# Finite sample distributional error bounds for empirical autocovariances and cross-covariances

Andreas Anastasiou* and Tobias Kley†

**Abstract**

The autocovariance and cross-covariance functions are a key ingredient for linear 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 in order to obtain the bound. As a worked example, we discuss how to compute the bound for 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.

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

## 1 Introduction

Assessing the quality of various asymptotic results has attracted a lot of interest in recent years. The error in distributional approximations is measured in terms of 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 and derive such a bound explicitly for the case of stationary, short-range dependent time series under a moment assumption. Our aim is to facilitate a bound where rate and explicit constants can be computed for a range of time series models.

Let $\{\boldsymbol{X}(t) : t \in \mathbb{Z}\}$ be a $d$-variate, stationary process, denoted by $\{\boldsymbol{X}(t)\}$, from which a sequence $\boldsymbol{X}(1), \ldots, \boldsymbol{X}(n)$ is available, with $\boldsymbol{X}(t)$ being $\mathbb{R}^d$-valued, $t = 1, \ldots, n$. The components of $\boldsymbol{X}(t)$ are denoted by $X_a(t)$, $a = 1, \ldots, d$. We are interested in the empirical cross-covariance and autocovariance, defined as

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

where $\bar{X}_j := \frac{1}{n} \sum_{t=1}^{n} X_j(t)$, $j = a, b$, denotes the empirical mean. For $k = -n+1, \ldots, -1$ we define $\tilde{\gamma}_{ab}(k) := \tilde{\gamma}_{ba}(-k)$. Other definitions, that are asymptotically equivalent under regularity conditions, also exist in the literature. For example, see Anderson (1971), Chapter 8, for

*Department of Mathematics and Statistics, University of Cyprus, P.O. Box: 20537, 1678, Nicosia, Cyprus, anastasiou.andreas@ucy.ac.cy

†Institute for Mathematical Stochastics,Georg-August-University of Göttingen, Goldschmidtstraße 7, 37077 Göttingen, Germany,
tobias.kley@uni-goettingen.de

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. Autocovariances are, for example, important for forecasting procedures (Kley et al., 2019).

In the case where the population means are known we may substitute the empirical means in (1.1) by their population counterparts

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

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

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

and $\hat{\gamma}_{ab}(k) := \hat{\gamma}_{ba}(-k)$, $k=-n+1,\dots,-1$. Under a summability condition for the cumulants, it can be shown that $\tilde{\gamma}_{ab}(k)$ and $\hat{\gamma}_{ab}(k)$ are consistent estimates for $\gamma_{ab}(k) := \mathbb{E}[X_a(t+k)X_b(t)]$. The asymptotic normality for the distribution of the estimator holds as well:

$$\sqrt{n}\big(\tilde{\gamma}_{ab}(k)-\gamma_{ab}(k)\big) \xrightarrow[n\to\infty]{} N, \quad N\sim\mathcal{N}(0,\Sigma_{ab}(k)); \tag{1.3}$$

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

$$\Sigma_{ab}(k) := \lim_{n\to\infty} \operatorname{var}\left(n^{-1/2}\sum_{t=1}^{n-k} X_a(t+k)X_b(t)\right) \tag{1.4}$$

depends on the second and fourth order moment structure of the underlying data; cf. eq. (7.6.11) in Brillinger (1975). In Section S5 (which as all other sections, lemmas, et cetera following the convention ‘S+number’ is from the supplementary material) we compute $\Sigma_{ab}(k)$ for the case of an AR(1) time series. It is straightforward to compute $\Sigma_{ab}(k)$ also for other time series models. For (1.3) to hold, assumptions limiting the intensity of the dependence structure and the moments of the random variables involved (such as the summability of cumulants) are necessary; cf. Horváth and Kokoszka (2008) for the discussion of cases where normality fails.

In this paper, we obtain upper bounds on the distributional distance between the true distribution of the empirical autocovariance and cross-covariance function and the limiting normal distribution. Before proceeding with an outline of our results, we provide, for ease of presentation, the general framework and the notation used throughout the paper. Firstly, for $\mathbb{R}^d$-valued random vectors $\boldsymbol{X}$ and $\boldsymbol{Y}$, we work on probability metrics of the form $d_{\mathcal{H}}(\boldsymbol{X},\boldsymbol{Y}) := \sup_{h\in\mathcal{H}}|\mathbb{E}[h(\boldsymbol{X})]-\mathbb{E}[h(\boldsymbol{Y})]|$, where $\mathcal{H}$ is a class of test functions. Examples of classes $\mathcal{H}$ of functions used in the paper are

$$\mathcal{H}_{\mathrm{K}} = \{\mathbf{1}(\cdot\le\boldsymbol{z})\,|\,\boldsymbol{z}\in\mathbb{R}^d\}, \quad \mathcal{H}_{\mathrm{W}} = \{h:\mathbb{R}^d\to\mathbb{R}\,|\,h \text{ is Lipschitz, } \|h\|_{\mathrm{Lip}}\le 1\}, \tag{1.5}$$

where $\|h\|_{\mathrm{Lip}} = \sup_{\boldsymbol{x}\neq\boldsymbol{y}}|h(\boldsymbol{x})-h(\boldsymbol{y})|/|\boldsymbol{x}-\boldsymbol{y}|$. In (1.5), $\mathcal{H}_{\mathrm{K}}$ gives the Kolmogorov distance and $\mathcal{H}_{\mathrm{W}}$ leads to the 1-Wasserstein metric to which we refer to as the Wasserstein metric from now on. The aforementioned metrics are denoted by $d_{\mathrm{K}}$ and $d_{\mathrm{W}}$, respectively. There is a remarkable relation between these two important metrics. Let $Y$ be any real-valued random variable and

$Z \sim \mathcal{N}(0,1)$. Then by Proposition 1.2 in Ross (2011) (see also Chen et al. (2011, Theorem 3.3)), we have that

$$d_{\mathrm{K}}(Y,Z) \leq \left(\frac{2}{\pi}\right)^{1/4}\sqrt{d_{\mathrm{W}}(Y,Z)}. \tag{1.6}$$

In Section 2.2 of Anastasiou and Gaunt (2020), a similar result is proven to hold for the multivariate case. Results for the Kolmogorov distance are particularly useful in the derivation of confidence intervals; see Remark 2.6. Our main purpose is to assess the quality of the distributional approximation in (1.3) through upper bounds on the Wasserstein distance between the actual distribution of the quantity of interest in the left-hand side of (1.3) and its limiting normal distribution; this is achieved in Theorem 2.3. Combining this result with Lemma S4.1 to bound the Wasserstein distance between $\sqrt{n}\big(\tilde{\gamma}_{ab}(k) - \gamma_{ab}(k)\big)$ and $\sqrt{n}\big(\hat{\gamma}_{ab}(k) - \gamma_{ab}(k)\big)$, we obtain a bound when the data are non-centered; more details can be found in Sections 2.4 and S4.

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. This advantage of Stein's method over other known techniques, such as characteristic functions, that can give answers on distributional distances, is what makes it suitable under the dependence setting promoted in the current paper.

There has been a lot of interest recently where $X_1, \ldots, X_n$ follow a specific dependence structure. Staying in the setting of explicit bounds but moving away from the $m$-dependence structure, Röllin (2018) provides bounds on the Wasserstein distance between the distribution of $W_n = \sum_{i=1}^n X_i$, where $X_i, \ldots, X_n$ 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 multivariate martingale difference sequence and the multivariate 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) and Sunklodas (2011) for the cases of strong and $\varphi$-mixing conditions, respectively.

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 $S_n = \sum_{i=1}^n c_{n,i}X_n$, 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. The obtained bounds are though not explicit, in the sense that they depend on a varying absolute constant not given explicitly.

The obtained distributional bounds in this paper are given in terms of the Wasserstein distance and they are fully explicit in terms of the sample size $n$, the lag $k$, as well as constants that are defined in the assumptions. Our results can be applied 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 $\mathcal{O}(n^{-1/2}\log n)$. Since the bounds are in terms of the Wasserstein distance, one can use the result in (1.6) and obtain explicit upper bounds in terms of the Kolmogorov distance that are of orders $\mathcal{O}(n^{-1/4})$ and $\mathcal{O}(n^{-1/4}(\log n)^{1/2})$, respectively. Such explicit bounds provide the opportunity to obtain confidence intervals and perform hypothesis tests that are legitimate in a non-asymptotic sense. The main result is given in Theorem 2.3 and a discussion on its order can be found in Remark 2.5 and, in more detail, in Section 2.5. Exact conservative confidence intervals for the parameter of interest are provided in Remark 2.6.

The paper is organized as follows. In Section 2.1 we state and discuss the key assumptions. In Section 2.2 we present our main result, an upper bound on the Wasserstein distance between the distribution of the autocovariance and cross-covariance functions $\hat{\gamma}_{ab}(k)$, defined in (1.2), and their limiting normal distribution. Sections 2.3 and 2.4 are devoted to computing the bounds in terms of moments of the $m$-dependent approximation or the original process, respectively. A detailed explanation of the order of the bound with respect to the sample size $n$ is given in Section 2.5. In Section 3, we apply our general results to the specific case of a causal autoregressive process of order 1. Section 4 concludes the paper with a brief discussion on our results. Technical details on the computation of the bound, step-by-step proofs for all results, as well as additional discussions, for example how our results can be used to obtain the respective results for the estimator defined in (1.1), are collected in the supplementary material.

# 2 Main results

## 2.1 Key assumptions

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

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

where the sum is with respect to all partitions $\nu := \{\nu_1, \ldots, \nu_p\}$ of $\{1, \ldots, r\}$; cf. Brillinger (1975). The assumptions used for the main result are collected below.

**Assumption 2.1.** *Let $\{\boldsymbol{X}(t)\}$ be $d$-variate, centered and weakly stationary, with $\|\boldsymbol{X}(t)\|_6 < \infty$.*

Further, we will use $m$-dependent processes to approximate the data. To this end, we assume

**Assumption 2.2.** *For every $m \in \mathbb{N}_0$ there exists an $m$-dependent, $d$-variate process $\{\boldsymbol{X}^{(m)}(t)\}$ with*

$$D_{a,q,m} := \|X_a(t) - X_a^{(m)}(t)\|_q < \infty, \quad a = 1, \ldots, d \tag{2.8}$$

*and not depending on $t$. The number $q \geq 1$ is specified whenever we refer to Assumption 2.2.*

We assume stationarity of $\{\boldsymbol{X}(t)\}$ in order to allow for a meaningful definition of $\gamma_{ab}(k)$. However, our method of proof does not require stationarity and in Section S1, we state and prove Theorem S1.3, a generalized version of Theorem 2.3 for non-stationary time series. 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; i. e., we require Assumption 2.2 with $q = 2$. For other results approximation in $L^q$ with $q > 2$ is required. In Section 2.4 we explain the general steps to obtaining a bound in terms of properties of the original process $\{\boldsymbol{X}(t)\}$. The precise results to pursue such a bound, Lemmas S3.3 and S3.4, are stated in Section S3. For Lemmas S3.3 and S3.4 we require Assumption 2.2 with $q = 4$ and $q = 6$, respectively. We do not require $\{\boldsymbol{X}^{(m)}(t)\}$ to be stationary or centered, though in applications this will often be the case. The assumption that $\|X_a(t) - X_a^{(m)}(t)\|_q$ is independent of $t$ is for notational convenience and satisfied in our example discussed in Section 3. In Section S5 we provide details on how to compute $D_{a,q,m}$ for the example. If $\{(\boldsymbol{X}(t), \boldsymbol{X}^{(m)}(t))\}$ is jointly stationary up to moments of order $q$, then Assumption 2.2 is satisfied. If (2.8) depends on $t$, define $D_{a,q,m} = \sup_{t=1,\dots,n} \|X_a(t) - X_a^{(m)}(t)\|_q$ and our results still hold; cf. Theorem S1.3. Our framework is similar in spirit to that in Aue et al. (2009), see their Assumption 2.1 and the examples provided in their Section 4 that illustrate how to apply the framework to several popular time series models. Note the following difference though. The quantity $D_{a,q,m}$ 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 $D_{a,q,m}$), 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, that are finite sample in nature, we only require that the quantity $D_{a,q,m}$ is finite.

## 2.2 The explicit upper bound

For the statement of the main result we denote the variance of the estimator obtained from a sample segment of length $n$ of the $m$-dependent approximation by

$$\tilde{\Sigma}_{ab}(k) := \operatorname{var}\left( n^{-1/2} \sum_{t=1}^{n-k} X_a^{(m)}(t+k) X_b^{(m)}(t) \right). \tag{2.9}$$

In addition, $\Sigma_{ab}(k)$ is as in (1.4) and the upper bound on the quantity of interest is given below.

**Theorem 2.3.** *Let $\{\boldsymbol{X}(t)\}$ be a $d$-variate, centered and weakly stationary process as in Assumption 2.1. For $m \in \mathbb{N}_0$, let the process $\{\boldsymbol{X}^{(m)}(t)\}$ be as in Assumption 2.2, which we assume holds with $q = 2$. For given $n \in \mathbb{N}$ and $k = 0, \dots, n-1$ assume that both $\Sigma_{ab}(k)$ and $\tilde{\Sigma}_{ab}(k)$, defined in* (1.4) *and* (2.9) *respectively, are positive. Denote $N \sim \mathcal{N}(0, \Sigma_{ab}(k))$ and let $A_t := \{\ell = 1, \dots, n-k : |\ell - t| \le m+k\}$ and $B_t := \{\ell = 1, \dots, n-k : |\ell - t| \le 2(m+k)\}$. Then,*

$$\begin{aligned} & d_{\mathrm{W}}\left(\sqrt{n}\left(\hat{\gamma}_{ab}(k) - \gamma_{ab}(k)\right), N\right) \\ & \le \frac{k}{\sqrt{n}} |\gamma_{ab}(k)| + \frac{\sqrt{2}}{\sqrt{\pi \Sigma_{ab}(k)}} \left| \Sigma_{ab}(k) - \tilde{\Sigma}_{ab}(k) \right| + \frac{2(n-k)}{\sqrt{n}} K_{2,m} + \frac{2}{\left(n\tilde{\Sigma}_{ab}(k)\right)^{3/2}} \sum_{t=1}^{n-k} \tilde{Q}_t \end{aligned} \tag{2.10}$$

*where $K_{2,m} := D_{a,2,m}\|X_b(0)\|_2 + D_{a,2,m}D_{b,2,m} + \|X_a(0)\|_2 D_{b,2,m}$, and*

$$\tilde{Q}_t := \mathbb{E}\left| \left( \tilde{Z}_t \sum_{j \in A_t} \tilde{Z}_j - \mathbb{E}\left( \tilde{Z}_t \sum_{j \in A_t} \tilde{Z}_j \right) \right) \sum_{j \in B_t} \tilde{Z}_j \right| + \frac{1}{2} \mathbb{E}\left| \tilde{Z}_t \left( \sum_{j \in A_t} \tilde{Z}_j \right)^2 \right| \tag{2.11}$$

*with $\tilde{Z}_t := X_a^{(m)}(t+k) X_b^{(m)}(t) - \mathbb{E}[X_a^{(m)}(t+k) X_b^{(m)}(t)]$.*

**Remark 2.4.** *The four terms that make up the right-hand side of* (2.10) *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 random variable* $N$ *is centered; (ii)* $\sqrt{2/(\pi\Sigma_{ab}(k))}|\Sigma_{ab}(k) - \tilde{\Sigma}_{ab}(k)|$, *is due to the fact that the variance of* $\hat{\gamma}_{ab}(k)$ *differs from* $n^{-1}\mathrm{var}(N)$, *(iii)* $2(n-k)n^{-1/2}K_{2,m}$, *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 S1.2.*

The following remark provides a brief discussion of the computation and of the order of the bound; detailed explanations are given in Section 2.5.

**Remark 2.5.** *At first glance, the bound might seem slightly complicated, especially due to the expression* $\tilde{Q}_t$. *In Section 2.3 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 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 optimization of the order of the bound with respect to* $n$ *depends on the underlying process* $\{\boldsymbol{X}(t)\}$. *In Section 2.5 we discuss two scenarios where the bound is of the order* $\mathcal{O}(n^{-1/2})$ *or* $\mathcal{O}(n^{-1/2}\log n)$, *respectively.*

Remark 2.6 below explains how one can employ our main result in (2.10) in order to obtain conservative $100(1-\alpha)\%$ confidence intervals, where $\alpha \in (0,1)$, for the unknown parameter $\gamma_{ab}(k)$.

**Remark 2.6.** *We have already discussed that the Kolmogorov distance relates directly to exact conservative confidence intervals. The result of Theorem 2.3 is expressed in terms of the Wasserstein distance. Using* (1.6) *leads to* $d_{\mathrm{K}}\left(\sqrt{n}\left(\hat{\gamma}_{ab}(k) - \gamma_{ab}(k)\right), N\right) \leq \left(\frac{2}{\pi}\right)^{1/4}\sqrt{B_W} =: B_K$, *where* $B_W$ *denotes the explicit bound in* (2.10) *for the Wasserstein distance. Therefore, for* $Z \sim \mathcal{N}(0,1)$ *and* $y \in \mathbb{R}$:

$$\begin{aligned}
&\left|\mathrm{pr}\left(\sqrt{\frac{n}{\Sigma_{ab}(k)}}\left(\hat{\gamma}_{ab}(k) - \gamma_{ab}(k)\right) \leq y\right) - \mathrm{pr}\left(Z \leq y\right)\right| \\
&= \left|\mathrm{pr}\left(\sqrt{n}\left(\hat{\gamma}_{ab}(k) - \gamma_{ab}(k)\right) \leq \sqrt{\Sigma_{ab}(k)}y\right) - \mathrm{pr}\left(N \leq \sqrt{\Sigma_{ab}(k)}y\right)\right| \leq B_K \\
&\Leftrightarrow -B_K \leq \mathrm{pr}\left(\sqrt{\frac{n}{\Sigma_{ab}(k)}}\left(\hat{\gamma}_{ab}(k) - \gamma_{ab}(k)\right) \leq y\right) - \mathrm{pr}\left(Z \leq y\right) \leq B_K. \qquad (2.12)
\end{aligned}$$

*For* $\Phi^{-1}(.)$ *the quantile function for the standard normal distribution, applying* (2.12) *to* $y = \Phi^{-1}\left(\frac{\alpha}{2} - B_K\right)$ *and to* $y = \Phi^{-1}\left(1 - \frac{\alpha}{2} + B_K\right)$ *yields*

$$\mathrm{pr}\left(\Phi^{-1}\left(\frac{\alpha}{2} - B_K\right) \leq \sqrt{\frac{n}{\Sigma_{ab}(k)}}\left(\hat{\gamma}_{ab}(k) - \gamma_{ab}(k)\right) \leq \Phi^{-1}\left(1 - \frac{\alpha}{2} + B_K\right)\right) \geq 1 - \alpha.$$

*Hence, if* $\Sigma_{ab}(k)$, *is known, then*

$$\left(\hat{\gamma}_{ab}(k) - \sqrt{\frac{\Sigma_{ab}(k)}{n}}\Phi^{-1}\left(1 - \frac{\alpha}{2} + B_K\right), \hat{\gamma}_{ab}(k) - \sqrt{\frac{\Sigma_{ab}(k)}{n}}\Phi^{-1}\left(\frac{\alpha}{2} - B_K\right)\right)$$

*is a conservative* $100(1-\alpha)\%$ *confidence interval for* $\gamma_{ab}(k)$.

## 2.3 The bound 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 $\boldsymbol{X}^{(m)}(t)$ such that $\tilde{\Sigma}_{ab}(k)$, $D_{a,2,m}$ and $D_{b,2,m}$ can be computed. Then, the only missing piece to obtain (2.10) is $\tilde{Q}_t$, defined in (2.11). 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. Firstly, if $\{\boldsymbol{X}^{(m)}(t)\}$ is stationarity (otherwise see below), we bound $\tilde{Q}_t$ in terms of $\|\boldsymbol{X}^{(m)}(0)\|_6$, which is finite under Assumptions 2.1 and 2.2 with $q = 6$. Secondly, 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 $\|\boldsymbol{X}^{(m)}(0)\|_8 < \infty$.

