Anastasiou, Andreas; Kley, Tobias

Wasserstein distance bounds on the normal approximation of empirical autocovariances and cross-covariances under non-stationarity and stationarity

Journal of Time Series Analysis

Provided in Cooperation with:
John Wiley & Sons

Suggested Citation: Anastasiou, Andreas; Kley, Tobias (2023) : Wasserstein distance bounds on the normal approximation of empirical autocovariances and cross-covariances under non-stationarity and stationarity, Journal of Time Series Analysis, ISSN 1467-9892, Wiley, Hoboken, NJ, Vol. 45, Iss. 3, pp. 361-375,
https://doi.org/10.1111/jtsa.12716

This Version is available at:
https://hdl.handle.net/10419/290181

JOURNAL OF TIME SERIES ANALYSIS
J. Time Ser. Anal. **45**: 361–375 (2024)
Published online 18 August 2023 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1111/jtsa.12716

ORIGINAL ARTICLE

WASSERSTEIN DISTANCE BOUNDS ON THE NORMAL APPROXIMATION OF EMPIRICAL AUTOCOVARIANCES AND CROSS-COVARIANCES UNDER NON-STATIONARITY AND STATIONARITY

ANDREAS ANASTASIOU[a] AND TOBIAS KLEY[b*]
[a]*Department of Mathematics and Statistics, University of Cyprus, Nicosia, Cyprus*
[b]*Institute for Mathematical Stochastics, Georg-August-University of Göttingen, Göttingen, Germany*

The autocovariance and cross-covariance functions naturally appear in many time series procedures (e.g. autoregression or prediction). Under assumptions, empirical versions of the autocovariance and cross-covariance are asymptotically normal with covariance structure depending on the second- and fourth-order spectra. Under non-restrictive assumptions, we derive a bound for the Wasserstein distance of the finite-sample distribution of the estimator of the autocovariance and cross-covariance to the Gaussian limit. An error of approximation to the second-order moments of the estimator and an m-dependent approximation are the key ingredients to obtain the bound. As a worked example, we discuss how to compute the bound for causal autoregressive processes of order 1 with different distributions for the innovations. To assess our result, we compare our bound to Wasserstein distances obtained via simulation.

Received 10 October 2022; Accepted 17 July 2023

Keywords: Autocovariance; time series; Wasserstein distance; Stein's method.

MOS subject classification: Primary 62E17; secondary 62F12.

1. INTRODUCTION

Assessing the quality of various asymptotic results has attracted a lot of interest in recent years. One way to measure the error in distributional approximations is to consider explicit upper bounds on the Wasserstein distance between the limiting and the actual distribution of the quantity of interest; to derive such bounds is undoubtedly a technically tedious task.

We consider the empirical autocovariance and cross-covariance

1. without assuming stationarity, and
2. for the case of weakly stationary time series.

Our aim is to facilitate a bound where the rate, but also explicit constants can be computed for a wide range of time series models.

We consider the case where a d-variate time series $\boldsymbol{X}(1), \dots, \boldsymbol{X}(n)$ is available, that is, $\boldsymbol{X}(t)$ are $\mathbb{R}^d$-valued, $t = 1, \dots, n$. The components of $\boldsymbol{X}(t)$ are denoted by $X_a(t)$, $a = 1, \dots, d$. We are interested in the empirical cross-covariance and autocovariance, defined as

$$\hat{\gamma}^*_{ab}(k) := \frac{1}{n}\sum_{t=1}^{n-k}(X_a(t+k) - \overline{X}_a)(X_b(t) - \overline{X}_b), \quad k = 0, \dots, n-1, \tag{1.1}$$

*Correspondence to: Tobias Kley, Institute for Mathematical Stochastics, Georg-August-University of Göttingen, Goldschmidtstraße 7, 37077 Göttingen, Germany. Email: tobias.kley@uni-goettingen.de

where $\overline{X}_j := \frac{1}{n}\sum_{t=1}^{n} X_j(t), j = a, b$, denotes the empirical mean. For $k = -n+1, \ldots, -1$ we define $\hat{\gamma}^*_{ab}(k) := \hat{\gamma}^*_{ba}(-k)$. Other definitions, which are asymptotically equivalent under regularity conditions, also exist in the literature. For example, see Anderson (1971), Chapter 8, for some common variants in the case of the autocovariance and in particular Corollary 8.4.1 in Anderson (1971) for a result asserting that these variants converge to the same Gaussian limit, under specific regularity conditions.

In the case of stationary data where the population means are known, we may substitute the empirical means in (1.1) by their population counterparts as below

$$\frac{1}{n}\sum_{t=1}^{n-k}(X_a(t+k) - \mathbb{E}X_a(t+k))(X_b(t) - \mathbb{E}X_b(t)), \quad k = 0, \ldots, n-1.$$

This corresponds to assuming that $\{\boldsymbol{X}(t)\}$ is centred (i.e. $\mathbb{E}\boldsymbol{X}(t) = 0$), and working with the following definition of the empirical cross-covariance:

$$\hat{\gamma}_{ab}(k) := \frac{1}{n}\sum_{t=1}^{n-k} X_a(t+k)X_b(t), \quad k = 0, \ldots, n-1, \tag{1.2}$$

and $\hat{\gamma}_{ab}(k) := \hat{\gamma}_{ba}(-k)$, $k = -n+1, \ldots, -1$. Autocovariances and cross-covariances are important for many time series methods; for example, autoregression (Jirak, 2012, 2014) and forecasting (Brockwell and Davis, 2006; Kley *et al.,* 2019).

Under conditions, it can be shown that $\hat{\gamma}^*_{ab}(k)$ and $\hat{\gamma}_{ab}(k)$ are consistent estimates for

$$\gamma_{ab}(k) := \mathbb{E}[X_a(t+k)X_b(t)]. \tag{1.3}$$

The asymptotic normality for the distribution of the estimator holds as well. We have that

$$\sqrt{n}\left(\hat{\gamma}^*_{ab}(k) - \gamma_{ab}(k)\right) \xrightarrow[n\to\infty]{} N, \quad N \sim \mathcal{N}(0, \Sigma_{ab}(k)); \tag{1.4}$$

see, for example, Exercise 7.10.36 in Brillinger (1975). The asymptotic variance

$$\Sigma_{ab}(k) := \lim_{n\to\infty} \operatorname{var}\left(n^{-1/2}\sum_{t=1}^{n-k} X_a(t+k)X_b(t)\right) \tag{1.5}$$

depends on the second- and fourth-order moment structure of the underlying data; cf. eq. (7.6.11) in Brillinger (1975). It is usually straightforward to compute $\Sigma_{ab}(k)$. Details for the case of an AR(1) time series that we consider in Section 3 are provided in Section D of the online supplement. To prove (1.4), it is common practice to make assumptions limiting the intensity of the dependence structure and the moments of the random variables involved (such as the summability of cumulants); cf. Horváth and Kokoszka (2008) for the discussion of cases where normality fails.

We now provide the general framework and notation used throughout the article. For $\mathbb{R}^d$-valued random vectors $\boldsymbol{U}$ and $\boldsymbol{V}$, we work with the 1-Wasserstein metric defined as

$$d_{\mathrm{W}}(\mathcal{L}(\boldsymbol{U}), \mathcal{L}(\boldsymbol{V})) := \sup_{h\in\mathcal{H}} |\mathbb{E}[h(\boldsymbol{U})] - \mathbb{E}[h(\boldsymbol{V})]|, \quad \mathcal{H} = \{h : \mathbb{R}^d \to \mathbb{R} \mid \|h\|_{\mathrm{Lip}} \le 1\}, \tag{1.6}$$

where $\mathcal{L}(\boldsymbol{U})$ is the law of $\boldsymbol{U}$. Furthermore, for any vector $\boldsymbol{x} = (x_1, \ldots, x_d)$, we denote its Euclidean norm by $|\boldsymbol{x}| := \left(\sum_{i=1}^{d} x_i^2\right)^{1/2}$ and $\|h\|_{\mathrm{Lip}} = \sup_{\boldsymbol{u}\neq\boldsymbol{v}} |h(\boldsymbol{u}) - h(\boldsymbol{v})|/|\boldsymbol{u} - \boldsymbol{v}|$. In this article, we refer short to the distance in

Journal of Time Series Analysis published by John Wiley & Sons Ltd. DOI: 10.1111/jtsa.12716

(1.6) as the Wasserstein distance. The main purpose of the article is to assess the quality of the distributional approximation in (1.4) through upper bounds on the Wasserstein distance between the actual distribution of the quantity of interest on the left-hand side of (1.4) and its limiting normal distribution; for centred data, this is achieved in Theorems 2.1 and 2.3 for the case of a non-stationary or weakly stationary sequence respectively. Combining Theorem 2.3 with Lemma C.1 to bound the Wasserstein distance between $\sqrt{n}\left(\hat{\gamma}^{*}_{ab}(k) - \gamma_{ab}(k)\right)$ and $\sqrt{n}\left(\hat{\gamma}_{ab}(k) - \gamma_{ab}(k)\right)$ in the stationary case, we obtain a bound when the data are non-centred; details can be found in Sections 2.5 and C.

