An Online Bootstrap for Time Series

Nicolai Palm
LMU Munich
Munich Center for Machine Learning
(MCML)

Thomas Nagler
LMU Munich
Munich Center for Machine Learning
(MCML)

Abstract

Resampling methods such as the bootstrap have proven invaluable in the field of machine learning. However, the applicability of traditional bootstrap methods is limited when dealing with large streams of dependent data, such as time series or spatially correlated observations. In this paper, we propose a novel bootstrap method that is designed to account for data dependencies and can be executed online, making it particularly suitable for real-time applications. This method is based on an autoregressive sequence of increasingly dependent resampling weights. We prove the theoretical validity of the proposed bootstrap scheme under general conditions. We demonstrate the effectiveness of our approach through extensive simulations and show that it provides reliable uncertainty quantification even in the presence of complex data dependencies. Our work bridges the gap between classical resampling techniques and the demands of modern data analysis, providing a valuable tool for researchers and practitioners in dynamic, data-rich environments.

1 INTRODUCTION

Uncertainty quantification (UQ) has become indispensable in statistics, machine learning, and numerous other scientific disciplines. It plays a pivotal role in assessing the reliability of predictions, parameter estimates, and models. Bootstrapping is a universal ad-hoc approach for UQ and a cornerstone of many approaches leveraging UQ. Especially in the context of theoretically unknown or difficult-to-compute uncertainty distributions, bootstrap methods have proven to be remarkably powerful.

A potential bottleneck in real applications is that computation of the bootstrapped distributions and storage of underlying data gets expensive in time and memory with increasing amount of data. Especially in the context of big data sets and/or streaming data settings, this limits the applicability of standard bootstrap methods. Yet, this is a common setup in modern data analysis.

Online algorithms attempt to address these challenges by performing continuous, cheap updates of a model/estimate — optimally processing only a fraction of the data within each iteration. This significantly decreases the associated costs and requires only a fraction of the data kept in memory.

Existing bootstrap schemes, however, are either not computable by an online algorithm or make restrictive assumptions on dependence in the data. This motivates the development of an online bootstrap scheme for general time series.

Our main contributions can be summarized as follows:

1. We propose a novel bootstrap procedure that (i) can be computed online and (ii) works for both independent and dependent data streams. To the best of our knowledge, this is the first such bootstrapping scheme.
2. We prove its theoretical validity under general conditions and provide theoretical insights into the optimal choice of hyperparameters.
3. We demonstrate validity and effectiveness through a number of simulations illustrating its advantages over the current state of the art.

The remainder of the paper is structured as follows. Section 2 provides some theoretical background and summarizes related work. Section 3 introduces the

Proceedings of the 27^{th} International Conference on Artificial Intelligence and Statistics (AISTATS) 2024, Valencia, Spain. PMLR: Volume 238.

new method and presents the main theoretical results, which is evaluated empirically in Section 4. Section 5 discusses applications and limitations. All proofs are provided in the supplementary material.

2 BACKGROUND AND RELATED WORK

2.1 Online learning

Online learning deals with problems where there is a continuous stream of data (Cesa-Bianchi and Lugosi, 2006). Such problems arise naturally if events are observed at the moment they occur. In other settings, a complete data set is available from the start, but it is computationally preferable to work through it sequentially or in batches. Online convex optimization methods like stochastic gradient descent are prime examples (Shalev-Shwartz et al., 2012).

In such situations, one could recompute a quantity of interest at every time step using all data observed so far. However, this is very inefficient if the number of parameters or observations is large. In fact, re-computing on the full data set is often infeasible, because there is limited memory or limited time to update. For example, computing the sample average at $\overline{X}_n = \frac{1}{n}\sum_{t=1}^n X_t$ naively requires $O(n)$ time at any given moment in time.