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

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

Employing the triangle inequality, a generalized version of Hölder's inequality, and the stationarity of $\boldsymbol{X}^{(m)}(t)$, the joint moments in the definition of $\tilde{Q}_t$ were broken up into moments of the marginals. If $\{\boldsymbol{X}^{(m)}(t)\}$ is not stationary we can bound with $\sup_{t=1,\ldots,n}\|\boldsymbol{X}^{(m)}(t)\|_6$ instead. The bound in (2.13) is particularly simple and straightforward to compute. In essence, we see a product of the marginal moments $\|X_j^{(m)}(0)\|_q$ 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 upon (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

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

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 $\{\boldsymbol{X}^{(m)}(t)\}$. These arguments are straightforward but tedious and therefore deferred to Section S2. 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 to 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 in Section S12.

## 2.4 The bound with respect to the original data

In Section 2.3 we explained computational details regarding a bound for the case when $\{\boldsymbol{X}^{(m)}(t)\}$, the $m$-dependent approximation, is known. The method described required the computation of joint moments of $\{\boldsymbol{X}^{(m)}(t)\}$ in order 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 (2.10) 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 $\{\boldsymbol{X}^{(m)}(t)\}$ might be inconvenient (e. g., when $\{\boldsymbol{X}^{(m)}(t)\}$ 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 (2.10) only depends on $\{\boldsymbol{X}(t)\}$, too. Therefore, the right-hand side of (2.10) being defined, amongst others, in terms of $\tilde{\Sigma}_{ab}(k)$ and $\tilde{Q}_t$, both depending on $\{\boldsymbol{X}^{(m)}(t)\}$, can be considered a discrepancy. The second reason is a practical one. In Section 2.5 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 $\boldsymbol{X}^{(m)}(t)$.

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

$$Q_t := \mathbb{E}\left|\left(Z_t\sum_{j\in A_t} Z_j - \mathbb{E}\left(Z_t\sum_{j\in A_t} Z_j\right)\right)\sum_{j\in B_t} Z_j\right| + \frac{1}{2}\mathbb{E}\left|Z_t\left(\sum_{j\in A_t} Z_j\right)^2\right|, \tag{2.15}$$

where $Z_t := X_a(t+k)X_b(t) - \mathbb{E}[X_a(t+k)X_b(t)]$. In Section S3 we provide results that can be used to derive a bound in terms of moments of $\{\boldsymbol{X}(t)\}$. Further, in Section S4, we discuss the case of non-centred data and provide a result to derive a bound for $d_{\mathrm{W}}\left(\sqrt{n}\left(\tilde{\gamma}_{ab}(k) - \gamma_{ab}(k)\right), N\right)$ in this case.

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

**Regime 1.** Let $\{\boldsymbol{X}(t)\}$ satisfy Assumption 2.1 and be $M$-dependent (for a fixed $M \in \mathbb{N}_0$).

**Regime 2.** Let $\{\boldsymbol{X}(t)\}$ satisfy Assumption 2.1, Assumption 2.2 with $q = 8$, and (2.16) and (2.17), below. We require summability of cumulants up to order 8; i. e., 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; i. e., there exist constants $K \geq 0$ and $\rho \in (0, 1)$ such that

$$D_{a,q,m} = \|X_a(0) - X_a^{(m)}(0)\|_q \leq K\rho^m, \quad \text{for } a = 1, \dots, d;\ q = 8. \tag{2.17}$$

**Remark 2.7.** *(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* $\boldsymbol{X}^{(m)}(t) = \boldsymbol{X}(t)$ *for* $m \geq M$*. Choosing the quantity* $m$ *in the bound* (2.10) *as* $m = \min\{M, n\}$ *we see that a stronger version of* (2.17) *is satisfied, where we have* $D_{a,q,m} = 0$ *for* $n \geq M$*.*

*(ii) As an example for Regime 2, consider a linear process* $\boldsymbol{X}(t) = \sum_{j=0}^{\infty} \boldsymbol{\Psi}(j)\boldsymbol{Z}(t-j)$ *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* $\|\boldsymbol{Z}(t)\|_q < \infty$*. Then, it can be shown that* (2.16) *holds and* (2.17) *holds with* $C := \|\boldsymbol{Z}(t)\|_q\rho/(1-\rho)$*. In particular, causal autoregressive processes are included.*

The following proposition gives the order of the bound (2.10) in Theorem 2.3. The proof is in Section S12.

**Proposition 2.8.** *(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* $m := C \log n$, $C \geq \frac{3}{2\log(1/\rho)}$*, 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 $X(t) = aX(t-1) + \varepsilon(t)$, with $|a| < 1$, where $\{\varepsilon(t)\}$ are i. i. d. and satisfy $\mathbb{E}|\varepsilon(t)|^8 < \infty$. We consider $a \in \{0, 0.1, 0.3, 0.5, 0.7\}$ and three cases for the distribution of the innovations:

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

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 8th moments, and the scaled $t_{14}$-distribution as an example in-between. Note that for each of these three cases we have standardized cumulants of orders 1 and 2; i. e., $\kappa_1 := \mathbb{E}(\varepsilon_t) = 0$ 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; i. e., $\kappa_p := \mathrm{cum}_p(\varepsilon_t) = 0$, 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 in this section 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 2.3, 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 S5. Note that, for given autoregressive parameter $a$, 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 emphasize 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 due to the fact that in this example of the causal AR(1) process, we are under Regime 2 explained in Section 2.5; recall the asymptotic considerations of Proposition 2.8 where $m^* \asymp \log n$, which

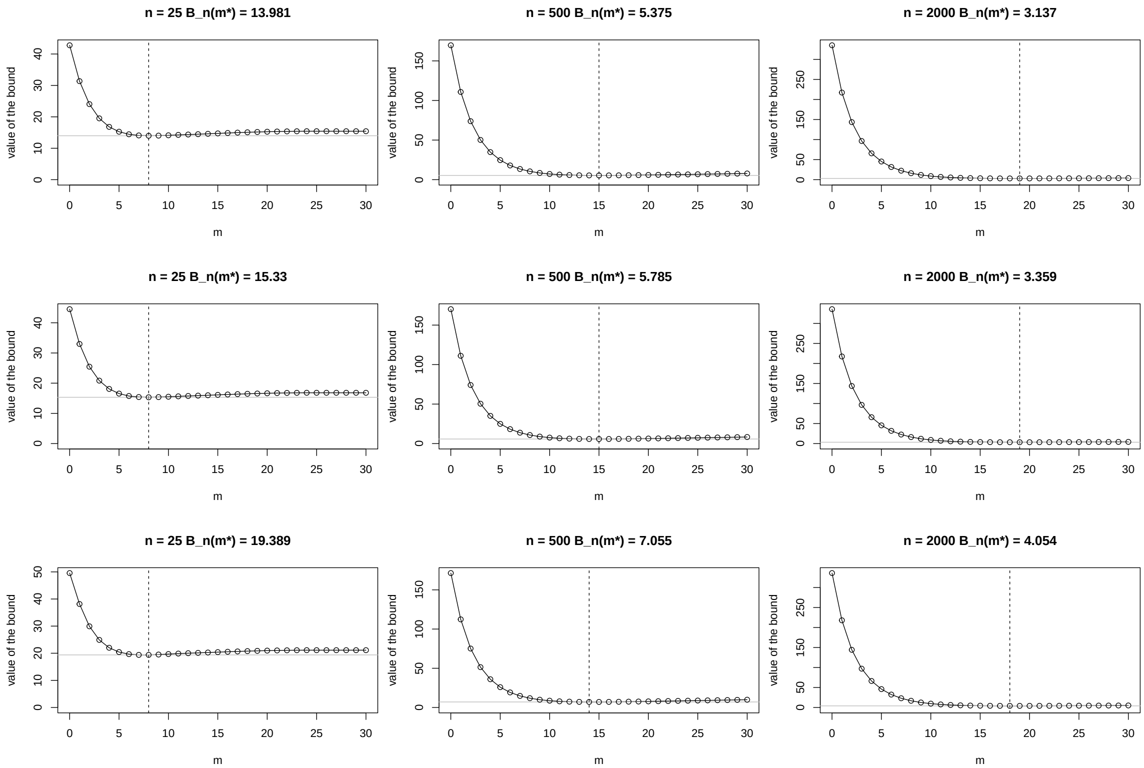


Figure 1: *Value of the bound from Theorem 2.3 for empirical autocovariances of lag* $k = 0$ *and for an AR(1) process with* $a = 0.7$ *as a function of* $m$. *The dashed vertical line indicates* $m^*$ *where the minimum is achieved. The gray horizontal line indicates the minimum value. Top, middle and 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. Left, center and right column show* $n = 25, 500, 2000$, *respectively.*

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 1 the values of the bound $B_n(m^*)$ for different values of $k$, $a$, 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 S3 and S5 in Section S13, respectively. 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 $a$, because the results are the same as for $-a$. It can be seen that the value of the bound increases as $|a|$ 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 1, S3, and S5, we also present simulated numbers for the true 1-Wasserstein distance in Tables 2, S4, and S6, with Tables S4 and S6 shown in Section S13. Additional details about simulation of the true 1-Wasserstein distance are deferred to Section S6. By inspection of the numbers, it can be seen that, as expected, our bound is always larger than the true 1-Wasserstein distance obtained by simulation.

| k | a \| 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 | 7.809 | 6.912 | 6.314 | 4.924 | 3.916 | 2.861 |
| | 0.3 | 16.855 | 13.964 | 12.635 | 11.375 | 9.806 | 8.930 | 8.328 | 6.304 | 4.813 | 3.645 |
| | 0.5 | 19.138 | 15.296 | 13.419 | 12.106 | 10.531 | 9.448 | 8.732 | 6.703 | 5.088 | 3.826 |
| | 0.7 | 19.389 | 15.910 | 13.973 | 12.683 | 11.021 | 9.941 | 9.176 | 7.055 | 5.370 | 4.054 |
| 1 | 0 | 2.564 | 1.818 | 1.485 | 1.286 | 1.050 | 0.909 | 0.813 | 0.574 | 0.406 | 0.287 |
| | 0.1 | 8.334 | 6.592 | 5.528 | 4.902 | 4.179 | 3.768 | 3.503 | 2.558 | 1.875 | 1.417 |
| | 0.3 | 12.834 | 9.713 | 8.334 | 7.544 | 6.387 | 5.671 | 5.192 | 3.938 | 2.958 | 2.199 |
| | 0.5 | 15.899 | 12.478 | 10.794 | 9.706 | 8.322 | 7.466 | 6.827 | 5.186 | 3.910 | 2.931 |
| | 0.7 | 18.138 | 15.038 | 13.175 | 11.952 | 10.343 | 9.299 | 8.548 | 6.543 | 4.961 | 3.734 |
| 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 | 3.985 | 2.882 | 2.104 | 1.579 |
| | 0.3 | 11.848 | 9.406 | 8.101 | 7.328 | 6.167 | 5.484 | 5.028 | 3.748 | 2.826 | 2.081 |
| | 0.5 | 15.405 | 12.092 | 10.412 | 9.304 | 7.999 | 7.107 | 6.507 | 4.922 | 3.701 | 2.769 |
| | 0.7 | 17.646 | 15.108 | 13.209 | 11.951 | 10.319 | 9.277 | 8.510 | 6.490 | 4.907 | 3.685 |

Table 1: 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 $n$. The data stems from an AR(1) process with $\varepsilon_t \sim \sqrt{7/9}\, t_9$ where $a$ takes a range of values.

| k | a \| n | 25 | 50 | 75 | 100 | 150 | 200 | 250 | 500 | 1000 | 2000 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0.288 | 0.218 | 0.184 | 0.163 | 0.137 | 0.120 | 0.109 | 0.079 | 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 | 0.132 | 0.096 | 0.069 | 0.050 |
| | 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 | 0.278 | 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 | 0.012 | 0.010 | 0.005 | 0.003 | 0.001 |
| | 0.1 | 0.088 | 0.048 | 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 | 0.196 | 0.170 | 0.153 | 0.109 | 0.077 | 0.055 |
| | 0.7 | 1.307 | 0.966 | 0.802 | 0.701 | 0.578 | 0.503 | 0.452 | 0.322 | 0.228 | 0.162 |

Table 2: Value of the true 1-Wasserstein distance considered in Theorem 2.3 for empirical autocovariances, for a range of lags $k$ and sample sizes $n$. The data stems from an AR(1) process with $\varepsilon_t \sim \sqrt{7/9}\, t_9$ where $a$ takes a range of values.

## 4 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. 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 paper are novel in two additional main aspects. Firstly, our focus is not only on rates of convergence (Dedecker et al., 2009; Fan, 2019), but the derived bound is explicit in terms of the sample size, the lag and the constants depending on the time series model. This allows to compute the bound in examples and to construct non-asymptotic confidence bands. Secondly, the assumptions that we have used are fully explicit, non-restrictive, and they are partly based on an $m$-dependence approximation of the original time series. 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 $\varphi$-mixing conditions (Sunklodas, 2007, 2011) or the discrete time martingales setting (Röllin, 2018; Fan and Ma, 2020).

## Supplementary material

Supplementary material includes the technical details on the computation of the bound for the example of the AR(1) process, additional discussions and explanations, as well as all the proofs.

## 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., K. Balasubramanian, and M. A. Erdogdu (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*, pp. 115–137.

Anastasiou, A. and R. E. Gaunt (2020). Wasserstein distance error bounds for the multivariate normal approximation of the maximum likelihood estimator. *arXiv:2005.05208*.

Anderson, T. W. (1971). *The Statistical Analysis of Time Series.* New York: Wiley.

Aue, A., S. Hörmann, L. Horváth, and M. Reimherr (2009, 12). Break detection in the covariance structure of multivariate time series models. *Ann. Statist. 37*(6B), 4046–4087.

Brillinger, D. R. (1975). *Time Series: Data Analysis and Theory.* New York: Holt, Rinehart and Winston, Inc.

Chen, L. H. Y., L. Goldstein, and Q.-M. Shao (2011). *Normal Approximation by Stein's Method.* Probability and its Applications. Springer.

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

Fan, X. (2019). Exact rates of convergence in some martingale central limit theorems. *Journal of Mathematical Analysis and Applications 469*, 1028–1044.

Fan, X. and X. Ma (2020). On the Wasserstein distance for a martingale central limit theorem. *Statistics and Probability Letters 167*, 108892.

Horváth, L. and P. Kokoszka (2008). Sample autocovariances of long-memory time series. *Bernoulli 14*(2), 405–418.

Jirak, M. (2016). Berry-Esseen theorems under weak dependence. *Annals of Probability 44*, 2024–2063.

Kley, T., P. Preuß, and P. Fryzlewicz (2019). Predictive, finite-sample model choice for time series under stationarity and non-stationarity. *Electronic Journal of Statistics 13*(2), 3710 – 3774.

Röllin, A. (2018). On quantitative bounds in the mean martingale central limit theorem. *Statistics and Probability Letters 138*, 171–176.

Ross, N. (2011). Fundamentals of Stein's method. *Probab. Surv. 8*, 210–293.

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*, Volume **2**, pp. 586–602. Berkeley: University of California Press.

Sunklodas, J. (2007). On Normal Approximation for Strongly Mixing Random Variables. *Acta Appl Math 97*, 251–260.

Sunklodas, J. (2011). Some estimates of the normal approximation for $\varphi$-mixing random variables. *Lithuanian Mathematical Journal 51*, 260–273.

# Supplement to "Finite sample distributional error bounds for empirical autocovariances and cross-covariances"

Andreas Anastasiou* and Tobias Kley†

**Abstract**

In this supplement we provide additional discussion and results. In Section S1 we state and prove Theorem S1.3 that generalizes Theorem 2.3 to the context of non-stationary time series. In Sections S2–S4 we provide technical details of the discussion and additional results regarding the general upper bound that were omitted from Section 2 of the main paper. In Sections S5 and S6 we provide technical details regarding computation and simulation that were omitted from Section 3 of the main paper. In Sections S7–S12 we provide proofs that were not included in the main paper. In Section S13 we provide tables with additional results for the example in Section 3 of the main paper.

## S1 Proof of Theorem 2.3

Before the main proof of this section, we discuss a useful lemma that summarizes the Stein's method result used in this paper which is based on a general local dependence condition. Consider a set of random variables $\{\xi_i, i \in J\}$, for a finite index set $J$. For any $A \subset J$ we denote

$$A^c = \{i \in J : i \notin A\}, \qquad \xi_A = \{\xi_i : i \in A\}.$$

Then, the local dependence condition is

(LD) For each $i \in J$ there exist $A_i \subset B_i \subset J$ such that $\xi_i$ is independent of $\xi_{A_i^c}$ and $\xi_{A_i}$ is independent of $\xi_{B_i^c}$.

Whenever this condition holds, we denote by

$$\eta_i = \sum_{j \in A_i} \xi_j, \qquad \tau_i = \sum_{j \in B_i} \xi_j. \tag{1}$$

**Remark S1.1** *Consider an $m$-dependent sequence of random variables $X_1, \ldots, X_n$. Then the sets of random variables $\{X_j : j \leq i\}$ and $\{X_j : j > i + m\}$ are independent for each $i = 1, \ldots, n$. Thus, (LD) is satisfied with $J := \{1, \ldots, n\}$, $A_i := \{\ell \in J : |\ell - i| \leq m\}$, and $B_i := \{\ell \in J : |\ell - i| \leq 2m\}$.*

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 S7.

*Department of Mathematics and Statistics, University of Cyprus, P.O. Box: 20537, 1678, Nicosia, Cyprus, anastasiou.andreas@ucy.ac.cy

†Institute for Mathematical Stochastics,Georg-August-University of Göttingen, Goldschmidtstraße 7, 37077 Göttingen, Germany,
tobias.kley@uni-goettingen.de

**Lemma S1.2** *Let $\{\xi_i, i \in J\}$ be an $\mathbb{R}$-valued random field with mean zero, satisfying Condition (LD). Denote $W := \sum_{i\in J} \xi_i$ and assume that $0 < \sigma^2 := \operatorname{var}(W) < \infty$. Then, with $N \sim \mathcal{N}(0, \sigma^2)$, and $\eta_i$ and $\tau_i$ as in* (1)*, we have for the Wasserstein-metric, $d_W$, that*

$$d_W(W, N) \leq \frac{2}{\sigma^3} \sum_{i\in J} \left\{ \mathbb{E}\left|(\xi_i\eta_i - \mathbb{E}(\xi_i\eta_i))\,\tau_i\right| + \frac{1}{2}\mathbb{E}\left|\xi_i\eta_i^2\right| \right\}.$$

We now state a result that generalizes Theorem 2.3 to the context of non-stationary time series, then explain why the more general result entails Theorem 2.3, and finally prove the more general result.

