcal{C})$ follows from the fact that the eigenvalues of a matrix depend continuously on the entries of the matrix (see, e.g., Zhang, 2011, Theorem 2.11) so that $\lambda_j(\Sigma_n^*)\xrightarrow{\mathcal{C}-p}\lambda_j(\Sigma)$ a.s. for all $1\le j\le d$.

Since the eigenvalues of Σ_n^* converge to the eigenvalues of Σ and the latter are a.s. positive it follows from Lemma B.9 that

$$d_K\Big(S_{1,n}^*,\|\Sigma^{1/2}\eta\|\mid\mathcal{G}_n,\mathcal{C}\Big)\xrightarrow{\mathcal{C}-p}0\quad\text{a.s.}$$

Finally, by Lemma B.3, $T_{1,n}\to\|\Sigma^{1/2}\eta\|$ $\mathcal{C}$-weakly and, hence, the result follows from Corollary 3.1.

Consider the second assertion. First, for any $c\in\mathbf{R}^v$ s.t. $\|c\|=1$ and $\epsilon>0$, we get

$$\begin{aligned}\mathsf{P}\Big(|c^\top(\tilde{Y}_n-\mu_n^*)|>\epsilon\mid\mathcal{C}\Big)&\le\frac{1}{(K_n\epsilon)^2}\mathsf{E}\left[\left(\sum_{k=1}^{K_n}c^\top\tilde{Z}_{n,k}^*\right)^2\mid\mathcal{C}\right]\\&=\frac{1}{K_n\epsilon^2}\mathsf{E}[c^\top\Sigma_n^*c\mid\mathcal{C}]\to 0\quad\text{a.s.},\end{aligned}$$

where the convergence follows from the consistency of Σ_n^*. Therefore, $\tilde{Y}_n^*-\mu_n^*\xrightarrow{\mathcal{C}-p}0$ a.s. Also since the $\mathcal{C}$-asymptotic tightness of a vector follows from that of its elements,

$\sqrt{n}(\tilde{Y}_n^* - \mu_n^*)$ is $\mathcal{C}$-asymptotically tight due to the same reason (i.e., the convergence of Σ_n^*). Write

$$\nabla\phi(\mu_n^*)^\top \sqrt{n}(\tilde{Y}_n^* - \mu_n^*) = \frac{1}{\sqrt{K_n}} \sum_{k=1}^{K_n} (\delta_n(s_n))^{-1/2} \nabla\phi(\mu_n^*)^\top W_{n,k}^*.$$

Then by Lemma B.8, letting $T_{2,n}'^* = \nabla\phi(\mu_n^*)^\top \sqrt{n}(\tilde{Y}_n^* - \mu_n^*)$ and $S_{2,n}'^* = \nabla\phi(\mu_n^*)^\top Q_n$, we have

$$d_K\big(T_{2,n}'^*, S_{2,n}'^* \mid \mathcal{G}_n\big) \le \frac{C}{(\sigma/2)^3} \times \frac{\|\nabla\phi(\mu_n^*)\|^3 \mathsf{E}[\|W_{n,1}^*\|^3 \mid \mathcal{G}_n]}{\sqrt{n}\delta_n(s_n)}$$

a.s. on $\{\sigma_n^* \ge \sigma/2\}$, where $\sigma_n^{*2} = \nabla\phi(\mu_n^*)^\top \Sigma_n^* \nabla\phi(\mu_n^*)$ and $\sigma^2 = \nabla\phi(\mu)^\top \Sigma \nabla\phi(\mu)$. Since $x \mapsto \nabla\phi(x)$ is continuous and μ_n^* is a consistent estimator of μ, $\nabla\phi(\mu_n^*) \xrightarrow{\mathcal{C}-p} \nabla\phi(\mu)$ and $\sigma_n^* \xrightarrow{\mathcal{C}-p} \sigma$ a.s. Consequently, as in the previous case,

$$d_K\big(T_{2,n}'^*, S_{2,n}'^* \mid \mathcal{G}_n\big) \xrightarrow{\mathcal{C}-p} 0 \quad \text{a.s.}$$

and by Lemma B.10,

$$d_K\Big(S_{2,n}'^*, \nabla\phi(\mu)^\top \Sigma^{1/2}\eta \mid \mathcal{G}_n, \mathcal{C}\Big) \xrightarrow{\mathcal{C}-p} 0 \quad \text{a.s.}$$

The result then follows from Corollary 3.2. ■

Proof of Proposition 3.5. The proof is similar to one for Proposition 3.3, and so is omitted. ■

Proof of Proposition 3.6. By Lemma B.11,

$$\begin{aligned} &d_K\big(\|T_{1,n}^*\|, \|S_{1,n}^*\| \mid \mathcal{G}_n\big) \xrightarrow{\mathcal{C}-p} 0 \quad \text{a.s. and} \\ &d_K\big(T_{2,n}'^*, S_{2,n}'^* \mid \mathcal{G}_n\big) \xrightarrow{\mathcal{C}-p} 0 \quad \text{a.s.,} \end{aligned}$$

where $S_{1,n}^* = \|Q_n\|$, $S_{2,n}'^* = \nabla\phi(\bar{Y}_n)^\top Q_n$ and Q_n is conditionally normal given $\mathcal{G}_n$ with zero mean and variance Σ_n^*. The rest is similar to the proof of Proposition 3.3. ■

B.2 Network HAC estimator

Although the HAC estimator (3.5) is consistent in the sense that $\hat{\Sigma}_n - \Sigma_n \xrightarrow{\mathcal{C}-p} 0$ a.s. it does not necessarily yields positive semi-definite covariance matrix. There exist a number methods of approximating a symmetric matrix by a positive definite matrix (see, e.g., Higham, 1988, 2002). Borrowing some ideas from that literature we suggest a simple way of obtaining a positive definite estimate.

Let $Q_n \Lambda_n Q_n^\top$ be the eigendecomposition of $\hat{\Sigma}_n$ (since $\hat{\Sigma}_n$ is symmetric all its eigenvalues are real). Also let $\underline{\lambda}(A)$ denote the smallest eigenvalue of A, e.g., $\underline{\lambda}(\hat{\Sigma}_n) = \min_{1 \le k \le v} \Lambda_n$. Consider a sequence of small positive real numbers $c_n \searrow 0$. We approximate $\hat{\Sigma}_n$ by

$$\hat{V}_n^+ := Q_n(\Lambda_n \vee c_n I_v) Q_n^\top,$$

where the maximum is taken element-wise. By construction, the matrix $\hat{\Sigma}_n^+$ is positive definite. Moreover, in the case when the smallest eigenvalue of Σ_n is bounded from below by some positive constant, it is also a consistent estimator of the true variance as follows from the next result.

Proposition B.1. *Suppose that* $\hat{\Sigma}_n - \Sigma_n \xrightarrow{\mathcal{C}-p} 0$ *a.s. and there exists a constant* $c > 0$ *s.t.* $\mathsf{P}(\underline{\lambda}(\Sigma_n) \ge c \text{ ev.}) = 1$. *Then*

$$\hat{\Sigma}_n^+ - \Sigma_n \xrightarrow{\mathcal{C}-p} 0 \quad a.s.$$

Proof. Fix $\epsilon > 0$. Then

$$\begin{aligned} \mathsf{P}\Big(\|\hat{\Sigma}_n^+ - \Sigma_n\| > \epsilon \mid \mathcal{C}\Big) \le \mathsf{P}\Big(\|\hat{\Sigma}_n - \Sigma_n\| > \epsilon \mid \mathcal{C}\Big) \\ + \mathsf{P}\Big(\underline{\lambda}(\hat{\Sigma}_n) < c_n \mid \mathcal{C}\Big) \quad \text{a.s.} \end{aligned} \tag{B.10}$$

The first term on the RHS of (B.10) trivially converges to 0 a.s. As for the second term, using the properties of the Rayleigh quotient,

$$\underline{\lambda}(\hat{\Sigma}_n) = \min_{x:\|x\|=1} x^\top \hat{\Sigma}_n x \ge \underline{\lambda}(\Sigma_n) + \underline{\lambda}(\hat{\Sigma}_n - \Sigma_n).$$

Therefore, noticing that $|\underline{\lambda}(A)| \le \|A\|$,

$$\begin{aligned} \mathsf{P}\Big(\underline{\lambda}(\hat{\Sigma}_n) < c_n \mid \mathcal{C}\Big) \le \mathsf{P}\Big(\underline{\lambda}(\hat{\Sigma}_n - \Sigma_n) < c_n - c \mid \mathcal{C}\Big) \\ + \mathbf{1}\{\underline{\lambda}(\Sigma_n) < c\} \to 0 \quad \text{a.s.} \end{aligned}$$

■

If Σ_n converges a.s. to an a.s. positive definite matrix Σ, then we may relax the assumptions of the preceding result.

Proposition B.2. *Suppose that* $\hat{\Sigma}_n - \Sigma \xrightarrow{\mathcal{C}-p} 0$ *a.s., where* Σ *is a.s. positive definite. Then*

$$\hat{\Sigma}_n^+ - \Sigma \xrightarrow{\mathcal{C}-p} 0 \quad a.s.$$

Proof. As in the proof of Proposition B.1 for any $\epsilon > 0$,

$$\limsup_{n\to\infty} \mathsf{P}\Big(\|\hat{\Sigma}_n^+ - \Sigma\| > \epsilon \mid \mathcal{C}\Big) \le \operatorname*{ess\,inf}_{c>0} \mathbf{1}\{\underline{\lambda}(\Sigma) < c\} = 0 \quad \text{a.s.}$$

■

B.3 Auxiliary results

In the following, we assume that all random elements are defined on a common probability space $(\Omega, \mathsf{P}, \mathcal{H})$. Also for a vector $x \in \mathbf{R}^v$ let $\|x\|$ denote the Euclidean norm of x and let $\|\cdot\|_{e,p}$ be the element-wise p-norm in $\mathbf{R}^{a\times b}$, i.e., $\|A\|_{e,p} := \|\mathrm{vec}(A)\|_p$.

Lemma B.1. *Let A_n be a sequence of symmetric matrices in $\mathbf{R}^{v\times v}$ and $\mathcal{F} \subset \mathcal{H}$. Then the following are equivalent:*

(a) $\mathsf{E}[\|A_n\|_{e,1} \mid \mathcal{C}] \to 0$ *a.s.*
(b) $\mathsf{E}[\|A_n\|_F \mid \mathcal{C}] \to 0$ *a.s.*
(c) $\mathsf{E}[|c^\top A_n c| \mid \mathcal{C}] \to 0$ *a.s. for any* $c \in \mathbf{R}^v$ *s.t.* $\|c\| = 1$.

Proof. (a) is equivalent to (b) because

$$\|A_n\|_F \le \|A_n\|_{e,1} \le v^2 \|A_n\|_F.$$

The equivalence of (a) and (c) follows from the next inequalities:

$$|c^\top A_n c| \le \|c\|_\infty^2 \|A_n\|_{e,1}.$$

and, letting $z_{ij}^+ = (e_i + e_j)/\sqrt{2}$ and $z_{ij}^- = (e_i - e_j)/\sqrt{2}$, where $\{e_1, \dots, e_v\}$ is a standard basis for $\mathbf{R}^v$,

$$\|A_n\|_{e,1} \le \frac{1}{2} \sum_{i,j=1}^{v} \Big(|z_{ij}^{+\top} A_n z_{ij}^+| + |z_{ij}^{-\top} A_n z_{ij}^-| \Big). \qquad \blacksquare$$

The following is a simple extension of Lemma A.3. in Crimaldi (2009) to the multidimensional case. For a random vector $X \in \mathbf{R}^v$ and $\mathcal{F} \subset \mathcal{H}$ let $\mathsf{Q}_X^{\mathcal{F}}$ denote the regular conditional distribution of X given $\mathcal{F}$, and let $\hat{\varphi}_X$ be the corresponding characteristic functions, i.e., for $t \in \mathbf{R}^v$,

$$\hat{\varphi}_X(\omega, t) = \int \exp(it^\top x) \mathsf{Q}_X^{\mathcal{F}}(\omega, dx).$$

Also the conditional characteristic function of X given $\mathcal{F}$ is given by

$$\varphi_X(t \mid \mathcal{F}) := \mathsf{E}[\exp(it^\top X) \mid \mathcal{F}]$$

and for a fixed $t \in \mathbf{R}^v$ and almost all $\omega \in \Omega$, $\hat{\varphi}_X(\omega, t) = \varphi_X(t \mid \mathcal{F})(\omega)$.

Lemma B.2. *Let $\{X_n\}$ be a sequence of random vectors in $\mathbf{R}^v$ and $\mathcal{F} \subset \mathcal{H}$. Then $X_n \to X$ $\mathcal{F}$-weakly, i.e., for almost all $\omega \in \Omega$, $\mathsf{Q}_{X_n}^{\mathcal{F}}(\omega, \cdot) \to \mathsf{Q}_X^{\mathcal{F}}(\omega, \cdot)$ weakly, iff for every $t \in \mathbf{R}^v$, $\hat{\varphi}_{X_n}(\cdot, t) \to \hat{\varphi}_X(\cdot, t)$ a.s.*

The next lemma provides a number of useful properties of the almost sure conditional convergence, which are typical of the usual weak convergence.

Lemma B.3. *Let $\{X_n\}$ and $\{Y_n\}$ be sequences of random vectors in $\mathbf{R}^v$ and $\mathbf{R}^w$, respectively, and $\mathcal{F} \subset \mathcal{H}$. Then*

(a) *If $X_n \to X$ $\mathcal{F}$-weakly and $g : \mathbf{R}^v \to \mathbf{R}^d$ is continuous, then $g(X_n) \to g(X)$ $\mathcal{F}$-weakly.*
(b) *$X_n \to X$ $\mathcal{F}$-weakly iff $s^\top X_n \to t^\top X$ $\mathcal{F}$-weakly for all $s \in \mathbf{R}^v$.*
(c) *If $Y_n \xrightarrow{\mathcal{F}-p} Y$ a.s., where Y is $\mathcal{F}$-measurable, then $Y_n \to Y$ $\mathcal{F}$-weakly.*
(d) *Let $v = w$. If $X_n \to X$ $\mathcal{F}$-weakly and $X_n - Y_n \xrightarrow{\mathcal{F}-p} 0$ a.s., then $Y_n \to X$ $\mathcal{F}$-weakly.*
(e) *If $X_n \to X$ $\mathcal{F}$-weakly, $Y_n \xrightarrow{\mathcal{F}-p} Y$ a.s., where Y is $\mathcal{F}$-measurable, then $(X_n^\top, Y_n^\top) \to (X^\top, Y^\top)$ $\mathcal{F}$-weakly.*

Proof. (a) This follows from Lemma B.2 and the fact that $x \mapsto \exp(it^\top g(x))$ is a bounded, continuous function.
(b) The sufficiency follow from part (a) because $x \mapsto s^\top x$ is continuous. For the necessity, suppose that all linear combinations converge $\mathcal{F}$-weakly. Then

$$\varphi_{X_n}(t \mid \mathcal{F}) = \varphi_{t^\top X_n}(1 \mid \mathcal{F}) \to \varphi_{t^\top X}(1 \mid \mathcal{F}) = \varphi_X(t \mid \mathcal{F}) \quad \text{a.s.},$$

and the result follows from Lemma B.2.
(c) Since for any $t \in \mathbf{R}^w$ and $\epsilon > 0$, $|e^{it^\top(Y_n - Y)} - 1| \le \epsilon$ on $|t^\top(Y_n - Y)| \le \epsilon$, we have

$$\begin{aligned} |\varphi_{Y_n}(t \mid \mathcal{F}) - \varphi_Y(t \mid \mathcal{F})| &\le \mathsf{E}\Big[\Big|e^{it^\top(Y_n - Y)} - 1\Big| \mid \mathcal{F}\Big] \\ &\le \epsilon + \mathsf{P}\Big(|t^\top(Y_n - Y)| > \epsilon \mid \mathcal{F}\Big) \quad \text{a.s.} \end{aligned}$$

Therefore,

$$\limsup_{n \ge 1} |\varphi_{Y_n}(t \mid \mathcal{F}) - \varphi_Y(t \mid \mathcal{F})| \le \epsilon \quad \text{a.s.}$$

The result follows by considering a sequence $\epsilon_m \searrow 0$ and Lemma B.2.
(d) Similarly to part (c), for any $t \in \mathbf{R}^v$,

$$|\varphi_{Y_n}(t \mid \mathcal{F}) - \varphi_X(t \mid \mathcal{F})| \le \mathsf{E}\Big[\Big|e^{it^\top(Y_n - X_n)} - 1\Big| \mid \mathcal{F}\Big] \to 0 \quad \text{a.s.}$$

(e) Since $(X_n^\top, Y^\top) \to (X^\top, Y^\top)$ $\mathcal{F}$-weakly, the result follows from part (d). ■

Lemma B.4. *Let $\{X_n\}$ be a sequence of random variables, $\mathcal{F} \subset \mathcal{H}$, and let X be a random variable with (a.s.) continuous conditional cdf given $\mathcal{F}$ (i.e., the map $t \mapsto F_X^{\mathcal{F}}(\omega, t)$ is continuous for (almost) all $\omega \in \Omega$). Then $X_n \to X$ $\mathcal{F}$-weakly iff $d_K(X_n, X \mid \mathcal{F}) \to 0$ a.s.*

Proof. The necessity holds by Theorem 3.1.2 in Shiryaev (2016) because $d_K(X_n, X \mid \mathcal{F}) \to 0$ a.s. implies that for almost all $\omega \in \Omega$ the regular conditional cdfs converge and the sufficiency follows from the ω-wise application of Pólya's theorem (e.g. Athreya and Lahiri, 2006, Theorem 9.1.4). ■

Lemma B.5. *Suppose that $f : \mathbf{R}^v \to \mathbf{R}^w$ is continuous and $\{X_n\}$ and $\{Y_n\}$ are sequences of random vectors in $\mathbf{R}^v$ such that $Y_n - X_n \xrightarrow{\mathcal{F}-p} 0$ a.s. for some $\mathcal{F} \subset \mathcal{H}$ and $\{X_n\}$ is $\mathcal{F}$-asymptotically tight. Then*

$$f(Y_n) - f(X_n) \xrightarrow{\mathcal{F}-p} 0 \quad a.s.$$

Proof. For any $z > 0$, the restriction $f|_{\overline{B(0,z)}}$ is uniformly continuous, i.e., $\forall \epsilon > 0$, $\exists \delta_\epsilon > 0$ s.t. for all $x, y \in \overline{B(0,z)}$, $\|f(x) - f(y)\| < \epsilon$ whenever $\|x - y\| < \delta_\epsilon$. Fix $\epsilon > 0$. Then

$$\begin{aligned}
\mathsf{P}(\|f(Y_n) - f(X_n)\| > \epsilon \mid \mathcal{F}) &\le \mathsf{P}(\|Y_n - X_n\| > \delta_\epsilon \mid \mathcal{F}) \\
&\quad + \mathsf{P}(\|Y_n\| > x \mid \mathcal{F}) + \mathsf{P}(\|X_n\| > z \mid \mathcal{F}) \\
&\le 2\mathsf{P}(\|Y_n - X_n\| > \delta_\epsilon \wedge z/2 \mid \mathcal{F}) \\
&\quad + 2\mathsf{P}(\|X_n\| > z/2 \mid \mathcal{F}) \quad \text{a.s.}
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\limsup_{n\to\infty} \mathsf{P}(\|f(Y_n) - f(X_n)\| > \epsilon \mid \mathcal{F}) \\
&\quad \le 2 \operatorname*{ess\,inf}_{z>0} \limsup_{n\to\infty} \mathsf{P}(\|X_n\| > z \mid \mathcal{F}) = 0 \quad \text{a.s.}
\end{aligned}$$

■

Lemma B.6. *Suppose that $\{X_n\}$ and $\{Y_n\}$ are sequences of random vectors in $\mathbf{R}^v$ such that X_n is $\mathcal{F}$-measurable for all $n \ge 1$ and some $\mathcal{F} \subset \mathcal{H}$, $\sup_n \|X_n\| < \infty$ a.s., and $Y_n - X_n \xrightarrow{\mathcal{F}-p} 0$ a.s. Then $\{Y_n\}$ is $\mathcal{F}$-asymptotically tight.*

Proof. For any $y > 0$,

$$\begin{aligned}
\mathsf{P}(\|Y_n\| > y \mid \mathcal{F}) &\le \mathsf{P}(\|Y_n - X_n\| > y/2 \mid \mathcal{F}) \\
&\quad + \mathbf{1}\{\sup_n \|X_n\| > y/2\} \quad \text{a.s.}
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\operatorname*{ess\,inf}_{y>0} \limsup_{n\to\infty} \mathsf{P}(\|Y_n\| > y \mid \mathcal{F}) \\
&\quad \le \operatorname*{ess\,inf}_{y>0} \mathbf{1}\{\sup_n \|X_n\| > y\} = 0 \quad \text{a.s.}
\end{aligned}$$

■

In the following, for $r, \epsilon \ge 0$ let

$$S_{r,\epsilon} := \{x \in \mathbf{R}^v : r \le \|x\| \le r + \epsilon\}.$$

Lemma B.7. *Suppose that Z is a standard normal random vector in $\mathbf{R}^v$ with $v \ge 2$ and $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_v > 0$ are constants. Let $\Lambda := \operatorname{diag}(\lambda_1, \ldots, \lambda_v)$ and $N = \Lambda^{1/2} Z$. Then for*

all $\epsilon \geq 0, r \geq 0$,

$$Q(\epsilon, \|N\|) = \sup_{r \geq 0} \mathsf{P}\big(N \in S^1_{r,\epsilon}\big) \leq \frac{C_d \epsilon}{\sqrt{\lambda_2}},$$

where $C_v \equiv \sqrt{v-1}$.