Online algorithms strive to perform continuous updates that are cheap, optimally processing only a single observation at every step. Continuing the example above, the sample average at time n adheres to the cheap update rule $\overline{X}_n = (1-1/n)\overline{X}_{n-1} + X_n/n$, which scales as $O(1)$ in time and memory. Such algorithms and their corresponding theoretical properties are currently subject of interest in numerous fields, such as optimization (Fang et al., 2018; Godichon-Baggioni, 2019; Zhong et al., 2023), multi-armed bandits (Dimakopoulou et al., 2021; Wan et al., 2023) and reinforcement learning (Ramprasad et al., 2022).

2.2 Bootstrapping

Bootstrapping is considered as one of the fundamental achievements of statistics (Kotz and Johnson, 1992). Generally speaking, bootstrapping is a form of resampling: From a given set of samples $x_1, \dots, x_n$ according to some random variables $X_1, \dots, X_n$, we generate synthetic samples $x_1^*, \dots, x_n^*$, or more generally, synthetic random variables $X_1^*, \dots, X_n^*$ depending on $X_1, \dots, X_n$. Efron (1979) proposed the first bootstrap, based on sampling with replacement, nowadays called the empirical bootstrap.

Example 2.1 (Empirical bootstrap)**.** *The empirical bootstrap generates a new sample $X_1^*, \dots, X_n^*$ by drawing uniformly at random from the observed sample $\{X_1, ..., X_n\}$ (with replacement).*

The multiplier bootstrap (Van der Vaart and Wellner, 1996) offers a general class of bootstrapping schemes based on perturbations of the original observations with suitable weights. The empirical bootstrap cannot be computed online, because it requires keeping track of the entire observed sample $\{X_1, \dots, X_n\}$. The multiplier bootstrap does not suffer from this issue.

Example 2.2 (Multiplier bootstrap)**.** *Let $V_1, \dots, V_n$ be iid real valued random variables with*

$$\mathbb{E}(V_i) = \mathbb{V}\mathrm{ar}(V_i) = 1.$$

Then we obtain the multiplier bootstrap for iid data by

$$X_i^* = \frac{V_i}{\overline{V}_n} X_i, \quad \textit{where} \quad \overline{V}_n = \frac{1}{n}\sum_{i=1}^n V_i.$$

Popular special cases include the Gaussian bootstrap (Burke, 1998) and Bayesian bootstrap (Rubin, 1981) where the weights V_i are drawn from the standard normal and standard exponential distributions, respectively. The *iid* multiplier bootstrap can be computed online. Indeed, each X_i^* depends only on the i-th observation X_i, a new weight V_i and a running average of V_i's. Fang et al. (2018) and Zhong et al. (2023) used this insight to design an online method for computing bootstrap confidence intervals for the SGD estimator. However, this method is only valid for *iid* data.

Recall the random variables X_i^* depend on X_i. We abbreviate the probabilities and variances given realizations of X_i by

$$\mathbb{P}^*(\cdot) = \mathbb{P}(\cdot \mid X_1, X_2, \dots),$$
$$\mathbb{V}\mathrm{ar}^*(\cdot) = \mathbb{V}\mathrm{ar}(\cdot \mid X_1, X_2, \dots).$$

These quantities depend on $X_1, ..., X_n$ and are, therefore, itself random variables. Now if the average over $(X_i)_{i\in\mathbb{N}}$ satisfies a central limit theorem, then, we expect the average over synthetic random variables to satisfy a similar central limit theorem.

Definition 2.3 (Bootstrap consistency)**.** *Let $(X_i)_{i\in\mathbb{N}}$ and $(X_i^*)_{i\in\mathbb{N}}$ be sequences of $\mathbb{R}^d$-valued random variables and X_i^* depending on X_i. Define*

$$T_n = \frac{1}{n}\sum_{i=1}^n X_i \textit{ and } T_n^* = \frac{1}{n}\sum_{i=1}^n X_i^*$$

The sequence $(X_i^*)_{i\in\mathbb{N}}$ is a consistent resampling scheme for $(X_i)_{i\in\mathbb{N}}$ *if*

$$\sup_{x\in\mathbb{R}^d}\Big|\mathbb{P}^*\left\{\sqrt{n}(T_n^*-T_n)\le x\right\} - \mathbb{P}\left\{\sqrt{n}(T_n-\mathbb{E}(T_n))\le x\right\}\Big| \overset{n\to\infty}{\to} 0,$$

in probability with respect to $(X_i)_{i\in\mathbb{N}}$.