**Theorem S1.3** *Let $\boldsymbol{X}(1), \ldots, \boldsymbol{X}(n)$ and $\boldsymbol{X}^{(m)}(1), \ldots, \boldsymbol{X}^{(m)}(n)$ be $d$-variate random vectors with $\mathbb{E}(\boldsymbol{X}(t)) = 0$, $\|\boldsymbol{X}(t)\|_4 < \infty$ for $t = 1, \ldots, n$, and assume that $D_{a,2,m}(t) := \|X_a(t) - X_a^{(m)}(t)\|_2 < \infty$, $t = 1, \ldots, n$, $a = 1, \ldots, d$. For $n \in \mathbb{N}$ and $k = 0, \ldots, n-1$ assume that $\tilde{\Sigma}_{ab}(k)$ as in* (2.9) *is positive. Let $\gamma \in \mathbb{R}$ and $\Sigma > 0$ and denote $N \sim \mathcal{N}(0, \Sigma)$. Let $A_t$ and $B_t$ be as in Theorem 2.3. With $K_{2,m}(t,k) := D_{a,2,m}(t+k)\|X_b(t)\|_2 + D_{a,2,m}(t+k)D_{b,2,m}(t) + \|X_a(t+k)\|_2 D_{b,2,m}(t)$, and $\tilde{Q}_t$ defined in* (2.11)*, we have that*

$$\begin{aligned}
&d_W\left(\sqrt{n}\left(\hat{\gamma}_{ab}(k) - \gamma\right), N\right) \\
&\leq \frac{1}{\sqrt{n}} \sum_{t=1}^{n-k} K_{2,m}(t,k) + \frac{1}{\sqrt{n}} \sum_{t=1}^{n-k} \left| \frac{n}{n-k}\gamma - \mathbb{E}[X_a^{(m)}(t+k)X_b^{(m)}(t)] \right| \\
&\quad + \sqrt{\frac{2}{\pi\Sigma}} \left|\Sigma - \tilde{\Sigma}_{ab}(k)\right| + \frac{2n^{-3/2}}{\left(\tilde{\Sigma}_{ab}(k)\right)^{3/2}} \sum_{t=1}^{n-k} \tilde{Q}_t. \qquad (2)
\end{aligned}$$

To see that Theorem 2.3 follows from Theorem S1.3, choose $\gamma := \gamma_{ab}(k)$ and $\Sigma := \Sigma_{ab}(k)$. Note that the conditions of Theorem 2.3 imply that the conditions of Theorem S1.3 are satisfied. The bound in the stationary case then follows from $K_{2,m} = K_{2,m}(t,k)$ for all $t = 1, \ldots, n-k$ and

$$\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{t=1}^{n-k} \left| \frac{n}{n-k}\gamma - \mathbb{E}[X_a^{(m)}(t+k)X_b^{(m)}(t)] \right| \\
&\leq \frac{1}{\sqrt{n}} \sum_{t=1}^{n-k} \left| \frac{n}{n-k}\gamma_{ab}(k) - \gamma_{ab}(k) \right| + \frac{1}{\sqrt{n}} \sum_{t=1}^{n-k} \mathbb{E}\left| X_a(t+k)X_b(t) - X_a^{(m)}(t+k)X_b^{(m)}(t) \right| \\
&\leq \frac{n-k}{\sqrt{n}} \left| \frac{k}{n-k}\gamma_{ab}(k) \right| + \frac{n-k}{\sqrt{n}} K_{2,m},
\end{aligned}$$

where we have used $\mathbb{E}(X_a(t+k)X_b(t)) = \gamma_{ab}(k)$ and Lemma S3.1 with $\alpha = 1$ and $p = 2$.
***Proof of Theorem S1.3***. Firstly, for $Z_t^* = X_a(t+k)X_b(t) - \frac{n}{n-k}\gamma$ we have that

$$\sqrt{n}\left(\hat{\gamma}_{ab}(k) - \gamma\right) = \sqrt{n}\left(\frac{1}{n}\sum_{t=1}^{n-k} X_a(t+k)X_b(t) - \gamma\right) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n-k} Z_t^*. \qquad (3)$$

For $\tilde{Z}_t = X_a^{(m)}(t+k)X_b^{(m)}(t) - \mathbb{E}[X_a^{(m)}(t+k)X_b^{(m)}(t)]$, the triangle inequality and (3) yield

$$d_{\mathrm{W}}\left(\sqrt{n}\left(\hat{\gamma}_{ab}(k)-\gamma\right), N\right) = d_{\mathrm{W}}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n-k} Z_t^*, N\right) \leq d_{\mathrm{W}}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n-k} Z_t^*, \frac{1}{\sqrt{n}}\sum_{t=1}^{n-k}\tilde{Z}_t\right) \tag{4}$$

$$+\, d_{\mathrm{W}}\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n-k}\tilde{Z}_t, \tilde{N}\right) \tag{5}$$

$$+\, d_{\mathrm{W}}\left(\tilde{N}, N\right), \tag{6}$$

with $\tilde{N} \sim \mathcal{N}\left(0, \tilde{\Sigma}_{ab}(k)\right)$, where $\tilde{\Sigma}_{ab}(k)$ is as in (2.9). We now proceed to find upper bounds for (4), (5) and (6).

**Bound for** (4): With $h \in \mathcal{H}_W$, then

$$\left|h\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n-k} Z_t^*\right) - h\left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n-k}\tilde{Z}_t\right)\right| \leq \frac{\|h\|_{\mathrm{Lip}}}{\sqrt{n}}\sum_{t=1}^{n-k}\mathbb{E}\left|Z_t^* - \tilde{Z}_t\right|$$

$$\leq \frac{\|h\|_{\mathrm{Lip}}}{\sqrt{n}}\sum_{t=1}^{n-k}\mathbb{E}\left|X_a(t+k)X_b(t) - X_a^{(m)}(t+k)X_b^{(m)}(t)\right| \tag{7}$$

$$+ \frac{\|h\|_{\mathrm{Lip}}}{\sqrt{n}}\sum_{t=1}^{n-k}\left|\frac{n}{n-k}\gamma - \mathbb{E}[X_a^{(m)}(t+k)X_b^{(m)}(t)]\right|.$$

To bound (7), we employ Lemma S8.1, which is a generalized version of Lemma S3.1 without the assumed stationarity. Applying its result, with $\alpha = 1$, $p = 2$, $X_1 := X_a(t+k)$, $X_2 := X_b(t)$, $Y_1 := X_a^{(m)}(t+k)$, and $Y_2 := X_b^{(m)}(t)$ yields (7) $\leq \frac{\|h\|_{\mathrm{Lip}}}{\sqrt{n}}\sum_{t=1}^{n-k} K_{2,m}(t,k)$. Recalling the definition of $d_{\mathrm{W}}$ we have

$$(4) \leq \frac{1}{\sqrt{n}}\sum_{t=1}^{n-k} K_{2,m}(t,k) + \frac{1}{\sqrt{n}}\sum_{t=1}^{n-k}\left|\frac{n}{n-k}\gamma - \mathbb{E}[X_a^{(m)}(t+k)X_b^{(m)}(t)]\right|, \tag{8}$$

since $h \in \mathcal{H}_{\mathrm{W}}$ of (1.5) implies that $\|h\|_{\mathrm{Lip}} \leq 1$.

**Bound for** (5): Here we use Stein's method. Let $W := \sum_{t=1}^{n-k}\tilde{Z}_t/\sqrt{n}$ and note that $\mathbb{E}(\tilde{Z}_t/\sqrt{n}) = 0$ and $\mathrm{var}(W) = \tilde{\Sigma}_{ab}(t)$. Lemma S1.2 then yields that

$$(5) \leq \frac{2}{\left(\tilde{\Sigma}_{ab}(k)\right)^{3/2}}\sum_{t=1}^{n-k}\tilde{Q}_t, \tag{9}$$

**Bound for** (6): Using the results in pages 69 and 70 of Nourdin and Peccati (2012), it is straightforward to conclude that

$$(6) \leq \frac{\sqrt{2}\|h\|_{\mathrm{Lip}}}{\sqrt{\pi\Sigma_{ab}(k)}}\left|\Sigma_{ab}(k) - \tilde{\Sigma}_{ab}(k)\right|. \tag{10}$$

Combining now the results of (8), (9) and (10), yields the result of Theorem S1.3 as in (2). ■

## S2 Technical Details regarding Section 2.3

We now explain how the right-hand side of (2.14) can be computed from cumulants, up to order 8, of $\{\boldsymbol{X}^{(m)}(t)\}$ that is stationary. First, note that

$$\mathbb{E}\Big(\tilde{Z}_t \sum_{j \in A_t} \tilde{Z}_j\Big)^2 = \text{var}\Big(\tilde{Z}_t \sum_{j \in A_t} \tilde{Z}_j\Big) + \Big[\mathbb{E}\Big(\tilde{Z}_t \sum_{j \in A_t} \tilde{Z}_j\Big)\Big]^2,$$

and

$$\begin{aligned}
&\text{var}\Big(\tilde{Z}_t \sum_{j \in A_t} \tilde{Z}_j\Big) = \sum_{j_1 \in A_t} \sum_{j_2 \in A_t} \text{cov}\big(\tilde{Z}_t \tilde{Z}_{j_1}, \tilde{Z}_t \tilde{Z}_{j_2}\big) \\
&= \sum_{j_1 \in A_t} \sum_{j_2 \in A_t} \Big[\text{cum}(\tilde{Z}_t, \tilde{Z}_{j_1}, \tilde{Z}_t, \tilde{Z}_{j_2}) + \text{cum}(\tilde{Z}_t, \tilde{Z}_t)\text{cum}(\tilde{Z}_{j_1}, \tilde{Z}_{j_2}) + \text{cum}(\tilde{Z}_t, \tilde{Z}_{j_2})\text{cum}(\tilde{Z}_{j_1}, \tilde{Z}_t)\Big] \\
&= \sum_{j_1 \in A_t} \sum_{j_2 \in A_t} \text{cum}(\tilde{Z}_t, \tilde{Z}_t, \tilde{Z}_{j_1}, \tilde{Z}_{j_2}) + \tilde{C}(0)\text{var}\Big(\sum_{j \in A_t} \tilde{Z}_j\Big) + \Big[\mathbb{E}\Big(\tilde{Z}_t \sum_{j \in A_t} \tilde{Z}_j\Big)\Big]^2,
\end{aligned}$$

where

$$\begin{aligned}
\tilde{C}(u) := \text{cum}(\tilde{Z}_0, \tilde{Z}_u) = \text{cum}(&X_a^{(m)}(k), X_b^{(m)}(0), X_a^{(m)}(u+k), X_b^{(m)}(u)) \\
&+ \gamma_{aa}^{(m)}(u)\gamma_{bb}^{(m)}(u) + \gamma_{ab}^{(m)}(k-u)\gamma_{ab}^{(m)}(k+u),
\end{aligned} \tag{11}$$

with $\gamma_{ab}^{(m)}(k) := \text{cum}(X_a^{(m)}(k), X_b^{(m)}(0))$.

Noting that $A_t$ and $B_t$ in (2.11) are finite sets of consecutive integers of the form $\{s, \ldots, e\}$, we conclude that it suffices to compute or bound $\tilde{C}(0)$ and the following three quantities

$$\sum_{j_1=s}^{e} \sum_{j_2=s}^{e} \text{cum}(\tilde{Z}_t, \tilde{Z}_t, \tilde{Z}_{j_1}, \tilde{Z}_{j_2}), \quad \text{var}\Big(\sum_{j=s}^{e} \tilde{Z}_j\Big), \quad \text{and} \quad \mathbb{E}\Big(\tilde{Z}_t \sum_{j=s}^{e} \tilde{Z}_j\Big), \tag{12}$$

where $s \le t \le e$. The second and the third quantity in (12) only depend on second and fourth order moments of $\boldsymbol{X}^{(m)}(t)$ and, in this sense, are easier to obtain. We have that

$$\begin{aligned}
\text{var}\Big(\sum_{j=s}^{e} \tilde{Z}_j\Big) &= \sum_{j_1=s}^{e} \sum_{j_2=s}^{e} \text{cov}(\tilde{Z}_0, \tilde{Z}_{j_2-j_1}) = \sum_{|u| \le e-s} (e-s+1-|u|)\text{cov}(\tilde{Z}_0, \tilde{Z}_u) \\
&= (e-s+1) \sum_{|u| \le e-s} \Big(1 - \frac{|u|}{e-s+1}\Big)\tilde{C}(u),
\end{aligned}$$

with $\tilde{C}(u)$ as defined in (11), and similarly

$$\mathbb{E}\Big(\tilde{Z}_t \sum_{j=s}^{e} \tilde{Z}_j\Big) = \sum_{j=s}^{e} \text{cov}(\tilde{Z}_0, \tilde{Z}_{j-t}) = \sum_{u=s-t}^{e-t} \text{cov}(\tilde{Z}_0, \tilde{Z}_u) = \sum_{u=s-t}^{e-t} \tilde{C}(u).$$

For the first term in (12), we have, by Theorem 2.3.2 in Brillinger (1975), that

$$\sum_{j_1=s}^{e} \sum_{j_2=s}^{e} \text{cum}(\tilde{Z}_t, \tilde{Z}_t, \tilde{Z}_{j_1}, \tilde{Z}_{j_2}) = \sum_{j_1=s}^{e} \sum_{j_2=s}^{e} \sum_{\nu} \prod_{r=1}^{R} \text{cum}(\tilde{Y}_{ij} : (ij) \in \nu_r) \tag{13}$$

| | |
|---|---|
| (1,1) | (1,2) |
| (2,1) | (2,2) |
| (3,1) | (3,2) |
| (4,1) | (4,2) |

(a) Table of indices to be partitioned

| | |
|---|---|
| $\tilde{Y}_{1,1} := X_a^{(m)}(t+k)$ | $\tilde{Y}_{1,2} := X_b^{(m)}(t)$ |
| $\tilde{Y}_{2,1} := X_a^{(m)}(t+k)$ | $\tilde{Y}_{2,2} := X_b^{(m)}(t)$ |
| $\tilde{Y}_{3,1} := X_a^{(m)}(j_1+k)$ | $\tilde{Y}_{3,2} := X_b^{(m)}(j_1)$ |
| $\tilde{Y}_{4,1} := X_a^{(m)}(j_2+k)$ | $\tilde{Y}_{4,2} := X_b^{(m)}(j_2)$ |

(b) Variables of which the cumulants are considered

Table S1: Table of indices and variables to be partitioned for the addends of the sum in (13).

where $\tilde{Y}_{ij}$ are defined in Table S1(b) and the sum is with respect to all indecomposable partitions $\nu := \{\nu_1, \ldots, \nu_R\}$, where $|\nu_r| \geq 2$, of Table S1(a).

The number of ways to partition a table with 8 entries is the 8th Bell number $B_8 = 4140$. It is straight forward to enumerate all partitions algorithmically, for example using the R function `all_indecomposable_partitio` that we provide as part of the replication package, and to verify that only 545 of them are indecomposable and have at least two elements for each of the sets of the partitions and are therefore the ones to consider for distributions with $\mathbb{E}\left(\boldsymbol{X}^{(m)}(t)\right) = 0$. For the case where all cumulants of odd orders vanish, such as in the examples of Section 3, we only consider partitions where the number of elements in the sets is even and there are 249 such partitions. Further, for the case where $\boldsymbol{X}^{(m)}(t)$ is centered and normally distributed, only indecomposable partitions with sets of exactly 2 elements will be considered and there are 48 of them.

## S3 Technical Details regarding Section 2.4

Next, we provide Lemmas S3.3 and S3.4 that can be used to yield a bound on (2.10) in terms of quantities defined only via $\{\boldsymbol{X}(t)\}$ and the approximation errors $D_{j,q,m}$ from Assumption 2.2. For the statement of the results in this section, we will define, in (14), the quantities $K_{q,m}^{(\alpha)}$. Next, we state Lemma S3.1 which asserts that $K_{p,m}^{(\alpha)}$ is a bound for the approximation error (measured in $L^\alpha$) of products of $\{\boldsymbol{X}(t)\}$ by the corresponding products of $\{\boldsymbol{X}^{(m)}(t)\}$.

**Lemma S3.1** *Let $\alpha \geq 1$, $p \in \mathbb{N}$, $a_1, \ldots, a_p \in \{1, \ldots, d\}$, and $t_1, \ldots, t_p \in \mathbb{Z}$. Grant Assumption 2.1 and assume $\|\boldsymbol{X}(0)\|_{\alpha p} < \infty$. Grant Assumption 2.2 with $q = \alpha p$. Define*

$$K_{p,m}^{(\alpha)} = K_{p,m}^{(\alpha)}(\{a_1, \ldots, a_p\}) := \sum_{(\ell_1,\ldots,\ell_p)\in\Lambda_p} \prod_{i=1}^{p} D_{a_i,\alpha p,m}^{\ell_i} \|X_{a_i}(0)\|_{\alpha p}^{1-\ell_i}, \tag{14}$$

*where $D_{a_i,\alpha p,m}$ is as in* (2.8)*, $\Lambda_1 := \{1\}$, $\Lambda_2 := \{(1,0),(1,1),(0,1)\}$ and*

$$\Lambda_p := \{1\} \times \{0,1\}^{p-1} \cup \bigcup_{j=2}^{p-1} \{0\}^{j-1} \times \{1\} \times \{0,1\}^{p-j} \cup \{0\}^{p-1} \times \{1\}, \quad p = 3, 4, \ldots.$$

*Then, we have that*

$$\Big\| \prod_{i=1}^{p} X_{a_i}(t_i) - \prod_{i=1}^{p} X_{a_i}^{(m)}(t_i) \Big\|_\alpha \leq K_{p,m}^{(\alpha)}. \tag{15}$$

The proof of Lemma S3.1 is in Section S8. The case where $p := 2$, $a_1 := a$, and $a_2 := b$ with

$$K_{2,m}^{(\alpha)} := K_{2,m}^{(\alpha)}(\{a,b\}) := D_{a,2\alpha,m}\|X_b(0)\|_{2\alpha} + D_{a,2\alpha,m}D_{b,2\alpha,m} + D_{b,2\alpha,m}\|X_a(0)\|_{2\alpha},$$

appears repeatedly throughout the proofs, lemmas that follow, and it is also in the main theorem.

**Remark S3.2** *(i) The quantity* $K^{(\alpha)}_{p,m}$ *in* (14) *depends on the set of indices, but it does not depend on their order. This can be seen from the fact that the set* $\Lambda_p$ *is closed under permutations of the indices, which is the case as* $\Lambda_p$ *is closed under transpositions (swapping any two indices yields another element of* $\Lambda_p$*). Thus, the bound remains the same regardless of the order of the indices* $a_i$ *on the left-hand side of* (15)*.*
*(ii) In applications,* $D_{a,\alpha p,m}$ *will typically be of order* $o(1)$ *and* $\|X_a(0)\|_{\alpha p} = \mathcal{O}(1)$*, as* $m \to \infty$*. Therefore,* $K^{(\alpha)}_{p,m} = \mathcal{O}(\max_{a\in\{a_1,\ldots,a_q\}} D_{a,\alpha p,m})$ *which can be seen from the fact that each element of* $\Lambda_p$ *has at least one index 1.*
*(iii) To obtain a bound on the closeness of joint moments of* $\{\boldsymbol{X}(t)\}$ *to the joint moments of the* $m$*-dependent sequence* $\{\boldsymbol{X}^{(m)}(t)\}$ *we may employ Lemma S3.1 with* $\alpha = 1$

$$\left|\mathbb{E}\Big[\prod_{i=1}^{p} X_{a_i}(t_i)\Big] - \mathbb{E}\Big[\prod_{i=1}^{p} X^{(m)}_{a_i}(t_i)\Big]\right| \le \Big\|\prod_{i=1}^{p} X_{a_i}(t_i) - \prod_{i=1}^{p} X^{(m)}_{a_i}(t_i)\Big\|_1 \le K^{(1)}_{p,m}.$$

The second term of (2.10) depends on $\{\boldsymbol{X}^{(m)}(t)\}$ via $\tilde{\Sigma}_{ab}(k)$. Lemma S3.3, below, quantifies the error made by replacing $\tilde{\Sigma}_{ab}(k)$, defined in terms of $\{\boldsymbol{X}^{(m)}(t)\}$, by $n\text{var}\big(\hat{\gamma}_{ab}(k)\big)$ defined in terms of $\{\boldsymbol{X}(t)\}$. The proof of Lemma S3.3 is in Section S9.