Proof. Let $X := \sum_{i=1}^2 N_i^2$, $Y := \sum_{i=3}^d N_i^2$, and note that $\|N\| = \sqrt{X+Y}$. Then letting f_X denote the density of X we have

$$\begin{aligned} f_X(x) &= \frac{1}{2\pi\sqrt{\lambda_1\lambda_2}} \int_0^x e^{-\frac{z}{2\lambda_2} - \frac{x-z}{2\lambda_1}} (z(x-z))^{-1/2} dz \\ &\leq \frac{1}{2\pi\sqrt{\lambda_1\lambda_2}} B(1/2, 1/2) e^{-\frac{x}{2\lambda_1}}. \end{aligned} \tag{B.11}$$

For $y \geq 0$ the density of $\sqrt{X+y}$ is zero on $(-\infty, \sqrt{y})$ and using (B.11) it can be bounded on $[\sqrt{y}, \infty)$ by

$$f_{\sqrt{X+y}}(x) = 2x f_X(x^2 - y) \leq \frac{\sqrt{y+\lambda_1}}{\sqrt{\lambda_1\lambda_2}},$$

so that for all $r \geq 0$,

$$g(y) = \mathsf{P}\Big(r \leq \sqrt{X+y} \leq r+\epsilon\Big) \leq \frac{\sqrt{y+\lambda_1}}{\sqrt{\lambda_1\lambda_2}} \epsilon.$$

Hence, for $d \geq 3$, noticing that X is independent of Y, we find that

$$\begin{aligned} \mathsf{P}\Big(r \leq \sqrt{X+Y} \leq r+\epsilon\Big) &= \mathsf{E}[g(Y)] \\ &\leq \frac{\epsilon}{\sqrt{\lambda_1\lambda_2}} \mathsf{E}[Y+\lambda_1]^{1/2} \leq \frac{\epsilon}{\sqrt{\lambda_2}} \left(\sum_{i=3}^d \frac{\lambda_i}{\lambda_1} + 1 \right)^{1/2}, \end{aligned}$$

which proves the result. ■

Let $\phi(w) := \|w\|$. This function is trice continuously differentiable on $\mathbf{R}^v \setminus \{0\}$ and the following bounds on the derivatives of ϕ hold:

$$\begin{aligned} \big|\phi'(w)(x)\big| &\leq \|x\| \\ \big|\phi''(w)(x,y)\big| &\leq 2\|w\|^{-1}\|x\|\|y\| \\ \big|\phi'''(w)(x,y,z)\big| &\leq 5\|w\|^{-2}\|x\|\|y\|\|z\|. \end{aligned} \tag{B.12}$$

For a real symmetric matrix B we denote the j-th order statistic of its eigenvalues by $\lambda_{(j)}(B)$. Finally, we say that a random vector X in $\mathbf{R}^v$ is conditionally normal given $\mathcal{F} \subset \mathcal{H}$ with zero mean and the conditional covariance matrix V, denoted by $X \mid \mathcal{F} \sim \mathcal{N}(0, V)$, if V is $\mathcal{F}$-measurable, a.s. *finite and positive semi-definite*, and the conditional characteristic

function of X is given by

$$\mathsf{E}[e^{it^\top X} \mid \mathcal{F}] = \exp\left(-\frac{1}{2}t^\top V t\right) \quad \text{a.s.}$$

for all $t \in \mathbf{R}^v$.

Theorem B.1. *Let $X_1, \ldots, X_n$ be random vectors in $\mathbf{R}^v$ that are conditionally independent given $\mathcal{F} \subset \mathcal{H}$ with $\mathsf{E}[X_i \mid \mathcal{F}] = 0$ and $\mathsf{E}[\|X_i\|^3 \mid \mathcal{F}] < \infty$ a.s. Let $T \coloneqq \sum_{i=1}^n X_i$ and let N be a random vector in $\mathbf{R}^v$ s.t. $N \mid \mathcal{F} \sim \mathcal{N}(0, V)$, where $V = \mathsf{E}[TT^\top \mid \mathcal{F}]$ a.s. Then, assuming that $\upsilon \equiv \lambda_{(d\vee 2-1)}(V) > 0$ a.s.,*

$$d_K(\|T\|, \|N\| \mid \mathcal{F}) \leq C_d \left(\upsilon^{-3/2} \sum_{i=1}^n \mathsf{E}[\|X_i\|^3 \mid \mathcal{F}]\right)^{1/4} \quad \text{a.s.},$$

where $C_d > 0$ is a constant depending only on d.

Proof. Let f be a trice continuously differential function, s.t. $f(x) = 1$ if $x \leq 0$, $f = 0$ if $x \geq \epsilon > 0$, and $\left|f^{(j)}(x)\right| \leq D\epsilon^{-j}\mathbf{1}_{(0,\epsilon)}(x)$ for some constant $D > 0$ and $1 \leq j \leq 3$. Also define

$$g_r(s) \coloneqq f(\|s\| - r).$$

First,

$$\begin{aligned}\mathsf{P}(\|T\| \leq r \mid \mathcal{F}) &\leq \mathsf{E}[g_r(T) \mid \mathcal{F}] \\ &\leq \mathsf{P}(\|N\| \leq r + \epsilon \mid \mathcal{F}) + \mathsf{E}[g_r(T) - g_r(N) \mid \mathcal{F}]\end{aligned}$$

and

$$\begin{aligned}\mathsf{P}(\|T\| > r \mid \mathcal{F}) &\leq 1 - \mathsf{E}[g_{r-\epsilon}(T) \mid \mathcal{F}] \\ &\leq \mathsf{P}(\|N\| > r - \epsilon \mid \mathcal{F}) + \mathsf{E}[g_{r-\epsilon}(N) - g_{r-\epsilon}(T) \mid \mathcal{F}]\end{aligned}$$

a.s. for all $r \geq 0$ and $\epsilon > 0$. Therefore, w.p.1,

$$\begin{aligned}&d_K(\|T\|, \|N\| \mid \mathcal{F}) \\ &\quad = \sup_{q \in \mathbb{Q}_{\geq 0}} |\mathsf{P}(\|T\| \leq q \mid \mathcal{F}) - \mathsf{P}(\|N\| \leq q \mid \mathcal{F})| \\ &\quad \leq \sup_{q \in \mathbb{Q}_{>0}} |\mathsf{E}[g_q(T) - g_q(N) \mid \mathcal{F}]| + \sup_{q \in \mathbb{Q}_{\geq 0}} \mathsf{P}(N \in S_{q,\epsilon} \mid \mathcal{F}),\end{aligned} \tag{B.13}$$

Consider the first term on the third line of (B.13).

Claim B.3. *There exists a constant $B > 0$ s.t. for any $q > 0$,*

$$|\mathsf{E}[g_q(T) - g_q(N) \mid \mathcal{F}]| \leq \frac{B}{\epsilon^3} \sum_{i=1}^{n} \mathsf{E}[\|X_i\|^3 \mid \mathcal{F}] \quad a.s.$$

Proof. Let $Z_1, \ldots, Z_n$ be i.i.d. standard normal random vectors in $\mathbf{R}^v$ independent of $X_1, \ldots, X_n$ and $\mathcal{F}$, and let $Y_i \coloneqq V_i^{1/2} Z_i$, where V_i is a version of $\mathsf{E}[X_i X_i^\top \mid \mathcal{F}]$. Define

$$U_i \coloneqq \sum_{k=1}^{i-1} X_k + \sum_{k=i+1}^{n} Y_k$$

and

$$W_i \coloneqq g_q(U_i + X_i) - g_q(U_i + Y_i).$$

Then $g_q(T) - g_q(N) = \sum_{i=1}^n W_i$ and

$$|\mathsf{E}[g_q(T) - g_q(N) \mid \mathcal{F}]| \leq \sum_{i=1}^{n} |\mathsf{E}[W_i \mid \mathcal{F}]| \quad \text{a.s.}$$

Let $\mathcal{G}_i \coloneqq \mathcal{F} \bigvee \sigma(X_1, \ldots, X_{i-1}, Z_{i+1}, \ldots, Z_n)$ and let $h_{i1}(\lambda) \coloneqq g_q(U_i + \lambda X_i)$ and $h_{i2}(\lambda) \coloneqq g_q(U_i + \lambda Y_i)$. Using Taylor expansion up to the third order, we find that

$$W_i = \sum_{j=0}^{2} \frac{1}{j!} \Big(h_{i1}^{(j)}(0) - h_{i2}^{(j)}(0) \Big) + \frac{1}{3!} \Big(h_{i1}^{(3)}(\lambda_1) - h_{i2}^{(3)}(\lambda_2) \Big),$$

where $|\lambda_1|, |\lambda_2| < 1$. The tower property of conditional expectations and the fact that X_i and Y_i are conditionally independent of $\mathcal{G}_i$ given $\mathcal{F}$ imply that

$$\mathsf{E}[h_{i1}^{(j)}(0) - h_{i2}^{(j)}(0) \mid \mathcal{F}] = 0 \quad \text{a.s.}$$

for $j = 1, 2$. Finally, using the bounds in (B.12) and noticing that $|f^{(j)}(x-q)| \leq D\epsilon^{-3} x^{3-j} \times \mathbf{1}_{(q,q+\epsilon)}(x)$ for $1 \leq j \leq 3$, we get

$$|\mathsf{E}[h_{i1}^{(3)} - h_{i2}^{(3)} \mid \mathcal{F}]| \leq \frac{B}{\epsilon^3} \Big(\mathsf{E}[\|X_i\|^3 \mid \mathcal{F}] + \mathsf{E}[\|Y_i\|^3 \mid \mathcal{F}] \Big) \quad \text{a.s.}$$

for some constant $B > 0$. The result then follows from Lemma 4 in Rhee and Talagrand (1986), i.e., there is a constant $M > 0$ s.t. $\mathsf{E}[\|Y_i\|^3 \mid \mathcal{F}] \leq M \mathsf{E}[\|X_i\|^3 \mid \mathcal{F}]$ a.s. □

Using Lemma B.7 when $d \geq 2$ it follows that

$$d_K(\|T\|, \|N\| \mid \mathcal{G}) \leq \frac{B}{\epsilon^3} \sum_{i=1}^{n} \mathsf{E}[\|X_i\|^3 \mid \mathcal{F}] + \frac{C'_d}{\sqrt{\upsilon}} \epsilon \quad \text{a.s.} \tag{B.14}$$

For $d = 1$ we have $\mathsf{P}(N \in S_{q,\epsilon} \mid \mathcal{F}) \leq \epsilon/\sqrt{2\pi\upsilon}$ and the same bound holds. Finally, since (B.14) holds for any $\epsilon > 0$ a.s., it holds for random ϵ a.s. on $\{\epsilon \in (0,\infty)\}$. Then the result follows by taking $\epsilon = \left(\sqrt{\upsilon}\sum_{i=1}^n \mathsf{E}[\|X_i\|^3 \mid \mathcal{F}]\right)^{1/4}$. ■

Corollary B.1. *Let $X_1, \ldots, X_n$ be conditionally i.i.d. given $\mathcal{F} \subset \mathcal{H}$ with $\mathsf{E}[X_1 \mid \mathcal{F}] = 0$ and $\mathsf{E}[\|X_1\|^3 \mid \mathcal{F}] < \infty$ a.s. Let $T := n^{-1/2}\sum_{i=1}^n X_i$ and let $N \mid \mathcal{F} \sim \mathcal{N}(0, V)$, where $V = \mathsf{E}[X_1 X_1^\top \mid \mathcal{F}]$ a.s. Then, assuming that $\upsilon \equiv \lambda_{(d\vee 2-1)}(V) > 0$ a.s.*

$$d_K(\|T\|, \|N\| \mid \mathcal{F}) \leq C_d \left(\frac{\mathsf{E}[\|X_1\|^3 \mid \mathcal{F}]}{\upsilon^{3/2}\sqrt{n}}\right)^{1/4} \quad a.s.,$$

where $C_d > 0$ is a constant depending only on d.

Lemma B.8. *Let $X_1, \ldots, X_n$ be random variables that are conditionally i.i.d. given $\mathcal{F} \subset \mathcal{H}$ with $\mathsf{E}[X_1 \mid \mathcal{F}] = 0$ and $\mathsf{E}[|X_1|^3 \mid \mathcal{F}] < \infty$ a.s. Let $T := n^{-1/2}\sum_{i=1}^n X_i$ and $N \mid \mathcal{F} \sim \mathcal{N}\big(0, \sigma^2\big)$, where $\sigma^2 = \mathrm{Var}(X_1 \mid \mathcal{F})$ a.s. Then, assuming that $\sigma > 0$ a.s.,*

$$d_K(T, N \mid \mathcal{F}) \leq C \frac{\mathsf{E}[|X_1|^3 \mid \mathcal{F}]}{\sigma^3\sqrt{n}} \quad a.s.,$$

where $C > 0$ is a constant.

Proof. The proof is similar to the proof of Theorem 11.4.1 in Athreya and Lahiri (2006) (*for the unconditional case*) and so is omitted. ■

Lemma B.9. *Suppose that $\mathcal{G}$ and $\mathcal{F}$ are σ-fields s.t. $\mathcal{F} \subset \mathcal{G} \subset \mathcal{H}$, X and Y are random vectors in $\mathbf{R}^d$ s.t. $X \mid \mathcal{G} \sim \mathcal{N}(0, \Sigma_X)$ and $Y \mid \mathcal{F} \sim \mathcal{N}(0, \Sigma_Y)$. Then, assuming that $\upsilon \equiv \lambda_{(d\vee 2-1)}(\Sigma_Y) > 0$ a.s.,*

$$d_K(\|X\|, \|Y\| \mid \mathcal{G}, \mathcal{F}) \leq C_d\Big(\upsilon^{-1}\|\Lambda_X - \Lambda_Y\|_{e,\infty}\Big)^{1/3} \quad a.s., \tag{B.15}$$

where C_d is a constant depending only on d and $\Lambda_{(\,\cdot\,)}$ is the matrix of eigenvalues corresponding to $\Sigma_{(\,\cdot\,)}$.

Proof. Let f be a twice continuously differential function s.t. $f(x) = 1$ if $x \leq 0$, $f(x) = 0$ if $x \geq \epsilon > 0$ and $\left|f^{(j)}\right| \leq D\epsilon^{-j}\mathbf{1}_{(0,\epsilon)}(x)$ for some constant $D > 0$ and $1 \leq j \leq 2$. Further, set

$$g_r(s) := f(\|s\| - r).$$

As in the proof of Theorem B.1 for any $\epsilon > 0$ w.p.1,

$$d_K(\|X\|, \|Y\| \mid \mathcal{G}, \mathcal{F})$$

$$\leq \sup_{q\in\mathbb{Q}_{>0}} |\mathsf{E}[g_q(X) \mid \mathcal{G}] - \mathsf{E}[g_q(Y) \mid \mathcal{F}]| + \sup_{q\in\mathbb{Q}_{\geq 0}} \mathsf{P}(Y \in S_{q,\epsilon} \mid \mathcal{F}).$$

Let Z_1 and Z_2 be independent standard normal random vectors in $\mathbf{R}^d$ that are independent of $\mathcal{G}$ and $\mathcal{F}$, respectively. Then

$$\begin{aligned}\mathsf{E}[g_q(X) \mid \mathcal{G}] - \mathsf{E}[g_q(Y) \mid \mathcal{F}] &= \mathsf{E}[g_q(\Lambda_X^{1/2} Z_1) \mid \mathcal{G}] - \mathsf{E}[g_q(\Lambda_Y^{1/2} Z_2) \mid \mathcal{F}] \\ &= h_{q,1}(\Lambda_X) - h_{q,2}(\Lambda_Y) \quad \text{a.s.},\end{aligned}$$

where $h_{q,1}(\lambda) \coloneqq \mathsf{E}g_q(\lambda^{1/2} Z_1)$ and $h_{q,2}(\lambda) \coloneqq \mathsf{E}g_q(\lambda^{1/2} Z_2)$ (see, e.g., Durrett, 2010, Lemma 6.2.1).