Our approach depends on the existence of an m-dependent sequence, which allows us to use Stein's method, a powerful probabilistic technique first introduced in Stein (1972), under a local dependence structure. Stein's method is particularly powerful in assessing whether a given random variable has a distribution close to a target distribution in the presence of such dependence structures between the random variables. The bounds obtained through Stein's method are explicit in terms of the constants and in terms of the sample size; see, for example, Anastasiou (2017), where bounds for the normal approximation of the maximum likelihood estimator are provided under a local dependence structure between the random variables.

There has been a lot of interest recently on the assessment of the quality of the normal approximation related to the sum $\sum_{i=1}^{n} X_i$, where $X_1, \ldots, X_n$ are centred and follow a specific dependence structure. While at first sight, it seems that the empirical autocovariance and cross-covariance fit into this framework (replace n by $n-k$ and X_i by $n^{-1}X_a(t+k)X_b(t)$), the results in the literature for $\sum_{i=1}^{n} X_i$ do not immediately provide us with the result that we are interested in: an explicit finite-sample bound assessing the quality of the approximation in (1.4). Amongst other reasons, this is due to the fact that the empirical autocovariance and cross-covariance are biased. We consider the empirical autocovariance and cross-covariance to be of such fundamental importance for applications that results to assess their finite-sample distributional approximation, fully explicit in terms of the underlying process/model parameters, segment size n and lag k, should be available.

We now continue to discuss work related to assessing the quality of the normal approximation for sums of dependent data. Staying in the setting of explicit bounds but moving away from the m-dependence structure that we use, Röllin (2018) provides bounds on the Wasserstein distance between the distribution of $\sum_{i=1}^{n} X_i$, where $X_1, \ldots, X_n$ is a discrete time martingale difference sequence, and the standard normal distribution. The bound is of the order $\mathcal{O}(n^{-1/2}\log n)$ and the strategy followed to obtain the upper bounds consists of a combination of Stein's method and Lindeberg's argument. In their work related to the Polyak–Ruppert averaged stochastic gradient descent, Anastasiou *et al.* (2019) derive an explicit upper bound on the distributional distance between the distribution of the summation of a multi-variate martingale difference sequence and the multi-variate normal distribution. In their recent work, Fan and Ma (2020) extend the results of Röllin (2018) by relaxing conditions used in the latter. Apart from the setting of discrete time martingales, work has been done on assessing the normal approximation of a sum of random variables when these satisfy specific mixing conditions; see Sunklodas (2007, 2011) for the cases of strong and φ-mixing conditions respectively. Dedecker and Rio (2008) provide bounds for the Wasserstein distance between the distribution of $\sum_{i=1}^{n} X_i$ and the normal distribution, when either strong mixing assumptions are satisfied or when $(X_i)_{i\in\mathbb{Z}}$ is either an ergodic martingale difference sequence or an ergodic stationary sequence that satisfies specific projective criteria.

Moving away from the scenario of explicit constants in the bounds, Dedecker *et al.* (2009) provide, in the case of $X_1, \ldots, X_n$ being a martingale difference sequence, rates of convergence for minimal distances between linear statistics of the form $\sum_{i=1}^{n} c_{n,i}X_n$, where $c_{n,i} \in \mathbb{R}$, and their limiting Gaussian distribution. Fan (2019) gives rates of convergence for the Central Limit Theorem of a martingale difference sequence with conditional moment assumptions. For $X_1, \ldots, X_n$ a stationary sequence with finite $p \in (2, 3]$ moments, Jirak (2016) proves under a weak dependence condition a Berry–Esseen theorem and shows convergence rates in L^q-norm, where $q \geq 1$. The obtained bounds are though not explicit, in the sense that they depend on a varying absolute constant not given explicitly.

Apart from the machinery employed, the proof methodology followed, and the focus to the specific statistics of the empirical autocovariance and cross-covariance functions, the results presented in this article are novel in three additional main aspects. First, our results are applicable to non-stationary data sequences. Second, our focus

DOI: 10.1111/jtsa.12716 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd.

is not only on rates of convergence, but the Wasserstein distance bounds derived in the article are fully explicit in terms of the sample size n, the lag k, as well as constants that are related to the underlying data; this makes the bound completely computable in examples. Third, the assumptions that we have used are non-restrictive, and they are partly based on an m-dependence approximation of the original time series, which is convenient to work with in applications, making our results applicable in a wide range of scenarios. In the case where the range of dependence is finite, for example, independent observations or a moving average process of fixed order, the order of our bound is $\mathcal{O}(n^{-1/2})$. In more general cases where the serial dependence vanishes quickly at large lags and moments of order eight exist, the order of our bound is $\mathcal{O}(n^{-1/2}\log n)$. A discussion on the order of the bound can be found in Remark 2.5 and, in more detail, in Section 2.6.

The article is organised as follows. In Section 2.1, we give our main result in the general case; this is an upper bound on the Wasserstein distance between the distribution of the empirical autocovariance and cross-covariance functions $\hat{\gamma}_{ab}(k)$, defined in (1.2), and their limiting normal distribution. We highlight that the data are not necessarily obtained from a stationary process. In Section 2.2, we state and discuss the key assumption for the weakly stationary case. In Section 2.3, the main result under stationarity is given. Sections 2.4 and 2.5 are devoted to computing bounds in terms of moments of the m-dependent approximation or the original centred process, respectively; details on the computation of bounds in terms of moments of the original uncentred process are deferred to Section C. A detailed explanation of the order of the bound with respect to the sample size n is given in Section 2.6. In Section 3, we apply our general results to the specific case of a causal autoregressive process of order 1. In Section 4, our main result is proven. Section 5 concludes the article with a brief discussion on the results. Technical details on the computation of the bound from Section 2.4, step-by-step proofs that were not included in the main text, technical details regarding computation and simulation, as well as additional tables for the example in Section 3 are provided in a supplement, which is available online. Sections, results, etc. that are numbered with letters from the Latin alphabet are always to be found in the supplement.

2. MAIN RESULTS

2.1. The Explicit Upper Bound for the General Case

Here we present a general result that does not require the data to be from a stationary process. To apply the result, an m-dependent sequence of the same length, n, and dimension, d, with finite sixth moments needs to exist. The closeness of the data to the m-dependent approximation will determine the size of the bound.

For ease of presentation, some notation is in order. For any vector $\boldsymbol{x} = (x_1, \dots, x_d)$, we denote its Euclidean norm by $|\boldsymbol{x}| := \left(\sum_{i=1}^{d} x_i^2\right)^{1/2}$, while for a random vector $\boldsymbol{X}$, its L^q-norm is denoted by $\|\boldsymbol{X}\|_q := (\mathbb{E}[|\boldsymbol{X}|^q])^{1/q}$, $q \geq 1$. We denote $\mathbb{N} := \{1, 2, \dots\}$ and $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$. Recall, the r^{th} order joint cumulant of a random vector $(\zeta_1, \dots, \zeta_r)$ is defined as

$$\operatorname{cum}(\zeta_1, \dots, \zeta_r) := \sum_{\nu} (-1)^{p-1}(p-1)! \left(\mathbb{E}\prod_{j\in\nu_1}\zeta_j\right)\cdots\left(\mathbb{E}\prod_{j\in\nu_p}\zeta_j\right), \tag{2.7}$$

where the sum is with respect to all partitions $\nu := \{\nu_1, \dots, \nu_p\}$ of $\{1, \dots, r\}$; cf. Brillinger (1975). The general result for the case of a not necessarily stationary sequence is given in Theorem 2.1 below. Its proof is deferred to Section 4.