**Lemma S3.3** *Grant Assumption 2.2 with* $q = 4$*, then*

$$|\tilde{\Sigma}_{ab}(k) - \Sigma_{ab}(k)| \le \Big|n\text{var}\big(\hat{\gamma}_{ab}(k)\big) - \Sigma_{ab}(k)\Big| + 2\frac{(n-k)^2}{n}\Big(\max_{i=a,b} D_{i,4,m}\Big)F_m,$$

*where*

$$\begin{aligned} F_m :=\; & |\gamma_{ab}(k)|\big(\|X_a(0)\|_2 + \|X_b(t)\|_2 + \min_{i=a,b} D_{i,2,m}\big) \\ & + \frac{1}{2}\Big(\max_{i=a,b} D_{i,2,m}\Big)\big(\|X_a(0)\|_2 + \|X_b(t)\|_2 + \min_{i=a,b} D_{i,2,m}\big)^2 \\ & + (\|X_a(0)\|_4 + D_{a,4,m})(\|X_b(0)\|_4 + D_{b,4,m})\Big(\|X_a(0)\|_4 + \|X_b(0)\|_4 + D_{a,4,m} + D_{b,4,m}\Big). \end{aligned} \tag{16}$$

To bound the fourth term in (2.10), which depends on $\{\boldsymbol{X}^{(m)}(t)\}$ via $\tilde{Q}_t$ and $\tilde{\Sigma}_{ab}(k)$, we will use Lemma S3.3, above, to replace $\tilde{\Sigma}_{ab}(k)$ by $\Sigma_{ab}(k)$ and Lemma S3.4, below, to replace $\tilde{Q}_t$, defined in terms of $\{\boldsymbol{X}^{(m)}(t)\}$, by

$$Q_t := \mathbb{E}\left|\left(Z_t\sum_{j\in A_t} Z_j - \mathbb{E}\bigg(Z_t\sum_{j\in A_t} Z_j\bigg)\right)\sum_{j\in B_t} Z_j\right| + \frac{1}{2}\mathbb{E}\left|Z_t\bigg(\sum_{j\in A_t} Z_j\bigg)^2\right|, \tag{17}$$

defined in terms of $\{\boldsymbol{X}(t)\}$ via $Z_t := X_a(t+k)X_b(t) - \mathbb{E}[X_a(t+k)X_b(t)]$. We state Lemma S3.4 first and defer the detailed discussion regarding a bound for the fourth term in (2.10) to the end of this section. The proof of Lemma S3.4 is in Section S10.

**Lemma S3.4** *Grant Assumption 2.2 with* $q = 6$*. Let* $\tilde{Q}_t$ *and* $\tilde{Z}_t$ *be as in and below* (2.11) *and* $Q_t$ *and* $Z_t$ *be as in and below* (17)*. Then,*

$$|Q_t - \tilde{Q}_t| \le K^{(3)}_{2,m}\Big(|A_t||B_t| + \frac{1}{2}|A_t|^2\Big)C_{1,m} + K^{(2)}_{2,m}|A_t||B_t|C_{2,m} + K^{(1)}_{2,m}|B_t|C_{3,m},$$

*with* $K^{(\alpha)}_{2,m} := K^{(\alpha)}_{2,m}(\{a,b\})$*,* $\alpha = 1,2,3$*, as defined in Lemma S3.1, and*

$$\begin{aligned} C_{1,m} &:= 6\big(2\|X_a(0)\|_6\|X_b(0)\|_6 + K^{(3)}_{2,m}\big)^2 + 2(K^{(3)}_{2,m})^2, \\ C_{2,m} &:= 8\Big(2\|X_a(0)\|_4\|X_b(0)\|_4 + K^{(2)}_{2,m}\Big)\Big(\|X_a(0)\|_2\|X_b(0)\|_2 + K^{(1)}_{2,m}\Big), \\ C_{3,m} &:= \Big|\sum_{j\in A_t}\mathbb{E}(Z_tZ_j)\Big|. \end{aligned}$$

**Remark S3.5** *Both* $|Q_t - \tilde{Q}_t|$ *and the bound in Lemma S3.4 depend on* $m$*. The bound has three addends, each of which is factored into three terms according to behavior with respect to* $m$*:*
*(i)* $K^{(\alpha)}_{2,m}$ *typically decreases as* $m$ *increases;*
*(ii) polynomials in the variables* $|A_t|$ *and* $|B_t|$ *that typically increase as* $m$ *increases;*
*(iii)* $C_{\alpha,m}$ *is bounded from below and above: for* $C_{1,m}$ *and* $C_{2,m}$ *this is the case if* $|K^{(\alpha)}_{2,m}|$ *is small enough, for* $C_{3,m}$ *if cumulants of* $\{\boldsymbol{X}(t)\}$ *are sumable; cf. discussion of the third term in* (12)*.*
*In applications, where* $K^{(\alpha)}_{2,m} = o(m^2)$*, the bound will vanish, uniformly with respect to* $t$*.*

Using Lemmas S3.3 and S3.4, we now bound $2(n\tilde{\Sigma}_{ab}(k))^{-3/2}\sum_{t=1}^{n-k}\tilde{Q}_t$ in (2.10) by a multiple of $2(n\Sigma_{ab}(k))^{-3/2}\sum_{t=1}^{n-k}Q_t$, where the last expression relies only on $\{\boldsymbol{X}(t)\}$. We have

$$\begin{aligned} \frac{2n^{-3/2}}{\Big(\tilde{\Sigma}_{ab}(k)\Big)^{3/2}}\sum_{t=1}^{n-k}\tilde{Q}_t &\le \frac{2n^{-3/2}}{\Big(\tilde{\Sigma}_{ab}(k)\Big)^{3/2}}\sum_{t=1}^{n-k}Q_t + \frac{2n^{-3/2}}{\Big(\tilde{\Sigma}_{ab}(k)\Big)^{3/2}}\sum_{t=1}^{n-k}|Q_t - \tilde{Q}_t| \qquad (18) \\ &= \left(\frac{2n^{-3/2}}{(\Sigma_{ab}(k))^{3/2}}\sum_{t=1}^{n-k}Q_t + \frac{2n^{-3/2}}{(\Sigma_{ab}(k))^{3/2}}\sum_{t=1}^{n-k}|Q_t - \tilde{Q}_t|\right)\left(\frac{\Sigma_{ab}(k)}{\tilde{\Sigma}_{ab}(k)}\right)^{3/2}. \end{aligned}$$

If for a specific example $\tilde{\Sigma}_{ab}(k)$ is tractable, but $\tilde{Q}_t$ is not, then it suffices to employ Lemma S3.4 and then, the bound in (18) is usable. In the case where $\tilde{\Sigma}_{ab}(k)$ is not tractable, but $\Sigma_{ab}(k)$ is tractable, we additionally employ Lemma S3.3 to find $n$ and $m$ large enough such that we have $|\tilde{\Sigma}_{ab}(k) - \Sigma_{ab}(k)| \le C\cdot\Sigma_{ab}(k)$, for some fixed $C \in (0,1)$, as this implies that $\big(\Sigma_{ab}(k)/\tilde{\Sigma}_{ab}(k)\big)^{3/2} \le 1/(1-C)^{3/2}$. Finally, note that we can bound $Q_t$ using either of the two strategies described in Section 2.3 to bound $\tilde{Q}_t$ with a computable quantity. The only adaptation required to employ these methods is to replace $\{\boldsymbol{X}^{(m)}(t)\}$ in the computations by $\{\boldsymbol{X}(t)\}$.

# S4 The bound for the non-centred sequence

The bound in Theorem 2.3 is for $\hat{\gamma}_{ab}(k)$, in (1.2), for data assumed to be centered. In practice, though, we often see $\tilde{\gamma}_{ab}(k)$, defined in (1.1), with the empirical centering added for the case when the data are not centered. In this section, we explain how an upper bound for $d_{\mathrm{W}}\left(\sqrt{n}\left(\tilde{\gamma}_{ab}(k) - \gamma_{ab}(k)\right), N\right)$, where $N \sim \mathcal{N}\left(0, \Sigma_{ab}(k)\right)$, can be obtained. Lemma S4.1 below, provides a bound on the Wasserstein distance between the distributions of centered and non-centered random variables and it can be employed to get an upper bound for the quantity of interest when the data are non-centered; this is explained in Remark S4.2(i), below. The proof of Lemma S4.1 is in Section S11.

**Lemma S4.1** *Assume that* $\mathbb{E}\boldsymbol{X}(t) = 0$ *and* $\sum_{u=-\infty}^{\infty} |u||\gamma_{jj}(u)| < \infty$ *for* $j = a, b$*, then*

$$d_{\mathrm{W}}\big(n^{1/2}(\tilde{\gamma}_{ab}(k)-\gamma_{ab}(k)), n^{1/2}(\hat{\gamma}_{ab}(k)-\gamma_{ab}(k))\big) \leq n^{-1/2}\Big(\sum_{u=-\infty}^{\infty} |\gamma_{aa}(u)|\Big)^{1/2}\Big(\sum_{u=-\infty}^{\infty} |\gamma_{bb}(u)|\Big)^{1/2}$$
$$+ \frac{|k|}{n^{3/2}}\left(\sum_{u=-\infty}^{\infty} |\gamma_{aa}(u)| + \frac{1}{n}\sum_{u=-\infty}^{\infty} |u||\gamma_{aa}(u)|\right)^{1/2}\left(\sum_{u=-\infty}^{\infty} |\gamma_{bb}(u)| + \frac{1}{n}\sum_{u=-\infty}^{\infty} |u||\gamma_{bb}(u)|\right)^{1/2}$$

**Remark S4.2** *(i) Lemma S4.1 facilitates use of Theorem 2.3 in the context of stationary time series, say* $\boldsymbol{Y}(t)$*, where the expectation is not assumed to be zero. This can be seen as follows. Define* $\boldsymbol{X}(t) := \boldsymbol{Y}(t) - \mathbb{E}(\boldsymbol{Y}(t))$ *and note that* $X_j(t) - \bar{X}_j = Y_j(t) - \bar{Y}_j$*. Thus,*

$$\tilde{\gamma}_{ab}^{Y}(k) := \frac{1}{n}\sum_{t=1}^{n-k}(Y_a(t+k) - \bar{Y}_a)(Y_b(t) - \bar{Y}_b) = \tilde{\gamma}_{ab}^{X}(k) := \frac{1}{n}\sum_{t=1}^{n-k}(X_a(t+k) - \bar{X}_a)(X_b(t) - \bar{X}_b),$$

*where* $\tilde{\gamma}_{ab}^{X}(k)$ *is defined in terms of the centered process to which we can apply our results. An upper bound on* $d_{\mathrm{W}}\left(\sqrt{n}\left(\tilde{\gamma}_{ab}^{Y}(k) - \gamma_{ab}(k)\right), N\right)$*, is obtain through the triangle inequality and*

$$\begin{aligned} &d_{\mathrm{W}}\left(\sqrt{n}\left(\tilde{\gamma}_{ab}^{Y}(k) - \gamma_{ab}(k)\right), N\right) \\ &\quad\leq d_{\mathrm{W}}\left(\sqrt{n}(\tilde{\gamma}_{ab}^{X}(k) - \gamma_{ab}(k)), \sqrt{n}(\hat{\gamma}_{ab}^{X}(k) - \gamma_{ab}(k))\right) + d_{\mathrm{W}}\left(\sqrt{n}\left(\hat{\gamma}_{ab}^{X}(k) - \gamma_{ab}(k)\right), N\right). \end{aligned} \quad (19)$$

*The results of Lemma S4.1 and Theorem 2.3 are applied to the terms in* (19) *to get the desired bound.*
*(ii) Lemma S4.1 asserts a bound of the order* $\mathcal{O}(n^{-1/2})$*, uniformly with respect to* $k$*.*
*(iii) Note the following, interesting, special cases: (a) if* $k = 0$*, the bound reduces to*

$$n^{-1/2}\Big(\sum_{u=-\infty}^{\infty} |\gamma_{aa}(u)| \sum_{u=-\infty}^{\infty} |\gamma_{bb}(u)|\Big)^{1/2}.$$

*(b) if the data are componentwise uncorrelated (i. e.,* $\gamma_{aa}(u) = \gamma_{bb}(u) = 0$ *for* $u \neq 0$*), then the bound reduces to*

$$\Big(n^{-1/2} + k\, n^{-3/2}\Big)\Big(\gamma_{aa}(0)\gamma_{bb}(0)\Big)^{1/2}.$$

# S5 Technical Details on the computation of the bound for causal AR(1)

Here, we provide the technical details on how to obtain the bound in the AR(1) example which has been discussed in Section 3.1. The bound in Theorem 2.3 is

$$\frac{k}{\sqrt{n}}|\gamma_{ab}(k)| + \frac{\sqrt{2}}{\sqrt{\pi\Sigma_{ab}(k)}}\left|\Sigma_{ab}(k) - \tilde{\Sigma}_{ab}(k)\right| + \frac{2(n-k)}{\sqrt{n}}K_{2,m} + \frac{2n^{-3/2}}{\left(\tilde{\Sigma}_{ab}(k)\right)^{3/2}}\sum_{t=1}^{n-k}\tilde{Q}_t.$$

Since we are in the setting of univariate time series we drop the indices $a$ and $b$.

In this section we treat the case of a univariate (i. e., $a = b$) AR(1) time series

$$X(t) = aX(t-1) + \varepsilon_t = \sum_{j=0}^{\infty} a^j \varepsilon_{t-j},$$

with parameter $a \in (-1,1)$ and $\{\varepsilon_t\}$ i. i. d. such that $\kappa_p := \mathrm{cum}_p(\varepsilon_t)$ exists for $p = 8$ and $\kappa_p = 0$ when $p$ is odd. We assume that $\kappa_1 = \mathbb{E}\varepsilon_t = 0$ and use the following moving average process as the $m$-dependent approximation

$$X^{(m)}(t) = \sum_{j=0}^{m} a^j \varepsilon_{t-j}.$$

We first treat the first and third term in the bound and the second and fourth term after that.

For the first and third term it suffices to compute

- $D_{q,m} := \left[\mathbb{E}\left(X(0) - X^{(m)}(0)\right)^q\right]^{1/q}$, for $q = 2, 4, 6, 8$, and
- $\mathbb{E}\left(X(t+k)X(t)\right)^q$, for $q = 1, 2, 3, 4$.

Note that $\|X(t)\|_{2q}$ follows from $\mathbb{E}\left(X(t+k)X(t)\right)^q$ with $k = 0$

For $D_{q,m}$, using Theorem 2.3.2 in Brillinger (1975) about cumulants of products, we obtain

$$\begin{aligned}
&\mathbb{E}\left(X(0) - X^{(m)}(0)\right)^q = \mathbb{E}\left(\sum_{j=m+1}^{\infty} a^j \varepsilon_{-j}\right)^q \\
&= \sum_{j_1=m+1}^{\infty} \cdots \sum_{j_q=m+1}^{\infty} a^{j_1+\cdots+j_q} \sum_{\{\nu_1,\ldots,\nu_R\}} \prod_{r=1}^{R} \mathrm{cum}(\varepsilon(-j_\ell) : \ell \in \nu_r) \\
&= \sum_{\{\nu_1,\ldots,\nu_R\}} \prod_{r=1}^{R} \sum_{j=m+1}^{\infty} a^{|\nu_r| j} \kappa_{|\nu_r|} = \sum_{\{\nu_1,\ldots,\nu_R\}} \prod_{r=1}^{R} \kappa_{|\nu_r|} a^{|\nu_r|(m+1)} \sum_{j=0}^{\infty} a^{|\nu_r| j} \\
&= a^{q(m+1)} \sum_{\{\nu_1,\ldots,\nu_R\}} \prod_{r=1}^{R} \frac{\kappa_{|\nu_r|}}{1 - a^{|\nu_r|}}
\end{aligned}$$

where the sum is with respect to partitions $\{\nu_1, \ldots, \nu_R\}$ of $\{1, \ldots, q\}$. In particular, for $q = 2, 4, 8$, and since cumulants of odd orders vanish for the distributions considered, we have

$$\begin{aligned}
D_{2,m} &= a^{m+1}\Big(\frac{\kappa_2}{1-a^2}\Big)^{1/2}, \\
D_{4,m} &= a^{m+1}\Big(\frac{\kappa_4}{1-a^4} + 3\Big(\frac{\kappa_2}{1-a^2}\Big)^2\Big)^{1/4}, \\
D_{8,m} &= a^{m+1}\Big(\frac{\kappa_8}{1-a^8} + 28\frac{\kappa_6}{1-a^6}\frac{\kappa_2}{1-a^2} + 35\Big(\frac{\kappa_4}{1-a^4}\Big)^2 \\
&\qquad + 210\frac{\kappa_4}{1-a^4}\Big(\frac{\kappa_2}{1-a^2}\Big)^2 + 105\Big(\frac{\kappa_2}{1-a^2}\Big)^4\Big)^{1/8}
\end{aligned}$$

Similarly, we obtain that for $a \neq 0$, we have

$$\mathbb{E}\left(X(t+k)X(t)\right)^q = a^{q|k|} \sum_{\{\nu_1,\ldots,\nu_R\}} \prod_{r=1}^{R} \frac{\kappa_{|\nu_r|}}{1 - a^{|\nu_r|}},$$

where the sum is with respect to partitions $\{\nu_1, \ldots, \nu_R\}$ of $\{1, \ldots, 2q\}$.

In particular, for $q = 1, 2, 3, 4$ we have,

$$\begin{aligned}
\mathbb{E}\left(X(t+k)X(t)\right) &= a^{|k|}\frac{\kappa_2}{1-a^2} =: \gamma(k) \\
\mathbb{E}\left(X(t+k)X(t)\right)^2 &= a^{2|k|}\Big(\frac{\kappa_4}{1-a^4} + 3\Big(\frac{\kappa_2}{1-a^2}\Big)^2\Big) \\
\mathbb{E}\left(X(t+k)X(t)\right)^3 &= a^{3|k|}\Big(\frac{\kappa_6}{1-a^6} + 15\frac{\kappa_4}{1-a^4}\frac{\kappa_2}{1-a^2} + 30\Big(\frac{\kappa_2}{1-a^2}\Big)^3\Big) \\
\mathbb{E}\left(X(t+k)X(t)\right)^4 &= a^{4|k|}\Big(\frac{\kappa_8}{1-a^8} + 28\frac{\kappa_6}{1-a^6}\frac{\kappa_2}{1-a^2} + 35\Big(\frac{\kappa_4}{1-a^4}\Big)^2 \\
&\qquad + 210\frac{\kappa_4}{1-a^4}\Big(\frac{\kappa_2}{1-a^2}\Big)^2 + 105\Big(\frac{\kappa_2}{1-a^2}\Big)^4\Big)
\end{aligned} \tag{20}$$

Further, for $a = 0$, we have

$$\mathbb{E}\left(X(t+k)X(t)\right)^q = \begin{cases} \mathbb{E}\varepsilon_t^{2q} & k = 0 \\ (\mathbb{E}\varepsilon_t^q)^2 & k \neq 0. \end{cases}$$

Thus, for the case of $a = 0$ and $k = 0$ the expressions in (20) are correct when we apply the convention that $0^0 = 1$. For $a = 0$ and $k \neq 0$ we have

$$\begin{aligned}
\mathbb{E}\left(X(t+k)X(t)\right) &= 0 \\
\mathbb{E}\left(X(t+k)X(t)\right)^2 &= (\kappa_2)^2 \\
\mathbb{E}\left(X(t+k)X(t)\right)^3 &= 0 \\
\mathbb{E}\left(X(t+k)X(t)\right)^4 &= \Big(\kappa_4 + 3(\kappa_2)^2\Big)^2
\end{aligned}$$

This covers the relevant pieces for the first and third term in the bound and we now turn our attention to the second and fourth term. To this end we first discuss how to compute joint cumulants of the $m$-dependent approximation in our example and then turn our attention to the second term of the bound.