In practice, one generates a few hundred resampled data sets and approximates the distribution $\mathbb{P}^*\{\sqrt{n}(T_n^*-T_n)\le x\}$ by empirical quantities of the bootstrap replicates. From this, other measures of uncertainty, such as confidence intervals or mappings of the underlying statistic, can be derived (e.g., Van der Vaart, 2000, Chapter 23).

Example 2.1 and Example 2.2 are consistent resampling schemes provided that the underlying data is independent, but fail otherwise.

2.3 Time series bootstrap

Künsch (1989) proposed the blockwise bootstrap as a general resampling scheme for time series. Roughly, the idea is to draw overlapping blocks of observations. Since observations appear in blocks, the resampled observations naturally inherit dependencies from the original samples. Increasing the block lengths with the sample sizes make the blockwise bootstrap a consistent resampling scheme.

Later, extensions (Politis and Romano, 1993) and further investigations (Hall et al., 1995) of the blockwise bootstrap arose. Recently, Liu et al. (2023) use the blockwise bootstrap for inference in SGD estimators. Bühlmann (1993) proposed a multiplier variant of the blockwise bootstrap that includes regular block bootstrap methods as a special case. Here, the multiplier weights are itself a dependent time series. Increasing the multiplier weights' serial dependence then allows to capture the dependencies in the original observations.

The above methods are not fit for the online setting, however. For the method to work, all blocks have to increase in size with n. To compute the bootstrap in practice, the entire data set $(X_i)_{i=1}^n$ needs to be kept in memory and processed fully, every time the block size changes. This quickly becomes prohibitively expensive when n is large.

3 NEW BOOTSTRAP PROCEDURE

3.1 Proposed method

Example 2.2 gives rise to a general class of bootstrapping schemes by constructing synthetic random variables

$$X_i^* = \frac{V_i}{\overline{V}_n}X_i.$$

A key insight from the block multiplier bootstrap is the following: to remain valid for time series $(X_i)_{i\in\mathbb{N}}$, the dependencies between weights V_i and V_j must increase with the sample size n, but at the same time remain almost independent when the time gap $|i-j|$ is sufficiently large compared to n. In the non-iid case, a scaling of the weights by their arithmetic mean is also necessary.

As a general construction of such $(V_i)_{i\in\mathbb{N}}$ we propose the following autoregressive sequence of weights:

Construction 3.1. *Let* $(\zeta_i)_{i\in\mathbb{N}}$ *be an iid sequence such that* $\zeta_i\sim\mathcal{N}(0,1)$. *Define*

$$V_0 = 0$$
$$V_i = 1+\rho_i(V_{i-1}-1)+\sqrt{1-\rho_i^2}\zeta_i$$

where $\rho_i = 1-i^{-\beta}$, $\beta\in(0,\frac{1}{2})$, *and the bootstrapping scheme*

$$X_i^* = \frac{V_i}{\overline{V}_n}X_i \quad \textit{with} \quad \overline{V}_n = \frac{1}{n}\sum_{i=1}^n V_i.$$

The proposed weight sequence $(V_i)_{i\in\mathbb{N}}$ and the corresponding bootstrap average

$$\overline{X}_n^* = \frac{1}{n}\sum_{i=1}^n \frac{V_i}{\overline{V}_n}X_i$$

can naturally be computed with cheap online updates. Indeed, the i-th bootstrapped variable $X_i^* = (V_i/\overline{V}_i)X_i$ depends only on the predecessor weight V_{i-1}, a single random perturbation ζ_i and the scaling average $\overline{V}_n$. The latter can be obtained recursively by

$$\overline{V}_n = \frac{(n-1)\overline{V}_{n-1}+V_n}{n}.$$

In summary, the bootstrap average can be updated via the relation

$$\overline{X}_n^* = \frac{(n-1)\overline{V}_{n-1}\overline{X}_{n-1}^* + X_nV_n}{(n-1)\overline{V}_{n-1}+V_n},$$

which scales as $O(1)$ in memory and time. This is a huge computational advantage compared to (multiplier) block bootstrap methods and opens new application areas for bootstrapping methods in online settings.