Claim B.4. *There exists a constant B_d depending only on d s.t. for any $q > 0$,*

$$|h_{q,1}(\lambda_X) - h_{q,2}(\lambda_Y)| \leq \frac{B_d}{\epsilon^2} \|\lambda_X - \lambda_Y\|_{e,\infty}.$$

Proof. For $t \in [0,1]$ let $Z(t) \coloneqq \sqrt{t}\lambda_X^{1/2} Z_1 + \sqrt{1-t}\lambda_Y^{1/2} Z_2$ and $\phi(t) \coloneqq \mathsf{E}g_q(Z(t))$. Then

$$h_{q,1}(\lambda_X) - h_{q,2}(\lambda_Y) = \phi(1) - \phi(0) = \int_0^1 \phi'(t)dt.$$

Using the integration by parts formula (see Equation A.17 in Talagrand, 2011, Section A.6) for $t \in (0,1)$,

$$\begin{aligned}\phi'(t) &= \frac{1}{2}\mathsf{E}\left[\left(\lambda_X^{1/2} Z_1/\sqrt{t} - \lambda_Y^{1/2} Z_2/\sqrt{1-t}\right)^\top \nabla g_q(Z(t))\right] \\ &= \frac{1}{2}\mathsf{E}\left[\mathbf{i}^\top (\lambda_X - \lambda_Y) \circ \nabla^2 g_q(Z(t))\mathbf{i}\right],\end{aligned}$$

where $\mathbf{i}$ is the vector of ones. Therefore,

$$\left|\int_0^1 \phi'(t)dt\right| \leq \|\lambda_X - \lambda_Y\|_{e,\infty} \int_0^1 \mathsf{E}\left|\mathbf{i}^\top \nabla^2 g_q(Z(t))\mathbf{i}\right| dt.$$

Since $|f^{(j)}(x-q)| \leq D\epsilon^{-2}x^{2-j} \times \mathbf{1}_{(q,q+\epsilon)}(x)$ for $1 \leq j \leq 2$, the (k,l)-th element of the Hessian of g_q is bounded by $D'\epsilon^{-2}$ for some constant $D' > 0$. Therefore,

$$|h_{q,1}(\lambda_X) - h_{q,2}(\lambda_Y)| \leq \frac{D'd^2}{\epsilon^2} \|\lambda_X - \lambda_Y\|_{e,\infty}. \qquad \square$$

Using Lemma B.7 when $d \geq 2$ it follows that

$$d_K(\|X\|, \|Y\| \mid \mathcal{G}) \leq \frac{B_d}{\epsilon^2} \|\Lambda_X - \Lambda_Y\|_{e,\infty} + \frac{C'_d}{\sqrt{\upsilon}}\epsilon \quad \text{a.s.} \tag{B.16}$$

For $d = 1$, $\mathsf{P}(N \in S_{q,\epsilon} \mid \mathcal{F}) \leq \epsilon/\sqrt{2\pi\upsilon}$ a.s. so that the same bound holds. Finally, since (B.16) holds for any $\epsilon > 0$ a.s., it holds for random ϵ a.s. on $\{\epsilon \in (0, \infty)\}$. Consequently, the result follows by taking $\epsilon = (\sqrt{\upsilon}\|\Lambda_X - \Lambda_Y\|_{e,\infty})^{1/3}$ and noticing that (B.15) holds trivially on $\{\|\Lambda_X - \Lambda_Y\|_{e,\infty} = 0\}$. ■

Lemma B.10. *Suppose that $\mathcal{G}$ and $\mathcal{F}$ are σ-fields s.t. $\mathcal{F} \subset \mathcal{G} \subset \mathcal{H}$, and let $X \mid \mathcal{G} \sim \mathcal{N}\big(0, \sigma_X^2\big)$ and $Y \mid \mathcal{F} \sim \mathcal{N}\big(0, \sigma_Y^2\big)$. Then, assuming that $\sigma_Y > 0$ a.s.,*

$$d_K(X, Y \mid \mathcal{G}, \mathcal{F}) \leq C\big|\sigma_X^2/\sigma_Y^2 - 1\big|^{1/3} \quad a.s.,$$

where $C > 0$ is a constant.

Proof. The proof is similar to one for Lemma B.9, and so is omitted. ■

Lemma B.11. *Let $(G, (Y, X))$ be a network dependent process in $\mathbf{R} \times \mathbf{R}^d$ and let $\mathcal{F}$ be a sub-σ-field of $\mathcal{H}$ such that:*

(a) *Y is conditionally independent of X given $\mathcal{F}$,*
(b) *Y_i and Y_j are conditionally independent given $\mathcal{F}$ if $j \notin B_i := N(i; s)$ for some $s > 0$,*
(c) *$\mathcal{D}(G)$ is $\mathcal{F}$-measurable.*

Let $\mathcal{G} := \sigma(\mathcal{F} \cup \sigma(X))$, $T := \sum_{i \in N} Y_i X_i$, and $Z \mid \mathcal{G} \sim \mathcal{N}(0, V)$, where $V = \mathsf{E}[TT^\top \mid \mathcal{G}]$ a.s. Then, assuming that $\upsilon \equiv \lambda_{(d \vee 2-1)}(V) > 0$ a.s.,

$$d_K(\|T\|, \|Z\| \mid \mathcal{G}) \leq C_d\Big(\upsilon^{-3/2}\beta\Big)^{1/4} \quad a.s.,$$

where $C_d > 0$ is a constant depending only on d and

$$\beta := \sum_{i \in N} \sum_{j \in B_i} \sum_{k \in B_i \cup B_j} \prod_{l \in \{i,j,k\}} \|Y_l\|_{\mathcal{F},3} \|X_l\|_\infty.$$

In addition, when $d = 1$,

$$d_K(T, Z \mid \mathcal{G}) \leq C_1\Big(\upsilon^{-3/2}\beta\Big)^{1/4} \quad a.s.$$

Proof. We use the notation from the proof of Theorem B.1. First, for any $\epsilon > 0$ w.p.1,

$$\begin{aligned} &d_K(\|T\|, \|Z\| \mid \mathcal{G}) \\ &\quad \leq \sup_{q \in Q_{>0}} |\mathsf{E}[g_q(T) - g_q(Z) \mid \mathcal{G}]| + \sup_{q \geq 0} \mathsf{P}(Z \in S_{q,\epsilon} \mid \mathcal{G}). \end{aligned}$$

Let $Y' \mid \mathcal{F} \sim \mathcal{N}(0, \Sigma)$ conditionally independent of (Y, X) given $\mathcal{F}$, where $\Sigma = \mathrm{Var}(Y \mid \mathcal{F})$ a.s., and let $Z' := \sum_{i \in N} Y_i' X_i$. Note that $\mathsf{E}[g_q(Z) \mid \mathcal{G}] = \mathsf{E}[g_q(Z') \mid \mathcal{G}]$ a.s. Also let Q_Y and

$\mathsf{Q}_{Y'}$ be the regular conditional distributions of Y and Y' given $\mathcal{F}$ and $\mathsf{Q} \coloneqq \mathsf{Q}_Y \otimes \mathsf{Q}_{Y'}$. Since X is $\mathcal{G}$-measurable, for almost all $\omega \in \Omega$,

$$\mathsf{E}[g_q(T) - g_q(Z) \mid \mathcal{G}](\omega) = h_q(\omega),$$

where

$$h_q(\omega) \coloneqq \int_{R^{n\times 2}} g_q\left(\sum_{i\in N} y_i X_i(\omega)\right) - g_q\left(\sum_{i\in N} y_i' X_i(\omega)\right) \mathsf{Q}(\omega, d(y, y'))$$

(see, e.g., Kallenberg, 2002, Theorem 5.4).

Claim B.5. *There exists a constant $B_d > 0$ depending only on d s.t. for any $q > 0$,*

$$|h_q(\omega)| \leq \frac{B_d}{\epsilon^3} \sum_{i\in N} \sum_{j\in B_i} \sum_{k\in B_i\cup B_j} \prod_{l\in\{i,j,k\}} (\chi_l(\omega))^{1/3} \|X_l(\omega)\|_\infty,$$

where $\chi_i(\omega) \coloneqq \int_{\mathbf{R}^n} y_i^3 \mathsf{Q}_Y(\omega, d(y))$.

Proof. For $y \equiv \{y_i\}_{i\in N}$ let $\phi(y) \coloneqq g_q(\sum_{i\in N} y_i X_i(\omega))$. Then the result follows from Theorem 3.4 in Röllin (2013) by observing that

$$\|\phi_{ijk}\|_\infty \leq \frac{B_d'}{\epsilon^3} \prod_{l\in\{i,j,k\}} \|X_l(\omega)\|_\infty$$

for some constant $B_d' > 0$ depending only on d, where ϕ_{ijk} is the third order partial derivative of ϕ w.r.t. the coordinates i, j, and k. □

As in the proof of Theorem B.1 there exists a constant $C_d' > 0$ depending only on d s.t.

$$\sup_{q\geq 0} \mathsf{P}(Z \in S_{q,\epsilon} \mid \mathcal{F}) \leq \frac{C_d'}{\sqrt{\upsilon}}\epsilon \quad \text{a.s.}$$

Therefore, noticing that $\chi_i = \mathsf{E}[|Y_i|^3 \mid \mathcal{F}]$ a.s., the result follows by taking $\epsilon = (\sqrt{\upsilon}\beta)^{1/4}$. The second assertion for $d = 1$ follows similarly. ■

Appendix C

Appendix to Chapter 4

C.1 Extension to multinomial action sets

In this appendix, we show how the results of the main text can be extended to the case with a multinomial action set, i.e., $k \equiv |A| > 2$. The formal results in this appendix include Theorems 1-4 as special cases.

C.1.1 Belief-free hindsight regrets

First, we extend the measure of strategic interdependence. For $i, j \in N$ and $a, a' \in A$ we define

$$\Delta_{ij}(a, a', t) \coloneqq V_j(b \mapsto u_i^{\Delta}(a, a', b; t)),$$

where $V_j(\,\cdot\,)$ is defined in (4.7) and

$$u_i^{\Delta}(a, a', b; t) \coloneqq u_i(a, b; t) - u_i(a', b; t)$$

is player i's payoff differential between choosing a and a' when other players choose $b \in A^{k-1}$. We make the following assumption, which extends Assumption 4.3 in the main text.

Assumption 3.3'. For each $i, j \in N$ with $i \neq j$ and $a, a' \in A$, the map $t \mapsto \Delta_{ij}(a, a', t)$ is $(\mathcal{E}_i, \mathcal{B}_{\mathbf{R}})$-measurable.

Definition C.1. Given an equilibrium $\mathcal{Y}$ and $\rho \in (0, 1)$, an $\mathcal{F}_i$-measurable random vector $\boldsymbol{\lambda} \in \mathbf{R}_{\geq 0}^{k-1}$ is a *ρ-hindsight regret* for player $i \in N$ if

$$\mathsf{Q}_i\big(\mathbf{u}_i^{\Delta}(Y_i, Y_{-i}; \xi_i) \geq -\boldsymbol{\lambda} \mid \mathcal{I}_i\big) \geq 1 - \rho \quad (\mathsf{Q}_i\text{-a.s.}), \tag{C.1}$$

where $\mathbf{u}_i^{\Delta}(a, b; t) \coloneqq [u_i^{\Delta}(a, a', b; t)]_{a' \in A \setminus \{a\}}$.

Let $\boldsymbol{\lambda_{i,\rho}}(a,t) \coloneqq [\lambda_{i,\rho}(a,a',t)]_{a'\in A\setminus\{a\}}$ be a vector in $\mathbf{R}^{k-1}$ whose elements are given by

$$\lambda_{i,\rho}(a,a',t) \coloneqq \sqrt{-\frac{1}{2}\ln\left(\frac{\rho}{k-1}\right)\cdot\Lambda_i(a,a',t)},$$

where

$$\Lambda_i(a,a',t) \coloneqq \sum_{j\in N\setminus\{i\}} (\Delta_{ij}(a,a',t))^2$$

The following theorem confirms that $\boldsymbol{\lambda_{i,\rho}}(Y_i,\xi_i)$ is a ρ-hindsight regret.

Theorem C.1. *A.1 Suppose that Assumptions 4.1 and 4.2 hold. Then for any pure strategy equilibrium $\mathcal{Y}$ and any $\rho \in (0,1)$, $\boldsymbol{\lambda}_{i,\rho}(Y_i,\xi_i)$ is a ρ-hindsight regret for player $i \in N$. Moreover, there exists $\rho \in (0,1)$ such that*

$$\mathsf{P}\big(\mathbf{u}_i^{\Delta}(Y_i,Y_{-i};\xi_i) \geq -\boldsymbol{\lambda}_{i,\rho}(Y_i,\xi_i) \mid \mathcal{F}_i\big) \geq 1-\rho_i(Y_i,\xi_i) \quad (\mathsf{P}\text{-}a.s.),$$

where $\rho_i(a,t) \coloneqq \rho\sum_{a'\in A\setminus\{a\}} 1\{\lambda_{i,\rho}(a,a',t)>0\}/(k-1)$.

C.1.2 Testable implications

First, for each $i \in N$ and $a \in A$ we construct $\pi_{i,L}(a) \equiv \pi_{i,L}(a,X_i,C,Y_{-i})$ and $\pi_{i,U}(a) \equiv \pi_{i,U}(a,X_i,C,Y_{-i})$ that can be simulated as follows:

$$\pi_{i,L}(a,x,c,b) = \mathsf{P}(\eta_i \in H_{i,L}(a,x,b) \mid X_i = x, C = c) \quad \text{and}$$
$$\pi_{i,U}(a,x,c,b) = \mathsf{P}(\eta_i \in H_{i,U}(a,x,b) \mid X_i = x, C = c)$$

where, letting $t \equiv (x,\eta)$,

$$H_{i,L}(a,x,b) \coloneqq \{\eta \in \mathbf{R}^w : \exists a' \in A\setminus\{a\} \text{ s.t. } \mathbf{u}_i^{\Delta}(a',b;t) \geq -\boldsymbol{\lambda}_{i,\rho}(a',t)\}$$

and

$$H_{i,U}(a,x,b) \coloneqq \{\eta \in \mathbf{R}^w : \mathbf{u}_i^{\Delta}(a,b;t) \geq -\boldsymbol{\lambda}_{i,\rho}(a,t)\}.$$

Then we define

$$\begin{aligned} e_{i,L}(a) &\coloneqq \mathsf{P}(Y_i = a \mid X_i) - \left(1 - \frac{\mathsf{E}_{\mathsf{P}}[\pi_{i,L}(a) \mid \mathcal{F}]}{1-\rho_{i,L}(a)}\right) \quad \text{and} \\ e_{i,U}(a) &\coloneqq \mathsf{P}(Y_i = a \mid X_i) - \frac{\mathsf{E}_{\mathsf{P}}[\pi_{i,U}(a) \mid \mathcal{F}]}{1-\rho_{i,U}(a)}, \end{aligned} \tag{C.2}$$

where

$$\rho_{i,L}(a) \coloneqq \rho\,\mathbf{1}\left\{\max_{c,a'\in A\setminus\{a\}} \sup_{t\in E_i} \lambda_{i,\rho}(c,a',t) > 0\right\} \quad \text{and}$$

$$\rho_{i,U}(a) \coloneqq \rho\,\mathbf{1}\Big\{\max_{a'\in A\setminus\{a\}}\sup_{t\in E_i}\lambda_{i,\rho}(a,a',t)>0\Big\}.$$

Let $\mathbf{e}_{i,L}$ and $\mathbf{e}_{i,U}$ be $(k-1)$-dimensional vectors whose elements are $e_{i,L}(a)$ and $e_{i,U}(a)$ with a running in $A_{-1}\equiv A\setminus\{a_1\}$ (a_1 is used as a reference action). Also let $\otimes$ denotes the Kronecker product. As before, by choosing a vector of non-negative measurable functions $\mathbf{g}_i\equiv[g_{i,1},\ldots,g_{i,m}]^\top:\mathbf{R}^v\to\mathbf{R}^m_{\geq 0}$, we consider the following moment inequalities:

$$\text{(C.3)}\qquad \begin{aligned}\boldsymbol{\mu}_L &\coloneqq n^{-1}\sum_{i=1}^n \mathbf{e}_{i,L}\otimes\mathbf{g}_i(X_i)\geq 0 \quad (\mathsf{P}\text{-a.s.}),\\ \boldsymbol{\mu}_U &\coloneqq n^{-1}\sum_{i=1}^n \mathbf{e}_{i,U}\otimes\mathbf{g}_i(X_i)\leq 0 \quad (\mathsf{P}\text{-a.s.}).\end{aligned}$$

The following result confirms that the inequality restrictions in (C.3) hold.

Proposition C.1. *Suppose that Assumptions 4.1-(3.3')-4.5 are satisfied. Then the inequality restrictions in* (C.3) *hold.*

C.1.3 Bootstrap inference

C.1.3.1 Test statistic

For each $a\in A$, we define

$$\begin{aligned}\hat{e}_{i,L}(a) &\coloneqq \mathbf{1}\{Y_i=a\}-\left(1-\frac{\pi_{i,L}(a)}{1-\rho_{i,L}(a)}\right) \quad \text{and}\\ \hat{e}_{i,U}(a) &\coloneqq \mathbf{1}\{Y_i=a\}-\frac{\pi_{i,U}(a)}{1-\rho_{i,U}(a)}.\end{aligned}$$

Let $\hat{\mathbf{e}}_{i,L}$ and $\hat{\mathbf{e}}_{i,U}$ be vectors of dimension $d\equiv m(k-1)>1$ whose elements are $\hat{e}_{i,L}(a)$ and $\hat{e}_{i,U}(a)$ with a running in A_{-1}. The sample analogues of $\boldsymbol{\mu}_L$ and $\boldsymbol{\mu}_U$ are given by

$$\hat{\boldsymbol{\mu}}_L \coloneqq n^{-1}\sum_{i=1}^n \hat{\mathbf{e}}_{i,L}\otimes\mathbf{g}_i(X_i) \quad\text{and}\quad \hat{\boldsymbol{\mu}}_U \coloneqq n^{-1}\sum_{i=1}^n \hat{\mathbf{e}}_{i,U}\otimes\mathbf{g}_i(X_i)$$

Using these sample moments, we take the following as our test statistic:

$$T\coloneqq \mathcal{T}\big(\sqrt{n}(\hat{\boldsymbol{\mu}}_L+\mathbf{w}_L),\sqrt{n}(\hat{\boldsymbol{\mu}}_U-\mathbf{w}_U)\big),$$

where $\mathcal{T} : \mathbf{R}^d \times \mathbf{R}^d \to \mathbf{R}$ is a function defined by $\mathcal{T}(x, y) := \|[x]_- + [y]_+\|_1$,[1] and $\mathbf{w}_L$ and $\mathbf{w}_U$ are constructed as follows. For $j \neq i$, $1 \leq l \leq m$, and $a \in A$ let

$$d_{l,L}(a) := n^{-1} \sum_{j \in N} \sum_{i \in N \setminus \{j\}} \frac{V_{\sigma(i,j)}(b \mapsto \pi_{i,L}(a, X_i, C, b)) g_{i,l}(X_i)}{1 - \rho_{i,L}(a)} \quad \text{and}$$
$$d_{l,U}(a) := n^{-1} \sum_{j \in N} \sum_{i \in N \setminus \{j\}} \frac{V_{\sigma(i,j)}(b \mapsto \pi_{i,U}(a, X_i, C, b)) g_{i,l}(X_i)}{1 - \rho_{i,U}(a)},$$

where $\sigma(i, j)$ is the index of the j-th element of $N = \{1, \ldots, n\}$ in N_{-i}. Then for a given $\varrho \in (0, 1)$ the elements of $\boldsymbol{w}_L$ are defined to be

$$\sqrt{-\frac{1}{2} \ln\Big(\frac{\varrho}{4d}\Big) \big([d_{l,L}(a)]^*\big)^2},$$

where $[d_{l,L}(a)]^*$ is the minimal measurable majorant of $d_{l,L}(a)$, with a running in A_{-1} and l running in $\{1, \ldots, m\}$, keeping the same order of elements as in the sample moments $\hat{\boldsymbol{\mu}}_L$ and $\hat{\boldsymbol{\mu}}_U$. The elements of $\mathbf{w}_U$ is defined similarly.