Theorem 2.1. Let $\boldsymbol{X}(1), \dots, \boldsymbol{X}(n)$ be a sequence of d-variate random vectors, and assume $\mathbb{E}(\boldsymbol{X}(t)) = 0$ for all $t \in \{1, \dots, n\}$. Fix $a, b \in \{1, \dots, d\}$, and $k \in \{0, \dots, n-1\}$, and let $\hat{\gamma}_{ab}(k)$ be defined as in (1.2). Fix $m \in \mathbb{N}_0$, and let $\boldsymbol{Y}(1), \dots, \boldsymbol{Y}(n)$ be a sequence of m-dependent d-variate random vectors, such that $\|\boldsymbol{Y}(t)\|_6 < \infty$ for all $t = 1, \dots, n$ and assume that

$$\tilde{\Sigma}_{ab}(k) := \operatorname{var}\left(n^{-1/2}\sum_{t=1}^{n-k} Y_a(t+k)Y_b(t)\right) > 0. \tag{2.8}$$

Journal of Time Series Analysis published by John Wiley & Sons Ltd. DOI: 10.1111/jtsa.12716

Let

$$\tilde{D}_{a,t} := \|X_a(t) - Y_a(t)\|_2 < \infty, \quad t \in \{1, \dots, n\},$$

and if $b \neq a$, let $\tilde{D}_{b,t}$ be defined analogously. Moreover, for $t = 1, \dots, n-k$, let

$$\tilde{Z}(t) := Y_a(t+k)Y_b(t) - \mathbb{E}[Y_a(t+k)Y_b(t)]$$

and define the quantities

$$\tilde{Q}_t := \mathbb{E}\left|\left(\tilde{Z}(t)\tilde{Z}(A_t) - \mathbb{E}\left(\tilde{Z}(t)\tilde{Z}(A_t)\right)\right)\tilde{Z}(B_t)\right| + \frac{1}{2}\mathbb{E}\left|\tilde{Z}(t)\left(\tilde{Z}(A_t)\right)^2\right|, \tag{2.9}$$

where $\tilde{Z}(A) = \sum_{j\in A}\tilde{Z}(j)$ for $A \subset \mathbb{N}$ and

$$A_t := \{\ell = 1, \dots, n-k : |\ell - t| \leq m + k\},$$
$$B_t := \{\ell = 1, \dots, n-k : |\ell - t| \leq 2(m+k)\}.$$

Finally, let

$$\tilde{K}_t := \tilde{D}_{a,t+k}\|X_b(t)\|_2 + \tilde{D}_{a,t+k}\tilde{D}_{b,t} + \|X_a(t+k)\|_2\tilde{D}_{b,t}.$$

Then, for any $\gamma \in \mathbb{R}$ and any $\sigma^2 > 0$,

$$d_{\mathrm{W}}\left(\mathcal{L}\left(n^{1/2}\left(\hat{\gamma}_{ab}(k) - \gamma\right)\right), \mathcal{N}(0,\sigma^2)\right) \leq \frac{1}{\sqrt{n}}\sum_{t=1}^{n-k}\tilde{K}_t + \frac{1}{\sqrt{n}}\sum_{t=1}^{n-k}\left|\frac{n}{n-k}\gamma - \mathbb{E}[Y_a(t+k)Y_b(t)]\right| + \sqrt{\frac{2}{\pi\sigma^2}}\left|\sigma^2 - \tilde{\Sigma}_{ab}(k)\right| + \frac{2}{n^{3/2}\left(\tilde{\Sigma}_{ab}(k)\right)^{3/2}}\sum_{t=1}^{n-k}\tilde{Q}_t. \tag{2.10}$$

In the theorem and throughout the rest of the article, we use the convention to distinguish notation related to the m-dependent approximation with the tilde symbol (e.g. $\tilde{D}_{a,t}, \tilde{Z}(t), \tilde{Q}_t$).

Due to the non-restrictive assumptions and explicitness of the constants, Theorem 2.1 can, for example, be used to show asymptotic normality of sequences of estimators where the underlying model or the lag k depends on the segment length n. It can also be applied to models with time-varying coefficients. Because of the countless situations in which the result can potentially be applied, but this article only offers limited space, we will focus on one of the most relevant situations for applications in the following sections: the case of weakly stationary data.

2.2. The Key Assumption for the Stationary Case

From this section onwards, we consider $\{\boldsymbol{X}(t) : t \in \mathbb{Z}\}$ to be a d-variate, centred and weakly stationary process, denoted by $\{\boldsymbol{X}(t)\}$, from which a sequence $\boldsymbol{X}(1), \dots, \boldsymbol{X}(n)$ is available, with $\boldsymbol{X}(t)$ being $\mathbb{R}^d$-valued, $t = 1, \dots, n$. The main assumption used for the result under stationarity is given below.

Assumption 2.2. For a given $m \in \mathbb{N}_0$ there exists an m-dependent, d-variate process $\{\boldsymbol{Y}(t)\}$ where, for $a \in \{1, \dots, d\}$, and $t \in \{1, \dots, n\}$,

$$\tilde{D}^{(q)}_{a,t} := \|X_a(t) - Y_a(t)\|_q < \infty.$$

The number $q \geq 1$ is specified whenever we refer to Assumption 2.2.

DOI: 10.1111/jtsa.12716 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd.

We denote by

$$\tilde{D}_a^{(q)} := \sup_{1\le t\le n} \tilde{D}_{a,t}^{(q)}. \tag{2.11}$$

We highlight that if there is a choice on the m-dependent sequence, then one could define $\tilde{D}_a^{(q)} := \inf_Y \sup_{1\le t\le n} \|X_a(t) - Y_a(t)\|_q$, with the infimum taken with respect to all possible choices of $\{\boldsymbol{Y}(t)\}$ that satisfy $\|X_a(t) - Y_a(t)\|_q < \infty$; in the interest of obtaining an easily computable bound, however, we state our result for a specific choice.

Even though in this section, we assume stationarity of $\{\boldsymbol{X}(t)\}$ to allow for a meaningful definition of $\gamma_{ab}(k)$, our method of proof, as already stated, does not require stationarity. We state and discuss the result for the stationary case in full detail because of its relevance for applications and because it sheds light on the more general result. Assumption 2.2 implies that the original process can be approximated in L^q by an m-dependent sequence. For Theorem 2.3, approximation in L^2 is sufficient; that is, we require Assumption 2.2 with $q = 2$. In Section 2.5 we explain the general steps to obtaining a bound in terms of properties of the original process $\{\boldsymbol{X}(t)\}$. Lemmas B.3 and B.4 can be used to pursue such a bound; Assumption 2.2 with $q = 4$ and $q = 6$ respectively, is then required. We do not require $\{\boldsymbol{Y}(t)\}$ to be stationary or centred, though in applications this will often be the case. In the example discussed in Section 3, $\|X_a(t) - Y_a(t)\|_q$ is actually independent of t; details on how to compute $\tilde{D}_a^{(q)}$ as in (2.11) for the example are available in Section D. If $\{(\boldsymbol{X}(t), \boldsymbol{Y}(t))\}$ is jointly stationary up to moments of order q, then $\tilde{D}_a^{(q)} = \tilde{D}_{a,0}^{(q)}$ and the supremum in (2.11) could be omitted. Our Assumption 2.2 is similar in spirit to Assumption 2.1 in Aue *et al.* (2009); also see the examples provided in their Section 4 that illustrate how to apply such a framework to several popular time series models. Note the following important difference though. The quantity $\tilde{D}_a^{(q)}$ gives a bound to the goodness of the m-dependent approximation measured in L^q and while a larger m will typically result in a better approximation (i.e. a smaller $\tilde{D}_a^{(q)}$), there is no requirement at the rate of decay that we would usually have if we were deriving an asymptotic result. For our main results, which are finite sample in nature, we only require that $X_a(t) - Y_a(t)$ is in L^q; that is, the quantity $\tilde{D}_{a,t}^{(q)}$ is finite.

2.3. The Explicit Upper Bound for Centred Stationary Data

The upper bound on the quantity of interest in the case of a weakly stationary sequence is given below.