**Joint cumulants of $\{X^{(m)}(t)\}$.** To compute the bound we will frequently require to know the joint cumulants of $\{X^{(m)}(t)\}$. For $u_1, \ldots, u_p \in \mathbb{Z}$ denote $M := \min\{u_1, \ldots, u_p\}$, $R := \max\{u_1, \ldots, u_p\} - M$, and $S := \sum_{i=1}^p (u_i - M)$. Then, we have that

$$\begin{aligned}
&\text{cum}(X^{(m)}(u_1), \ldots, X^{(m)}(u_p)) \\
&= \sum_{j_1=0}^{m} \cdots \sum_{j_p=0}^{m} a^{j_1+\ldots+j_p}\text{cum}(\varepsilon_{u_1-M-j_1}, \ldots, \varepsilon_{u_p-M-j_p}) \\
&= \begin{cases} \sum_{j_s=0}^{m-R} a^{pj_s+S}\text{cum}(\varepsilon_{-j_s}, \ldots, \varepsilon_{-j_s}) = \kappa_p a^S \dfrac{1-a^{p(m-R+1)}}{1-a^p} & R \leq m \\ 0 & R > m. \end{cases}
\end{aligned} \tag{21}$$

For the second equality, note that $u_\ell - M \geq 0$, for all $\ell = 1, \ldots, p$, and $u_s - M = 0$ for at least one $s = 1, \ldots, p$. We now argue that for the cumulants $\text{cum}(\varepsilon_{u_1-M-j_1}, \ldots, \varepsilon_{u_p-M-j_p})$ with $(j_1, \ldots, j_p) \in \{1, \ldots, m\}^p$ and $j_s$ such that $u_s = M$, at most one of these cumulants is non zero. To this end, note that we have to have $u_s - M - j_s = -j_s = u_\ell - M - j_\ell$, for all $\ell \in 1, \ldots, p$, including $\ell = s$, which implies that $j_\ell = j_s + u_\ell - M \in \{1, \ldots, m\}$ for the cumulants that do not vanish. Hence, to ensure that $j_\ell \leq m$

for all $\ell$, we have to exclude $j_s > m - R$. Further, we have that $u_{\ell_1} - j_{\ell_1} = u_{\ell_2} - j_{\ell_2}$, which implies that we have to require that

$$\max_{\ell_1,\ell_2\in\{1,\dots,p\}} |j_{\ell_1} - j_{\ell_2}| = \max_{\ell_1,\ell_2\in\{1,\dots,p\}} |u_{\ell_1} - u_{\ell_2}| = R \le m.$$

A very similar, but simpler, argument can be applied to obtain

$$\mathrm{cum}(X(u_1),\dots,X(u_p)) = \kappa_p \frac{a^S}{1-a^p}.$$

**Computation of $\Sigma_{ab}(k)$ and $\tilde{\Sigma}_{ab}(k)$.** To do so, we note that

$$\begin{aligned}\Sigma_{ab}(k) &= \sum_{u=-\infty}^{\infty}\gamma^2(u) + \sum_{u=-\infty}^{\infty}\gamma(u-k)\gamma(u+k) + \frac{\kappa_4}{(\kappa_2)^2}\gamma^2(k)\\ &= (\kappa_2)^2\frac{1+a^2+a^{2|k|}\big(1+a^2+2k(1-a^2)\big)}{(1-a^2)^3} + \kappa_4\frac{a^{2|k|}}{(1-a^2)^2}.\end{aligned}$$

Further, we have

$$\begin{aligned}\tilde{\Sigma}_{ab}(k) &= n^{-1}\mathrm{var}\Big(\sum_{t=1}^{n-k} X^{(m)}(t+k)X^{(m)}(t)\Big)\\ &= \frac{1}{n}\sum_{t_1=1}^{n-k}\sum_{t_2=1}^{n-k}\tilde{C}(t_1-t_2) = \frac{n-k}{n}\sum_{|u|\le\min\{n-k,m\}}\Big(1-\frac{|u|}{n-k}\Big)\tilde{C}(u),\end{aligned}$$

with $\tilde{C}(u)$ defined in (2.7). The cumulants of $\{X^{(m)}(t)\}$ that appear in the definition of $\tilde{C}(u)$ can be computed as described in (21) of the previous section.

It remains to discuss the fourth term of the bound. Recall that in Section 2.2 we discussed two methods to compute a bound for $\tilde{Q}_t$. In this section we use the second method and discuss how to compute the exact value of the right hand side of (2.14).

**Computation of the bound for $\tilde{Q}_t$.** We have

$$\begin{aligned}\tilde{Q}_t &\le \mathrm{var}\Big(\tilde{Z}_t\sum_{j\in A_t}\tilde{Z}_j\Big)^{1/2}\mathrm{var}\Big(\sum_{j\in B_t}\tilde{Z}_j\Big)^{1/2} + \frac{1}{2}\Big[\mathbb{E}\Big(\tilde{Z}_t\sum_{j\in A_t}\tilde{Z}_j\Big)^2\Big]^{1/2}\mathrm{var}\Big(\sum_{j\in A_t}\tilde{Z}_j\Big)^{1/2}\\ &= \Big(M_{1,t} + \tilde{C}(0)\times M_{2a,t} + M_{3,t}^2\Big)^{1/2}\times M_{2b,t}^{1/2} + \frac{1}{2}\Big(M_{1,t}+\tilde{C}(0)\times M_{2a,t} + 2M_{3,t}^2\Big)^{1/2}M_{2a,t}^{1/2},\end{aligned}$$

where

$$\begin{aligned}M_{1,t} &:= \sum_{j_1\in A_t}\sum_{j_2\in A_t}\sum_{\nu}\prod_{r=1}^{R}\mathrm{cum}(\tilde{Y}_{ij} : (ij)\in\nu_r) = \sum_{\ell_1\in A_t-t}\sum_{\ell_2\in A_t-t}\tilde{D}(\ell_1,\ell_2),\\ M_{2a,t} &:= |A_t|\sum_{|u|\le|A_t|-1}\Big(1-\frac{|u|}{|A_t|}\Big)\tilde{C}(u), \quad M_{2b,t} := |B_t|\sum_{|u|\le|B_t|-1}\Big(1-\frac{|u|}{|B_t|}\Big)\tilde{C}(u),\\ M_{3,t} &:= \sum_{u\in A_t}\tilde{C}(u-t), \qquad \tilde{D}(u_1,u_2) := \sum_{\nu}\prod_{r=1}^{R}\mathrm{cum}(\tilde{Y}_{ij} : (ij)\in\nu_r)\end{aligned}$$

where the sums in the definitions of $M_{1,t}$ and $\tilde{D}(u_1, u_2)$ are with respect to the indecomposable partitions $\nu$ of Table S1(a) and the terms $\tilde{Y}_{ij}$ are defined in Table S1(b). We use the original indexing for the definition of $M_{1,t}$ and an indexing shifted by $t$ for the definition of $\tilde{D}(u_1, u_2)$. More precisely, of the following definitions we use the first in $M_{1,t}$ and the second in $\tilde{D}(u_1, u_2)$:

$$\begin{aligned}
&\tilde{Y}_{1,1} := X^{(m)}(t+k) \stackrel{d}{=} X^{(m)}(k) && \tilde{Y}_{1,2} := X^{(m)}(t) \stackrel{d}{=} X^{(m)}(0)\\
&\tilde{Y}_{2,1} := X^{(m)}(t+k) \stackrel{d}{=} X^{(m)}(k) && \tilde{Y}_{2,2} := X^{(m)}(t) \stackrel{d}{=} X^{(m)}(0)\\
&\tilde{Y}_{3,1} := X^{(m)}(j_1+k) \stackrel{d}{=} X^{(m)}(u_1+k) && \tilde{Y}_{3,2} := X^{(m)}(j_1) \stackrel{d}{=} X^{(m)}(u_1)\\
&\tilde{Y}_{4,1} := X^{(m)}(j_2+k) \stackrel{d}{=} X^{(m)}(u_2+k) && \tilde{Y}_{4,2} := X^{(m)}(j_2) \stackrel{d}{=} X^{(m)}(u_2)
\end{aligned}$$

Naively computing $\tilde{Q}_t$ is very costly. For computational efficiency, we therefore note that

- $\tilde{C}(u) = \tilde{C}(-u)$,
- $\tilde{D}(u_1, u_2) = \tilde{D}(u_2, u_1)$,
- $M_{1,t}$, $M_{2a,t}$ and $M_{3,t}$ are constant on $t = m+k+1, \ldots, n-m-2k$, and
- $M_{2b,t}$ is constant on $t = 2(m+k)+1, \ldots, n-2m-3k$.

Further, we have that the addends $\tilde{C}(u)$ and $\tilde{D}(\ell_1, \ell_2)$ do not depend on $t$. Therefore, for a given $m$ and $k$, we first compute $\tilde{C}(u)$, for $u = 0, \ldots, 2(m+k)$, and $\tilde{D}(u_1, u_2)$ for $u_1, u_2 = -(m+k), \ldots, m+k$ that satisfy $u_1 \leq u_2$. Then, computing the bound can be done rather efficiently.

# S6 Simulated Wasserstein distance

We obtain the Wasserstein distance, to compare our bound against, via simulation. Denoting the cdf of $\sqrt{n}(\hat{\gamma}_{ab}(k) - \gamma_{ab}(k))$ by $F$ and the cdf of $N \sim \mathcal{N}(0, \Sigma_{ab}(k))$ by $G$, we have

$$W := d_{\mathrm{W}}\left(\sqrt{n}\left(\hat{\gamma}_{ab}(k) - \gamma_{ab}(k)\right), N\right) = \int_0^1 |F^{-1}(u) - G^{-1}(u)| \mathrm{d}u, \tag{22}$$

where $F^{-1}(u) := \inf\{x \in \mathbb{R} : F(x) \geq u\}$ is the quantile function of $F$; similarly, $G^{-1}$ is the quantile function for $G$. We proceed by sampling i. i. d. $Z_1, \ldots, Z_R \sim F$ and then replace $F$ in (22) by the empirical distribution $\mathbb{F}(x) := \frac{1}{R}\sum_{r=1}^{R} I\{Z_r \leq x\}$. More precisely,

$$\begin{aligned}
(22) \approx \int_0^1 |\mathbb{F}^{-1}(u) - G^{-1}(u)| \mathrm{d}u &= \sum_{r=1}^{R} \int_{(i-1)/R}^{i/R} |Z_{(r)} - G^{-1}(u)| \mathrm{d}u\\
&\approx \frac{1}{R}\sum_{r=1}^{R} \Big| Z_{(r)} - \sqrt{\Sigma_{ab}(k)}\Phi^{-1}\Big(\frac{2i-1}{2R}\Big)\Big| =: \hat{W}_i.
\end{aligned}$$

If $F \neq G$ (and additional regularity conditions hold), then the error in the first approximation is of order $O_P(R^{-1/2})$; cf. e. g. Munk and Czado (1998). The error in the second approximation is of order $O(R^{-1})$. Since we are considering $n$ to the order of thousands, we choose $R = 4 \times 10^6$. Note that we have chosen $R$ as $2000^2$ and $n = 2000$ is the maximum sample length. Further, to reduce the variance of the (pseudo) random $\hat{W}_i$ we simulate independent copies of it ($i = 1, \ldots, B = 50$) and then report the average.

## S7 Proof of Lemma S1.2

For $N \sim \mathcal{N}(0, \sigma^2)$ and $h \in \mathcal{H}_{\mathrm{W}}$, let first $f = f_h$ be the solution of the Stein equation

$$\sigma^2 f'(w) - w f(w) = h(w) - \mathbb{E}[h(N)] \tag{23}$$

for the $\mathcal{N}(0, \sigma^2)$ distribution. Because $\forall i \in J$, $\xi_i$ and $W - \eta_i$, with $\eta_i$ as in (1), are independent and since $\mathbb{E}(\xi_i) = 0$, we have that

$$\mathbb{E}(W f(W)) = \sum_{i \in J} \mathbb{E}(\xi_i f(W)) = \sum_{i \in J} \mathbb{E}(\xi_i (f(W) - f(W - \eta_i))).$$

Adding now and subtracting the quantity $\sum_{i \in J} \mathbb{E}(\xi_i \eta_i f'(W))$ yields

$$\mathbb{E}(W f(W)) = \sum_{i \in J} \mathbb{E}\left(\xi_i \left(f(W) - f(W - \eta_i) - \eta_i f'(W)\right)\right) + \mathbb{E}\left(\sum_{i \in J} \xi_i \eta_i f'(W)\right).$$

For $\mathbb{E}\left(W^2\right) = \mathrm{var}(W) = \sigma^2$, we have now that

$$\sigma^2 = \mathbb{E}\left(W^2\right) = \mathbb{E}\left(\sum_{i \in J} \xi_i \sum_{j \in J} \xi_j\right) = \sum_{i \in J} \mathbb{E}(\xi_i \eta_i)$$

because $\xi_i$ is independent of all the elements that are in $A_i^c$. Using now the Stein equation as in (23), we have that

$$\begin{aligned}
&\mathbb{E}\left(\sigma^2 f'(W) - W f(W)\right) \\
&= \mathbb{E}\left(\sum_{i \in J} \mathbb{E}(\xi_i \eta_i) f'(W)\right) \\
&\quad - \sum_{i \in J} \mathbb{E}\left(\xi_i \left(f(W) - f(W - \eta_i) - \eta_i f'(W)\right)\right) - \mathbb{E}\left(\sum_{i \in J} \xi_i \eta_i f'(W)\right) \\
&= -\sum_{i \in J} \mathbb{E}\left((\xi_i \eta_i - \mathbb{E}(\xi_i \eta_i)) f'(W)\right) - \sum_{i \in J} \mathbb{E}\left(\xi_i \left(f(W) - f(W - \eta_i) - \eta_i f'(W)\right)\right) \\
&= -\sum_{i \in J} \mathbb{E}\left((\xi_i \eta_i - \mathbb{E}(\xi_i \eta_i)) \left(f'(W) - f'(W - \tau_i)\right)\right) \\
&\quad - \sum_{i \in J} \mathbb{E}\left(\xi_i \left(f(W) - f(W - \eta_i) - \eta_i f'(W)\right)\right)
\end{aligned} \tag{24}$$

with the last equality being due to the fact that both $\xi_i$ and $\eta_i$ are independent with $W - \tau_i$ and therefore,

$$\sum_{i \in J} \mathbb{E}\left((\xi_i \eta_i - \mathbb{E}(\xi_i \eta_i)) f'(W - \tau_i)\right) = \sum_{i \in J} \mathbb{E}(\xi_i \eta_i - \mathbb{E}(\xi_i \eta_i)) \mathbb{E}\left(f'(W - \tau_i)\right) = 0.$$

Using a first order Taylor expansion,

$$f'(W) - f'(W - \tau_i) = \tau_i f''(W^*), \tag{25}$$

for $W^*$ between $W$ and $W - \tau_i$. Furthermore, using now a second order Taylor expansion of $f(W - \eta_i)$ about $W$, leads to

$$f(W) - f(W - \eta_i) - \eta_i f'(W) = -\frac{\eta_i^2}{2} f''(\tilde{W}), \tag{26}$$

where $\tilde{W}$ is between $W$ and $W - \eta_i$. We denote the sup-norm of a function by $\|\cdot\|$ and applying the results of (25) and (26) to (24) yields

$$\begin{aligned}
|\mathbb{E}[h(W)] - \mathbb{E}[h(N)]| &= \left|\mathbb{E}\left(\sigma^2 f'(W) - Wf(W)\right)\right| \\
&= \left|-\sum_{i\in J}\mathbb{E}\left((\xi_i\eta_i - \mathbb{E}(\xi_i\eta_i))\,\tau_i f''(W^*)\right) + \sum_{i\in J}\mathbb{E}\left(\xi_i\frac{\eta_i^2}{2}f''(\tilde{W})\right)\right| \\
&\le \sum_{i\in J}\mathbb{E}\left|(\xi_i\eta_i - \mathbb{E}(\xi_i\eta_i))\,\tau_i f''(W^*)\right| + \sum_{i\in J}\mathbb{E}\left|\xi_i\frac{\eta_i^2}{2}f''(\tilde{W})\right| \\
&\le \|f''\|\sum_{i\in J}\mathbb{E}\left|(\xi_i\eta_i - \mathbb{E}(\xi_i\eta_i))\,\tau_i\right| + \frac{1}{2}\|f''\|\sum_{i\in J}\mathbb{E}\left|\xi_i\eta_i^2\right| \\
&\le \|f''\|\sum_{i\in J}\left\{\mathbb{E}\left|(\xi_i\eta_i - \mathbb{E}(\xi_i\eta_i))\,\tau_i\right| + \frac{1}{2}\mathbb{E}\left|\xi_i\eta_i^2\right|\right\}. \qquad (27)
\end{aligned}$$

From Section 2.2 of Chen et al. (2011), we know that the solution of the Stein equation in (23) satisfies that $\|f''\| \le 2\frac{\|h'\|}{\sigma^3}$. Using this result in (27) yields

$$|\mathbb{E}[h(W)] - \mathbb{E}[h(N)]| \le 2\frac{\|h'\|}{\sigma^3}\sum_{i\in J}\left\{\mathbb{E}\left|(\xi_i\eta_i - \mathbb{E}(\xi_i\eta_i))\,\tau_i\right| + \frac{1}{2}\mathbb{E}\left|\xi_i\eta_i^2\right|\right\}.$$

Since we work for $h \in H_{\mathrm{W}}$ as in (1.5), we have that $\|h'\| \le 1$, which leads to the result of the lemma. ■

## S8 Proof of Lemma S3.1

Lemma S3.1 follows from the more general Lemma S8.1, below, if we apply it with $X_i := X_{a_i}(t_i)$ and $Y_i := X_{a_i}^{(m)}(t_i)$ such that $\|X_i - Y_i\|_{\alpha p} = D_{a_i,\alpha p,m}$.