To quantify uncertainties in practice, we have to keep several independent bootstrap 'chains' $\overline{X}_n^{*(1)}, \overline{X}_n^{*(2)}, \ldots$. From those, we can compute the empirical standard deviation or quantiles at any point in time. The whole procedure is summarized in Algorithm 1.

Algorithm 1 Online AR-bootstrap

Initialize: $\overline{X}^{*(b)} = 0$, $V^{(b)} = 0$, $\overline{V}^{(b)} = 0$, $b = 1, \ldots, B$.

For times $t = 1, 2, \ldots$:

1. Observe new datum X_t.
2. **For all** $b = 1, \ldots, B$:
   (i) Simulate $\zeta^{(b)} \sim \mathcal{N}(0, 1)$.
   (ii) With $\rho = 1 - t^{-\beta}$, update (in this order)
   $$V^{(b)} \leftarrow 1 + \rho(V^{(b)} - 1) + \sqrt{1-\rho^2}\zeta^{(b)},$$
   $$\overline{X}^{*(b)} \leftarrow \frac{(t-1)\overline{V}^{(b)}\overline{X}^{*(b)} + X_t V^{(b)}}{(t-1)\overline{V}^{(b)} + V^{(b)}},$$
   $$\overline{V}^{(b)} \leftarrow (1 - 1/t)\overline{V}^{(b)} + V^{(b)}/t.$$
3. Compute empirical variance and/or quantiles of $\{\overline{X}^{*(1)}, \ldots, \overline{X}^{*(B)}\}$ to quantify uncertainty in $\overline{X}_t$.

3.2 Theory

In the following, we provide rigorous mathematical guarantees for the validity of the proposed scheme. We first recall some common concepts from time series analysis.

Definition 3.2 (Stationarity). *A stochastic process $(X_i)_{i\in\mathbb{N}}$ is strictly stationary if*
$$P_{(X_{t_1},\ldots,X_{t_n})} = P_{(X_{t_1+\tau},\ldots,X_{t_n+\tau})}$$
for every $\tau, n, t_1, \ldots, t_n \in \mathbb{N}$.

A stationary time series does not change its fundamental behavior, at least on large time scales. Stationarity is a standard condition for statistical limit theorems. In applications, it is often ensured by appropriate preprocessing steps like detrending or differencing (see, Hamilton, 2020, and the discussion in Section 5).

Definition 3.3 (α-mixing). *Let $(X_i)_{i\in\mathbb{N}}$ be a strictly stationary stochastic process. Define the α-mixing coefficient of order h*
$$\alpha(h) = \sup\{|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| \colon s \in \mathbb{N}, A \in \sigma(X_i | i \leq s), B \in \sigma(X_i | i > s + h)\}.$$
Then, $(X_i)_{i\in\mathbb{N}}$ is α-mixing (or strong mixing) if
$$\alpha(h) \overset{h\to\infty}{\to} 0.$$

The α-mixing coefficient quantifies how quickly the influence of past events diminishes as one moves further into the sequence. Accordingly, α-mixing means that the events become close to independent when they are far apart in time. If the strong mixing coefficients converge fast enough to zero, the sequence satisfies a central limit theorem (Bosq, 2012, Theorem 1.7).

Now denote by
$$C^X(h) = \mathbb{C}\mathrm{ov}(X_i, X_{i+h})$$
the covariance of X_i and X_{i+h} for $h \geq 0$. Since $(X_i)_{i\in\mathbb{N}}$ is stationary, $C^X(h)$ is independent of i. Our main results require the following conditions on the observed sequence $(X_i)_{i=1}^n$.

(A1) $\mathbb{E}(X_i^8) < \infty$.

(A2) $\alpha(i) = O(i^{-\gamma})$ for some $\gamma > 2$.