Theorem 2.3. Let $\{X(t)\}$ be a d-variate, centred and weakly stationary process. Fix $a, b \in \{1, \dots, d\}$, and $k \in \{0, \dots, n-1\}$, and let $\hat{\gamma}_{ab}$ and $\gamma_{ab}(k)$ be defined as in (1.2) and (1.3) respectively. Fix $m \in \mathbb{N}_0$, and let $\{\boldsymbol{Y}(t)\}$ be a process as in Assumption 2.2, which we assume holds with $q = 2$, and also $\|\boldsymbol{Y}(t)\|_6 < \infty$ for all $t = 1, \dots, n$. For given $n \in \mathbb{N}$, assume that both $\Sigma_{ab}(k)$ and $\tilde{\Sigma}_{ab}(k)$, defined in (1.5) and (2.8) respectively, are positive. Finally, with $\tilde{D}_a^{(q)}$ as in (2.11), let,

$$\tilde{K} := \tilde{D}_a^{(2)} \|X_b(0)\|_2 + \tilde{D}_a^{(2)} \tilde{D}_b^{(2)} + \|X_a(0)\|_2 \tilde{D}_b^{(2)}.$$

For A_t and B_t as in Theorem 2.1, and $\tilde{Q}_t$ as in (2.9), we have that

$$\begin{aligned} &d_{\mathrm{W}}\left(\mathcal{L}\left(\sqrt{n}\left(\hat{\gamma}_{ab}(k) - \gamma_{ab}(k)\right)\right), \mathcal{N}\left(0, \Sigma_{ab}(k)\right)\right) \\ &\quad\le \frac{k}{\sqrt{n}}|\gamma_{ab}(k)| + \frac{\sqrt{2}}{\sqrt{\pi\Sigma_{ab}(k)}}\left|\Sigma_{ab}(k) - \tilde{\Sigma}_{ab}(k)\right| \\ &\qquad + \frac{2(n-k)}{\sqrt{n}}\tilde{K} + \frac{2}{\left(n\tilde{\Sigma}_{ab}(k)\right)^{3/2}}\sum_{t=1}^{n-k}\tilde{Q}_t. \end{aligned} \tag{2.12}$$

Journal of Time Series Analysis published by John Wiley & Sons Ltd. DOI: 10.1111/jtsa.12716

Proof of Theorem 2.3. Choose $\gamma := \gamma_{ab}(k)$ and $\sigma^2 := \Sigma_{ab}(k)$. Note that the conditions of Theorem 2.3 imply that the conditions of Theorem 2.1 are satisfied. The bound in the stationary case then follows from $\tilde{K}_t \leq \tilde{K}$ for all $t = 1, \dots, n-k$ and

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{t=1}^{n-k}\left|\frac{n}{n-k}\gamma - \mathbb{E}[Y_a(t+k)Y_b(t)]\right| &\leq \frac{1}{\sqrt{n}}\sum_{t=1}^{n-k}\left|\frac{n}{n-k}\gamma_{ab}(k) - \gamma_{ab}(k)\right| \\ &\quad + \frac{1}{\sqrt{n}}\sum_{t=1}^{n-k}\mathbb{E}\left|X_a(t+k)X_b(t) - Y_a(t+k)Y_b(t)\right| \\ &\leq \frac{n-k}{\sqrt{n}}\left|\frac{k}{n-k}\gamma_{ab}(k)\right| + \frac{n-k}{\sqrt{n}}\tilde{K},\end{aligned}$$

where we used (1.3) and a telescoping sum argument (Lemma B.1 with $\alpha = 1$ and $p = 2$). ∎

Remark 2.4. The four terms that make up the right-hand side of (2.12) can roughly be interpreted as follows: (i) $kn^{-1/2}|\gamma_{ab}(k)|$, is due to the fact that $\hat{\gamma}_{ab}(k)$ is a biased estimate for $\gamma_{ab}(k)$, but the limiting normal distribution has mean equal to zero; (ii) $\sqrt{2/(\pi\Sigma_{ab}(k))}|\Sigma_{ab}(k) - \tilde{\Sigma}_{ab}(k)|$, is related to the fact that the variance of $\hat{\gamma}_{ab}(k)$ may differ from $n^{-1}\Sigma_{ab}(k)$; (iii) $2(n-k)n^{-1/2}\tilde{K}$, is due to our method of proof where we use the m-dependent approximation, and (iv) $2(n\tilde{\Sigma}_{ab}(k))^{-3/2}\sum_{t=1}^{n-k}\tilde{Q}_t$, is due to an application of Stein's method; cf. Lemma 4.2.

The following remark provides a brief discussion of the computation and of the order of the bound; detailed explanations are given in Section 2.6.

Remark 2.5. At first glance, the bound might seem slightly complicated, especially due to the expression $\tilde{Q}_t$. In Section 2.4, we explain two methods that allow to bound $\tilde{Q}_t$ by expressions whose exact value can be computed in examples. In Section 3, we then calculate the exact value of such a bound term by term for the case of a causal autoregressive process. To obtain a rate, we choose m as a function of n. The choice that allows optimisation of the order of the bound with respect to n depends on the underlying process $\{\boldsymbol{X}(t)\}$. In Section 2.6, we discuss two general scenarios where the bound is of the order $\mathcal{O}(n^{-1/2})$ or $\mathcal{O}(n^{-1/2}\log n)$ respectively.

2.4. The Bound in Theorem 2.3 when the m-dependent Approximation is Known

Let $\boldsymbol{X}(t)$ be such that, for given n, k, and m, we can compute $\gamma_{ab}(k)$, $\Sigma_{ab}(k)$, $\|X_a(0)\|_2$ and $\|X_b(0)\|_2$. Assume further that we may choose $\boldsymbol{Y}(t)$ such that $\tilde{\Sigma}_{ab}(k)$, $\tilde{D}_a^{(2)}$ and $\tilde{D}_b^{(2)}$ can be computed. Then, the only missing piece to obtain the upper bound in (2.12) is $\tilde{Q}_t$, defined in (2.9). The absolute joint moments in the definition of $\tilde{Q}_t$ can be inconvenient. To address potential problems in the computation of $\tilde{Q}_t$, we now describe two ways to bound $\tilde{Q}_t$ by quantities that can be explicitly computed in examples. First, if $\{\boldsymbol{Y}(t)\}$ is stationary (otherwise see below, within Method 1), we bound $\tilde{Q}_t$ in terms of $\|\boldsymbol{Y}(0)\|_6$, which is finite from the statement of Theorem 2.3. Second, we obtain a better bound for $\tilde{Q}_t$ when m is large. The price we pay for the second method is a more complicated computation and the requirement that $\|\boldsymbol{Y}(0)\|_8 < \infty$.

Method 1 to bound $\tilde{Q}_t$. Denoting $\mu_{jq} := \|Y_j(0)\|_q$, we have

$$\begin{aligned}\tilde{Q}_t &\leq \sum_{j_1\in A_t}\sum_{j_2\in B_t}\left(\mathbb{E}\left|\tilde{Z}(t)\tilde{Z}(j_1)\tilde{Z}(j_2)\right| + \mathbb{E}\left|\tilde{Z}(t)\tilde{Z}(j_1)\right|\mathbb{E}\left|\tilde{Z}(j_2)\right|\right) + \frac{1}{2}\sum_{j_1\in A_t}\sum_{j_2\in A_t}\mathbb{E}\left|\tilde{Z}(t)\tilde{Z}(j_1)\tilde{Z}(j_2)\right| \\ &\leq |A_t||B_t|\left(\mu_{a6}^3\mu_{b6}^3 + \mu_{a4}^2\mu_{b4}^2\mu_{a2}\mu_{b2}\right) + \frac{1}{2}|A_t|^2\mu_{a6}^3\mu_{b6}^3 \leq \frac{5}{2}(4m+4k+1)^2\|\boldsymbol{Y}(0)\|_6^6. \qquad (2.13)\end{aligned}$$

Employing the triangle inequality, a generalised version of Hölder's inequality, and the stationarity of $\boldsymbol{Y}(t)$, the joint moments in the definition of $\tilde{Q}_t$ were broken up into moments of the marginals. If $\{\boldsymbol{Y}(t)\}$ is not stationary we

can use the $\sup_{t=1,\dots,n} \|\boldsymbol{Y}(t)\|_6$ instead. The bound in (2.13) is particularly simple and straightforward to compute. In essence, we see a product of the marginal moments $\|Y_j(0)\|_q$ for $j = a, b$ and $q = 2, 4, 6$ scaled by a multiple of m^2. A bound of the order m^2 is most useful when m is small. To improve (2.13) in the case when m is large, we next derive a bound for $\tilde{Q}_t$ in terms of joint (non-absolute) moments.