C.1.4 Bootstrap critical values

As for the bootstrap counterpart of T consider i.i.d. standard normal random variables $\{\varepsilon_1, \ldots, \varepsilon_n\}$ and let

$$\boldsymbol{\zeta}^* := n^{-1} \sum_{i=1}^{n} ((\mathbf{Y}_i - \boldsymbol{\mu}_i^*) \otimes \mathbf{g}_i(X_i))\, \varepsilon_i,$$

where $\mathbf{Y}_i$ and $\boldsymbol{\mu}_i^*$ are column vector formed by $\mathbf{1}\{Y_i = a\}$ and

$$\frac{1}{2}\left(1 - \frac{\pi_{i,L}(a) - \pi_{i,U}(a)}{1 - \rho_i(a)}\right) \vee 0 \wedge 1,$$

respectively, with a running in A_{-1}. In addition, given fixed $\kappa \in (0, 1)$ we define

$$\hat{\boldsymbol{\varphi}}_L := \big[\hat{\boldsymbol{\mu}}_L - \mathbf{w}_{\tau,L} - \mathbf{1}_d \cdot q^*(1 - \kappa/2)/\sqrt{n}\big]_+ \quad \text{and}$$
$$\hat{\boldsymbol{\varphi}}_U := \big[\hat{\boldsymbol{\mu}}_U + \mathbf{w}_{\tau,U} + \mathbf{1}_d \cdot q^*(1 - \kappa/2)/\sqrt{n}\big]_-,$$

where $q^*(z)$ is a certain random variable that dominates the z-quantile of the conditional distribution of the maximum of $\tilde{\boldsymbol{\mu}}_L - \boldsymbol{\mu}_L$ (or $\tilde{\boldsymbol{\mu}}_U - \boldsymbol{\mu}_U$) given $\mathcal{F}$. The computation of $q^*(z)$ is described in the end of this subsection.

We consider the following bootstrap test statistic:

$$T^* := \mathcal{T}\big(\sqrt{n}(\boldsymbol{\zeta}^* + \hat{\boldsymbol{\varphi}}_L \wedge \hat{\boldsymbol{\varphi}}_U), \sqrt{n}(\boldsymbol{\zeta}^* - \hat{\boldsymbol{\varphi}}_L \wedge \hat{\boldsymbol{\varphi}}_U)\big)$$

[1] For a vector $x \in \mathbf{R}^d$, $[x]_+ \equiv [x_j \vee 0]_{j=1}^d$ and $[x]_- \equiv -[x_j \wedge 0]_{j=1}^d$.

(the minimum between $\hat{\boldsymbol{\varphi}}_L$ and $\hat{\boldsymbol{\varphi}}_U$ is taken element-wise), and the confidence set for $\theta_0 \in \Theta$ at nominal level $1-\alpha$ is given by

$$CS_\epsilon := \{\theta \in \Theta : T \leq c^*(\gamma) \vee \epsilon\}, \tag{C.4}$$

where $\epsilon > 0$ is a fixed small number and $c^*(\gamma)$ is the $\gamma \equiv (1-\alpha+\varrho+\kappa)$-quantile of the bootstrap distribution of T^*.

Theorem C.2. *Suppose that Assumptions 1-4 hold and $\exists C_{\mathbf{g}} > 0$ s.t. $\|g_{il}\|_\infty \leq C_{\mathbf{g}}$ for all $i \geq 1$ and $1 \leq l \leq m$. Furthermore, assume that there exists a sequence $\{\lambda_n\}$ s.t. $\lambda^{-1} = O(n^{1/(3+\delta)})$ for some $\delta > 0$ and*

$$\lim_{n\to\infty} \sup_{\mathsf{P}\in\mathcal{P}_0} \mathsf{P}(\underline{\lambda}(V) < \lambda) = 0.$$

Then, for a sequence $\{\epsilon_n\}$ s.t. $\epsilon^{-1} = o(n^{1/8})$,

$$\limsup_{n\to\infty} \sup_{\mathsf{P}\in\mathcal{P}_0} \mathsf{P}(\theta_0 \notin CS_\epsilon) \leq \alpha.$$

Consider the construction of $q^*(z)$. Let $\mathscr{S}_d \subset \mathbf{R}^{d\times d}$ denote the space of semi-positive definite matrices, and for $x \equiv (x_1, \ldots, x_n)$ let

$$B(x) := n^{-1} \sum_{i=1}^{n} (\mathbf{1}_{k-1} \otimes \mathbf{g}_i(x_i))(\mathbf{1}_{k-1} \otimes \mathbf{g}_i(x_i))^\top.$$

The smallest eigenvalue of $A \in \mathscr{S}_d$ is denoted by $\underline{\lambda}(A)$. Then $q^*(z)$ is the minimal measurable majorant of $q^*(z, (X_1, \ldots, X_n))$, where

$$q^*(z, x) := \sup\{H^{-1}(z; A) : A \in \mathscr{S}_d,\ \underline{\lambda}(A) \geq \lambda, |A| \leq [B(x)]\},$$

$H(\,\cdot\,;\Sigma)$ is the cdf of the maximum of $Z \sim \mathcal{N}(0, \Sigma)$, $|A|$ is the element-wise absolute value of A, and λ is the parameter appearing in Theorem C.2.

C.1.4.1 Infeasible test statistic

Similarly to the previous section, we define infeasible moments. First, for $a \in A$ let

$$\begin{aligned}
\tilde{e}_{i,L}(a) :=& \mathbf{1}\{Y_i = a\} - \left(1 - \frac{\mathsf{E}_{\mathsf{P}}[\pi_{i,L}(a) \mid \mathcal{F}]}{1-\rho_i(a)}\right) \quad \text{and} \\
\tilde{e}_{i,U}(a) :=& \mathbf{1}\{Y_i = a\} - \frac{\mathsf{E}_{\mathsf{P}}[\pi_{i,U}(a) \mid \mathcal{F}]}{1-\rho_i(a)},
\end{aligned}$$

and let $\tilde{\mathbf{e}}_{i,L}$ and $\tilde{\mathbf{e}}_{i,U}$ be column vectors whose elements are $\tilde{e}_{i,L}(a)$ and $\tilde{e}_{i,U}(a)$ with a running in A_{-1}. The infeasible moments are given by

$$\tilde{\boldsymbol{\mu}}_L := n^{-1}\sum_{i=1}^{n}\tilde{\mathbf{e}}_{i,L}\otimes\mathbf{g}_i(X_i) \quad \text{and} \quad \tilde{\boldsymbol{\mu}}_U := n^{-1}\sum_{i=1}^{n}\tilde{\mathbf{e}}_{i,U}\otimes\mathbf{g}_i(X_i).$$

In addition, we define

$$\mathcal{M}(\mathbf{w}_L,\mathbf{w}_U) := \{|\hat{\boldsymbol{\mu}}_L-\tilde{\boldsymbol{\mu}}_L|\le\mathbf{w}_L, |\hat{\boldsymbol{\mu}}_U-\tilde{\boldsymbol{\mu}}_U|\le\mathbf{w}_U\}$$

and the confidence rectangles for the conditional expectations of $\tilde{\boldsymbol{\mu}}_L$ and $\tilde{\boldsymbol{\mu}}_U$ at nominal level $1-\kappa/2$:

$$R_L := \left\{\boldsymbol{\mu}\in\mathbf{R}^d : \min_j \sqrt{n}(\boldsymbol{\mu}_j-\tilde{\boldsymbol{\mu}}_{L,j})\ge q^*(1-\kappa/2)\right\} \quad \text{and}$$
$$R_U := \left\{\boldsymbol{\mu}\in\mathbf{R}^d : \max_j \sqrt{n}(\boldsymbol{\mu}_j-\tilde{\boldsymbol{\mu}}_{U,j})\le -q^*(1-\kappa/2)\right\}.$$

The following proposition shows that $\mathbf{w}_L$ and $\mathbf{w}_U$ control the difference between the feasible and infeasible moments with predetermined probability.

Proposition C.2. *Suppose that Assumptions 4.1-(3.3')-4.5 hold. Then for a given $\varrho\in(0,1)$,*

$$\mathsf{P}(\mathcal{M}(\mathbf{w}_L,\mathbf{w}_U))\ge 1-\varrho.$$

On the event $\mathcal{M}(\mathbf{w}_{\tau,L},\mathbf{w}_{\tau,U})$,

$$T\le\widetilde{T} := \mathcal{T}\big(\sqrt{n}\tilde{\boldsymbol{\mu}}_L,\sqrt{n}\tilde{\boldsymbol{\mu}}_U\big),$$

and

$$\hat{\boldsymbol{\varphi}}_L \le \big[\tilde{\boldsymbol{\mu}}_L-\mathbf{1}_d\cdot q^*(1-\kappa/2)/\sqrt{n}\big]_+,$$
$$\hat{\boldsymbol{\varphi}}_U \le \big[\tilde{\boldsymbol{\mu}}_U+\mathbf{1}_d\cdot q^*(1-\kappa/2)/\sqrt{n}\big]_-.$$

The latter inequalities imply that under $\mathsf{P}\in\mathcal{P}_0$, for which $\boldsymbol{\mu}_L\ge 0$ and $\boldsymbol{\mu}_U\le 0$ (P-a.s.),

$$T^*\ge\widetilde{T}^* := \mathcal{T}\big(\sqrt{n}(\boldsymbol{\zeta}^*+\boldsymbol{\mu}),\sqrt{n}(\boldsymbol{\zeta}^*-\boldsymbol{\mu})\big)$$

on $\mathcal{M}(\mathbf{w}_{\tau,L},\mathbf{w}_{\tau,U})\cap\{\boldsymbol{\mu}_L\in R_L\}\cap\{\boldsymbol{\mu}_U\in R_U\}$, where $\boldsymbol{\mu}:=\boldsymbol{\mu}_L\wedge(-\boldsymbol{\mu}_U)$. Consequently, letting $\tilde{c}^*(\gamma)$ denote the γ-quantile of the bootstrap distribution of $\widetilde{T}^*$,

$$\begin{aligned}\mathsf{P}(T>c^*(\gamma)\vee\epsilon) &\le \mathsf{P}\Big(\widetilde{T}>\tilde{c}^*(\gamma)\vee\epsilon\Big)+\mathsf{P}(\mathcal{M}(\mathbf{w}_{\tau,L},\mathbf{w}_{\tau,U})^c)\\ &\quad+\mathsf{P}(\boldsymbol{\mu}_L\notin R_L)+\mathsf{P}(\boldsymbol{\mu}_U\notin R_U).\end{aligned} \tag{C.5}$$

C.2 Proofs of the main results

We prove the results presented in the appendix. Theorems and propositions in the main text follow as corollaries for the special case of $A = \{0, 1\}$.

Proof of Theorem C.1. First, note that Y_i and ξ_i are $\mathcal{I}_i$-measurable and the elements of Y_{-i} are conditionally independent given $\mathcal{I}_i$ by Assumption 4.1. Then the implications of the theorem follow from Lemma C.1 and Assumption 4.2. ■

Proof of Proposition C.1. For $a \in A$ define the event

$$S_{i,U}(a) \coloneqq \{\mathbf{u}_i^{\Delta}(a, Y_{-i}; \xi_i) \geq -\boldsymbol{\lambda}_{i,\rho}(a, \xi_i)\}.$$

By the definition of $\boldsymbol{\lambda}_{i,\rho}$ and Theorem C.1 we have

$$\sum_{a \in A} \mathsf{P}(S_{i,U}(a) \mid \mathcal{I}_i)\mathbf{1}\{Y_i = a\} \geq 1 - \sum_{a \in A} \rho_i(a, \xi_i)\mathbf{1}\{Y_i = a\} \quad (\mathsf{P}\text{-a.s.}).$$

Therefore, noticing that $\rho_i(a, \xi_i) \leq \rho_{i,U}(a)$,

$$\mathbf{1}\{Y_i = a\} \leq \mathbf{1}\{\mathsf{P}(S_{i,U}(a) \mid \mathcal{I}_i) \geq 1 - \rho_{i,U}(a)\} \quad (\mathsf{P}\text{-a.s.}). \tag{C.6}$$

Similarly,

$$\mathbf{1}\{Y_i \neq a\} \leq \mathbf{1}\{\mathsf{P}(S_{i,L}(a) \mid \mathcal{I}_i) \geq 1 - \rho_{i,L}(a)\} \quad (\mathsf{P}\text{-a.s.}), \tag{C.7}$$

where $S_{i,L}(a) \coloneqq \bigcup_{c \in A \setminus \{a\}} S_{i,U}(c)$. Taking the conditional expectations w.r.t. $\mathcal{F}$ on both sides of (C.6) and (C.7), and using Markov's inequality, we find that

$$\mathsf{P}(Y_i = a \mid \mathcal{F}) \leq \frac{1}{1 - \rho_{i,U}(a)} \mathsf{P}(S_{i,U}(a) \mid \mathcal{F}) \quad (\mathsf{P}\text{-a.s.}).$$

and

$$\mathsf{P}(Y_i \neq a \mid \mathcal{F}) \leq \frac{1}{1 - \rho_{i,L}(a)} \mathsf{P}(S_{i,L}(a) \mid \mathcal{F}) \quad (\mathsf{P}\text{-a.s.}).$$ ■

Proof of Proposition C.2. Since the elements of Y_{-i} are conditionally independent given $\mathcal{F}$ by Assumption 4.1, the result follows from the definition of $\mathbf{w}_L$ and $\mathbf{w}_U$ and Lemma C.1. ■

Proof of Theorem C.2. Let $\mathcal{G} \coloneqq \mathcal{F} \vee \sigma(Y_1, \ldots, Y_n)$ and let

$$\boldsymbol{\zeta} \coloneqq n^{-1} \sum_{i=1}^{n} (\mathbf{Y}_i - \mathsf{E}_{\mathsf{P}}[\mathbf{Y}_i \mid \mathcal{F}]) \otimes \mathbf{g}_i(X_i),$$

Under Assumption 4.1 $Y_1, \ldots, Y_n$ are conditionally independent given $\mathcal{F}$. It follows that $\boldsymbol{\zeta}$ is the sum of conditionally independent random vectors given $\mathcal{F}$ and

$$\boldsymbol{\zeta} = \tilde{\boldsymbol{\mu}}_L - \boldsymbol{\mu}_L = \tilde{\boldsymbol{\mu}}_U - \boldsymbol{\mu}_U \quad (\mathsf{P}\text{-a.s.}).$$

Claim C.1. *For any* $\mathsf{P} \in \mathcal{P}_0$ *and* $\upsilon \in (0, \gamma)$,

$$\begin{aligned}
&\mathsf{P}\Big(\widetilde{T} > \tilde{c}^*(\gamma) \vee \epsilon\Big) - (1-\gamma) \\
&\qquad \lesssim h_{1,n} + \upsilon^{-1} h_{2,n} + (1 + \upsilon^{-1}) \mathsf{P}(\underline{\lambda}(V) < \lambda) + \upsilon,
\end{aligned}$$

where

$$h_{1,n} \equiv \frac{1}{(\lambda^3 n)^{1/8}} + \frac{1}{\epsilon^4 \sqrt{n}} \quad \textit{and} \quad h_{2,n} \equiv \frac{1}{(\lambda^2 n)^{1/6}} + \frac{1}{\epsilon^3 \sqrt{n}}.$$

Proof. Let Z be a standard normal random vector in $\mathbf{R}^d$ independent of $\mathcal{G}$. Define

$$\begin{aligned}
\widetilde{T}' &\coloneqq \mathcal{T}(\sqrt{n}(\boldsymbol{\zeta} + \boldsymbol{\mu}), \sqrt{n}(\boldsymbol{\zeta} - \boldsymbol{\mu})), \\
S' &\coloneqq \mathcal{T}(V^{1/2} Z + \sqrt{n}\boldsymbol{\mu}, V^{1/2} Z - \sqrt{n}\boldsymbol{\mu}), \quad \text{and} \\
S^* &\coloneqq \mathcal{T}(W^{1/2} Z + \sqrt{n}\boldsymbol{\mu}, W^{1/2} Z - \sqrt{n}\boldsymbol{\mu}),
\end{aligned}$$

where V and W are versions of $n\mathsf{E}[\boldsymbol{\zeta}\boldsymbol{\zeta}^\top \mid \mathcal{F}]$ and

$$n^{-1} \sum_{i=1}^{n} \mathsf{E}_{\mathsf{P}}\Big[((\mathbf{Y}_i - \boldsymbol{\mu}_i^*) \otimes g_i(X_i))((\mathbf{Y}_i - \boldsymbol{\mu}_i^*) \otimes g_i(X_i))^\top \mid \mathcal{F}\Big],$$

respectively. Since $W - V$ is positive semidefinite, and sets of the form $\{x \in \mathbf{R}^d : \mathcal{T}(x + a, x - a) \le t\}$ with $a \in \mathbf{R}^d_{\ge 0}$ and $t \ge 0$, are convex and symmetric under reflection, Theorem 1 in Jensen (1984) implies that for all $t \in \mathbf{R}$,

$$\mathsf{P}\big(S' \le t \mid \mathcal{F}\big) \ge \mathsf{P}(S^* \le t \mid \mathcal{F}) \quad (\mathsf{P}\text{-a.s.}).$$

Letting $c(z)$ denote the z-quantile of the conditional distribution of S' given $\mathcal{F}$, we get

$$\begin{aligned}
\mathsf{P}\Big(\widetilde{T} > \tilde{c}^*(\gamma) \vee \epsilon \mid \mathcal{F}\Big) - (1-\gamma) &\le d_K^\epsilon\Big(\widetilde{T}', S' \mid \mathcal{F}\Big) \\
&\quad + \mathsf{P}\Big(\widetilde{T}' \le c(\gamma) \vee \epsilon \mid \mathcal{F}\Big) - \mathsf{P}\Big(\widetilde{T}' \le \tilde{c}^*(\gamma) \vee \epsilon \mid \mathcal{F}\Big) \\
&\le 3 d_K^\epsilon\Big(\widetilde{T}', S' \mid \mathcal{F}\Big) + \mathsf{P}(c(\gamma - \upsilon) > \tilde{c}^*(\gamma) \vee \epsilon \mid \mathcal{F}) \\
&\quad + \mathsf{P}\big(c(\gamma - \upsilon) \vee \epsilon < S' \le c(\gamma) \vee \epsilon \mid \mathcal{F}\big) \quad (\mathsf{P}\text{-a.s.}),
\end{aligned}$$

where $d_K^\epsilon(\,\cdot\,,\,\cdot\,\mid\mathcal{F})$ is defined in (C.12). Let $\Delta \equiv d_K^\epsilon\Big(\widetilde{T}^*, S^* \mid \mathcal{G}, \mathcal{F}\Big)$. Then

$$\mathsf{P}(c(\gamma-\upsilon) > \tilde{c}^*(\gamma)\vee\epsilon \mid \mathcal{F}) \le \mathsf{P}(\Delta > \upsilon \mid \mathcal{F}) \quad (\mathsf{P}\text{-a.s.})$$

because on the event $\{\Delta \le \upsilon\}\cap\{c(\gamma-\upsilon) > \epsilon\}$,

$$\mathsf{P}\big(S' \le \tilde{c}^*(\gamma)\vee\epsilon \mid \mathcal{G}\big) \ge \mathsf{P}\big(S' \le c(\gamma-\upsilon) \mid \mathcal{G}\big) \quad (\mathsf{P}\text{-a.s.}).$$