(A3) $\lim_{n\to\infty} \frac{1}{n} \sum_{h=-n}^{n} |h|^{\frac{1}{\beta}} |C^X(|h|)| = 0$.

The first condition excludes extremely heavy tails in the variables X_i, the other two restrict the strength of dependence in the time series. Overall, the conditions should be considered rather mild in view of the time series bootstrap literature (cf., Künsch, 1989; Bühlmann, 1993).

To simplify our asymptotic analysis of the procedure, we first have a closer look at the role of the scaling average $\overline{V}_n$. The following result shows that scaling by $\overline{V}_n$ implicitly allows to assume that $\mathbb{E}[X_i] = 0$, but is otherwise negligible asymptotically.

Lemma 3.4. *If the time series $X_1, X_2, \ldots \in \mathbb{R}$ satisfies assumptions (A1)–(A3), it holds*
$$\frac{1}{n}\sum_{i=1}^{n} \frac{V_i}{\overline{V}_n} X_i - \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}\sum_{i=1}^{n} V_i(X_i - \mathbb{E}[X_i]) - \frac{1}{n}\sum_{i=1}^{n}(X_i - \mathbb{E}[X_i]) + o_P(n^{-1/2}).$$

Our main result establishes the validity of the bootstrap scheme.

Theorem 3.5. *If the time series $X_1, X_2, \ldots \in \mathbb{R}$ satisfies assumptions (A1)–(A3), Construction 3.1 provides a consistent resampling scheme.*

As a major part within the proof of the above theorem, we show that the proposed bootstrap procedure gives consistent estimates of the variance. This investigation provides fundamental insights into the influence of the procedure's hyperparameter β. In particular, we determine an optimal bias-variance trade-off.

Theorem 3.6. *Define the target variance*

$$\sigma_\infty^2 = \lim_{n\to\infty} \mathbb{V}\mathrm{ar}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i\right).$$

If $\mathbb{E}[X_i] = 0$ and assumptions (A1)–(A3) hold,

(a) $\mathbb{E}\left[\mathbb{V}\mathrm{ar}^*\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n V_i X_i\right)\right] - \sigma_\infty^2 = \mathcal{O}\big(n^{-\frac{\beta}{1+\beta}}\big),$

(b) $\mathbb{V}\mathrm{ar}\left[\mathbb{V}\mathrm{ar}^*\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n V_i X_i\right)\right] = \mathcal{O}\big(n^{\beta-1}\big),$

(c) $\mathbb{V}\mathrm{ar}^*\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n V_i X_i\right) \xrightarrow{P} \sigma_\infty^2,$

(d) the (asymptotically) optimal β minimizing

$$\mathbb{E}\left[\mathbb{V}\mathrm{ar}^*\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n V_i X_i\right) - \sigma_\infty^2\right]^2$$

is given by

$$\beta_{opt} = \sqrt{2} - 1.$$

In Appendix C, we prove this result for a larger class of bootstrap weights $V_i/\overline{V}_n$ (see Lemma C.3) under more involved assumptions. These assumptions are verified for our specific scheme from Construction 3.1 in Lemma C.4. Note also that we omitted the scaling average $\overline{V}_n$ in this result.

The optimal choice β_{opt} makes the mean-squared-error in the first display of (c) converge at rate $O(n^{\sqrt{2}-2}) \approx O(n^{-0.59})$. This is slightly slower than the rate $O(n^{-2/3})$ attained by the blockwise bootstrap (Bühlmann, 1993, Section 3.3). This is the statistical price we pay for the computational advantage of autoregressive bootstrap weights. In many applications, the latter can easily outweigh the small loss in statistical efficiency, see our experiments in Section 4.

3.3 Beyond the simple sample average

The preceding results were stated for simple sample averages $n^{-1}\sum_{i=1}^n X_i$ of real-valued random variables to simplify the exposition. The methodology is much more broadly applicable, however.