Method 2 to bound $\tilde{Q}_t$. We apply the Cauchy–Schwarz inequality and $\mathbb{E}(\tilde{Z}(t)) = 0$ to obtain

$$\tilde{Q}_t \leq \operatorname{var}\big(\tilde{Z}(t)\tilde{Z}(A_t)\big)^{1/2} \operatorname{var}\big(\tilde{Z}(B_t)\big)^{1/2} + \frac{1}{2}\Big[\mathbb{E}\big(\tilde{Z}(t)\tilde{Z}(A_t)\big)^2\Big]^{1/2} \operatorname{var}\big(\tilde{Z}(A_t)\big)^{1/2}. \tag{2.14}$$

A crucial difference between the right-hand side of (2.14) and the first bound in (2.13) is that the former is in terms of joint moments of $\tilde{Z}(t)$ and the later in terms of joint moments of $|\tilde{Z}(t)|$. Using standard combinatorial arguments (cf. Theorem 2.3.2 in Brillinger (1975)), the right-hand side in (2.14) can be computed from cumulants of $\{\boldsymbol{Y}(t)\}$. These arguments are straightforward but tedious and therefore deferred to Section A. Another important advantage of the second method is that, in the common situation where serial dependence is less pronounced at larger lags, such that cumulants are summable, the bound obtained by the second method is of the order $\mathcal{O}(m)$, which, compared with the $\mathcal{O}(m^2)$ bound obtained by Method 1, is much advantageous when m is large. Intuitively, this can be seen from the fact that the variance of a sum of m elements of a short-range-dependent sequence is of the order m. Additional details are available in the proof of Proposition 2.7 that can be found in Section F.5.

2.5. The Bound with Respect to the Original Data

In Section 2.4 we explained computational details regarding a bound for the case when $\{\boldsymbol{Y}(t)\}$, the m-dependent approximation, is known. The method described required the computation of joint moments of $\{\boldsymbol{Y}(t)\}$ to obtain $\tilde{\Sigma}_{ab}(k)$ and the right-hand side of either (2.13) or (2.14). If such computation is possible, then numerically evaluating the bound obtained from (2.12) in combination with (2.13) or (2.14), for fixed values of n and m, is the preferred method. The aim of this section is to facilitate our result for situations where a bound that depends on $\{\boldsymbol{Y}(t)\}$ might be inconvenient (e.g. when $\{\boldsymbol{Y}(t)\}$ is unknown).

There are, at least two, good reasons to pursue a bound that only depends on quantities defined in terms of $\{\boldsymbol{X}(t)\}$. The first reason is a philosophical one. Noting that the statistic of interest, $\hat{\gamma}_{ab}(k)$, is defined in terms of the original process $\{\boldsymbol{X}(t)\}$ we observe that the left-hand side of (2.12) only depends on $\{\boldsymbol{X}(t)\}$, too. Therefore, the right-hand side of (2.12) being defined, amongst others, in terms of $\tilde{\Sigma}_{ab}(k)$ and $\tilde{Q}_t$, both depending on $\{\boldsymbol{Y}(t)\}$, can be considered a discrepancy. The second reason is a practical one. In Sections 2.6 and F.5 it can be seen that the discussion of asymptotic properties of the bound can be simplified when the dependence on m is not via properties of $\boldsymbol{Y}(t)$.

To obtain a bound in terms of moments of $\{\boldsymbol{X}(t)\}$, it suffices to quantify the effect of replacing $\tilde{\Sigma}_{ab}(k)$ by $\Sigma_{ab}(k)$ and the effect of replacing $\tilde{Q}_t$ by

$$Q_t := \mathbb{E}\left|\big(Z(t)Z(A_t) - \mathbb{E}\left(Z(t)Z(A_t)\right)\big)\, Z(B_t)\right| + \frac{1}{2}\mathbb{E}\left|Z(t)\big(Z(A_t)\big)^2\right|, \tag{2.15}$$

where $Z(t) := X_a(t+k)X_b(t) - \mathbb{E}[X_a(t+k)X_b(t)]$, $t \in \mathbb{N}$, and $Z(A) = \sum_{j\in A} Z(j)$, $A \subset \mathbb{N}$.

In Section B, we provide results that can be used to derive a bound in terms of moments of $\{\boldsymbol{X}(t)\}$. Further, in Section C, we discuss the case of non-centred data and provide a result to derive a bound for $d_{\mathrm{W}}\left(\mathcal{L}\left(\sqrt{n}\left(\hat{\gamma}^*_{ab}(k) - \gamma_{ab}(k)\right)\right), \mathcal{N}(0, \Sigma_{ab}(k))\right)$ in this case.

2.6. Explanation on the Order of the Bound

In Remark 2.5, we have stated the outcomes of the asymptotic analysis of our bound. In this section, the details are provided. We begin by making the conditions we work under precise. For simplicity, we consider only the

Journal of Time Series Analysis published by John Wiley & Sons Ltd. DOI: 10.1111/jtsa.12716

case where the underlying process $\{\boldsymbol{X}(t)\}$ and the lag k are not allowed to change with n. The two regimes we consider are:

Regime 1. Let $\{\boldsymbol{X}(t)\}$ be d-variate, centred and stationary, with $\|\boldsymbol{X}(0)\|_6 < \infty$, and also be M-dependent (for a fixed $M \in \mathbb{N}_0$).

Regime 2. Let $\{\boldsymbol{X}(t)\}$ be d-variate, centred and stationary, with $\|\boldsymbol{X}(0)\|_8 < \infty$, and it also satisfies Assumption 2.2 with $q = 8$, and (2.16) and (2.17), below.

We assume stationarity up to moment of order 6 or 8 in Regimes 1 and 2 respectively. In Regime 2, we require summability of cumulants up to order 8; that is, for $p = 2, \dots, 8$, we have

$$\sum_{k_1,\dots,k_{p-1}=-\infty} (1 + |k_j|)|\mathrm{cum}(X_{a_1}(k_1), \dots, X_{a_{p-1}}(k_{p-1}), X_{a_p}(0)| < \infty, \tag{2.16}$$

for $j = 1, \dots, p-1$ and any p tuple $a_1, \dots, a_p$. Further, we require that the m-dependent approximation from Assumption 2.2 is good enough such that the L^q-error vanishes at an exponential rate; that is, there exist constants $K \geq 0$ and $\rho \in (0,1)$ such that for every $n \in \mathbb{N}$ we can choose $m = m_n \in \mathbb{N}_0$ and an m-dependent, d-variate process $\{\boldsymbol{Y}^{(m)}(t)\}$ that satisfies

$$\tilde{D}_a^{(q)} := \sup_{1\leq t\leq n} \|X_a(t) - Y_a^{(m)}(t)\|_q \leq K\rho^m, \quad \text{for } a = 1, \dots, d;\ q = 8. \tag{2.17}$$

Remark 2.6.

(i) Examples for Regime 1 include moving average processes of finite order and independent data. In this regime, (2.16) holds for $p = 2, \dots, 6$, as cumulants vanish if one of the variables is independent of the others. Further, for any M-dependent process, as in Regime 1, the canonical choice for the m-dependent approximation of Assumption 2.2 is $\boldsymbol{Y}^{(m)}(t) = \boldsymbol{X}(t)$ for $m \geq M$. Choosing the quantity m in the bound (2.12) as $m = \min\{M, n\}$ we see that a stronger version of (2.17) is satisfied, where we have $\tilde{D}_a^{(q)} = 0$ for $n \geq M$.

(ii) As an example for Regime 2, one can consider a linear process $\boldsymbol{X}(t) = \sum_{j=0}^{\infty} \boldsymbol{\Psi}(j)\boldsymbol{\epsilon}(t-j)$ where the spectral norms of the coefficients satisfy $\|\boldsymbol{\Psi}(j)\|_2 \leq \rho^j$ for some $\rho \in (0,1)$ and the innovations are i.i.d. with $\|\boldsymbol{\epsilon}(t)\|_8 < \infty$. Then, it can be shown that (2.16) holds and (2.17) holds with $C := \|\boldsymbol{\epsilon}(t)\|_8 \rho/(1-\rho)$. In particular, causal autoregressive processes are included.

The following proposition gives the order of the bound (2.12) in Theorem 2.3. The proof is in Section F.5.

Proposition 2.7.

(i) In Regime 1, with $m := \min\{M, n\}$, the order of the bound is $\mathcal{O}(n^{-1/2})$.

(ii) In Regime 2, with $m := C \log n$, $C \geq \frac{3}{2\log(1/\rho)}$, where ρ is as in (2.17), the order of the bound is $\mathcal{O}(n^{-1/2} \log n)$.