Consequently,

$$\text{(C.8)} \qquad \mathsf{P}\Big(\widetilde{T} > \tilde{c}^*(\gamma)\vee\epsilon\Big) - (1-\gamma) \le 3\,\mathsf{E}_\mathsf{P} d_K^\epsilon\Big(\widetilde{T}', S' \mid \mathcal{F}\Big) + \mathsf{P}(\Delta > \upsilon) + \upsilon.$$

Consider each term on the RHS of (C.8). First, the largest eigenvalue $\overline{\lambda}(V)$ of V is bounded, i.e.,

$$\overline{\lambda}(V) \le n\mathsf{E}_\mathsf{P}[\|\boldsymbol{\zeta}\|^2 \mid \mathcal{F}] \le C_\mathbf{g}^2 d \quad (\mathsf{P}\text{-a.s.}).$$

Therefore, using Lemmas C.4 and C.2 and setting

$$\Gamma \equiv n^{-3/2}\sum_{i=1}^N \mathsf{E}_\mathsf{P}\Big[\|(\mathbf{Y}_i - \mathsf{E}[\mathbf{Y}_i \mid \mathcal{F}])\otimes \mathbf{g}_i(X_i)\|_3^3 \mid \mathcal{F}\Big],$$

we find that

$$\begin{aligned}\mathsf{E}_\mathsf{P} d_K^\epsilon\Big(\widetilde{T}', S' \mid \mathcal{F}\Big) &\le \frac{C_d d^{3/4}}{\lambda^{3/8}}\mathsf{E}_\mathsf{P}\Gamma^{1/4} + \frac{C_\mathbf{g}\sqrt{d}}{\epsilon^4}\mathsf{E}_\mathsf{P}\Gamma + \mathsf{P}(\underline{\lambda}(V) < \lambda)\\ &\le \frac{C_1}{(\lambda^3 n)^{1/8}} + \frac{C_2}{\epsilon^4\sqrt{n}} + \mathsf{P}(\underline{\lambda}(V) < \lambda),\end{aligned}$$

where $C_d > 0$ is a constant depending only on d and the constants $C_1, C_2 > 0$ subsume factors that depend on d and $\mathbf{g}$. Similarly, since $\mathbf{Y}_i, \ldots, \mathbf{Y}_n$ are conditionally independent given $\mathcal{F}$,

$$\begin{aligned}\beta \equiv \mathsf{E}_\mathsf{P}[\|W - \mathsf{E}_\mathsf{P}[\boldsymbol{\zeta}^* \mid \mathcal{G}]\|_{e,\infty} \mid \mathcal{F}] &\le \mathsf{E}_\mathsf{P}[\|W - \mathsf{E}_\mathsf{P}[\boldsymbol{\zeta}^* \mid \mathcal{G}]\|_{e,1} \mid \mathcal{F}]\\ &\le n^{-1/2}d^2(C_\mathbf{g}^2 + 2C_\mathbf{g}) \quad (\mathsf{P}\text{-a.s.}),\end{aligned}$$

and by Lemmas C.5, C.2, and the conditional Jensen's inequality,

$$\begin{aligned}\upsilon\mathsf{P}(\Delta > \upsilon) \le \mathsf{E}_\mathsf{P}\Delta &\le \frac{C_d' d^{2/3}}{\lambda^{1/3}}\mathsf{E}_\mathsf{P}\beta^{1/3} + \frac{C_\mathbf{g}\sqrt{d}}{\epsilon^3}\mathsf{E}_\mathsf{P}\beta + \mathsf{P}(\underline{\lambda}(V) < \lambda)\\ &\le \frac{C_1'}{(\lambda^2 n)^{1/6}} + \frac{C_2'}{\epsilon^3\sqrt{n}} + \mathsf{P}(\underline{\lambda}(V) < \lambda).\end{aligned}$$

where $C_d' > 0$ is a constant depending only on d and the constants $C_1', C_2' > 0$ subsume

factors depending on d and $\mathbf{g}$. $\square$

Claim C.2. *For any* $\mathsf{P} \in \mathcal{P}_0$,

$$\mathsf{P}(\boldsymbol{\mu}_L \notin R_L) + \mathsf{P}(\boldsymbol{\mu}_U \notin R_U) - \kappa \lesssim h_{3,n} + \mathsf{P}(\underline{\lambda}(V) < \lambda),$$

where

$$h_{3,n} \equiv \frac{\ln^{7/8}(n/\sqrt{\lambda \wedge 1})}{(\lambda^3 n)^{1/8}}.$$

Proof. We reuse the notation from the proof of Claim C.1. In addition, for $x \in \mathbf{R}^d$ let $M(x) := \max_{1 \le j \le d}\{x_j\}$. Since $q^*(1-\kappa/2)$ is $\mathcal{F}$-measurable,

$$\begin{aligned}\mathsf{P}(\boldsymbol{\mu}_L \notin R_L \mid \mathcal{F}) &= \mathsf{P}\big(M(\sqrt{n}\boldsymbol{\zeta}) > q^*(1-\kappa/2) \mid \mathcal{F}\big) \\ &\le \mathsf{P}\Big(M(V^{1/2}Z) > q^*(1-\kappa/2) \mid \mathcal{F}\Big) \\ &\quad + d_K\Big(M(\sqrt{n}\boldsymbol{\zeta}), M(V^{1/2}Z) \mid \mathcal{F}\Big) \quad (\mathsf{P}\text{-a.s.}).\end{aligned}$$

By construction, $\mathsf{P}\big(M(V^{1/2}Z) > q^*(1-\kappa/2) \mid \mathcal{F}\big) \le \kappa/2$ (P-a.s.) on $\{\underline{\lambda}(V) \ge \lambda\}$, and by Corollary D.1 and Remark D.1,

$$d_K\Big(M(\sqrt{n}\boldsymbol{\zeta}), M(V^{1/2}Z) \mid \mathcal{F}\Big) \le C(\ln(dn\tau))^{7/8}(\tau^3\Gamma)^{1/4} quad(\mathsf{P}\text{-a.s.})$$

on $\{\underline{\lambda}(V) \ge \lambda\}$, where $C > 0$ is a constant, $\tau \equiv C_{\mathbf{g}}\sqrt{d/\lambda}$, and

$$\Gamma \equiv n^{-3/2}\sum_{i=1}^{n} \mathsf{E}_{\mathsf{P}}\Big[\|(\mathbf{Y}_i - \mathsf{E}_{\mathsf{P}}[\mathbf{Y}_i \mid \mathcal{F}]) \otimes \mathbf{g}_i(X_i)\|_2^3 \mid \mathcal{F}\Big].$$

Therefore,

$$\mathsf{P}(\boldsymbol{\mu}_L \notin R_L) \le \frac{\kappa}{2} + C'\frac{\ln^{7/8}(n/\sqrt{\lambda \wedge 1})}{(\lambda^3 n)^{1/8}} + 2\mathsf{P}(\underline{\lambda}(V) < \lambda),$$

where the constant $C' > 0$ subsumes all factors that depend on d and $\mathbf{g}$. Finally, the same bound holds for $\mathsf{P}(\boldsymbol{\mu}_U \notin R_U)$. $\square$

Combining Claims C.1, C.2, and Proposition C.2, in view of Equation (C.5), we find that for any $\upsilon \in (0, \gamma)$,

$$\limsup_{n\to\infty} \sup_{\mathsf{P}\in\mathcal{P}_0} \mathsf{P}(T > c^*(\gamma) \vee \epsilon) \le \alpha + \upsilon.$$

Since υ is arbitrary, the result follows. $\blacksquare$

C.3 Auxiliary results

In this section let $(\Omega, \mathcal{H}, \mathsf{P})$ denote the underlying probability space. First, we provide a McDiarmid-type bound for multivariate functions under the conditional independence assumption. Recall that the maximal variation of a function $f : \mathcal{X}^d \to \mathbf{R}$, $d \geq 1$, at the i-th coordinate is given by

$$V_i(f) = \sup_{x \in \mathcal{X}^d, x' \in \mathcal{X}} \big| f(x) - f(x_1, \ldots, x_{i-1}, x', x_{i+1}, \ldots, x_d) \big|.$$

Lemma C.1. *Let X be a random vector in $\mathbf{R}^d$ such that $X_1, \ldots, X_d$ are conditionally independent given $\mathcal{F} \subset \mathcal{H}$ and let Y be an $\mathcal{F}$-measurable random element taking values in $(E, \mathcal{E})$. Consider a measurable map $[f_1, \ldots, f_m]^\top : \mathbf{R}^d \times E \to \mathbf{R}^m$ such that $\mathsf{E}|f_l(X, Y)| < \infty$ and let $c_{il} \equiv V_i(x \mapsto f_l(x, Y))$. Then for any $\epsilon \in \mathbf{R}^m_{>0}$,*

$$\mathsf{P}\left(\bigcup_{l=1}^m \{ f_l(X, Y) - \mathsf{E}[f_l(X, Y) \mid \mathcal{F}] \geq \epsilon_l \} \mid \mathcal{F} \right) \leq \sum_{l=1}^m \exp\left(- \frac{2\epsilon_l^2}{\sum_{i=1}^d c_{il}^{*2}} \right) \quad a.s.,$$

where c_{il}^* is the minimal measurable majorant of c_{il}.

Proof. First,

$$\begin{aligned} &\mathsf{P}\left(\bigcup_{l=1}^m \{ f_l(X, Y) - \mathsf{E}[f_l(X, Y) \mid \mathcal{F}] \geq \epsilon_l \} \mid \mathcal{F} \right) \\ &\qquad \leq \sum_{l=1}^m \mathsf{P}(f_l(X, Y) - \mathsf{E}[f_l(X, Y) \mid \mathcal{F}] \geq \epsilon_l \mid \mathcal{F}) \quad \text{a.s.} \end{aligned}$$

Since $X_1, \ldots, X_d$ are conditionally indpendent given $\mathcal{F}$ and Y is $\mathcal{F}$-measurable,

$$\mathsf{P}(f_l(X, Y) - \mathsf{E}[f_l(X, Y) \mid \mathcal{F}] \geq \epsilon_l \mid \mathcal{F}) \leq \exp\left(- \frac{2\epsilon_l^2}{\sum_{i=1}^d c_{il}^{*2}} \right) \quad \text{a.s.}$$

by repeating the argument in the proof of Lemma (1.2) in McDiarmid (1989). ■

Next, we establish a number of results regarding Gaussian random vectors in $\mathbf{R}^d$ and their transformation $\mathcal{T} : \mathbf{R}^d \times \mathbf{R}^d \to \mathbf{R}$ given by

$$\mathcal{T}(x, y) := \|[x]_- + [y]_+\|_1.$$

Consider $X \sim \mathcal{N}(0, \Sigma)$, where Σ is a $d \times d$ positive definite covariance matrix. For $i \geq 1$ the marginal distribution of $(X_1, \ldots, X_i)^\top$ is $\mathcal{N}\big(0, \Sigma^{(i)}\big)$, where $\Sigma^{(i)}$ is a block of Σ corresponding to the first i rows and columns of the latter, and for $i > 1$ the conditional distribution of X_i given $X_1, \ldots, X_{i-1}$ is also normal with variance given by the Schur complement $\Sigma^{(i)}/\Sigma^{(i-1)}$.

Let Π denote the set of permutations of $\{1, \ldots, d\}$. We define

$$\psi(\Sigma) := \min_{\pi \in \Pi} \left\{ [\Sigma_{\pi,11}]^{-1/2} + \sum_{i=2}^{d} [\Sigma_\pi^{(i)} / \Sigma_\pi^{(i-1)}]^{-1/2} \right\},$$

where $\Sigma_\pi = P_\pi \Sigma P_\pi$, $\pi \in \Pi$, is the variance of $(X_{\pi(1)}, \ldots, X_{\pi(d)})^\top$ (P_π denotes the permutation matrix corresponding to π). When $d = 1$, we set $\psi(\Sigma) = \Sigma^{-1/2}$.

Lemma C.2. *Let λ_{min} and λ_{max} be the smallest and the largest eigenvalues of Σ. Then*

$$\frac{1}{\sqrt{\lambda_{max}}} \left(1 + \frac{d-1}{\sqrt{1+\kappa^2}} \right) \leq \psi(\Sigma) \leq \frac{d}{\sqrt{\lambda_{min}}},$$

where $\kappa = \lambda_{max}/\lambda_{min}$ is the condition number of Σ.

Proof. Fix $\pi \in \Pi$ and let

$$\psi_\pi(\Sigma) := [\Sigma_{\pi,11}]^{-1/2} + \sum_{i=2}^{d} [\Sigma_\pi^{(i)} / \Sigma_\pi^{(i-1)}]^{-1/2}. \tag{C.9}$$

In addition, let $\lambda_{\pi,\min}^{(i)}$ and $\lambda_{\pi,\max}^{(i)}$ denote the smallest and the largest eigenvalues of $\Sigma_\pi^{(i)}$. Notice that by the properties of the Rayleigh quotient $\lambda_{\min} \leq \lambda_{\pi,\min}^{(i)}$ and $\lambda_{\max} \geq \lambda_{\pi,\max}^{(i)}$.

For $i > 1$ consider the Schur complement $\Sigma_\pi^{(i)} / \Sigma_\pi^{(i-1)}$, i.e.,

$$\sigma_i^2 \equiv \Sigma_\pi^{(i)} / \Sigma_\pi^{(i-1)} = \Sigma_{\pi,ii}^{(i)} - v_i^\top [\Sigma_\pi^{(i-1)}]^{-1} v_i,$$

where v_i is the i-th column of $\Sigma_\pi^{(i)}$ without the last element, and let

$$A^{(i)} := \begin{bmatrix} \Sigma_\pi^{(i-1)} & 0 \\ 0 & \sigma_i^2 \end{bmatrix} \quad \text{and} \quad B^{(i)} := \begin{bmatrix} I & -[\Sigma_\pi^{(i-1)}]^{-1} v_i \\ 0 & 1 \end{bmatrix}.$$

Then

$$\sigma_i^2 = e_i^\top A^{(i)} e_i = (B^{(i)} e_i)^\top \Sigma_\pi^{(i)} (B^{(i)} e_i) \geq \lambda_{\pi,\min}^{(i)} \left\| B^{(i)} e_i \right\|^2 \geq \lambda_{\min}.$$

Moreover, $\Sigma_{\pi,11} = e_1^\top \Sigma_\pi e_1 \geq \lambda_{\min}$. Combining these inequalities, we get

$$\psi_\pi(\Sigma) \geq \frac{d}{\sqrt{\lambda_{\min}}}. \tag{C.10}$$

Similarly, $\Sigma_{\pi,11} \leq \lambda_{\max}$, and since $\|[\Sigma_\pi^{(i-1)}]^{-1}\| \leq \lambda_{\min}^{-1}$ and $\|v_i\| \leq \lambda_{\max}$,

$$\sigma_i^2 \leq \lambda_{\pi,\max}^{(i)} \left\| B^{(i)} e_i \right\|^2 \leq \lambda_{\max} \left(1 + \left\| [\Sigma_\pi^{(i-1)}]^{-1} v_i \right\|^2 \right) \leq \lambda_{\max}(1 + \kappa^2).$$

Therefore,

$$\psi_\pi(\Sigma) \le \frac{1}{\sqrt{\lambda_{\max}}}\left(1 + \frac{d-1}{\sqrt{1+\kappa^2}}\right). \tag{C.11}$$

The result follows by noticing that the bounds (C.10) and (C.11) do not depend on π. ■

Lemma C.3. *For any $\epsilon > 0$ and $a, b \in \mathbf{R}^d_{\ge 0}$,*

$$\sup_{r \ge 0} \mathsf{P}(r < \mathcal{T}(X+a, X-b) \le r + \epsilon) \le \psi(\Sigma)\epsilon.$$

Proof. For a given $\pi \in \Pi$ let $Y_i = [X_{\pi(i)} + a_{\pi(i)}]_- + [X_{\pi(i)} - b_{\pi(i)}]_+$, $W_0 = 0$, and $W_i = W_{i-1} + Y_i$, $1 \le i \le d$. Then since

$$\begin{aligned}\mathsf{P}(r < W_i \le r+\epsilon) \le \mathsf{P}(r < Y_i + W_{i-1} \le r+\epsilon, W_{i-1} \le t)\\ + \mathsf{P}(r < W_{i-1} \le r+\epsilon),\end{aligned}$$

we find that

$$\mathsf{P}(r < W_d \le r+\epsilon) \le \sum_{i=1}^{d} \mathsf{P}(r < Y_i + W_{i-1} \le r+\epsilon, W_{i-1} \le r).$$

The conditional distribution of X_i given $Z_{i-1} \equiv (X_{\pi(1)}, \ldots, X_{\pi(i-1)})^\top$ is normal, and its conditional variance is independent of the realization of Z_{i-1}. Thus, for $i > 1$ and $r \ge 0$,

$$\begin{aligned}&\mathsf{P}(r < Y_i + W_{i-1} \le r+\epsilon, W_{i-1} \le r)\\ &\quad = \mathsf{E}[\mathsf{P}(r < Y_i + W_{i-1} \le r+\epsilon \mid Z_{i-1})\mathbf{1}\{W_{i-1} \le r\}]\\ &\quad \le \mathsf{E}[\mathrm{Var}(X_{\pi(i)} \mid Z_{i-1})]^{-1/2}\epsilon \le [\Sigma_\pi^{(i)}/\Sigma_\pi^{(i-1)}]^{-1/2}\epsilon.\end{aligned}$$

In addition, $\mathsf{P}(r < Y_1 \le r+\epsilon) \le [\mathrm{Var}(X_{\pi(1)})]^{-1/2}\epsilon$. Therefore,

$$\mathsf{P}(r < W_d \le r+\epsilon) \le \psi_\pi(\Sigma)\epsilon,$$

where $\psi_\pi(\,\cdot\,)$ is given in (C.9). Since the probability on the RHS of the last inequality is independent of π,

$$\sup_{r \ge 0} \mathsf{P}(r < \mathcal{T}(X+a, X-b) \le r+\epsilon) \le \min_{\pi \in \Pi} \psi_\pi(\Sigma)\epsilon. \qquad ■$$

Remark C.1. In the preceding result, the distribution of $\mathcal{T}(X+a, X-b)$ has an atom at 0 when $(a+b) \in \mathbf{R}^d_{>0}$. Therefore, the uniform bound depending on ϵ can be established only over the non-negative reals.