**Lemma S8.1** *For $p \in \mathbb{N}$ and $\alpha \ge 1$, let $X_1, \dots, X_p, Y_1, \dots, Y_p$ be $\mathbb{R}$-valued random variables with $\|X_i\|_{\alpha p} < \infty$ and $\|Y_i\|_{\alpha p} < \infty$, $i = 1, \dots, p$. Then, with $\Lambda_p$ as in Lemma S3.1, we have*

$$\Big\|\prod_{i=1}^{p} X_i - \prod_{i=1}^{p} Y_i\Big\|_{\alpha} \le \sum_{(\ell_1,\dots,\ell_p)\in\Lambda_p}\prod_{i=1}^{p}\|X_i - Y_i\|_{\alpha p}^{\ell_i}\|X_i\|_{\alpha p}^{1-\ell_i}, \qquad (28)$$

**Remark S8.2** *Note that the important difference between Lemmas S8.1 and S3.1 is that the former can also be employed in the context of non-stationary data.*

**Proof of Lemma S8.1.** For $p = 1$ the assertion is obvious. For $p \ge 2$ the assertion follows from the following chain of inequalities, where we denote $D_i := \|X_i - Y_i\|_{\alpha p}$

$$
\begin{aligned}
&\Big\| \prod_{i=1}^{p} X_i - \prod_{i=1}^{p} Y_i \Big\|_\alpha = \Big\| \sum_{j=1}^{p} \Big( \prod_{\substack{i=1 \\ i<j}}^{p} X_i \Big) (X_j - Y_j) \Big( \prod_{\substack{i=1 \\ i>j}}^{p} Y_i \Big) \Big] \Big\|_\alpha \\
&\le \sum_{j=1}^{p} \Big\| \Big( \prod_{\substack{i=1 \\ i<j}}^{p} X_i \Big) (X_j - Y_j) \Big( \prod_{\substack{i=1 \\ i>j}}^{p} Y_i \Big) \Big\|_\alpha \\
&\le \sum_{j=1}^{p} \Big( \|X_j - Y_j\|_{\alpha p} \prod_{\substack{i=1 \\ i<j}}^{p} \|X_i\|_{\alpha p} \prod_{\substack{i=1 \\ i>j}}^{p} \big( \|X_i\|_{\alpha p} + \|X_i - Y_i\|_{\alpha p} \big) \Big) \\
&= D_1 \big( \|X_2\|_{\alpha p} + D_2 \big) \cdots \big( \|X_p\|_{\alpha p} + D_p \big) \\
&\quad + I\{p > 2\} \sum_{j=2}^{p-1} \|X_1\|_{\alpha p} \cdots \|X_{j-1}\|_{\alpha p} D_j \big( \|X_{j+1}\|_{\alpha p} + D_{j+1} \big) \cdots \big( \|X_p\|_{\alpha p} + D_p \big) \\
&\quad + D_p \|X_1\|_{\alpha p} \cdots \|X_{p-1}\|_{\alpha p} \\
&= D_1 \sum_{(\ell_2, \ldots, \ell_p) \in \{0,1\}^{p-1}} \|X_2\|_{\alpha p}^{1-\ell_2} D_2^{\ell_2} \cdots \|X_p\|_{\alpha p}^{1-\ell_p} D_p^{\ell_p} \\
&\quad + I\{p > 2\} \sum_{j=2}^{p-1} \|X_1\|_{\alpha p} \cdots \|X_{j-1}\|_{\alpha p} D_j \\
&\qquad \times \sum_{(\ell_{j+1}, \ldots, \ell_p) \in \{0,1\}^{p-j}} \|X_{j+1}\|_{\alpha p}^{1-\ell_{j+1}} D_{j+1}^{\ell_{j+1}} \cdots \|X_p\|_{p}^{1-\ell_p} D_p^{\ell_p} \\
&\quad + \|X_1\|_{\alpha p} \cdots \|X_{p-1}\|_{\alpha p} D_p,
\end{aligned}
$$

where the first inequality follows due to the triangle inequality that we have because $\alpha \geq 1$, and the second inequality follows by (a generalization of) Hölder's inequality. ■

## S9 Proof of Lemma S3.3

Note that, by the triangle inequality and the definition of $\tilde{\Sigma}_{ab}(k)$ in (2.9) of the main paper, we have

$$
|\tilde{\Sigma}_{ab}(k) - \Sigma_{ab}(k)| \le \Big| \text{var}\Big( n^{-1/2} \hat{\gamma}_{ab}^{(m)}(k) \Big) - \text{var}\Big( n^{-1/2} \hat{\gamma}_{ab}(k) \Big) \Big| + \Big| n \text{var}\big( \hat{\gamma}_{ab}(k) \big) - \Sigma_{ab}(k) \Big|,
$$

where

$$
\hat{\gamma}_{ab}^{(m)}(k) := \frac{1}{n} \sum_{t=1}^{n-k} X_a^{(m)}(t+k) X_b^{(m)}(t) \text{ for } k = 0, \ldots, n-1.
$$

We then have the following, more general, bound which we will subsequently use to derive the assertion of the lemma:

$$
\begin{aligned}
&\Big|\text{cov}\big(\sqrt{n}\hat\gamma_{a_1b_1}(k_1), \sqrt{n}\hat\gamma_{a_2b_2}(k_2)\big) - \text{cov}\big(\sqrt{n}\hat\gamma^{(m)}_{a_1b_1}(k_1), \sqrt{n}\hat\gamma^{(m)}_{a_2b_2}(k_2)\big)\Big| \\
&\le \frac{1}{n}\sum_{t_1=1}^{n-k_1}\sum_{t_2=1}^{n-k_2}\Big|\mathbb{E}\big[X_{a_1}(t_1+k_1)X_{b_1}(t_1)X_{a_2}(t_2+k_2)X_{b_2}(t_2)\big] \\
&\qquad\qquad - \mathbb{E}\big[X^{(m)}_{a_1}(t_1+k_1)X^{(m)}_{b_1}(t_1)X^{(m)}_{a_2}(t_2+k_2)X^{(m)}_{b_2}(t_2)\big]\Big| \\
&\quad + \frac{1}{n}\sum_{t_1=1}^{n-k_1}\sum_{t_2=1}^{n-k_2}\Big|\Big(\mathbb{E}\big[X_{a_1}(t_1+k_1)X_{b_1}(t_1)\big] - \mathbb{E}\big[X^{(m)}_{a_1}(t_1+k_1)X^{(m)}_{b_1}(t_1)\big]\Big) \\
&\qquad\qquad \times \mathbb{E}\big[X_{a_2}(t_2+k_2)X_{b_2}(t_2)\big] \\
&\qquad\qquad + \mathbb{E}\big[X^{(m)}_{a_1}(t_1+k_1)X^{(m)}_{b_1}(t_1)\big] \\
&\qquad\qquad \times \Big(\mathbb{E}\big[X_{a_2}(t_2+k_2)X_{b_2}(t_2)\big] - \mathbb{E}\big[X^{(m)}_{a_2}(t_2+k_2)X^{(m)}_{b_2}(t_2)\big]\Big)\Big| \\
&\le \frac{(n-k_1)(n-k_2)}{n}\Big(K_{4,m}(\{a_1,b_1,a_2,b_2\}) + K_{2,m}(\{a_1,b_1\})|\gamma_{a_2b_2}(k_2)| \\
&\qquad\qquad + \big(|\gamma_{a_1b_1}(k_1)| + K_{2,m}(\{a_1,b_1\})\big)K_{2,m}(\{a_2,b_2\}\Big) \\
&\le \frac{(n-k_1)(n-k_2)}{n}\Big(\max_{i=a_1,b_1,a_2,b_2} D_{i,4,m}\Big) \\
&\quad \times \Big(\sum_{j\in\{a_1,b_1,a_2,b_2\}}\ \prod_{i\in\{a_1,b_1,a_2,b_2\}\setminus\{j\}} (\|X_i(0)\|_4 + D_{i,4,m}) \\
&\qquad + (\|X_{a_1}(0)\|_2 + \|X_{b_1}(t)\|_2 + \min_{i=a_1,b_1} D_{i,2,m})|\gamma_{a_2b_2}(k_2)| \\
&\qquad + (\|X_{a_2}(0)\|_2 + \|X_{b_2}(t)\|_2 + \min_{i=a_2,b_2} D_{i,2,m})|\gamma_{a_1b_1}(k_1)| \\
&\qquad + \min\Big\{\max_{i=a_1,b_1} D_{i,2,m}, \max_{i=a_2,b_2} D_{i,2,m}\Big\} \\
&\qquad\quad \times (\|X_{a_1}(0)\|_2 + \|X_{b_1}(t)\|_2 + \min_{i=a_1,b_1} D_{i,2,m})(\|X_{a_2}(0)\|_2 + \|X_{b_2}(t)\|_2 + \min_{i=a_2,b_2} D_{i,2,m})\Big)
\end{aligned}
$$

where for the third inequality we have, by Lemma S3.1 with $\alpha = 1$ and $p = 2$, that

$$
\begin{aligned}
K_{2,m}(\{a,b\}) &= D_{a,2,m}\|X_b(t)\|_2 + D_{a,2,m}D_{b,2,m} + \|X_a(0)\|_2 D_{b,2,m} \\
&\le \Big(\max_{i=a,b} D_{i,2,m}\Big)(\|X_a(0)\|_2 + \|X_b(t)\|_2 + \min_{i=a,b} D_{i,2,m})
\end{aligned}
$$

and, by Lemma S3.1 with $\alpha = 1$ and $p = 4$, we have that

$$
\begin{aligned}
&K_{4,m}(\{a_1,b_1,a_2,b_2\}) \\
&= D_{a_1,4,m}(\|X_{b_1}(0)\|_4 + D_{b_1,4,m})(\|X_{a_2}(0)\|_4 + D_{a_2,4,m})(\|X_{b_2}(0)\|_4 + D_{b_2,4,m}) \\
&\quad + D_{b_1,4,m}\|X_{a_1}(0)\|_4(\|X_{a_2}(0)\|_4 + D_{a_2,4,m})(\|X_{b_2}(0)\|_4 + D_{b_2,4,m}) \\
&\quad + D_{a_2,4,m}\|X_{a_1}(0)\|_4\|X_{b_1}(0)\|_4(\|X_{b_2}(0)\|_4 + D_{b_2,4,m}) \\
&\quad + D_{b_2,4,m}\|X_{a_1}(0)\|_4\|X_{b_1}(0)\|_4\|X_{a_2}(0)\|_4 \\
&\le \Big(\max_{i=a_1,b_1,a_2,b_2} D_{i,4,m}\Big)\sum_{j\in\{a_1,b_1,a_2,b_2\}}\ \prod_{i\in\{a_1,b_1,a_2,b_2\}\setminus\{j\}} (\|X_i(0)\|_4 + D_{i,4,m}).
\end{aligned}
$$

For the third inequality we had further employed the fact that

$$
\max_{i=j_1,j_2} D_{i,2,m} \le \max_{i=a_1,b_1,a_2,b_2} D_{i,4,m}, \quad \text{for any } j_1, j_2 \in \{a_1,b_1,a_2,b_2\}.
$$

The assertion of the lemma follows, after some simplications applied to the special case where we have $k_1 = k_2$, $a_1 = a_2$ and $b_1 = b_2$. ■

## S10 Proof of Lemma S3.4

We obtain the assertion from the following chain of inequalities

$$
\begin{aligned}
&|Q_t - \tilde{Q}_t| \\
&\le \Bigg| \Bigg\| \Bigg( Z_t \sum_{j\in A_t} Z_j - \mathbb{E}\Bigg( Z_t \sum_{j\in A_t} Z_j \Bigg) \Bigg) \sum_{j\in B_t} Z_j \Bigg\|_1 \\
&\qquad - \Bigg\| \Bigg( \tilde{Z}_t \sum_{j\in A_t} \tilde{Z}_j - \mathbb{E}\Bigg( \tilde{Z}_t \sum_{j\in A_t} \tilde{Z}_j \Bigg) \Bigg) \sum_{j\in B_t} \tilde{Z}_j \Bigg\|_1 \Bigg| \\
&\quad + \Bigg| \frac{1}{2} \Bigg\| Z_t \Bigg( \sum_{j\in A_t} Z_j \Bigg)^2 \Bigg\|_1 - \frac{1}{2} \Bigg\| \tilde{Z}_t \Bigg( \sum_{j\in A_t} \tilde{Z}_j \Bigg)^2 \Bigg\|_1 \Bigg| \\
&\le \mathbb{E} \Bigg| \Bigg( Z_t \sum_{j\in A_t} Z_j - \mathbb{E}\Bigg( Z_t \sum_{j\in A_t} Z_j \Bigg) \Bigg) \sum_{j\in B_t} Z_j - \Bigg( \tilde{Z}_t \sum_{j\in A_t} \tilde{Z}_j - \mathbb{E}\Bigg( \tilde{Z}_t \sum_{j\in A_t} \tilde{Z}_j \Bigg) \Bigg) \sum_{j\in B_t} \tilde{Z}_j \Bigg| \\
&\quad + \frac{1}{2} \mathbb{E} \Bigg| Z_t \Bigg( \sum_{j\in A_t} Z_j \Bigg)^2 - \tilde{Z}_t \Bigg( \sum_{j\in A_t} \tilde{Z}_j \Bigg)^2 \Bigg| \\
&\le \mathbb{E} \Bigg| Z_t \sum_{j\in A_t} Z_j \sum_{j\in B_t} Z_j - \tilde{Z}_t \sum_{j\in A_t} \tilde{Z}_j \sum_{j\in B_t} \tilde{Z}_j \Bigg| \\
&\quad + \mathbb{E} \Bigg| \mathbb{E}\Bigg( Z_t \sum_{j\in A_t} Z_j \Bigg) \sum_{j\in B_t} Z_j - \mathbb{E}\Bigg( \tilde{Z}_t \sum_{j\in A_t} \tilde{Z}_j \Bigg) \sum_{j\in B_t} \tilde{Z}_j \Bigg| \\
&\quad + \frac{1}{2} \mathbb{E} \Bigg| Z_t \Bigg( \sum_{j\in A_t} Z_j \Bigg)^2 - \tilde{Z}_t \Bigg( \sum_{j\in A_t} \tilde{Z}_j \Bigg)^2 \Bigg| \\
&\le \sum_{j_1\in A_t} \sum_{j_2\in B_t} \mathbb{E} \left| Z_t Z_{j_1} Z_{j_2} - \tilde{Z}_t \tilde{Z}_{j_1} \tilde{Z}_{j_2} \right| + \frac{1}{2} \sum_{j_1\in A_t} \sum_{j_2\in A_t} \mathbb{E} \left| Z_t Z_{j_1} Z_{j_2} - \tilde{Z}_t \tilde{Z}_{j_1} \tilde{Z}_{j_2} \right| && (29) \\
&\quad + \mathbb{E} \Bigg| \mathbb{E}\Bigg( Z_t \sum_{j\in A_t} Z_j \Bigg) \sum_{j\in B_t} (Z_j - \tilde{Z}_j) \Bigg| && (30) \\
&\quad + \mathbb{E} \Bigg| \Bigg[ \mathbb{E}\Bigg( Z_t \sum_{j\in A_t} Z_j \Bigg) - \mathbb{E}\Bigg( \tilde{Z}_t \sum_{j\in A_t} \tilde{Z}_j \Bigg) \Bigg] \sum_{j\in B_t} (Z_j + \tilde{Z}_j - Z_j) \Bigg|, && (31)
\end{aligned}
$$

where we have used the the triangle inequality and reverse triangle inequality. The asserted bound for $|Q_t - \tilde{Q}_t|$ will be the sum of the bounds we now derive for (29), (30) and (31). First, we derive two preliminary bounds to control $\|Z_0\|_\alpha$ and $\|Z_0 - \tilde{Z}_0\|_\alpha$, $\alpha = 1, 2, 3$. Employing the triangle inequality and Hölder's inequality we have

$$\|Z_0\|_\alpha = \|X_a(t+k)X_b(t)\|_\alpha + |\mathbb{E}[X_a(t+k)X_b(t)]| \leq 2\|X_a(0)\|_{2\alpha}\|X_b(0)\|_{2\alpha} \tag{32}$$

and, using the triangle inequality and Lemma S3.1, we have

$$\|Z_0 - \tilde{Z}_0\|_\alpha \leq 2\|X_a(t+k)X_b(t) - X_a^{(m)}(t+k)X_b^{(m)}(t)\|_\alpha \leq 2K_{2,m}^{(\alpha)}(\{a,b\}) \tag{33}$$

Next, to bound (29), we use Lemma S8.1 with $\alpha = 1$, $p = 3$, $X_1, X_2, X_3$ equal to $Z_t, Z_{j_1}, Z_{j_2}$ and $Y_1, Y_2, Y_3$ equal to $\tilde{Z}_t, \tilde{Z}_{j_1}, \tilde{Z}_{j_2}$. Then, $\|X_i\|_\alpha = \|Z_0\|_\alpha$ and $\|X_i - Y_i\|_\alpha = \|Z_0 - \tilde{Z}_0\|_\alpha$, such that

$$\begin{aligned}
&\mathbb{E}\left|Z_t Z_{j_1} Z_{j_2} - \tilde{Z}_t \tilde{Z}_{j_1} \tilde{Z}_{j_2}\right| \\
&\leq \|Z_0 - \tilde{Z}_0\|_3\Big(3\|Z_0\|_3^2 + 3\|Z_0\|_3\|Z_0 - \tilde{Z}_0\|_3 + \|Z_0 - \tilde{Z}_0\|_3^2\Big) \\
&\leq 2K_{2,m}^{(3)}\Big(3\big(2\|X_a(0)\|_6\|X_b(0)\|_6 + K_{2,m}^{(3)}\big)^2 + (K_{2,m}^{(3)})^2\Big)
\end{aligned} \tag{34}$$

Clearly, this implies that

$$(29) \leq 2K_{2,m}^{(3)}\Big(3\big(2\|X_a(0)\|_6\|X_b(0)\|_6 + K_{2,m}^{(3)}\big)^2 + (K_{2,m}^{(3)})^2\Big)\Big(|A_t||B_t| + \frac{1}{2}|A_t|^2\Big) \tag{35}$$

For (30) we have

$$(30) = \Big|\sum_{j\in A_t}\mathbb{E}(Z_t Z_j)\Big| \times \mathbb{E}\left|\sum_{j\in B_t}(Z_j - \tilde{Z}_j)\right| \leq \Big|\sum_{j\in A_t}\mathbb{E}(Z_t Z_j)\Big| \times |B_t| \times 2K_{2,m}^{(1)}, \tag{36}$$

where we have used (33) with $\alpha = 1$. Finally, to bound (31), we use Lemma S8.1 similar to how we used it in (34), but with $p = 2$, and obtain

$$\begin{aligned}
\mathbb{E}\left|Z_t Z_j - \tilde{Z}_t\tilde{Z}_j\right| &\leq \|Z_0 - \tilde{Z}_0\|_2\Big(2\|Z_0\|_2 + \|Z_0 - \tilde{Z}_0\|_2\Big) \\
&\leq 2K_{2,m}^{(2)}\Big(4\|X_a(0)\|_4\|X_b(0)\|_4 + 2K_{2,m}^{(2)}\Big),
\end{aligned}$$

where the second inequality followed from the preliminary bounds (32) and (33), with $\alpha = 2$. Thus, we have

$$\begin{aligned}
(31) &= \left|\mathbb{E}\bigg(Z_t\sum_{j\in A_t}Z_j\bigg) - \mathbb{E}\bigg(\tilde{Z}_t\sum_{j\in A_t}\tilde{Z}_j\bigg)\right| \times \mathbb{E}\bigg[\Big|\sum_{j\in B_t}(Z_j + \tilde{Z}_j - Z_j)\Big|\bigg] \\
&\leq \sum_{j\in A_t}\mathbb{E}\Big|Z_t Z_j - \tilde{Z}_t\tilde{Z}_j\Big| \times \sum_{j\in B_t}\Big(\|Z_0\|_1 + \|\tilde{Z}_0 - Z_0\|_1\Big) \\
&\leq 2K_{2,m}^{(2)}\Big(4\|X_a(0)\|_4\|X_b(0)\|_4 + 2K_{2,m}^{(2)}\Big)|A_t| \\
&\qquad \times 2\Big(\|X_a(0)\|_2\|X_b(0)\|_2 + K_{2,m}^{(1)}\Big)|B_t|.
\end{aligned} \tag{37}$$

Summing the three bounds obtained in (35), (36) and (37) yields the assertion. ■

## S11 Proof of Lemma S4.1

Let $h \in \mathcal{H}_W$, as defined in (1.5). We have

$$\left| h\left(n^{1/2}(\tilde{\gamma}_{ab}(k) - \gamma_{ab}(k))\right) - h\left(n^{1/2}(\hat{\gamma}_{ab}(k) - \gamma_{ab}(k))\right) \right| \leq \|h\|_{\mathrm{Lip}} n^{1/2} |\tilde{\gamma}_{ab}(k) - \hat{\gamma}_{ab}(k)|. \tag{38}$$

Further, we see that

$$\begin{aligned} &n^{1/2}|\tilde{\gamma}_{ab}(k) - \hat{\gamma}_{ab}(k)| \\ &= \frac{n-k}{n^{1/2}} \left| \bar{X}_{a,1:n}\bar{X}_{b,1:n} - \bar{X}_{a,1:n}\bar{X}_{b,1:n-k} - \bar{X}_{a,k+1:n}\bar{X}_{b,1:n} \right| \\ &= \frac{n-k}{n^{1/2}} \left| (\bar{X}_{a,1:n} - \bar{X}_{a,k+1:n})(\bar{X}_{b,1:n} - \bar{X}_{b,1:n-k}) - \bar{X}_{a,k+1:n}\bar{X}_{b,1:n-k} \right| \end{aligned} \tag{39}$$

where $X_{a,u:v} = \frac{1}{v-u+1}\sum_{t=u}^{v} X_a(t)$ and analogously with $b$ instead of $a$.