3. EXAMPLES

3.1. Causal Autoregressive Processes of Order 1

As an example, for which we discuss the result of Theorem 2.3, we now consider the case where the data stem from a causal AR(1) process $\{X(t)\}$ that satisfies $X(t) = \alpha X(t-1) + \varepsilon(t)$, with $|\alpha| < 1$, where $\{\varepsilon(t)\}$ are i.i.d. and satisfy $\mathbb{E}|\varepsilon(t)|^8 < \infty$. We consider $\alpha \in \{0, 0.1, 0.3, 0.5, 0.7\}$ and three cases for the distribution of the innovations:

- $\varepsilon(t) \sim \mathcal{N}(0,1)$, or
- $\varepsilon(t) \sim \nu^{-1/2}(\nu-2)^{1/2} t_\nu$, where we choose $\nu \in \{9, 14\}$.

DOI: 10.1111/jtsa.12716 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd.

We have chosen the normal distribution as an example with light tails, the scaled t_9-distribution as a distribution with heavier tails that still satisfies the condition of existence of the eighth moments, and the scaled t_{14}-distribution as an example in-between. Note that for each of these three cases we have standardised cumulants of orders 1 and 2; that is, $\kappa_1 := \mathbb{E}(\varepsilon(t)) = 0$ and $\kappa_2 := \mathrm{var}(\varepsilon(t)) = 1$. Cumulants of higher order depend on the distribution of the innovations. If $\varepsilon(t) \sim \mathcal{N}(0,1)$, then cumulants of order higher than or equal to 3 vanish; that is, $\kappa_p := \mathrm{cum}_p(\varepsilon(t)) = 0$, for $p = 3, 4, \ldots$. In the case when $\varepsilon(t) \sim \nu^{-1/2}(\nu-2)^{1/2} t_\nu$, $\nu > 8$, we have that cumulants of orders $p = 3, 5, 7$ vanish due to symmetry, and that cumulants of order 4, 6 and 8 are

$$\kappa_4 = \frac{6}{\nu - 4}, \quad \kappa_6 = \frac{240}{(\nu-4)(\nu-6)}, \quad \text{and} \quad \kappa_8 = \frac{5040(5\nu - 22)}{(\nu-4)^2(\nu-6)(\nu-8)}.$$

R code and instructions to replicate the results of Section 3 are available on https://github.com/tobiaskley/ccf_bounds_replication_package.

3.2. Computing the Bound

We compute the bound from Theorem 2.3 in combination with (2.14) of the second method to bound $\tilde{Q}_t$, described in Section 2.4, where the data stem from an AR(1) process as described in Section 3.1. Details of how the bound is obtained in the case of the example are deferred to Section D. Note that, for given autoregressive parameter α, distribution of $\varepsilon(t)$, segment length n, and lag k the bound is still a function of m. We denote the bound by $B_n(m)$ to emphasise that it can be computed for different values of m. Further, we denote by $m^* := \arg\min_{m=0,1,\ldots,m_{\max}} B_n(m)$ the value of m for which the minimum is achieved. We have introduced the upper bound $m_{\max}$ as a stopping rule for computations, which we chose large enough such that $m^* < m_{\max}$ was satisfied in all cases of our example, meaning that the minimum is not obtained for $m = m_{\max}$. We chose $m_{\max} = 30$. In Figure 1, values of the bound $B_n(m)$ are shown as they depend on m, for different n and different distributions of $\varepsilon(t)$. Comparing the plots in Figure 1 from left to right, it can be seen that m^* increases very slowly as n increases. This is unsurprising because in this example of the causal AR(1) process, we are under Regime 2 explained in Section 2.6; recall the asymptotic considerations of Proposition 2.7 where $m^* = \mathcal{O}(\log n)$, which leads to $B_n(m^*) = \mathcal{O}(n^{-1/2}\log n)$. Comparing the plots in Figure 1 from top to bottom, it can be seen that the value of the bound gets larger as the tails get heavier. We expect this as well, as the cumulants of the distribution of the innovations $\varepsilon(t)$ become larger when we have distributions with heavier tails.

In Table I the values of the bound $B_n(m^*)$ for different values of k, α and n are shown for the case where $\varepsilon(t) \sim \sqrt{7/9}\, t_9$. The numbers for the cases where $\varepsilon(t) \sim \mathcal{N}(0,1)$ or $\varepsilon(t) \sim \sqrt{12/14}\, t_{14}$ are shown in Tables 5 and 7 respectively, in Section G. We chose to present the case with the heaviest tails in the main paper, because in this case the convergence of the estimator of the autocovariance and cross-covariance functions to the Gaussian limit is the slowest. We have omitted considering negative α, because in the case considered the results are the same as for $-\alpha$. It can be seen that the value of the bound increases as $|\alpha|$ increases. Comparing the bounds across tables, we see that for most cases the value of the bound is larger for heavier tails. It can be seen that the value of the bound decreases as n increases.

For comparison with our bound, as displayed in Tables I, 5 and 7, we also present simulated numbers for the true Wasserstein distance in Tables II, 6 and 8, with Tables 5–8 shown in Section G. Additional details about simulation of the true Wasserstein distance are deferred to Section G. By inspection of the numbers, it can be seen that, as expected, our bound is always larger than the true Wasserstein distance obtained by simulation.

4. PROOF OF THEOREM 2.1

Before the main proof of this section, we discuss a useful lemma that summarises the Stein's method result used in this article, which is applicable to a general local dependence condition. Consider a set of random variables $\{\xi_i, i \in J\}$, for a finite index set J. Then, the local dependence condition is

Journal of Time Series Analysis published by John Wiley & Sons Ltd. DOI: 10.1111/jtsa.12716

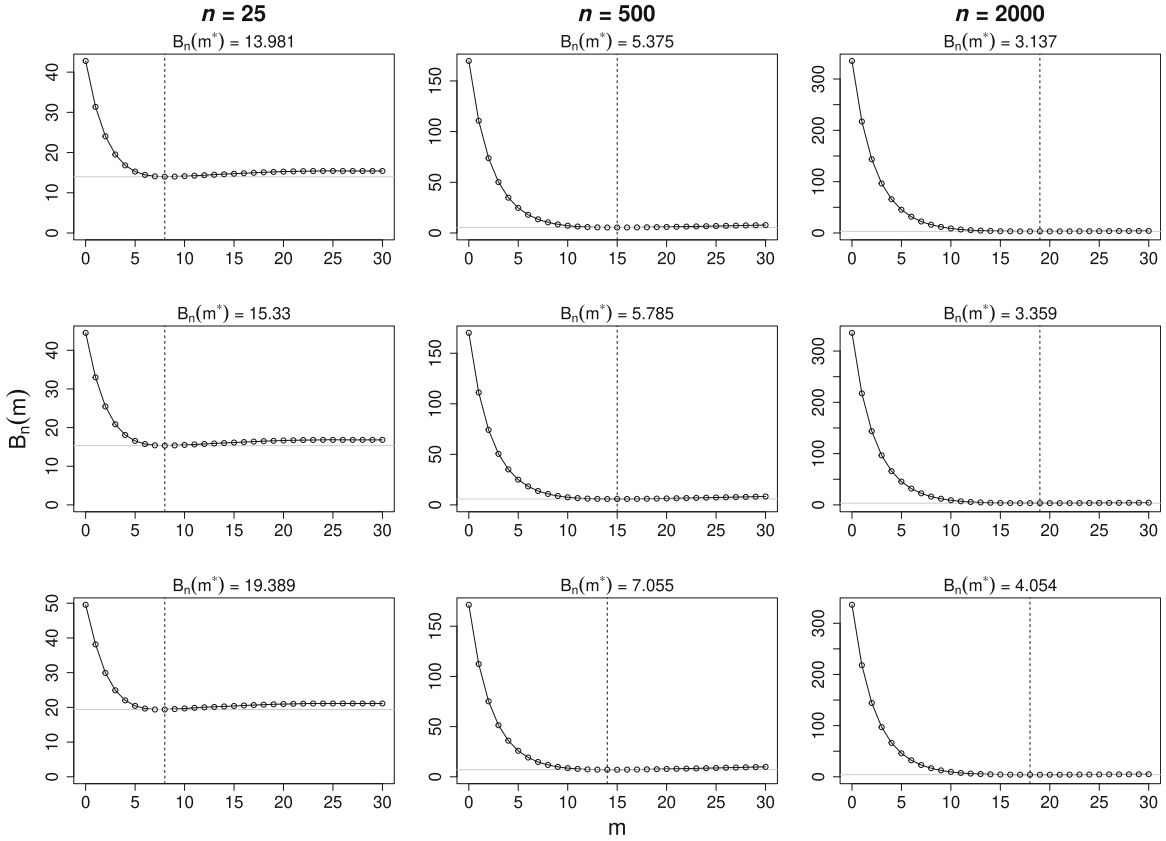

Figure 1. Value of the bound from Theorem 2.3 for empirical autocovariances of lag $k = 0$ and for an AR(1) process with $\alpha = 0.7$ as a function of m. The dashed vertical line indicates m^* where the minimum is achieved. The grey horizontal line indicates the minimum value. Top, middle and bottom row show $\varepsilon(t) \sim \mathcal{N}(0,1)$, $\varepsilon(t) \sim \sqrt{12/14}\, t_{14}$ and $\varepsilon(t) \sim \sqrt{7/9}\, t_9$ respectively. The left, centre and right columns show $n = 25$,500, and 2000 respectively

(LD) For each $i \in J$ there exist $A_i \subset B_i \subset J$ such that ξ_i is independent of $\{\xi_j : j \notin A_i\}$ and $\{\xi_j : j \in A_i\}$ is independent of $\{\xi_k : k \notin B_i\}$.