The next results establish bounds on the conditional Kolmogorov distance between $\mathcal{T}$-transformations of certain random vectors. For random variables X and Y and sub-σ-fields $\mathcal{F}, \mathcal{G} \subset \mathcal{H}$ we define

$$d_K^z(X, Y \mid \mathcal{G}, \mathcal{F}) := \sup_{x \geq z} \big| F_X^{\mathcal{F}}(\,\cdot\,, x) - F_Y^{\mathcal{F}}(\,\cdot\,, x) \big|, \tag{C.12}$$

where $F_X^{\mathcal{F}}(\,\cdot\,, x)$ and $F_Y^{\mathcal{F}}(\,\cdot\,, x)$ are the conditional cdfs of X and Y given $\mathcal{F}$ and $\mathcal{G}$, respectively (when $\mathcal{F} = \mathcal{G}$ we denote this measure by $d_K^z(X, Y \mid \mathcal{F})$; also we drop the superscript z when the supremum is taken over $\mathbf{R}$). Since the function $\mathcal{T}(\,\cdot\,, \cdot\,)$ is not differentiable we use its smooth approximation $\widetilde{\mathcal{T}}_\kappa : \mathbf{R}^d \times \mathbf{R}^d \to \mathbf{R}$, $\kappa > 0$ defined by

$$\widetilde{\mathcal{T}}_\kappa(x, y) := \|\boldsymbol{\varphi}_\kappa(-x) + \boldsymbol{\varphi}_\kappa(y)\|_1,$$

where $\boldsymbol{\varphi}_\kappa : \mathbf{R}^d \to \mathbf{R}^d$ is a function of the form $\boldsymbol{\varphi}_\kappa(x) = [\varphi_\kappa(x_1), \ldots, \varphi_\kappa(x_d)]^\top$ with $\varphi_\kappa(x) = \kappa^{-1} \ln(e^{\kappa x} + 1)$. Note that $0 \leq \varphi_\kappa(x) - (x \vee 0) \leq \kappa^{-1} \ln(2)$ for all $x \in \mathbf{R}$.

Lemma C.4. *Let $X_1, \ldots, X_n$ be random vectors in $\mathbf{R}^d$ that are conditionally independent given $\mathcal{F} \subset \mathcal{H}$ with $\mathsf{E}[X_i \mid \mathcal{F}] = 0$ and $\mathsf{E}[\|X_i\|_3^3 \mid \mathcal{F}] < \infty$ a.s. Let $S := \sum_{i=1}^n X_i$ and let N be a random vector in $\mathbf{R}^d$ s.t. $N \mid \mathcal{F} \sim \mathcal{N}(0, V)$, where $V = \mathsf{E}[SS^\top \mid \mathcal{F}]$ a.s. Then, assuming that V is a.s. positive definite, for any $\epsilon > 0$ and $\mathcal{F}$-measurable random vectors $a, b \in \mathbf{R}_{\geq 0}^d$,*

$$\begin{aligned} &d_K^\epsilon(\mathcal{T}(S + a, S - b), \mathcal{T}(N + a, N - b) \mid \mathcal{F}) \\ &\qquad \leq C_d \Gamma^{1/4} [\psi(V)]^{3/4} \quad \text{a.s. on } \{\delta^* \leq \epsilon^4\}, \end{aligned}$$

where $\Gamma := \sum_{i=1}^n \mathsf{E}[\|X_i\|_3^3 \mid \mathcal{F}]$, $\delta^* := \Gamma / \psi(V)$, and $C_d > 0$ is a constant depending only on d.

Proof. Let f be a trice continuously differential function, s.t. $f(x) = 1$ if $x \leq 0$, $f = 0$ if $x \geq \delta > 0$, and $\big| f^{(j)}(x) \big| \leq D \delta^{-j} \mathbf{1}_{(0,\delta)}(x)$ for some constant $D > 0$ and $1 \leq j \leq 3$. Further, for $\kappa > 0$ set

$$g_r(s) := f(\widetilde{\mathcal{T}}_\kappa(s + a, s - b) - r).$$

First, letting $\nu := 2 \ln(2) d \kappa^{-1}$, we find that

$$\begin{aligned} \mathsf{P}(\mathcal{T}(S + a, S - b) \leq r \mid \mathcal{F}) &\leq \mathsf{P}\Big(\widetilde{\mathcal{T}}_\kappa(S + a, S - b) \leq r + \nu \mid \mathcal{F}\Big) \\ &\leq \mathsf{E}[g_{r+\nu}(S) \mid \mathcal{F}] \\ &\leq \mathsf{P}(\mathcal{T}(N + a, N - b) \leq t + \delta + \nu \mid \mathcal{F}) + \mathsf{E}[g_{r+\nu}(S) - g_{r+\nu}(N) \mid \mathcal{F}] \end{aligned}$$

and

$$\begin{aligned} \mathsf{P}(\mathcal{T}(S + a, S - b) > r \mid \mathcal{F}) &\leq \mathsf{P}\Big(\widetilde{\mathcal{T}}_\kappa(S + a, S - b) > r \mid \mathcal{F}\Big) \\ &\leq 1 - \mathsf{E}[g_{r-\delta}(S) \mid \mathcal{F}] \\ &\leq \mathsf{P}(\mathcal{T}(N + a, N - b) > r - \delta - \nu \mid \mathcal{F}) + \mathsf{E}[g_{r-\delta}(S) - g_{r-\delta}(N) \mid \mathcal{F}] \end{aligned}$$

a.s. for all $r \geq 0$. Hence, for $0 < \delta + \nu \leq \epsilon$ w.p.1,

$$
\begin{aligned}
&d_K^{\epsilon}(\mathcal{T}(S+a, S-b), \mathcal{T}(N+a, N-b) \mid \mathcal{F}) \\
&\quad \leq \sup_{q \in \mathbb{Q}_{\geq 0}} |\mathsf{E}[g_q(S) - g_q(N) \mid \mathcal{F}]| \\
&\qquad + \sup_{q \in \mathbb{Q}_{\geq 0}} \mathsf{P}(q < \mathcal{T}(N+a, N-b) \leq q + \delta + \nu \mid \mathcal{F}).
\end{aligned} \tag{C.13}
$$

Consider the first term on the RHS of (C.13).

Claim C.3. *There is a constant $B_d > 0$ depending only on d s.t. for any $q \geq 0$,*

$$
|\mathsf{E}[g_q(S) - g_q(N) \mid \mathcal{F}]| \leq B_d \left(\frac{1}{\delta^3} + \frac{\kappa}{\delta^2} + \frac{\kappa^2}{\delta} \right) \Gamma \quad \textit{a.s.}
$$

Proof. Let $Z_1, \ldots, Z_n$ be i.i.d. standard normal random vectors in $\mathbf{R}^d$ independent of $X_1, \ldots, X_n$ and $\mathcal{F}$, and let $Y_i := V_i^{1/2} Z_i$, where V_i is a version of $\mathsf{E}[X_i X_i^\top \mid \mathcal{F}]$. Define

$$
U_i := \sum_{k=1}^{i-1} X_k + \sum_{k=i+1}^{n} Y_k
$$

and

$$
W_i := g_q(U_i + X_i) - g_q(U_i + Y_i).
$$

Then $g_q(S) - g_q(N) = \sum_{i=1}^n W_i$ and

$$
|\mathsf{E}[g_q(S) - g_q(N) \mid \mathcal{F}]| \leq \sum_{i=1}^{n} |\mathsf{E}[W_i \mid \mathcal{F}]| \quad \text{a.s.}
$$

Let $\mathcal{G}_i := \mathcal{F} \vee \sigma(X_1, \ldots, X_{i-1}, Z_{i+1}, \ldots, Z_n)$ and let $h_{i1}(\lambda) := g_q(U_i + \lambda X_i)$ and $h_{i2}(\lambda) := g_q(U_i + \lambda Y_i)$. Using Taylor expansion up to the third order, we find that

$$
W_i = \sum_{j=0}^{2} \frac{1}{j!} \Big(h_{i1}^{(j)}(0) - h_{i2}^{(j)}(0) \Big) + \frac{1}{3!} \Big(h_{i1}^{(3)}(\lambda_1) - h_{i2}^{(3)}(\lambda_2) \Big),
$$

where $|\lambda_1|, |\lambda_2| < 1$. Then, since U_i is $\mathcal{G}_i$-measurable,

$$
\mathsf{E}[\mathsf{E}[h_{i1}^{(j)}(0) - h_{i2}^{(j)}(0) \mid \mathcal{G}_i] \mid \mathcal{F}] = 0 \quad \text{a.s.}
$$

for $j \leq 2$. Also since $|\varphi_\kappa^{(j)}(x)| \leq \kappa^{j-1}$, $1 \leq j \leq 3$, we get

$$
\begin{aligned}
|h_{i1}^{(3)}(\lambda_1) - h_{i2}^{(3)}(\lambda_2)| \leq B \Big(&\left\| f^{(3)} \right\|_\infty \Big(\|X_i\|_1^3 + \|Y_i\|_1^3 \Big) \\
&+ \left\| f'' \right\|_\infty \kappa \Big(\|X_i\|_1 \|X_i\|_2^2 + \|Y_i\|_1 \|Y_i\|_2^2 \Big)
\end{aligned}
$$

$$+ \|f'\|_\infty \kappa^2 \Big(\|X_i\|_3^3 + \|Y_i\|_3^3 \Big) \Big),$$

where $B > 0$ is a constant. Finally, since $\mathsf{E}[\|Y_i\|_3^3 \mid \mathcal{F}] \le 2\sqrt{2/\pi}\mathsf{E}[\|X_i\|_3^3 \mid \mathcal{F}]$ a.s.,

$$\begin{aligned} |\mathsf{E}[h_{i1}^{(3)}(\lambda_1) - h_{i2}^{(3)}(\lambda_2) \mid \mathcal{F}]| &\le \mathsf{E}[|h_{i1}^{(3)}(\lambda_1) - h_{i2}^{(3)}(\lambda_2)| \mid \mathcal{F}] \\ &\le B_d \bigg(\frac{1}{\delta^3} + \frac{\kappa}{\delta^2} + \frac{\kappa^2}{\delta} \bigg) \mathsf{E}[\|X_i\|_3^3 \mid \mathcal{F}] \quad \text{a.s.} \end{aligned}$$

□

Using Lemma C.3 it follows that

$$\begin{aligned} &d_K^\epsilon(\mathcal{T}(S+a, S-b), \mathcal{T}(N+a, N-b) \mid \mathcal{F}) \\ &\quad \le B_d \bigg(\frac{1}{\delta^3} + \frac{\kappa}{\delta^2} + \frac{\kappa^2}{\delta} \bigg) \Gamma + \psi(V)(\delta + \nu) \quad \text{a.s.} \end{aligned} \tag{C.14}$$

We set $\nu = \delta$. The since (C.14) holds for any δ a.s., it holds for random δ on $\{\delta \in (0, \epsilon/2)\}$. Consequently, the result follows by taking $\delta = (\delta^*)^{1/4}/2$ and noticing that $0 < \psi(V) < \infty$ a.s. by Lemma C.2. ■

Lemma C.5. *Suppose that $\mathcal{G}$ and $\mathcal{F}$ are σ-fields s.t. $\mathcal{F} \subset \mathcal{G} \subset \mathcal{H}$, X and Y are random vectors in $\mathbf{R}^d$ s.t. $X \mid \mathcal{G} \sim \mathcal{N}(0, \Sigma_X)$ and $Y \mid \mathcal{F} \sim \mathcal{N}(0, \Sigma_Y)$. Then, assuming that Σ_Y is a.s. positive definite, for any $\epsilon > 0$ and $\mathcal{F}$-measurable random vectors $a, b \in \mathbf{R}^d_{\ge 0}$,*

$$\begin{aligned} &d_K^\epsilon(\mathcal{T}(X+a, X-b), \mathcal{T}(Y+a, Y-b) \mid \mathcal{G}, \mathcal{F}) \\ &\quad \le C_d \|\Sigma_X - \Sigma_Y\|_{e,\infty}^{1/3} \psi(\Sigma_Y)^{2/3} \quad \textit{a.s. on } \{\delta^* \le \epsilon^3\}, \end{aligned} \tag{C.15}$$

where $\delta^* := \|\Sigma_X - \Sigma_Y\|_{e,\infty}/\psi(\Sigma_Y)$ and $C_d > 0$ is a constant depending only on d.

Proof. Let f be a twice continuously differential function s.t. $f(x) = 1$ if $x \le 0$, $f(x) = 0$ if $x \ge \delta > 0$ and $\left|f^{(j)}\right| \le D\delta^{-j}\mathbf{1}_{(0,\epsilon)}(x)$ for some constant $D > 0$ and $1 \le j \le 2$. Further, set

$$g_r(s) := f(\mathcal{T}(s+a, s-b) - r).$$

As in the proof of Lemma C.4 for any $0 < \delta \le \epsilon$ w.p.1,

$$\begin{aligned} &d_K^\epsilon(\mathcal{T}(X+a, X-b), \mathcal{T}(Y+a, Y-b) \mid \mathcal{G}, \mathcal{F}) \\ &\quad \le \sup_{q \in \mathbb{Q}_{\ge 0}} |\mathsf{E}[g_q(X) \mid \mathcal{G}] - \mathsf{E}[g_q(Y) \mid \mathcal{F}]| \\ &\qquad + \sup_{q \in \mathbb{Q}_{\ge 0}} \mathsf{P}(q < \mathcal{T}(N+a, N-b) \le q + \delta \mid \mathcal{F}). \end{aligned}$$

Let Z_1 and Z_2 be independent standard normal random vectors in $\mathbf{R}^d$ that are independent

of $\mathcal{G}$. Then

$$\begin{aligned}\mathsf{E}[g_q(X) \mid \mathcal{G}] - \mathsf{E}[g_q(Y) \mid \mathcal{F}] &= \mathsf{E}[g_q(\Sigma_X^{1/2} Z_1) \mid \mathcal{G}] - \mathsf{E}[g_q(\Sigma_Y^{1/2} Z_2) \mid \mathcal{F}] \\ &= h_{q,1}(\Sigma_X) - h_{q,2}(\Sigma_Y) \quad \text{a.s.},\end{aligned}$$

where $h_{q,1}(\sigma) \coloneqq \mathsf{E} g_q(\sigma^{1/2} Z_1)$ and $h_{q,2}(\sigma) \coloneqq \mathsf{E} g_q(\sigma^{1/2} Z_2)$ (the functions $h_{q,1}$ and $h_{q,2}$ implicitly depend on a and b; however, since they are $\mathcal{F}$-measurable we treat them as constants).

Claim C.4. *There exists a constant B_d depending only on d s.t. for any $q \geq 0$,*

$$|h_{q,1}(\sigma_X) - h_{q,2}(\sigma_Y)| \leq \frac{B_d}{\delta^2} \|\sigma_X - \sigma_Y\|_{e,\infty} \big(\bar{\Phi}(\underline{a}/s) + \bar{\Phi}(\underline{b}/s)\big),$$

where $s^2 \equiv \max_i\{[\sigma_X]_{ii} \vee [\sigma_Y]_{ii}\}$, $\underline{a} \equiv \min_i\{a_i\}$, $\underline{b} \equiv \min_i\{b_i\}$.

Proof. Let $\tilde{g}_q(x) \coloneqq f(\widetilde{\mathcal{T}}_\kappa(x+a, x-b) - q)$ with $\kappa > 0$ and let

$$\tilde{h}_{q,1}(\sigma) \coloneqq \mathsf{E} \tilde{g}_q(\sigma^{1/2} Z_1) \quad \text{and} \quad \tilde{h}_{q,2}(\sigma) \coloneqq \mathsf{E} \tilde{g}_q(\sigma^{1/2} Z_2).$$

For $t \in [0,1]$ define $Z(t) \coloneqq \sqrt{t} \sigma_X^{1/2} Z_1 + \sqrt{1-t} \sigma_Y^{1/2} Z_2$ and $\phi(t) \coloneqq \mathsf{E} \tilde{g}_q(Z(t))$. Then

$$\tilde{h}_{q,1}(\sigma_X) - \tilde{h}_{q,2}(\sigma_Y) = \phi(1) - \phi(0) = \int_0^1 \phi'(t) dt.$$

Using the integration by parts formula (see Equation A.17 in Talagrand, 2011, Section A.6) for $t \in (0,1)$,

$$\begin{aligned}\phi'(t) &= \frac{1}{2} \mathsf{E}\Big[\Big(\sigma_X^{1/2} Z_1/\sqrt{t} - \sigma_Y^{1/2} Z_2/\sqrt{1-t}\Big)^\top \nabla \tilde{g}_q(Z(t))\Big] \\ &= \frac{1}{2} \mathsf{E}\Big[\mathbf{i}^\top (\sigma_X - \sigma_Y) \circ \nabla^2 \tilde{g}_q(Z(t)) \mathbf{i}\Big],\end{aligned}$$

where $\mathbf{i}$ is the vector of ones. Therefore,

$$\left| \int_0^1 \phi'(t) dt \right| \leq \|\sigma_X - \sigma_Y\|_{e,\infty} \int_0^1 \mathsf{E}\Big|\mathbf{i}^\top \nabla^2 \tilde{g}_q(Z(t)) \mathbf{i}\Big| dt.$$

The (r,s)-th element of the Hessian of $\tilde{g}_q$ can be bounded by

$$|D_{r,s}(\tilde{g}_q)(Z(t))| \leq \big\|f''\big\|_\infty (\varphi'_\kappa(-Z_r(t) - a_r) + \varphi'_\kappa(Z_r(t) - b_r)).$$

Therefore, noticing that $\varphi'_\kappa(x) \leq e^{\kappa x} \wedge 1$ and $Z_i(t) \sim N(0, \sigma_i^2(t))$, where $\sigma_i^2(t) = t[\sigma_X]_{ii} + (1-t)[\sigma_Y]_{ii}$, we get

$$\mathsf{E}|D_{r,s}(h)(Z(t))| \leq \big\|f''\big\|_\infty \big(\bar{\Phi}(a_r/\sigma_r(t)) + \bar{\Phi}(b_r/\sigma_r(t))$$

$$+ \xi_\kappa(a_r, \sigma_r(t)) + \xi_\kappa(b_r, \sigma_r(t))\big),$$

where $\xi_\kappa(x,s) := e^{\kappa(\kappa s^2/2 - x)}\bar{\Phi}(\kappa s - x/s)$. Combining these inequalities we find that

$$\begin{aligned}|\tilde{h}_{q,1}(\sigma_X) - \tilde{h}_{q,2}(\sigma_Y)| \leq \frac{B_d}{\delta^2}\|\sigma_X - \sigma_Y\|_{e,\infty} \\ \times \big(\bar{\Phi}(\underline{a}/s) + \bar{\Phi}(\underline{b}/s) + R(\kappa)\big),\end{aligned}$$

where

$$R(\kappa) = d^{-1}\sum_{i=1}^{d}\int_0^1 (\xi_\kappa(a_i, \sigma_i(t)) + \xi_\kappa(b_i, \sigma_i(t)))dt.$$

The result follows by the bounded convergence theorem because $R(\kappa) \to 0$ as $\kappa \to \infty$ in which case $\varphi_\kappa(x) \to x \vee 0$. □

Using Lemma C.3 it follows that

$$\begin{aligned}&d_K^\epsilon(\mathcal{T}(X+a, X-b), \mathcal{T}(Y+a, Y-b) \mid \mathcal{G}, \mathcal{F}) \\ &\quad \leq \frac{B_d}{\delta^2}\|\Sigma_X - \Sigma_Y\|_{e,\infty} + \psi(\Sigma_Y)\delta \quad \text{a.s.}\end{aligned} \tag{C.16}$$

Finally, since (C.16) holds for any $0 < \delta \leq \epsilon$ a.s., it holds for random δ a.s. on $\{\delta \in (0, \epsilon]\}$. Consequently, the result follows by taking $\delta = (\delta^*)^{1/3}$ and noticing that (C.15) holds trivially on $\{\|\Sigma_X - \Sigma_Y\|_{e,\infty} = 0\}$, and $0 < \psi(\Sigma_Y) < \infty$ a.s. by Lemma C.2. ■

Appendix D

Berry-Esseen Type Bounds for Maxima of Vector-valued Martingales

D.1 Introduction

Let $(\Omega, \mathcal{F}, \mathsf{P})$ be a probability space and let $\{X_n\}_{n=1}^{\infty}$ be an $\mathbf{R}^d$-valued martingale difference sequence with $d > 1$ adapted to a filtration $\{\mathcal{F}_n\}_{n=1}^{\infty} \subset \mathcal{F}$, i.e., each X_n is $\mathcal{F}_n$-measurable and $\mathsf{E}[X_n \mid \mathcal{F}_{n-1}] = 0$ a.s. In addition, suppose that we are given a sub-σ-field $\mathcal{F}_0 \subset \mathcal{F}_1$, not necessarily trivial. In what follows, we assume that the random vectors $\{X_n\}$ have a.s. finite conditional third moment given $\mathcal{F}_0$, i.e., $\mathsf{E}[\|X_n\|_{\infty}^3 \mid \mathcal{F}_0] < \infty$ a.s.