From (38) and the second line of (39) we obtain, for $k = 1, \ldots, n-1$, that

$$\begin{aligned} &\left| \mathbb{E}\left[ h\left(n^{1/2}(\tilde{\gamma}_{ab}(k) - \gamma_{ab}(k))\right)\right] - \mathbb{E}\left[ h\left(n^{1/2}(\hat{\gamma}_{ab}(k) - \gamma_{ab}(k))\right)\right] \right| \\ &\leq \frac{n-k}{n^{1/2}} \|h\|_{\mathrm{Lip}} \Big( \|\bar{X}_{a,1:n} - \bar{X}_{a,k+1:n}\|_2 \times \|\bar{X}_{b,1:n} - \bar{X}_{b,1:n-k}\|_2 + \|\bar{X}_{a,k+1:n}\|_2 \times \|\bar{X}_{b,1:n-k}\|_2 \Big) \\ &\leq \frac{n-k}{n^{1/2}} \|h\|_{\mathrm{Lip}} \Big(\frac{k}{n}\Big)^2 \left( \frac{n}{k(n-k)} \sum_{u=-\infty}^{\infty} |\gamma_{aa}(u)| + \frac{1}{k(n-k)} \sum_{u=-\infty}^{\infty} |u||\gamma_{aa}(h)| \right)^{1/2} \\ &\qquad \times \left( \frac{n}{k(n-k)} \sum_{u=-\infty}^{\infty} |\gamma_{bb}(u)| + \frac{1}{k(n-k)} \sum_{u=-\infty}^{\infty} |u||\gamma_{bb}(u)| \right)^{1/2} \\ &\quad + \frac{n-k}{n^{1/2}} \|h\|_{\mathrm{Lip}} \frac{1}{n-k} \Big( \sum_{u=-\infty}^{\infty} |\gamma_{aa}(u)| \Big)^{1/2} \Big( \sum_{u=-\infty}^{\infty} |\gamma_{bb}(u)| \Big)^{1/2} \\ &= \frac{k}{n^{3/2}} \|h\|_{\mathrm{Lip}} \left( \sum_{u=-\infty}^{\infty} |\gamma_{aa}(u)| + \frac{1}{n} \sum_{u=-\infty}^{\infty} |u||\gamma_{aa}(u)| \right)^{1/2} \\ &\qquad \times \left( \sum_{u=-\infty}^{\infty} |\gamma_{bb}(u)| + \frac{1}{n} \sum_{u=-\infty}^{\infty} |u||\gamma_{bb}(u)| \right)^{1/2} \\ &\quad + \frac{1}{n^{1/2}} \|h\|_{\mathrm{Lip}} \Big( \sum_{u=-\infty}^{\infty} |\gamma_{aa}(u)| \Big)^{1/2} \Big( \sum_{u=-\infty}^{\infty} |\gamma_{bb}(u)| \Big)^{1/2} \end{aligned}$$

where we have used

$$\begin{aligned} \mathbb{E}\bar{X}_{j,u:v}^2 &= \frac{1}{(v-u+1)^2} \sum_{s=u}^{v} \sum_{t=u}^{v} \mathbb{E}(X_j(s-t)X_j(0)) \\ &= \frac{1}{(v-u+1)^2} \sum_{\ell=-(v-u+1)}^{v-u+1} (v-u+1-|\ell|)\gamma_{jj}(\ell) \\ &= \frac{1}{v-u+1} \sum_{\ell=-(v-u+1)}^{v-u+1} \Big(1 - \frac{|\ell|}{v-u+1}\Big)\gamma_{jj}(\ell) \leq \frac{1}{v-u+1} \sum_{\ell=-\infty}^{\infty} |\gamma_{jj}(\ell)|. \end{aligned}$$

and the fact that for $k = 1, \dots, n-1$, we have

$$\begin{aligned}\|\bar{X}_{a,1:n} - \bar{X}_{a,k+1:n}\|_2^2 &= \mathbb{E}\Big(\frac{1}{n}\sum_{i=1}^{n} X_a(i) - \frac{1}{n-k}\sum_{i=k+1}^{n} X_a(i)\Big)^2\\ &= \mathbb{E}\Big(\frac{k}{n}\frac{1}{n-k}\sum_{i=k+1}^{n} X_a(i) - \frac{k}{n}\frac{1}{k}\sum_{i=1}^{k} X_a(i)\Big)^2\\ &= \Big(\frac{k}{n}\Big)^2\Big(\|\bar{X}_{a,k+1:n}\|_2^2 + \|\bar{X}_{a,1:k}\|_2^2 - 2\mathrm{cov}(\bar{X}_{a,k+1:n}, \bar{X}_{a,1:k})\Big)\\ &\le \Big(\frac{k}{n}\Big)^2\Bigg(\Big(\frac{1}{n-k} + \frac{1}{k}\Big)\Big(\sum_{u=-\infty}^{\infty} |\gamma_{aa}(u)|\Big) + 2\frac{1}{k(n-k)}\sum_{u=1}^{\infty} u|\gamma_{aa}(u)|\Bigg)\\ &\le \Big(\frac{k}{n}\Big)^2\Bigg(\frac{n}{k(n-k)}\sum_{u=-\infty}^{\infty} |\gamma_{aa}(u)| + \frac{1}{k(n-k)}\sum_{u=-\infty}^{\infty} |u||\gamma_{aa}(u)|\Bigg)\end{aligned}$$

and similarly we derive

$$\begin{aligned}\|\bar{X}_{b,1:n} - \bar{X}_{b,1:n-k}\|_2^2 &= \Big(\frac{k}{n}\Big)^2\|\bar{X}_{b,n-k+1:n} - \bar{X}_{b,1:n-k}\|_2^2\\ &\le \Big(\frac{k}{n}\Big)^2\Bigg(\frac{n}{k(n-k)}\sum_{u=-\infty}^{\infty} |\gamma_{bb}(u)| + \frac{1}{k(n-k)}\sum_{u=-\infty}^{\infty} |u||\gamma_{bb}(u)|\Bigg).\end{aligned}$$

■

## S12 Proof of Proposition 2.8

The bound given in (2.10) consists of four terms that we now discuss one by one, for Regimes 1 and 2 stated in Section 2.5. For Regime 1, we assume that $\boldsymbol{X}^{(m)}(t)$ is chosen as described in Remark 2.7. In this proof, we will refer to the first, second, third and fourth term of the bound, respectively, meaning the respective addends in (2.10). For the first term of the bound, it is obvious that $kn^{-1/2}|\gamma_{ab}(k)| = \mathcal{O}(n^{-1/2})$, in both regimes. With respect to the second term of the bound we now show that it is of the order $\mathcal{O}(n^{-1/2})$, in both regimes, if $m$ is chosen appropriately. The expression $|\Sigma_{ab}(k) - \tilde{\Sigma}_{ab}(k)|$ can be controlled by employing Lemma S3.3. Therefore,

$$\begin{aligned}\frac{\sqrt{2}}{\sqrt{\pi\Sigma_{ab}(k)}}\Big|\Sigma_{ab}(k) - \tilde{\Sigma}_{ab}(k)\Big| \le{}& \frac{\sqrt{2}}{\sqrt{\pi\Sigma_{ab}(k)}}\Big|n\mathrm{var}\big(\hat{\gamma}_{ab}(k)\big) - \Sigma_{ab}(k)\Big|\\ &+ 2\frac{\sqrt{2}}{\sqrt{\pi\Sigma_{ab}(k)}}\frac{(n-k)^2}{n}\Big(\max_{i=a,b} D_{i,4,m}\Big)F_m,\end{aligned} \tag{40}$$

where $F_m$ is given in (16). In Regime 1, $D_{j,q,m} = 0$ for $n \ge M$ and thus (40) vanishes for $n$ large enough. In Regime 2, since fourth moments are finite, from (2.17) and Jensen's inequality, we have that $\min_{i=a,b} D_{i,2,m} \le \max_{i=a,b} D_{i,2,m} \le \max_{i=a,b} D_{i,4,m} = o(1)$, as $m \to \infty$, which implies $F_m = \mathcal{O}(1)$. Now, since $m = C\log n$, with $C \ge -3/(2\log(\rho))$, we have that (40) is of the order $\mathcal{O}(n\rho^m) = \mathcal{O}(n^{-1/2})$, in Regime 2 and we had already seen that this is the case in Regime 1. It remains to assess the order of $|n\mathrm{var}\big(\hat{\gamma}_{ab}(k)\big) - \Sigma_{ab}(k)|$, the difference of the finite sample variance and the asymptotic variance of the empirical cross-covariance. This quantity is of the order $\mathcal{O}(n^{-1})$ under condition (2.16), which we have argued holds in both regimes. This can, for example, be concluded from the arguments

Table S2: Table of indices and variables to be partitioned for the addends of the sum in (43)

| | |
|---|---|
| (1,$a$) | (1,$b$) |
| (2,$a$) | (2,$b$) |
| (3,$a$) | (3,$b$) |
| (4,$a$) | (4,$b$) |

(a) Table of indices to be partitioned

| | |
|---|---|
| $Y_{1,a} := X_a(t+k)$ | $Y_{1,b} := X_b(t)$ |
| $Y_{2,a} := X_a(t+k)$ | $Y_{2,b} := X_b(t)$ |
| $Y_{3,a} := X_a(j_1+k)$ | $Y_{3,b} := X_b(j_1)$ |
| $Y_{4,a} := X_a(j_2+k)$ | $Y_{4,b} := X_b(j_2)$ |

(b) Variables of which the cumulants are considered

provided in Section 7.6 in Brillinger (1975). In conclusion, the second term of the bound is of the order $\mathcal{O}(n^{-1/2})$, in Regimes 1 and 2.

With respect to the third term of the bound, $2(n-k)n^{-1/2}K_{2,m}$, note that in Regime 1 we have that $K_{2,m}$ vanishes eventually, as $D_{j,q,m} = 0$ for $n \geq M$ and in Regime 2 we have $K_{2,m} = \mathcal{O}(\rho^m)$, by (2.17). Therefore, the third term in the bound vanishes for $n$ large enough in Regime 1 and is of order $\mathcal{O}(\sqrt{n}\rho^m) = \mathcal{O}(n^{-1/2})$ in Regime 2, if we choose $m = C\log n$, this time it suffices that we have $C \geq -1/\log(\rho)$. In conclusion, the third term in the bound is of the order $\mathcal{O}(n^{-1/2})$, in both regimes.

We now proceed to discuss the rate of the fourth term of the bound, $2n^{-3/2}(\tilde{\Sigma}_{ab}(k))^{-3/2}\sum_{t=1}^{n-k}\tilde{Q}_t$. For this term, the regime makes a difference in the outcome of our analysis. From the discussion of the second term of the bound we have, $\tilde{\Sigma}_{ab}(k) \to \Sigma_{ab}(k) > 0$, as $n \to \infty$. By (18), Lemma S3.4 and condition (2.17) (which holds in both regimes according to Remark 2.7), it suffices to show that $\sup_{t=1,\ldots,n-k} Q_t$ is of the order $\mathcal{O}(1)$ in Regime 1 and of the order $\mathcal{O}(\log n)$ in Regime 2, respectively.

We first consider Regime 1. Bounding $Q_t$ as in (2.13), but with $\{\boldsymbol{X}^{(m)}(t)\}$ replaced by $\{\boldsymbol{X}(t)\}$, we have $Q_t \leq 3(4m+4k+1)^2\|\boldsymbol{X}(0)\|_6^6 = \mathcal{O}(m^2) = \mathcal{O}(1)$, since $m = \min\{n, M\} \leq M$.

Next consider Regime 2. Here we use a version of (2.14), with $\{\boldsymbol{X}^{(m)}(t)\}$ replaced by $\{\boldsymbol{X}(t)\}$:

$$Q_t \leq \text{var}\Big(Z_t \sum_{j\in A_t} Z_j\Big)^{1/2} \text{var}\Big(\sum_{j\in B_t} Z_j\Big)^{1/2} + \frac{1}{2}\Big[\mathbb{E}\Big(Z_t\sum_{j\in A_t} Z_j\Big)^2\Big]^{1/2} \text{var}\Big(\sum_{j\in A_t} Z_j\Big)^{1/2} \quad (41)$$

It suffices to show that the three quantities defined in (12), with $\boldsymbol{X}^{(m)}(t)$ replaced by $\boldsymbol{X}(t)$, are $\mathcal{O}(m)$. We now show that this is implied by condition (2.16), the sumability of cumulants up to order 8. Following arguments from Section S2, we have

$$\text{var}\Big(\sum_{j\in A_t} Z_j\Big) \leq |A_t| \sum_{u=-\infty}^{\infty} |C(u)|, \quad \text{and} \quad \mathbb{E}\Big(Z_t \sum_{j\in A_t} Z_j\Big) \leq \sum_{u=-\infty}^{\infty} |C(u)|, \quad (42)$$

where

$$C(u) = \text{cum}(X_a(k), X_b(0), X_a(u+k), X_b(u)) + \gamma_{aa}(u)\gamma_{bb}(u) + \gamma_{ab}(k-u)\gamma_{ab}(k+u),$$

which is summable according to condition (2.16). Finally, with notation from Table S2(b), we have

$$\sum_{j_1\in A_t}\sum_{j_2\in A_t} \text{cum}(Z_t, Z_t, Z_{j_1}, Z_{j_2}) \leq \sum_{\nu} S(\nu), \quad S(\nu) := \sum_{j_1=-\infty}^{\infty}\sum_{j_2=-\infty}^{\infty}\prod_{r=1}^{R} |\text{cum}(Y_{ij} : (ij) \in \nu_r)|, \quad (43)$$

where the sum extends over all indecomposable partitions $\nu := \{\nu_1, \ldots, \nu_R\}$ of the indices in Table S2(a). To complete our argument it suffices to show that for every indecomposable partition $\nu$ of Table S2(a), with $|\nu_r| \geq 2$,

$$\exists C_\nu : |S(\nu)| \leq C_\nu, \quad (44)$$

with $C_\nu$ being independent of $t$. Then, since there are only finitely many $\nu$, we have shown that the left-hand side in (43) is $\mathcal{O}(1)$ uniformly in $t$. The rigorous proof of (44) is straightforward but tedious. It is given below, after the remaining steps to the proof of Proposition 2.8.

In conclusion, by (41) in combination with the discussion of how to compute (2.14), together with (42), (43), (44), $|A_t| \leq (4(m+k)+1)$ and the choice of $m = C \log n$ for the first, second and third term of the bound, we have shown that in Regime 2 we have $\sup_{t=0,\ldots,n-k} Q_t = \mathcal{O}(\log n)$, which implies that the fourth term of the bound under Regime 2 is of the order $\mathcal{O}(n^{-1/2} \log n)$.

We now complete the proof by showing that for every $\nu$ as in (44), one of the two following assertions holds true:

(a) there exists one set $\nu_\ell$ in $\nu$ such that

$$\sum_{j_1=-\infty}^{\infty} \sum_{j_2=-\infty}^{\infty} |\text{cum}(Y_{ij} : (ij) \in \nu_\ell)| \leq c_\nu,$$

(b) or there exist two distinct sets $\nu_{\ell_1}$ and $\nu_{\ell_2}$ in $\nu$ such that

$$\sum_{j_1=-\infty}^{\infty} \sum_{j_2=-\infty}^{\infty} |\text{cum}(Y_{ij} : (ij) \in \nu_{\ell_1})||\text{cum}(Y_{ij} : (ij) \in \nu_{\ell_2})| \leq c_\nu,$$

with the constant $c_\nu$ not depending on $t$. The desired $S(\nu) \leq C_\nu$ then follows from applying

$$|\text{cum}(Y_{ij} : (ij) \in \nu_r)| \leq |\nu_r|^{|\nu_r|}(|\nu_r| - 1)! \max\{1, \|\boldsymbol{X}(0)\|_{|\nu_r|}\}^{|\nu_r|} =: B(\nu_r), \tag{45}$$

to the remaining factors in the definition of $S(\nu)$. Note that $B(\nu_r)$ does not depend on $t$. The inequality in (45) follows from definition (2.7), the triangle inequality, the fact that the number of partitions of $\nu_r$ can be bounded by $|\nu_r|^{|\nu_r|}$, the fact that the maximum number of sets in a partition of $\nu_r$ is $|\nu_r|$, generalised Hölder inequality, stationarity and Jensen's inequality. In the following, we refer to (3,1) and (3,2) from Table S2(a) as the $j_1$-indices and we refer to (4,1) and (4,2) as the $j_2$-indices. By $j$-indices we will refer to (3,1), (3,2), (4,1) and (4,2), and by non-$j$-indices we will refer to the remaining (1,1), (1,2), (2,1) and (2,2). Now, assume that $\nu := \{\nu_1, \ldots, \nu_R\}$ is an indecomposable partition of Table S2(a), where each $\nu_r$ contains at least two elements. To show that either (a) or (b) holds, it clearly suffices to focus on the possible arrangements of the $j$-indices. To see this, note that by (45) we can ignore sets that don't contain at least one $j$-index.

Before we discuss all possible arrangements of $j$-indices in the indecomposable partitions $\nu$ with $|\nu_r| \geq 2$, we now explain how we show that (a) or (b) holds: To show that (a) holds, for a given partition $\nu := \{\nu_1, \ldots, \nu_R\}$, it suffices to *identify a set $\nu_\ell$ in $\nu$ that contains at least one $j_1$-index, at least one $j_2$-index and at least one non-$j$-index.* We then have $|\nu_\ell| =: p \geq 3$. Say that the elements in $\nu_\ell := \{(u_1, v_1), \ldots, (u_p, v_p)\}$ are ordered such that a $j_1$-index is the first, a $j_2$-index is second and a non-$j$-index is $p$th. Then, by (45) and the stationarity of $\{\boldsymbol{X}(t)\}$ we have

$$\begin{aligned} S(\nu) &\leq \left( \prod_{\substack{r=1 \\ r \neq \ell}}^{R} B(\nu_r) \right) \sum_{j_1=-\infty}^{\infty} \sum_{j_2=-\infty}^{\infty} |\text{cum}(X_{v_1}(j_1 + h_1 - s), X_{v_2}(j_2 + h_2 - s), \ldots, X_{v_p}(0))| \\ &\leq \left( \prod_{\substack{r=1 \\ r \neq \ell}}^{R} B(\nu_r) \right) \sum_{i_1=-\infty}^{\infty} \cdots \sum_{i_{p-1}=-\infty}^{\infty} |\text{cum}(X_{v_1}(i_1), \ldots, X_{v_{p-1}}(i_{p-1}), X_{v_p}(0))| =: c_\nu, \end{aligned}$$

where $h_1, h_2 \in \{0, k\}$ (depending on $v_1$ and $v_2$), $s \in \{t, t+k\}$ (depending on $v_p$), and the ... in the first line represents $|\nu_\ell| - 3$ variables from Table S2(a) with their time index shifted by $s$.

Now, to show that (b) holds, for a given partition $\nu := \{\nu_1, \dots, \nu_R\}$, it suffices to *identify sets $\nu_{\ell_1}$ and $\nu_{\ell_2}$ that satisfy the following conditions*:

(I) $\nu_{\ell_1} := \{(u_{1,1}, v_{1,1}), \dots, (u_{1,p_1}, v_{1,p_1})\}$ contains at least one $j_1$-index;

(II) $\nu_{\ell_2} := \{(u_{2,1}, v_{2,1}), \dots, (u_{2,p_2}, v_{2,p_2})\}$ contains at least one $j_2$-index;

(III) either $\nu_{\ell_1}$ or $\nu_{\ell_2}$ contains at least one non-$j$-index;

(IV) if $\nu_{\ell_1}$ contains both $j_1$-indices, then it must contain at least three elements and also if $\nu_{\ell_2}$ contains both $j_2$-indices, then it must contain at least three elements.