For any $A \subset J$, we now denote by

$$\xi(A) = \sum_{j \in A} \xi_j. \tag{4.18}$$

Remark 4.1. Consider an m-dependent sequence of random variables $X_1, \ldots, X_n$. Then the sets of random variables $\{X_j : j \le i\}$ and $\{X_j : j > i + m\}$ are independent for each $i = 1, \ldots, n$. Thus, (LD) is satisfied with $J := \{1, \ldots, n\}$, $A_i := \{\ell \in J : |\ell - i| \le m\}$, and $B_i := \{\ell \in J : |\ell - i| \le 2m\}$.

The following lemma gives an upper bound on the Wasserstein distance between the distribution of a sum of random variables satisfying condition (LD) above and the normal distribution. The random variables are assumed to have mean zero and the variance is not necessarily equal to one. The proof is in Section F.1 and is based on the steps followed for the proof of Theorem 4.13 in p. 134 of Chen *et al.* (2011).

Table I. Value of the bound from Theorem 2.3 in combination with (2.14), with $m = m^*$ to minimise the bound as described in Section 3.2, for empirical autocovariances, for a range of lags k and sample sizes n

k	$\alpha \mid n$	25	50	75	100	150	200	250	500	1000	2000
0	0	0.912	0.645	0.527	0.456	0.372	0.322	0.288	0.204	0.144	0.102
	0.1	11.003	9.294	8.822	8.707	7.509	6.658	6.091	4.779	3.770	2.773
	0.3	16.088	12.932	11.481	10.287	8.937	8.192	7.484	5.751	4.386	3.375
	0.5	16.952	13.518	11.760	10.689	9.245	8.365	7.706	5.955	4.579	3.514
	0.7	16.871	14.042	12.434	11.367	9.945	9.018	8.343	6.531	5.087	3.961
1	0	2.564	1.818	1.485	1.286	1.050	0.909	0.813	0.574	0.406	0.287
	0.1	7.711	5.701	4.808	4.285	3.686	3.350	3.118	2.283	1.708	1.326
	0.3	9.811	7.567	6.601	5.939	5.063	4.554	4.217	3.213	2.491	1.908
	0.5	12.512	9.980	8.716	7.861	6.825	6.133	5.671	4.394	3.402	2.641
	0.7	14.968	12.828	11.376	10.387	9.092	8.247	7.644	5.983	4.668	3.647
2	0	4.088	2.916	2.385	2.067	1.688	1.462	1.308	0.924	0.653	0.462
	0.1	10.398	7.659	6.409	5.667	4.804	4.309	3.925	2.833	2.074	1.561
	0.3	10.801	8.273	7.175	6.405	5.424	4.850	4.467	3.353	2.557	1.913
	0.5	12.211	9.739	8.459	7.632	6.592	5.921	5.471	4.210	3.233	2.486
	0.7	14.392	12.610	11.215	10.236	8.954	8.111	7.509	5.866	4.565	3.556

Note: The data stem from an AR(1) process with $\varepsilon(t) \sim \sqrt{7/9}\, t_9$ where α takes a range of values.

Table II. Value of the true 1-Wasserstein distance considered in Theorem 2.3 for empirical autocovariances, for a range of lags k and sample sizes n

k	$\alpha \mid n$	25	50	75	100	150	200	250	500	1000	2000
0	0	0.288	0.218	0.184	0.163	0.136	0.120	0.109	0.080	0.058	0.041
	0.1	0.294	0.222	0.188	0.166	0.139	0.123	0.111	0.081	0.059	0.042
	0.3	0.354	0.266	0.224	0.198	0.165	0.145	0.131	0.095	0.069	0.049
	0.5	0.536	0.401	0.336	0.296	0.246	0.216	0.194	0.140	0.101	0.072
	0.7	1.185	0.891	0.746	0.655	0.544	0.475	0.428	0.307	0.219	0.156
1	0	0.072	0.040	0.028	0.021	0.015	0.011	0.009	0.005	0.002	0.001
	0.1	0.103	0.069	0.055	0.047	0.038	0.032	0.029	0.020	0.014	0.010
	0.3	0.256	0.187	0.155	0.135	0.111	0.097	0.087	0.062	0.044	0.031
	0.5	0.524	0.384	0.319	0.279	0.230	0.200	0.180	0.128	0.091	0.065
	0.7	1.282	0.951	0.791	0.693	0.572	0.499	0.448	0.320	0.227	0.161
2	0	0.083	0.045	0.031	0.024	0.016	0.013	0.010	0.005	0.003	0.001
	0.1	0.088	0.049	0.034	0.026	0.018	0.014	0.012	0.006	0.004	0.002
	0.3	0.167	0.113	0.091	0.078	0.063	0.055	0.049	0.034	0.024	0.017
	0.5	0.449	0.329	0.272	0.237	0.195	0.170	0.152	0.109	0.077	0.055
	0.7	1.307	0.966	0.802	0.701	0.578	0.504	0.452	0.322	0.229	0.162

Note: The data stem from an AR(1) process with $\varepsilon(t) \sim \sqrt{7/9}\, t_9$ where α takes a range of values.

Lemma 4.2. Let $\{\xi_i, i \in J\}$ be an $\mathbb{R}$-valued random field with mean zero, satisfying Condition (LD). Denote $S := \sum_{i\in J} \xi_i$ and assume that $0 < \sigma^2 := \mathrm{var}(S) < \infty$. Then, with $\xi(A)$ as in (4.18), we have for the Wasserstein distance, d_{W}, defined in (1.6), that

$$d_{\mathrm{W}}(\mathcal{L}(S), \mathcal{N}(0,\sigma^2)) \le \frac{2}{\sigma^3} \sum_{i\in J} \left\{ \mathbb{E}\left| \left(\xi_i \xi(A_i) - \mathbb{E}(\xi_i \xi(A_i))\right) \xi(B_i) \right| + \frac{1}{2} \mathbb{E}\left| \xi_i (\xi(A_i))^2 \right| \right\}.$$

Journal of Time Series Analysis published by John Wiley & Sons Ltd. DOI: 10.1111/jtsa.12716

Proof of Theorem 2.1. First, for $Z^*(t) = X_a(t+k)X_b(t) - \frac{n}{n-k}\gamma$ we have that

$$\sqrt{n}\left(\hat{\gamma}_{ab}(k) - \gamma\right) = \sqrt{n}\left(\frac{1}{n}\sum_{t=1}^{n-k} X_a(t+k)X_b(t) - \gamma\right) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n-k} Z^*(t). \tag{4.19}$$

Recall that we use the convention to distinguish notation related to the m-dependent approximation with the tilde symbol (e.g. $\tilde{D}_{a,t}, \tilde{Z}(t), \tilde{Q}_t$). For $\tilde{Z}(t) = Y_a(t+k)Y_b(t) - \mathbb{E}[Y_a(t+k)Y_b(t)]$, the triangle inequality and (4.19) yield

$$d_{\mathrm{W}}\left(\mathcal{L}\left(\sqrt{n}\left(\hat{\gamma}_{ab}(k) - \gamma\right)\right), \mathcal{N}(0,\sigma^2)\right) = d_{\mathrm{W}}\left(\mathcal{L}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n-k} Z^*(t)\right), \mathcal{N}(0,\sigma^2)\right) \le D_1 + D_2 + D_3,$$

where

$$D_1 := d_{\mathrm{W}}\left(\mathcal{L}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n-k} Z^*(t)\right), \mathcal{L}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n-k} \tilde{Z}(t)\right)\right), \tag{4.20}$$

$$D_2 := d_{\mathrm{W}}\left(\mathcal{L}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n-k} \tilde{Z}(t)\right), \mathcal{N}\left(0, \tilde{\Sigma}_{ab}(k)\right)\right), \tag{4.21}$$

$$D_3 := d_{\mathrm{W}}\left(\mathcal{N}\left(0, \tilde{\Sigma}_{ab}(k)\right), \mathcal{N}\left(0, \sigma^2\right)\right), \tag{4.22}$$

where $\tilde{\Sigma}_{ab}(k)$ is as in (2.8). We now proceed to find upper bounds for (4.20), (4.21) and (4.22).