Consider the normalized sum

$$S_n \coloneqq \frac{1}{B_n} \sum_{i=1}^{n} X_i,$$

where $B_n \in (0, \infty)$ a.s. is a $\mathcal{F}_0$-measuralbe normalizing coefficient, which we do not specify explicitly. The goal of this appendix is to establish a uniform distributional approximation of $\max_{1 \le j \le d} S_{n,j}$ conditionally on $\mathcal{F}_0$ by a suitably chosen Gaussian analog. A sequence of the coefficients $\{B_n\}$ is used to prevent the variance of S_n form tending to 0 or diverging to infinity as $n \to \infty$.

The main difficulty regarding such an approximation is that the function $M : \mathbf{R}^d \to \mathbf{R}$ defined by $M(x) = \max_{1 \le j \le d} x_j$ is non-differentiable. In order to overcome this issue we employ a smooth version of M used in Chernozhukov et al. (2013). Namely, for $\kappa > 0$, we consider a function $G_\kappa : \mathbf{R}^d \to \mathbf{R}$ of the form

$$G_\kappa(x) = \kappa^{-1} \ln \left(\sum_{j=1}^{d} e^{\kappa x_j} \right). \tag{D.1}$$

Note that for all $x \in \mathbf{R}^d$,

$$M(x) \le G_\kappa(x) \le M(x) + \kappa^{-1} \ln d. \tag{D.2}$$

Therefore, the parameter κ controls the degree of approximation.

As for the Gaussian approximation of $M(S_n)$ we consider a random vector T_n whose conditional distribution given $\mathcal{F}_0$ is $\mathcal{N}(0, V_n)$, where the covariance matrix V_n is a version of $\mathsf{E}[S_n S_n^\top \mid \mathcal{F}_0]$. Namely, the conditional characteristic function of T_n is given by

$$\mathsf{E}[e^{it^\top T_n} \mid \mathcal{F}_0] = \exp\left(-\frac{1}{2} t^\top V_n t\right) \quad \text{a.s.}$$

for all $t \in \mathbf{R}^d$. Then we establish a bound on the conditional Kolmogorov distance between $M(S_n)$ and $M(T_n)$ given $\mathcal{F}_0$. Recall that for random variables X and Y and sub-σ-field $\mathcal{G} \subset \mathcal{F}$ this distance is defined by

$$d_K(X, Y \mid \mathcal{G}) := \sup_{r \in \mathbf{R}} |F_X(\,\cdot\,, r) - F_Y(\,\cdot\,, r)|,$$

where F_X and F_Y are the regular conditional cdfs of X and Y given $\mathcal{G}$.

D.2 Preliminary results

The following lemma summarizes some useful properties of the derivatives of the "smooth-max" function G_κ defined in (D.1).

Lemma D.1. *For $v \in \mathbf{R}^d$ let $p(v) := \sum_{i=1}^d e^{\kappa v_i}$. Then*

$$\begin{aligned} G'_\kappa(v)(x) &= \frac{1}{p(v)} \sum_{1 \le i \le d} e^{\kappa v_i} x_i, \\ G''_\kappa(v)(x, y) &= \frac{\kappa}{[p(v)]^2} \sum_{1 \le i, j \le d} e^{\kappa v_i} b_{ij}(v) x_i y_j, \\ G'''_\kappa(v)(x, y, z) &= \frac{\kappa^2}{[p(v)]^3} \sum_{1 \le i, j, k \le d} e^{\kappa v_i} c_{ijk}(v) x_i y_j z_k, \end{aligned} \tag{D.3}$$

where

$$b_{ij}(v) = \begin{cases} p(v) - e^{\kappa v_i}, & i = j, \\ -e^{\kappa v_j}, & i \ne j, \end{cases}$$

and

$$c_{ijk}(v) = \begin{cases} b_{ik}(v)(p(v) - 2e^{\kappa v_j}), & i = j \text{ or } j = k, \\ b_{ij}(v)(p(v) - 2e^{\kappa v_k}), & i = k \ne j, \\ 2 b_{ij}(v) b_{ik}(v), & i \ne j \ne k \ne i. \end{cases}$$

Moreover,

$$\begin{aligned} \left|G'_\kappa(v)(x)\right| &\le \|x\|_\infty, \\ \left|G''_\kappa(v)(x,y)\right| &\le 2\kappa\|x\|_\infty\|y\|_\infty, \quad \textit{and} \\ \left|G'''_\kappa(v)(x,y,z)\right| &\le 6\kappa^2\|x\|_\infty\|y\|_\infty\|z\|_\infty \end{aligned} \tag{D.4}$$

Proof. Properties (D.3) follow from an application of the chain rule and (D.4) follows from the fact that

$$\sum_{i,j} e^{\kappa v_i}|b_{ij}(v)| \le 2[p(v)]^2, \quad \text{and} \quad \sum_{i,j,k} e^{\kappa v_i}|c_{ijk}(v)| \le 6[p(v)]^3. \qquad \blacksquare$$

For the next two results we define function $\phi : \mathbf{R}^2_{>0} \to \mathbf{R}$ and $\varphi_r : \mathbf{R}^2_{>0} \to \mathbf{R}$, where r is a real-valued parameter, given by

$$\begin{aligned} \phi(x,y) &= (x/y^2)\sqrt{1 \vee \ln y} \quad \text{and} \\ \varphi_r(x,y) &= (1+\ln(x/y))^{r/2}x^r. \end{aligned}$$

Lemma D.2 (Anti-concentration bound, Chernozhukov et al., 2015)**.** *Suppose that Y is a zero-mean Gaussian random vector in $\mathbf{R}^d$ with $\sigma_j^2 := \mathsf{E}Y_j^2 > 0$ for all $1 \le j \le d$. Let $\underline{\sigma} := \min_{1\le j\le d}\{\sigma_j\}$ and $\overline{\sigma} := \max_{1\le j\le d}\{\sigma_j\}$. Then for every $\epsilon > 0$ there exists a constant $C > 0$ such that*

$$\sup_{r\in\mathbf{R}} \mathsf{P}(r-\epsilon \le M(Y) \le r+\epsilon) \le C\epsilon\phi(\overline{\sigma},\underline{\sigma})\sqrt{1 \vee \ln(d/\epsilon)}. \tag{D.5}$$

Lemma D.3. *Let Y be a zero-mean normal vector in $\mathbf{R}^d$ with $d \ge 2$ and $\sigma_j^2 := \mathsf{E}Y_j^2 > 0$ for all $1 \le j \le d$. Let $\underline{\sigma} := \min_{1\le j\le d}\{\sigma_j\}$ and $\overline{\sigma} := \max_{1\le j\le d}\{\sigma_j\}$. Then for any $r \ge 2$,*

$$\mathsf{E}\|Y\|_\infty^r \le C_r\varphi_r(\overline{\sigma},\underline{\sigma})(\ln d)^{r/2}, \tag{D.6}$$

where $C_r > 0$ is a constant depending only on r.

Proof. Let $f : [a,\infty) \to \mathbf{R}$, $a \ge 0$ be a strictly increasing convex function. Using Jensen's inequality we have

$$\mathsf{E}\|Y\|_\infty^r \le \mathsf{E}[a \vee \|Y\|_\infty^r] \le f^{-1}(\mathsf{E}[f(a \vee \|Y\|_\infty^r)]).$$

First, for $r > 2$ consider $f(x) = \exp\left(c_r(x/a)^{2/r}\right)$, where $c_r := r/2 - 1$, which is convex on

$[a, \infty)$. Letting $a = \left(2\sqrt{c_r}\overline{\sigma}\right)^r$, we find that

$$\mathsf{E}[f(a \vee \|Y\|_\infty^r) \le e^{c_r} \sum_{j=1}^d \mathsf{E} \exp\Big(|Y_j|^2/(4\overline{\sigma}^2)\Big) \le \sqrt{2} e^{c_r} d \times \frac{\overline{\sigma}}{\underline{\sigma}},$$

and, noticing that $d \ge 2$,

$$\mathsf{E}\|Y\|_\infty^r \le \left[\ln\left(\sqrt{2} e^{c_r} d \times \frac{\overline{\sigma}}{\underline{\sigma}}\right)\right]^{r/2} (2\overline{\sigma})^r \le C_r \varphi_r(\overline{\sigma}, \underline{\sigma})(\ln d)^{r/2} \tag{D.7}$$

for some $C_r > 0$ depending on r. For $r = 2$ the result follows by taking $f(x) = \exp(x/(2\overline{\sigma})^2)$ and $a = 0$. ■

Let $(D_1, \mathcal{D}_1), \ldots, (D_n, \mathcal{D}_n)$ be measurable spaces, $D = \prod_{j=1}^n D_j$ with the corresponding product σ-field $\bigotimes_{j=1}^n \mathcal{D}_j$, and $\pi_j : D \to D_j$ the coordinate maps, i.e., $\pi_j(d_1, \ldots, d_n) = d_j$. For a measurable function f on D_j, the composition $f \circ \pi_j$ is a measurable function from D to the codomain of f. In addition, let Σ_n denote a version of $\mathsf{E}[X_n X_n^\top \mid \mathcal{F}_0]$ and let

$$\underline{\sigma}_n^2 := \min_{1 \le j \le d} [\Sigma_n]_{jj} \quad \text{and} \quad \overline{\sigma}_n^2 := \max_{1 \le j \le d} [\Sigma_n]_{jj}.$$

The next result is a modification of Theorem 5 given in Rhee and Talagrand (1986). Specifically, we construct a new probability space and a sequence of sub-σ-fields (not necessary nested) which will replace the original probability space and the filtration $\{\mathcal{F}_n\}$ in the proof of the main result.

Lemma D.4. *Let $\Omega^* := \Omega \times \mathbf{R}^{d \times n}$ and let π_k, $1 \le k \le n+1$, be the k-th coordinate map on Ω^*. There exists a probability space $(\Omega^*, \mathcal{H}, \mathsf{Q})$, sub-$\sigma$-fields $\mathcal{H}_0, \ldots, \mathcal{H}_n$, $\mathbf{R}^d$-valued random vectors $X_1^*, \ldots, X_n^*$ and $Y_1^*, \ldots, Y_n^*$ defined on Ω^* such that the following properties hold:*

(a) *X_i^* and Y_i^* are F_k-measurable for $i < k$ and $1 < i \ne k$, respectively,*

$$\mathsf{E}[X_i^* \mid \mathcal{H}_i] = \mathsf{E}[Y_i^* \mid \mathcal{H}_i] = 0 \quad a.s., \quad \text{and} \tag{D.8}$$

$$\mathsf{E}[Y_i^* Y_i^{*\top} \mid \mathcal{H}_i] = \mathsf{E}[X_i^* X_i^{*\top} \mid \mathcal{H}_0] \quad a.s. \tag{D.9}$$

In addition, $\mathsf{E}[g(X_i) \mid \mathcal{F}_{(k-1)\vee 0}] \circ \pi_1 = \mathsf{E}[g(X_i^*) \mid \mathcal{H}_k]$ a.s. for any measurable function g such that $g(X_i) \in L^1$ and $i \ge k$.

(b) *Suppose that $\underline{\sigma}_i > 0$ a.s and $\overline{\sigma}_i < \infty$ a.s. Then for any $r \ge 2$,*

$$\mathsf{E}[\|Y_i^*\|_\infty^r \mid \mathcal{H}_0] \le C_r \varphi_r(\overline{\sigma}_i^*, \underline{\sigma}_i^*)(\ln d)^{r/2} \quad a.s., \tag{D.10}$$

where $\underline{\sigma}_n^* := \underline{\sigma}_n \circ \pi_1$, $\overline{\sigma}_n^* := \overline{\sigma}_n \circ \pi_1$, and $C_r > 0$ is a constant depending only on r.

(c) *Let* $S_n^* := B_n^{*-1}\sum_{i\le n} X_i^*$ *and* $T_n^* := B_n^{*-1}\sum_{i\le n} Y_i^*$*, where* $B_n^* := B_n \circ \pi_1$*. Then, given a measurable function* g *such that* $g(S_n) \in L^1$*,* $\mathsf{E}[g(S_n) \mid \mathcal{F}_0] \circ \pi_1 = \mathsf{E}[g(S_n^*) \mid \mathcal{H}_0]$ *a.s. Similarly, if* $g(T_n) \in L^1$*, then* $\mathsf{E}[g(T_n) \mid \mathcal{F}_0] \circ \pi_1 = \mathsf{E}[g(T_n^*) \mid \mathcal{H}_0]$ *a.s.*

Proof. First, the probability space $(\Omega^*, \mathcal{H}, \mathsf{Q})$ is constructed as follows. Let $\Omega^* = \Omega \times \mathbf{R}^{d\times n}$ and $\mathcal{H} = \mathcal{F} \otimes \mathcal{B}_{\mathbf{R}^{d\times n}}$. $\mathcal{H}_i, i = 1, \ldots, n$ are generated by the sets of the form $F \times \prod_{k=1}^n D_k$, where $F \in \mathcal{F}_{i-1}$, $D_k \in \mathcal{B}_{\mathbf{R}^d}$ for $k \neq i$ and $D_i = \mathbf{R}^d$. Finally, $\mathcal{H}_0$ is generated by the sets of the form $F \times \mathbf{R}^{d\times n}$ with $F \in \mathcal{F}_0$.

Letting μ_i be a zero-mean Gaussian probability kernel on $\Omega \times \mathbf{R}^d$ with variance Σ_i, we construct the product kernel $\mu = \bigotimes_{i=1}^n \mu_i$ and take Q to be the unique measure on $(\Omega^*, \mathcal{H})$ such that for every $F \in \mathcal{F}$ and $D \in \mathcal{B}_{\mathbf{R}^{d\times n}}$,

$$\mathsf{Q}(F \times D) = \int_F \mu(\omega, D)\mathsf{P}(d\omega).$$

Then for $1 \le i \le n$ we define

$$X_i^* := X_i \circ \pi_1 \quad \text{and} \quad Y_i^* := \pi_{i+1}.$$

Consequently, the implications of the lemma follow by construction and the properties of conditional expectation. Specifically, measurability of X_i^* w.r.t. $\mathcal{H}_k$ for $i < k$ in part (a) follows from the facts that X_i is $\mathcal{F}_i$-measurable and π_1 is a measurable function from Ω^* to Ω. Similarly, if $i \neq k$, $\{\pi_{i+1}^{-1}(D) : D \in \mathcal{B}_{\mathbf{R}^d}\} \subset \mathcal{H}_k$, which implies that Y_i^* is $\mathcal{H}_k$-measurable.

Let $A_F \in \mathcal{H}_0$ denote a set of the form $F \times \mathbf{R}^{d\times n}$ with $F \in \mathcal{F}_0$. Given a measurable function g, let $f_{1i} := \mathsf{E}[g(X_i) \mid \mathcal{F}_0] \circ \pi_1$. The function f_{1i} is F_0-measurable and $\forall F \in \mathcal{F}_0$,

$$\mathsf{E}[g(X_i^*); A_F] = \int_F g(X_i)\, d\,\mathsf{P} = \mathsf{E}[f_{1i}; A_F].$$

Hence, $f_{1i} = \mathsf{E}[g(X_i^*) \mid \mathcal{H}_0]$ a.s.