Note that for any indecomposable partition, (IV) is always satisfied, because a partition with a set of exactly two elements that are $j_1$-indices or exactly two elements that are $j_2$-indices is decomposable.

First, consider the case where the non-$j$-index which we know exists due to (III) is in $\nu_{\ell_1}$. The case where it is in $\nu_{\ell_2}$ can be treated analogously. We can number the sets and indices such that $(u_{1,1}, v_{1,1})$ is a $j_1$-index, $(u_{2,1}, v_{2,1})$ is a $j_2$-index, and $(u_{1,p_1}, v_{1,p_1})$ is a non-$j$-index. Further, due to (IV) and $|\nu_{\ell_2}| \geq 2$, $(u_{2,p_2}, v_{2,p_2})$ is not a $j_2$-index. Thus, for $\nu_{\ell_1}$ and $\nu_{\ell_2}$ satisfying (I)–(IV) we have

$$\begin{aligned} S(\nu) &\leq \left( \prod_{\substack{r=1 \\ r \notin \{\ell_1, \ell_2\}}}^{R} B(\nu_r) \right) \sum_{j_1=-\infty}^{\infty} \sum_{j_2=-\infty}^{\infty} |\text{cum}(X_{v_{1,1}}(j_1 + h_1 - s_1), \dots, X_{v_{1,p_1}}(0))| \\ &\qquad\qquad \times |\text{cum}(X_{v_{2,1}}(j_2 + h_2 - s_2), \dots, X_{v_{2,p_2}}(0))| \\ &\leq \left( \prod_{\substack{r=1 \\ r \notin \{\ell_1, \ell_2\}}}^{R} B(\nu_r) \right) \sum_{i_{1,1}=-\infty}^{\infty} \cdots \sum_{i_{1,p_1-1}=-\infty}^{\infty} |\text{cum}(X_{v_{1,1}}(i_{1,1}), \dots, X_{v_{1,p_1-1}}(i_{1,p_1-1}), X_{v_{1,p_1}}(0))| \\ &\qquad \times \sum_{i_{2,1}=-\infty}^{\infty} \cdots \sum_{i_{2,p_2-1}=-\infty}^{\infty} |\text{cum}(X_{v_{2,1}}(i_{2,1}), \dots, X_{v_{2,p_2-1}}(i_{2,p_2-1}), X_{v_{2,p_2}}(0))| =: c_\nu, \end{aligned}$$

where $h_1, h_2 \in \{0, k\}$ (depending on the values of $v_{1,1}$ and $v_{2,1}$), $s_1 \notin \{j_1, j_1 + k, j_2, j_2 + k\}$, and $s_2 \notin \{j_2, j_2 + k\}$. For the second inequality we have substituted $j_1 + h_1 - s_1$ by $i_{1,1}$, which corresponds to a shift in the index of the $j_1$-sum, since $(u_{1,p_1}, v_{1,p_1})$ is a non-$j$-index, and then bounded in two steps

$$\begin{aligned} &|\text{cum}(X_{v_{1,1}}(i_{1,1}), \dots, X_{v_{1,p_1}}(0))| \\ &\qquad \leq \sum_{i_{1,2}=-\infty}^{\infty} \cdots \sum_{i_{1,p_1-1}=-\infty}^{\infty} |\text{cum}(X_{v_{1,1}}(i_{1,1}), \dots, X_{v_{1,p_1-1}}(i_{1,p_1-1}), X_{v_{1,p_1}}(0))|, \end{aligned} \tag{46}$$

where the right-hand side in (46) does not depend on $j_2$ (anymore). Then, $S(\nu) \leq c_\nu$ follows from

$$\begin{aligned} &\sum_{j_2=-\infty}^{\infty} |\text{cum}(X_{v_{2,1}}(j_2 + h_2 - s_2), \dots, X_{v_{2,p_2}}(0))| \\ &\qquad \leq \sum_{i_{2,1}=-\infty}^{\infty} \cdots \sum_{i_{2,p_2-1}=-\infty}^{\infty} |\text{cum}(X_{v_{2,1}}(i_{2,1}), \dots, X_{v_{2,p_2-1}}(i_{2,p_2-1}), X_{v_{2,p_2}}(0))|, \end{aligned}$$

which we have since $(u_{2,p_2}, v_{2,p_2})$ is not a $j_2$-index. To conclude the proof it remains to discuss all possible arrangements of the $j$-indices. To cover all cases, we organise the following according to the number of sets in the partition that contain at least one $j$-index.

1. There exists one $\nu_\ell$ in the partition that contains all $j$-indices. In this case $\nu_\ell$ has to also contain at least one more non-$j$-index, because otherwise the partition would be decomposable. Thus, from the above, we have that (a) holds.

2. There exist two sets $\nu_{\ell_1}$ and $\nu_{\ell_2}$ such that each contains at least one $j$-index and their union contains all $j$-indices. The situation where $|\nu_{\ell_1}| = |\nu_{\ell_2}| = 2$ would imply a decomposable partition and is therefore not possible. Thus, we either have $|\nu_{\ell_1}| \geq 3$ and $|\nu_{\ell_2}| \geq 2$ or $|\nu_{\ell_1}| \geq 2$ and $|\nu_{\ell_2}| \geq 3$. Two sub-cases are possible: (i) one of the sets contains three $j$-indices and the other contains only one $j$-index, or (ii) $\nu_{\ell_1}$ and $\nu_{\ell_2}$ both contains exactly two $j$-indices.

   If (i), then the set containing only one $j$-index has to also contain at least one non-$j$-index (i. e., (III) is satisfied). Then (b) holds, since if the set with only one $j$-index contains, say, a $j_1$-index then the other set contains a $j_2$-index (or vice versa; i. e., (I) and (II) are satisfied).

   If (ii), then the set with at least three elements contains a non-$j$-index (i. e., (III) is satisfied). Further, since each set contains exactly two $j$-indices and there are exactly two $j_1$-indices and two $j_2$-indices we have that if one set contains a $j_1$-index, the other has to contain a $j_2$-index (i. e., (I) and (II) are satisfied). Thus, (b) holds.

3. There exist a set with exactly two $j$-indices and two sets with exactly one $j$-index each. Then, because of $|\nu_r| \geq 2$, the two sets with one $j$-index also contain at least one non-$j$-indices. Two sub-cases are possible: the $j$-indices in the sets with exactly one $j$-index are either (i) both $j_1$-indices or both $j_2$-indices, or (ii) we have an $j_1$-index in one and a $j_2$-index in the other set.

   If (i), then the $j$-indices in the set with exactly two $j$-indices are either both $j_2$-indices or both $j_1$-indices, respectively. The set with exactly two $j$-indices (either two $j_1$- or two $j_2$-indices) then has to have another non-$j$-index, as the partition would otherwise be decomposable. Taking $\nu_{\ell_1}$ as the set with two $j$-indices and $\nu_{\ell_2}$ as one of the two sets with exactly one $j$-index, we see that (b) holds: the indices are $j_1$- and $j_2$ indices in the two sets (i. e., (I) and (II) are satisfied) and each set has at least one non-$j$-index (i. e., (III) is satisfied).

   If (ii), then we take $\nu_{\ell_1}$ as one of the sets with exactly one $j_1$-index and $\nu_{\ell_2}$ as the set with exactly one $j_2$-index (i. e., (I) and (II) are satisfied). Then (b) is satisfied, as each of these sets also has one non-$j$-index (i. e., (III) is satisfied).

4. There exist four sets with exactly one $j$-index each. We take $\nu_{\ell_1}$ as one of the sets with exactly one $j_1$-index and $\nu_{\ell_2}$ as the set with exactly one $j_2$-index (i. e., (I) and (II) are satisfied). Each of these sets also has one non-$j$-index (i. e., (III) is satisfied), thus (b) is satisfied.

This finishes the proof of (44) and also clarified how to find $C_\nu$ in terms of the sum in (2.16) and the quantities $B(\nu_r)$ defined in (45), which are all independent of $t$. ■

## S13 Additional tables for the examples in the paper

In this section we provide the values of the bound from Theorem 2.3 and the values of the true 1-Wasserstein distances considered in Theorem 2.3 for the case of the example in Section 3.1 the cases that were committed in the main paper: $\varepsilon_t \sim \mathcal{N}(0,1)$, and $\varepsilon_t \sim \sqrt{12/14}t_{14}$.

## References

Brillinger, D. R. (1975). *Time Series: Data Analysis and Theory*. New York: Holt, Rinehart and Winston, Inc.

Chen, L. H. Y., L. Goldstein, and Q.-M. Shao (2011). *Normal Approximation by Stein's Method*. Probability and its Applications. Springer.

Munk, A. and C. Czado (1998). Nonparametric validation of similar distributions and assessment of goodness of fit. *Journal of the Royal Statistical Society. Series B (Statistical Methodology) 60*(1), 223–241.

Nourdin, I. and G. Peccati (2012). *Normal Approximations with Malliavin Calculus : From Stein's Method to Universality.* Number Vol. 192 in Cambridge Tracts in Mathematics. Cambridge University Press.

| k | a \| 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 | 4.438 | 3.784 | 3.219 | 2.894 | 2.533 | 2.340 | 2.222 | 1.661 | 1.239 | 0.967 |
| | 0.3 | 7.005 | 5.701 | 5.033 | 4.673 | 3.984 | 3.585 | 3.324 | 2.564 | 1.986 | 1.488 |
| | 0.5 | 9.842 | 8.080 | 7.164 | 6.484 | 5.676 | 5.110 | 4.715 | 3.650 | 2.798 | 2.128 |
| | 0.7 | 13.981 | 11.712 | 10.379 | 9.485 | 8.289 | 7.511 | 6.931 | 5.375 | 4.124 | 3.137 |
| 1 | 0 | 2.564 | 1.818 | 1.485 | 1.286 | 1.050 | 0.909 | 0.813 | 0.574 | 0.406 | 0.287 |
| | 0.1 | 6.778 | 5.045 | 4.259 | 3.801 | 3.279 | 2.988 | 2.752 | 1.996 | 1.477 | 1.135 |
| | 0.3 | 9.065 | 6.998 | 6.105 | 5.498 | 4.658 | 4.170 | 3.849 | 2.907 | 2.226 | 1.648 |
| | 0.5 | 11.388 | 9.075 | 7.926 | 7.126 | 6.179 | 5.522 | 5.084 | 3.886 | 2.947 | 2.221 |
| | 0.7 | 14.561 | 12.279 | 10.808 | 9.820 | 8.531 | 7.699 | 7.091 | 5.452 | 4.152 | 3.138 |
| 2 | 0 | 4.088 | 2.916 | 2.385 | 2.067 | 1.688 | 1.462 | 1.308 | 0.924 | 0.653 | 0.462 |
| | 0.1 | 8.272 | 6.135 | 5.159 | 4.582 | 3.917 | 3.540 | 3.249 | 2.348 | 1.726 | 1.312 |
| | 0.3 | 10.061 | 7.882 | 6.848 | 6.167 | 5.216 | 4.658 | 4.288 | 3.197 | 2.429 | 1.792 |
| | 0.5 | 12.843 | 10.204 | 8.804 | 7.902 | 6.803 | 6.068 | 5.575 | 4.234 | 3.196 | 2.400 |
| | 0.7 | 15.174 | 13.119 | 11.527 | 10.442 | 9.040 | 8.137 | 7.485 | 5.726 | 4.342 | 3.270 |

Table S3: 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 $n$. The data stems from an AR(1) process with $\varepsilon_t \sim \mathcal{N}(0, 1)$ where $a$ takes a range of values.

| k | a \| n | 25 | 50 | 75 | 100 | 150 | 200 | 250 | 500 | 1000 | 2000 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0.129 | 0.091 | 0.074 | 0.065 | 0.053 | 0.046 | 0.041 | 0.029 | 0.020 | 0.014 |
| | 0.1 | 0.135 | 0.096 | 0.078 | 0.068 | 0.055 | 0.048 | 0.043 | 0.030 | 0.022 | 0.015 |
| | 0.3 | 0.191 | 0.137 | 0.112 | 0.097 | 0.079 | 0.069 | 0.062 | 0.044 | 0.031 | 0.022 |
| | 0.5 | 0.360 | 0.261 | 0.215 | 0.187 | 0.153 | 0.133 | 0.119 | 0.084 | 0.060 | 0.042 |
| | 0.7 | 0.970 | 0.718 | 0.595 | 0.520 | 0.428 | 0.372 | 0.334 | 0.237 | 0.168 | 0.119 |
| 1 | 0 | 0.051 | 0.026 | 0.018 | 0.014 | 0.009 | 0.007 | 0.006 | 0.003 | 0.001 | 0.001 |
| | 0.1 | 0.079 | 0.052 | 0.041 | 0.035 | 0.028 | 0.024 | 0.022 | 0.015 | 0.011 | 0.008 |
| | 0.3 | 0.209 | 0.149 | 0.122 | 0.106 | 0.087 | 0.075 | 0.067 | 0.048 | 0.034 | 0.024 |
| | 0.5 | 0.440 | 0.317 | 0.260 | 0.226 | 0.185 | 0.161 | 0.144 | 0.102 | 0.072 | 0.051 |
| | 0.7 | 1.132 | 0.828 | 0.685 | 0.596 | 0.490 | 0.426 | 0.382 | 0.271 | 0.192 | 0.136 |
| 2 | 0 | 0.062 | 0.032 | 0.022 | 0.016 | 0.011 | 0.008 | 0.007 | 0.003 | 0.002 | 0.001 |
| | 0.1 | 0.067 | 0.035 | 0.024 | 0.019 | 0.013 | 0.010 | 0.008 | 0.005 | 0.003 | 0.002 |
| | 0.3 | 0.144 | 0.098 | 0.078 | 0.067 | 0.054 | 0.047 | 0.042 | 0.029 | 0.021 | 0.015 |
| | 0.5 | 0.406 | 0.293 | 0.241 | 0.210 | 0.172 | 0.149 | 0.133 | 0.095 | 0.067 | 0.047 |
| | 0.7 | 1.206 | 0.881 | 0.728 | 0.634 | 0.520 | 0.452 | 0.405 | 0.287 | 0.203 | 0.144 |

Table S4: Value of the true 1-Wasserstein distance considered in Theorem 2.3 for empirical autocovariances, for a range of lags $k$ and sample sizes $n$. The data stems from an AR(1) process with $\varepsilon_t \sim \mathcal{N}(0,1)$ where $a$ takes a range of values.

| k | a \| 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 | 6.056 | 5.796 | 5.004 | 4.443 | 3.800 | 3.438 | 3.205 | 2.551 | 1.870 | 1.413 |
| | 0.3 | 9.466 | 7.848 | 6.795 | 6.203 | 5.430 | 4.840 | 4.448 | 3.447 | 2.611 | 1.966 |
| | 0.5 | 11.987 | 9.754 | 8.544 | 7.774 | 6.736 | 6.089 | 5.592 | 4.307 | 3.285 | 2.487 |
| | 0.7 | 15.330 | 12.753 | 11.273 | 10.290 | 8.973 | 8.106 | 7.479 | 5.785 | 4.427 | 3.359 |
| 1 | 0 | 2.564 | 1.818 | 1.485 | 1.286 | 1.050 | 0.909 | 0.813 | 0.574 | 0.406 | 0.287 |
| | 0.1 | 7.560 | 5.787 | 4.868 | 4.329 | 3.710 | 3.362 | 3.131 | 2.264 | 1.667 | 1.269 |
| | 0.3 | 10.485 | 8.022 | 6.946 | 6.290 | 5.308 | 4.735 | 4.354 | 3.293 | 2.502 | 1.853 |
| | 0.5 | 12.840 | 10.173 | 8.881 | 7.958 | 6.886 | 6.148 | 5.645 | 4.303 | 3.255 | 2.448 |
| | 0.7 | 15.722 | 13.175 | 11.577 | 10.512 | 9.119 | 8.224 | 7.564 | 5.805 | 4.413 | 3.331 |
| 2 | 0 | 4.088 | 2.916 | 2.385 | 2.067 | 1.688 | 1.462 | 1.308 | 0.924 | 0.653 | 0.462 |
| | 0.1 | 9.329 | 6.892 | 5.780 | 5.121 | 4.358 | 3.922 | 3.621 | 2.612 | 1.913 | 1.444 |
| | 0.3 | 10.951 | 8.618 | 7.453 | 6.725 | 5.673 | 5.055 | 4.644 | 3.461 | 2.622 | 1.930 |
| | 0.5 | 13.933 | 11.008 | 9.489 | 8.499 | 7.312 | 6.511 | 5.972 | 4.527 | 3.411 | 2.556 |
| | 0.7 | 16.124 | 13.891 | 12.176 | 11.024 | 9.533 | 8.577 | 7.880 | 6.020 | 4.560 | 3.430 |

Table S5: 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 $n$. The data stems from an AR(1) process with $\varepsilon_t \sim \sqrt{12/14}\, t_{14}$ where $a$ takes a range of values.

| k | a \| n | 25 | 50 | 75 | 100 | 150 | 200 | 250 | 500 | 1000 | 2000 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0.206 | 0.151 | 0.125 | 0.109 | 0.090 | 0.078 | 0.070 | 0.050 | 0.035 | 0.025 |
| | 0.1 | 0.213 | 0.155 | 0.128 | 0.112 | 0.092 | 0.080 | 0.072 | 0.051 | 0.037 | 0.026 |
| | 0.3 | 0.270 | 0.197 | 0.163 | 0.142 | 0.117 | 0.102 | 0.091 | 0.065 | 0.046 | 0.033 |
| | 0.5 | 0.443 | 0.325 | 0.270 | 0.235 | 0.194 | 0.169 | 0.151 | 0.108 | 0.076 | 0.054 |
| | 0.7 | 1.072 | 0.798 | 0.664 | 0.581 | 0.479 | 0.417 | 0.375 | 0.267 | 0.190 | 0.134 |
| 1 | 0 | 0.062 | 0.034 | 0.023 | 0.017 | 0.012 | 0.009 | 0.007 | 0.004 | 0.002 | 0.001 |
| | 0.1 | 0.092 | 0.061 | 0.048 | 0.041 | 0.033 | 0.028 | 0.025 | 0.018 | 0.012 | 0.009 |
| | 0.3 | 0.234 | 0.169 | 0.139 | 0.121 | 0.099 | 0.086 | 0.077 | 0.055 | 0.039 | 0.027 |
| | 0.5 | 0.482 | 0.350 | 0.289 | 0.252 | 0.207 | 0.179 | 0.161 | 0.114 | 0.081 | 0.057 |
| | 0.7 | 1.207 | 0.888 | 0.736 | 0.643 | 0.529 | 0.460 | 0.412 | 0.294 | 0.208 | 0.147 |
| 2 | 0 | 0.074 | 0.039 | 0.027 | 0.020 | 0.014 | 0.010 | 0.008 | 0.004 | 0.002 | 0.001 |
| | 0.1 | 0.078 | 0.042 | 0.029 | 0.022 | 0.016 | 0.012 | 0.010 | 0.006 | 0.003 | 0.002 |
| | 0.3 | 0.156 | 0.106 | 0.085 | 0.073 | 0.059 | 0.051 | 0.045 | 0.032 | 0.022 | 0.016 |
| | 0.5 | 0.428 | 0.311 | 0.257 | 0.224 | 0.184 | 0.160 | 0.143 | 0.101 | 0.072 | 0.051 |
| | 0.7 | 1.257 | 0.923 | 0.764 | 0.667 | 0.548 | 0.477 | 0.427 | 0.304 | 0.216 | 0.153 |

Table S6: Value of the true 1-Wasserstein distance considered in Theorem 2.3 for empirical autocovariances, for a range of lags $k$ and sample sizes $n$. The data stems from an AR(1) process with $\varepsilon_t \sim \sqrt{12/14}\, t_{14}$ where $a$ takes a range of values.