Bound for (4.20): With $h \in \mathcal{H}$, since $\|h\|_{\mathrm{Lip}} \le 1$, then

$$\begin{aligned}\left|\mathbb{E}\left[h\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n-k} Z^*(t)\right) - h\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n-k} \tilde{Z}(t)\right)\right]\right| &\le \frac{1}{\sqrt{n}}\sum_{t=1}^{n-k}\mathbb{E}\left|Z^*(t) - \tilde{Z}(t)\right| \\ &\le \frac{1}{\sqrt{n}}\sum_{t=1}^{n-k}\mathbb{E}\left|X_a(t+k)X_b(t) - Y_a(t+k)Y_b(t)\right| \\ &\quad + \frac{1}{\sqrt{n}}\sum_{t=1}^{n-k}\left|\frac{n}{n-k}\gamma - \mathbb{E}[Y_a(t+k)Y_b(t)]\right|. \end{aligned} \tag{4.23}$$

Next, we bound (4.23) $\le \frac{1}{\sqrt{n}}\sum_{t=1}^{n-k}\tilde{K}_t$, by a telescoping sum argument, made precise by Lemma F.1 with $\alpha = 1$, $p = 2$, $X_1 := X_a(t+k)$, $X_2 := X_b(t)$, $Y_1 := Y_a(t+k)$, and $Y_2 := Y_b(t)$. Therefore, we have

$$(4.20) \le \frac{1}{\sqrt{n}}\sum_{t=1}^{n-k}\tilde{K}_t + \frac{1}{\sqrt{n}}\sum_{t=1}^{n-k}\left|\frac{n}{n-k}\gamma - \mathbb{E}[Y_a(t+k)Y_b(t)]\right|. \tag{4.24}$$

Bound for (4.21): Here we use Stein's method. Let $S := \frac{1}{\sqrt{n}}\sum_{t=1}^{n-k}\tilde{Z}(t)$ and note that $\mathbb{E}(\tilde{Z}(t)) = 0$ and $\mathrm{var}(S) = \tilde{\Sigma}_{ab}(t)$. Lemma 4.2 then yields that

$$(4.21) \le \frac{2}{\left(\tilde{\Sigma}_{ab}(k)\right)^{3/2}}\sum_{t=1}^{n-k}\tilde{Q}_t. \tag{4.25}$$

Bound for (4.22): Using the results in pages 69 and 70 of Nourdin and Peccati (2012), it is straightforward to conclude that

$$(4.22) \leq \frac{\sqrt{2}}{\sqrt{\pi \Sigma_{ab}(k)}} \left| \Sigma_{ab}(k) - \tilde{\Sigma}_{ab}(k) \right|. \tag{4.26}$$

Combining (4.24), (4.25) and (4.26), yields the result of Theorem 2.1 as in (2.10). ■

5. DISCUSSION

In this article, we have obtained upper bounds on the Wasserstein distance between the true distribution of the estimator of the autocovariance and cross-covariance functions and their limiting Gaussian distribution for non-stationary and stationary data. Compared with existing results in the literature for general linear statistics and apart from the machinery employed (partly based on Stein's method) and the proof methodology followed, the results presented in this article are novel in three main aspects. First, the results of the article are applicable to non-stationary data sequences; see Theorem 2.1. Second, our focus is not only on rates of convergence, but the derived bounds are fully explicit in terms of the sample size, the lag and the constants depending on the time series model. This allows us to compute the bound in examples. Third, the assumptions that we have used are non-restrictive, and they are partly based on an m-dependence approximation of the original time series, which is convenient to work with in applications. In contrast, existing results are focused on rather more restrictive structures that are often probabilistic in nature and difficult to verify in practice, such as the case of strong and φ-mixing conditions or the discrete time martingales setting.

ACKNOWLEDGEMENT

The authors thank the Editor in Chief, Robert Taylor, the Co-Editor and two anonymous referees, whose reports helped to improve the paper. Open Access funding enabled and organized by Projekt DEAL.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

REFERENCES

Anastasiou A. 2017. Bounds for the normal approximation of the maximum likelihood estimator from m-dependent random variables. *Stat. Probab. Lett.* **129**:171–181.

Anastasiou A, Balasubramanian K, Erdogdu MA. 2019. *Normal Approximation for Stochastic Gradient Descent via Non-Asymptotic Rates of Martingale CLT*. In Proceedings of the Thirty-Second Conference on Learning Theory, *volume 99 of* Proceedings of Machine Learning Research, PMLR; 115–137.

Anderson TW. 1971. *The Statistical Analysis of Time Series*. Wiley, New York.

Aue A, Hörmann S, Horváth L, Reimherr M. 2009. Break detection in the covariance structure of multivariate time series models. *Ann. Statist.* **37**(6B):4046–4087.

Brillinger DR. 1975. *Time Series: Data Analysis and Theory*. Holt, Rinehart and Winston, Inc., New York.

Brockwell PJ, Davis RA. 2006. *Time Series: Theory and Methods*, 2nd ed. Springer, New York.

Chen LHY, Goldstein L, Shao Q-M. 2011. *Normal Approximation by Stein's Method*. Springer, Probability and its Applications. Springer-Verlog, Berlin, Heidelberg.

Dedecker J, Merlevède F, Rio E. 2009. Rates of convergence for minimal distances in the central limit theorem under projective criteria. *Electronic Journal of Probability* **14**:978–1011.

Journal of Time Series Analysis published by John Wiley & Sons Ltd. DOI: 10.1111/jtsa.12716

Dedecker J, Rio E. 2008. On mean central limit theorems for stationary sequences. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* **44**(4):693–726.
Fan X. 2019. Exact rates of convergence in some martingale central limit theorems. *Journal of Mathematical Analysis and Applications* **469**:1028–1044.
Fan X, Ma X. 2020. On the Wasserstein distance for a martingale central limit theorem. *Statistics and Probability Letters* **167**:108892.
Horváth L, Kokoszka P. 2008. Sample autocovariances of long-memory time series. *Bernoulli* **14**(2):405–418.
Jirak M. 2012. Simultaneous confidence bands for Yule-Walker estimators and order selection. *Annals of Statistics* **40**(1):494–528.
Jirak M. 2014. Simultaneous confidence bands for sequential autoregressive fitting. *Journal of Multivariate Analysis* **124**:130–149.
Jirak M. 2016. Berry-Esseen theorems under weak dependence. *Annals of Probability* **44**:2024–2063.
Kley T, Preuß P, Fryzlewicz P. 2019. Predictive, finite-sample model choice for time series under stationarity and non-stationarity. *Electronic Journal of Statistics* **13**(2):3710–3774.
Nourdin I, Peccati G. 2012. *Normal Approximations with Malliavin Calculus: From Stein's Method to Universality*. Cambridge Tracts in Mathematics, Vol. 192, Cambridge University Press, Cambridge.
Röllin A. 2018. On quantitative bounds in the mean martingale central limit theorem. *Statistics and Probability Letters* **138**:171–176.
Stein C. 1972. *A bound for the error in the normal approximation to the distribution of a sum of dependent random variables*. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 2, University of California Press, Berkeley; 586–602.
Sunklodas J. 2007. On normal approximation for strongly mixing random variables. *Acta Applicandae Mathematicae* **97**:251–260.
Sunklodas J. 2011. Some estimates of the normal approximation for φ-mixing random variables. *Lithuanian Mathematical Journal* **51**:260–273.

DOI: 10.1111/jtsa.12716 *Journal of Time Series Analysis* published by John Wiley & Sons Ltd.