Transformed random variables. The most immediate generalization results from a simple relabeling. Suppose there is a sequence of random variables $(Z_i)_{i\in\mathbb{N}}$ and some function f. To quantify uncertainty in the statistic

$$T_n = \frac{1}{n}\sum_{i=1}^n f(Z_i),$$

we can use its bootstrapped version

$$T_n^* = \frac{1}{n}\sum_{i=1}^n \frac{V_i}{\overline{V}_n} f(Z_i).$$

All results of the Section 3.1 apply naturally upon defining $X_i := f(Z_i)$. This also reveals the following, more fundamental interpretation of the multiplier bootstrap: each observation is assigned a weight $V_i/\overline{V}_n$, and these weights are used when computing bootstrapped averages — irrespective of what exactly we're averaging.

Multidimensional vectors. Theorem 3.5 immediately extends to averages of random vectors through the Cramér-Wold device (e.g., Van der Vaart, 2000, p. 16).

Corollary 3.7. *Let $Z_1, Z_2, \ldots \in \mathbb{R}^d$ and assume that (A1)–(A3) hold for $X_i := \gamma^\top Z_i$ and every $\gamma \in \mathbb{R}^d$. Then Construction 3.1 provides a consistent resampling scheme.*

A similar generalization of Theorem 3.6 also holds for vectors, but is omitted for brevity.

Transformations of the sample average. The delta method for bootstrap (e.g. Theorem 23.5 in Van der Vaart (2000)) generalizes the consistency results of Theorem 3.5 to transformations of the sample average.

Corollary 3.8. *Let $\phi : \mathbb{R}^d \to \mathbb{R}^k$ be continuously differentiable and assume that (A1)–(A3) hold for $\gamma^T X_i$ with $X_i \in \mathbb{R}^d$ and all $\gamma \in \mathbb{R}^d$. Then,*

$$\sup_{x\in\mathbb{R}^d}\Big|\mathbb{P}^*\left\{\sqrt{n}(\phi(T_n^*) - \phi(T_n)) \le x\right\} - \mathbb{P}\left\{\sqrt{n}(\phi(T_n) - \phi(\mathbb{E}(T_n))) \le x\right\}\Big| \overset{n\to\infty}{\to} 0,$$

in probability with respect to $(X_i)_{i\in\mathbb{N}}$.

This enables a broader application of the proposed methodology, e.g. to the sample variance (Example 23.6. in Van der Vaart (2000)).

4 NUMERICAL VALIDATION

In this section, we assess the performance and applicability of the proposed bootstrap via simulations. In particular, we verify the theoretical results in a finite sample setting, illustrate the necessity of tailored bootstrap schemes for time series, and the computational benefits of our new method.

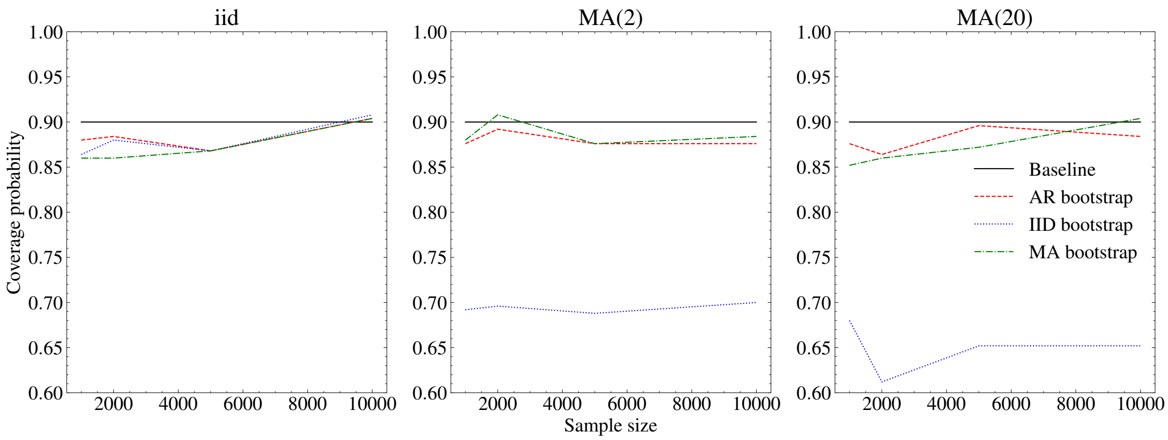

Figure 1: Estimated coverage probability of the bootstrap procedures. The target level of 90% is shown as solid line.