In order to prove (D.9) consider the function $f_{2i} := \mathsf{E}[X_i X_i^\top \mid \mathcal{F}_0] \circ \pi_1$, which is $\mathcal{H}_i$-measurable ($\because \mathcal{H}_0 \subset \mathcal{H}_i$). Then for $A = F \times \prod_{k=1}^n D_k$ with $F \in \mathcal{F}_{i-1}, D_k \in \mathcal{B}_{\mathbf{R}^d}$, $k \neq i$, and $D_i = \mathbf{R}^d$,

$$\mathsf{E}[Y_i^* Y_i^{*\top}; A] = \int_F \prod_{k\neq i} \mu_k(\omega, D_k) \int_{\mathbf{R}^d} yy^\top \mu_i(\omega, dy)\mathsf{P}(d\omega) = \mathsf{E}[f_{2i}; A].$$

On the other hand, $\forall F \in \mathcal{F}_0$,

$$\mathsf{E}[X_i^* X_i^{*\top}; A_F] = \int_F \mathsf{E}[X_i X_i^\top \mid \mathcal{F}_0]\, d\,\mathsf{P} = \mathsf{E}[f_{2i}; A_F] = \mathsf{E}[Y_i^* Y_i^{*\top}; A_F].$$

Other implications of part (a) follow from similar arguments. The inequality (D.10) follows

from Lemma D.3 by noticing that under the condition stated in part (b) $\forall F \in \mathcal{F}_0$,

$$\begin{aligned}\mathsf{E}[\|Y_i^*\|_\infty^r; A_F] &= \int_F \int_{\mathbf{R}^d} \|y\|_\infty^r \mu(\omega, dy)\mathsf{P}(d\omega) \\ &\le C_r(\ln d)^{r/2}\mathsf{E}[\varphi_r(\overline{\sigma}_i^*, \underline{\sigma}_i^*); A_F].\end{aligned}$$

As for part (c), note that for any $\omega \in \Omega$, the convolution $\mu_1 * \cdots * \mu_n(\omega, \cdot)$ is a Gaussian measure on $\mathbf{R}^d$ with covariance matrix $\sum_{i=1}^n \Sigma_i(\omega)$. Then, letting $f_3 := \mathsf{E}[g(T_n) \mid \mathcal{F}_0] \circ \pi_1$, $\forall F \in \mathcal{F}_0$,

$$\begin{aligned}\mathsf{E}[g(T_n^*); A_F] &= \int_F \int_{\mathbf{R}^{d\times n}} g\left(B_n^{-1}(\omega)\sum_{i=1}^n y_i\right)\mu(\omega, dy)\mathsf{P}(d\omega) \\ &= \int_F \mathsf{E}[g(T_n) \mid \mathcal{F}_0]\, d\,\mathsf{P} = \mathsf{E}[f_3; A_F],\end{aligned}$$

where we used the fact that the conditional covariance matrix of T_n given $\mathcal{F}_0$, $V_n = B_n^{-2}\sum_{i=1}^n \Sigma_i$ a.s. The second implication of part (c) follows from similar arguments. ■

D.3 Main results

In this section we derive a number of Berry-Esseen type bounds for the maximum of S_n. Let $\|\cdot\|_{e,p}$ denote the element-wise p-norm in $\mathbf{R}^{k\times l}$, i.e., for a $k \times l$ matrix A, $\|A\|_{e,p} = \|\mathrm{vec}(A)\|_p$, and let

$$\underline{v}_n^2 := \min_{1\le j\le d}[V_n]_{jj} \quad \text{and} \quad \overline{v}_n^2 := \max_{1\le j\le d}[V_n]_{jj}.$$

Lemma D.5. *Let $g : \mathbf{R} \to \mathbf{R}$ be a function in $C_b^3(\mathbf{R})$ and suppose that $\min_{i\le n}\underline{\sigma}_i^2 > 0$ a.s. and $\max_{i\le n}\overline{\sigma}_i^2 < \infty$ a.s. Then for all $\kappa > 0$ there exists a constant $C > 0$ such that*

$$\begin{aligned}&|\mathsf{E}[g(G_\kappa(S_n)) - g(G_\kappa(T_n)) \mid \mathcal{F}_0]| \\ &\quad \le C\Big[\big(\|g''\| + \kappa\|g'\|\big)\beta_n + \Big(\|g^{(3)}\| + \kappa\|g''\| + \kappa^2\|g'\|\Big)\Gamma_n\Big] \quad a.s.,\end{aligned} \tag{D.11}$$

where $\|\cdot\|$ denotes the sup norm,

$$\begin{aligned}\beta_n &:= B_n^{-2}\sum_{i=1}^n \mathsf{E}[\|\mathsf{E}[X_iX_i^\top \mid \mathcal{F}_{i-1}] - \Sigma_i\|_{e,1} \mid \mathcal{F}_0], \quad \textit{and} \\ \Gamma_n &:= B_n^{-3}\sum_{i=1}^n \Big(\mathsf{E}[\|X_i\|_\infty^3 \mid \mathcal{F}_0] + \varphi_3(\overline{\sigma}_i, \underline{\sigma}_i)(\ln d)^{3/2}\Big).\end{aligned} \tag{D.12}$$

Proof. Let $\hat{g} := g \circ G_\kappa$. Using Lemma D.4 we consider a probability space $(\Omega^*, \mathcal{H}, \mathsf{Q})$,

sub-σ-fields $\{\mathcal{H}_i\}_{i=0}^n \subset \mathcal{H}$, random vectors $\{X_i^*, Y_i^*\}_{i=1}^n$ such that

$$\mathsf{E}[\hat{g}(S_n) - \hat{g}(T_n) \mid \mathcal{F}_0] \circ \pi_1 = \mathsf{E}[\hat{g}(S_n^*) - \hat{g}(T_n^*) \mid \mathcal{H}_0] \quad \text{a.s.}$$

Setting $Z_i \coloneqq \hat{g}(U_i + X_i^*/B_n^*) - \hat{g}(U_i + Y_i^*/B_n^*)$ with

$$U_i \coloneqq B_n^{*-1}\left(\sum_{k=1}^{i-1} X_k^* + \sum_{k=i+1}^{n} Y_k^*\right),$$

where $B_n^* = B_n \circ \pi_1$, it follows that $\hat{g}(S_n^*) - \hat{g}(T_n^*) = \sum_{i=1}^n Z_i$ and, therefore,

$$|\mathsf{E}[\hat{g}(S_n^*) - \hat{g}(T_n^*) \mid \mathcal{H}_0]| \le \sum_{i=1}^n |\mathsf{E}[Z_i \mid \mathcal{H}_0]| \quad \text{a.s.}$$

Let $h_{i1}(\lambda) \coloneqq \hat{g}(U_i + \lambda X_i^*/B_n^*)$ and $h_{i2}(\lambda) \coloneqq \hat{g}(U_i + \lambda Y_i^*/B_n^*)$. Using Taylor expansion up to terms of the third order, we find that

$$Z_i = \sum_{j=0}^{2} \frac{1}{j!}\Big(h_{i1}^{(j)}(0) - h_{i2}^{(j)}(0)\Big) + \frac{1}{3!}\Big(h_{i1}^{(3)}(\lambda_1) - h_{i2}^{(3)}(\lambda_2)\Big),$$

where $|\lambda_1|, |\lambda_2| \le 1$. Noticing that U_i is $\mathcal{H}_i$-measurable, by Lemma D.1, equations (D.8) and (D.9), we have

$$\mathsf{E}\big[\mathsf{E}\big[h_{i1}'(0) - h_{i2}'(0) \mid \mathcal{H}_i\big] \mid \mathcal{H}_0\big] = 0 \quad \text{a.s.}$$

and

$$\begin{aligned}\big|\mathsf{E}[h_{i1}''(0) - h_{i2}''(0) \mid \mathcal{H}_0]\big| &\le \mathsf{E}\big[\big|\mathsf{E}\big[h_{i1}''(0) - h_{i2}''(0) \mid \mathcal{H}_i\big]\big| \mid \mathcal{H}_0\big] \\ &\le C_1\big(\|g''\| + \kappa\|g'\|\big)\Big(\mathsf{E}[\|\mathsf{E}[X_i^* X_i^{*\top} \mid \mathcal{H}_i] - \Sigma_i^*\|_{e,1} \mid \mathcal{H}_0]\Big)/B_n^{*2} \quad \text{a.s.},\end{aligned}$$

where Σ_i^* is a version of $\mathsf{E}[X_i^* X_i^{*\top} \mid \mathcal{H}_0]$. Using Lemma D.4(b),

$$\begin{aligned}|\mathsf{E}[h_{i1}^{(3)}(\lambda_1) - h_{i2}^{(3)}(\lambda_2) \mid \mathcal{H}_0]| &\le \mathsf{E}[|h_{i1}^{(3)}(\lambda_1) - h_{i2}^{(3)}(\lambda_2)| \mid \mathcal{H}_0] \\ &\le C_2\Big(\|g^{(3)}\| + \kappa\|g''\| + \kappa^2\|g'\|\Big) \\ &\quad \times \Big(\mathsf{E}[\|X_i^*\|_\infty^3 \mid \mathcal{H}_0] + \varphi_3(\overline{\sigma}_i^*, \underline{\sigma}_i^*)(\ln d)^{3/2}\Big)/B_n^{*3} \quad \text{a.s.}\end{aligned}$$

for some constants $C_1 > 0$ and $C_2 > 0$. Finally, (D.11) follows from Lemma D.4 and by noticing that

$$\begin{aligned}&\mathsf{E}[\|\mathsf{E}[X_i X_i^\top \mid \mathcal{F}_i] - \Sigma_i\|_{e,1} \mid \mathcal{F}_0] \circ \pi_1 \\ &\qquad = \mathsf{E}[\|\mathsf{E}[X_i^* X_i^{*\top} \mid \mathcal{H}_i] - \Sigma_i^*\|_{e,1} \mid \mathcal{H}_0] \quad \text{a.s.}\end{aligned}$$

■

In the case when $\underline{v}_n \in (0,\infty)$ a.s. we define

$$\tau_n := \overline{v}_n/\underline{v}_n, \quad \beta'_n := \beta_n/\underline{v}_n^2, \quad \text{and} \quad \Gamma'_n := \Gamma_n/\underline{v}_n^3,$$

where β_n and Γ_n are given in (D.12).

Theorem D.1. *Suppose that* $\min_{1\le i\le n}\{\underline{\sigma}_i\} \wedge \underline{v}_n > 0$ *a.s. and* $\max_{1\le i\le n}\{\overline{\sigma}_i\} \vee \overline{v}_n < \infty$ *a.s. Then there exists a constant* $C > 0$ *such that*

$$\begin{aligned} &d_K(M(S_n), M(T_n) \mid \mathcal{F}_0) \\ &\quad \le C\left[\beta'_n\sqrt{\tau_n/\Gamma'_n} + (\tau_n^3\Gamma'_n)^{1/4}\sqrt{(\ln d)(1\vee \ln(d[\tau_n/\Gamma'_n]^{1/4}))}\right] \quad \textit{a.s.} \end{aligned}$$

Proof. First, notice that the supremum in the definition of $d_K(\,\cdot\,,\,\cdot\mid\mathcal{F}_0)$ can be replaced by the supremum over rationals and for a positive, $\mathcal{F}_0$-measurable r.v. ξ,

$$\begin{aligned} &|\mathsf{P}(M(S_n)\le r\mid\mathcal{F}_0) - \mathsf{P}(M(T_n)\le r\mid\mathcal{F}_0)| \\ &\quad = |\mathsf{P}(M(\xi S_n)\le \xi r\mid\mathcal{F}_0) - \mathsf{P}(M(\xi T_n)\le \xi r\mid\mathcal{F}_0)| \\ &\quad \le d_K(M(\xi S_n), M(\xi T_n)\mid\mathcal{F}_0) \end{aligned}$$

a.s. for all $r\in\mathbf{R}$ (see, e.g., Kallenberg, 2002, Theorem 5.4). Therefore, we may assume that $\underline{v}_n = 1$ a.s. (i.e., we replace B_n with $\underline{v}_n B_n$).

Let f be a monotone, trice continuously differentiable function such that $f(x)=1$ if $x\le 0$, $f=0$ if $x\ge\epsilon>0$, and $|f^{(j)}(x)|\le D\epsilon^{-j}\mathbf{1}_{(0,\epsilon)}(x)$ for some constant $D>0$ and $1\le j\le 3$. Further, for $\kappa>0$ set

$$g_r(s) := f(G_\kappa(s)-r).$$

Using (D.2) and letting $\delta := \kappa^{-1}\ln d$, we find that

$$\begin{aligned} &\mathsf{P}(M(S_n)\le r\mid\mathcal{F}_0) \le \mathsf{P}(G_\kappa(S_n)\le r+\delta\mid\mathcal{F}_0) \\ &\quad \le \mathsf{E}[g_{r+\delta}(S_n)\mid\mathcal{F}_0] \\ &\quad \le \mathsf{P}(M(T_n)\le r+\delta+\varepsilon\mid\mathcal{F}_0) + \mathsf{E}[g_{r+\delta}(S_n)-g_{r+\delta}(T_n)\mid\mathcal{F}_0] \end{aligned}$$

and

$$\begin{aligned} &\mathsf{P}(M(S_n)> r\mid\mathcal{F}_0) \le \mathsf{P}(G_\kappa(S_n)> r\mid\mathcal{F}_0) \\ &\quad \le 1-\mathsf{E}[g_{r-\varepsilon}(S_n)\mid\mathcal{F}_0] \\ &\quad \le \mathsf{P}(M(T_n)> r-\delta-\varepsilon\mid\mathcal{F}_0) + \mathsf{E}[g_{r-\varepsilon}(T_n)-g_{r-\varepsilon}(S_n)\mid\mathcal{F}_0] \end{aligned}$$

a.s. for all $r \in \mathbf{R}$. Then w.p.1,

$$
\begin{aligned}
&d_K(M(S_n), M(T_n) \mid \mathcal{F}_0) \\
&\quad = \sup_{q\in\mathbb{Q}} |\mathsf{P}(M(S_n) \le q \mid \mathcal{F}_0) - \mathsf{P}(M(T_n) \le q \mid \mathcal{F}_0)| \\
&\quad \le \sup_{q\in\mathbb{Q}} |\mathsf{E}[g_q(S_n) - g_q(T_n) \mid \mathcal{F}_0]| + \sup_{q\in\mathbb{Q}} \mathsf{P}(q \le M(T_n) \le q + \delta + \varepsilon \mid \mathcal{F}_0).
\end{aligned} \tag{D.13}
$$

Consider the first term in the third line of (D.13). Using Lemma D.5 and the fact that $|f^{(j)}(x)| \le D\epsilon^{-j}\mathbf{1}_{(0,\epsilon)}(x)$ it follows that

$$
\begin{aligned}
&\sup_{q\in\mathbb{Q}} |\mathsf{E}[g_q(S_n) - g_q(T_n) \mid \mathcal{F}_0]| \\
&\qquad \le C_1 \left[\left(\frac{1}{\epsilon^2} + \frac{\kappa}{\epsilon} \right) \beta_n + \left(\frac{1}{\epsilon^3} + \frac{\kappa}{\epsilon^2} + \frac{\kappa^2}{\epsilon} \right) \Gamma_n \right] \quad \text{a.s.}
\end{aligned}
$$

As for the last term of (D.13), applying the anti-concentration bound (D.5), we get

$$
\begin{aligned}
&\sup_{q\in\mathbb{Q}} \mathsf{P}(q \le M(T_n) \le q + \delta + \epsilon \mid \mathcal{F}_0) \\
&\qquad \le \frac{C_2(\delta+\epsilon)}{2} \phi(\overline{v}_n, 1) \sqrt{1 \vee \ln\left(\frac{2d}{\delta+\epsilon} \right)} \quad \text{a.s.}
\end{aligned}
$$

Thus, setting $\delta = \epsilon$,

$$
\begin{aligned}
&d_K(M(S_n), M(T_n) \mid \mathcal{F}_0) \\
&\qquad \le 3C_1 \left(\frac{(\ln d)\beta_n}{\epsilon^2} + \frac{(\ln d)^2 \Gamma_n}{\epsilon^3} \right) + C_2 \epsilon \overline{v}_n \sqrt{1 \vee \ln\left(\frac{d}{\epsilon} \right)} \quad \text{a.s.}
\end{aligned} \tag{D.14}
$$

Since the last inequality holds for all $\epsilon > 0$ a.s., it also holds for random ϵ a.s. on $\{\epsilon \in (0, \infty)\}$. Consequently, the result follows by taking $\epsilon = (\ln d)^{1/2}(3\Gamma_n/\overline{v}_n)^{1/4}$. ■

The bound in the preceding theorem can be simplified if $B_n = \sqrt{n}$, which is useful when $\inf_{n\ge1} \underline{\sigma}_n > 0$ a.s. and $\sup_{n\ge1} \overline{\sigma}_n < \infty$ a.s. (e.g., the random vectors $\{X_n\}$ are i.i.d. conditionally on $\mathcal{F}_0$). In this case

$$
\Gamma_n \ge \left(\frac{\ln d}{n} \right)^{3/2} \sum_{i=1}^n \varphi_3(\overline{\sigma}_i, \underline{\sigma}_i) \ge \frac{(\ln d)^{3/2}}{\sqrt{n}} \times \overline{v}_n^3 \quad \text{a.s.},
$$

where we used the fact that $\{X_n\}$ is a martingale difference sequence and, therefore,

$$
V_n = \frac{1}{n} \sum_{i=1}^n \Sigma_i \quad \text{a.s.}
$$

Theorem D.2. *Suppose that the assumptions of Theorem D.1 hold and $B_n = \sqrt{n}$. Then there exists a constant $C > 0$ such that*

$$\begin{aligned} &d_K(M(S_n), M(T_n) \mid \mathcal{F}_0) \\ &\quad \le C\Big[(\ln d)^{1/4}\beta_n'\sqrt{\tau_n/\Gamma_n'} + (\ln(dn))^{7/8}(\tau_n^3\Gamma_n')^{1/4}\Big] \quad a.s. \end{aligned}$$

Proof. As in the proof of Theorem D.1 we may assume that $\underline{v_n} = 1$ a.s. Then the result follows by setting $\epsilon = (\ln d)^{3/8}(\Gamma_n/\overline{v}_n)^{1/4}$ in the inequality (D.14). ■

Moreover, if the conditional variances $\mathsf{E}[X_iX_i^\top \mid \mathcal{F}_{i-1}]$ are $\mathcal{F}_0$-measurable for all $i \le n$, then $\beta_n = 0$ a.s. and so we can state the following simple corollary.

Corollary D.1. *Suppose that the assumptions of Theorem D.1 hold and $\mathsf{E}[X_iX_i^\top \mid \mathcal{F}_{i-1}]$ is $\mathcal{F}_0$-measurable for all $1 \le i \le n$. Then there exists a constant $C > 0$ such that*

$$\begin{aligned} &d_K(M(S_n), M(T_n) \mid \mathcal{F}_0) \\ &\quad \le C(\tau_n^3\Gamma_n')^{1/4}\sqrt{(\ln d)(1 \vee \ln(d[\tau_n/\Gamma_n']^{1/4}))} \quad a.s. \end{aligned}$$

If, in addition, $B_n = \sqrt{n}$, then

$$d_K(M(S_n), M(T_n) \mid \mathcal{F}_0) \le C(\ln(dn))^{7/8}(\tau_n^3\Gamma_n')^{1/4} \quad a.s.$$

A trivial application of Corollary D.1 would be the case in which the random vectors $\{X_n\}$ are conditionally independent given $\mathcal{F}_0 \subset \mathcal{F}$ and $\mathcal{F}_n = \mathcal{F}_0 \vee \sigma(X_1, \ldots, X_n)$.

Remark D.1 (*Alternative Bounds*)**.** Since the maximum norm on $\mathbf{R}^d$ is bounded by the Euclidean norm, and for a Gaussian random vector Y and $r \ge 2$, $\mathsf{E}\|Y\|_2^r \le C_r(\mathsf{E}\|Y\|_2^2)^{r/2}$, where $C_r > 0$ is a constant depending on r (see, e.g., Rhee and Talagrand, 1986, Lemma 4), we may replace Γ_n in the results of this section with

$$\widetilde{\Gamma}_n := B_n^{-3}\sum_{i=1}^n \mathsf{E}[\|X_i\|_2^3 \mid \mathcal{F}_0].$$

In addition, when $B_n = \sqrt{n}$,

$$\widetilde{\Gamma}_n \ge \frac{1}{n^{3/2}}\sum_{i=1}^n\left(\sum_{j=1}^d[\Sigma_i]_{jj}\right)^{3/2} \ge \frac{d^{3/2}}{\sqrt{n}} \times \underline{v}_n^3 \quad \text{a.s.}$$

Therefore, the term $\ln(dn)$ in the bounds established in Theorem D.2 and Corollary D.1 should be replaced by $\ln(dn\tau_n)$. In this case the assumption regarding $\min_{1\le i\le n}\{\underline{\sigma}_i\}$ appearing in Theorem D.1 becomes redundant.