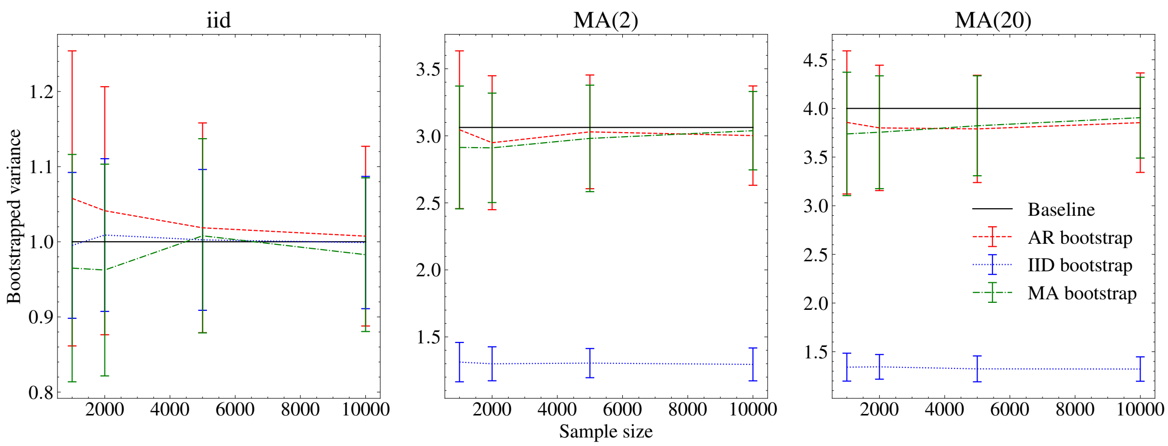

Figure 2: Average plus/minus standard deviation of the estimated variances. The target level σ_∞ is shown as solid line.

4.1 Experimental design

Data generating processes. We simulate X_i from a moving average process of order $q \in \mathbb{N}$, $MA(q)$ for short, i.e., a stochastic process of the form

$$X_i = \mu + \epsilon_i + \sum_{j=1}^{q} \theta_j \epsilon_{i-j},$$

with model parameters $\theta_1, ..., \theta_q \in \mathbb{R}$ and *iid* noise $\epsilon_i \sim \mathcal{N}(0,1)$. In addition, we simulate from a nonlinear transformation of such a process

$$Y_i = \exp(X_i)$$

and consider a nonlinear function of the sample averages

$$\ln\big[n^{-1} \sum_{i=1}^{n} \exp(X_i)\big]$$

for $q = 2$ (referred to as $LogMeanExp$). The bootstrapped distribution reflects the correct distribution according to Corollary 3.8. Further, we simulate from a nonlinear process $MA(2) - GARCH(1,1)$ reflecting the volatility clustering typical for financial time series. We realize the latter by

$$Z_i = \mu + Z_i + \theta_1 \gamma_{i-1} + \theta_2 \gamma_{i-2}$$

for $\gamma_i = \sigma_i \xi_i$, $\sigma_i^2 = \alpha_0 + \alpha \gamma_{i-1}^2 + \beta \sigma_{i-1}^2$ and $\xi_i \sim \mathcal{N}(0,1)$ iid.

All processes, except the latter, are q-dependent[1] and, hence, satisfy the assumptions of our theoretical results. For the latter we refer to Lindner (2009).

[1] $(X_i)_{i \in \mathbb{N}}$ is called q-dependent if $(\dots, X_{i-1}, X_i)$ is independent from $(X_{i+q+1}, X_{i+q+2}, \dots)$ for every i. Then,

$$C^X(|h|), \alpha(i) = 0$$

for all $i, |h| > q$, from which assumptions (A2) and (A3) follow immediately.

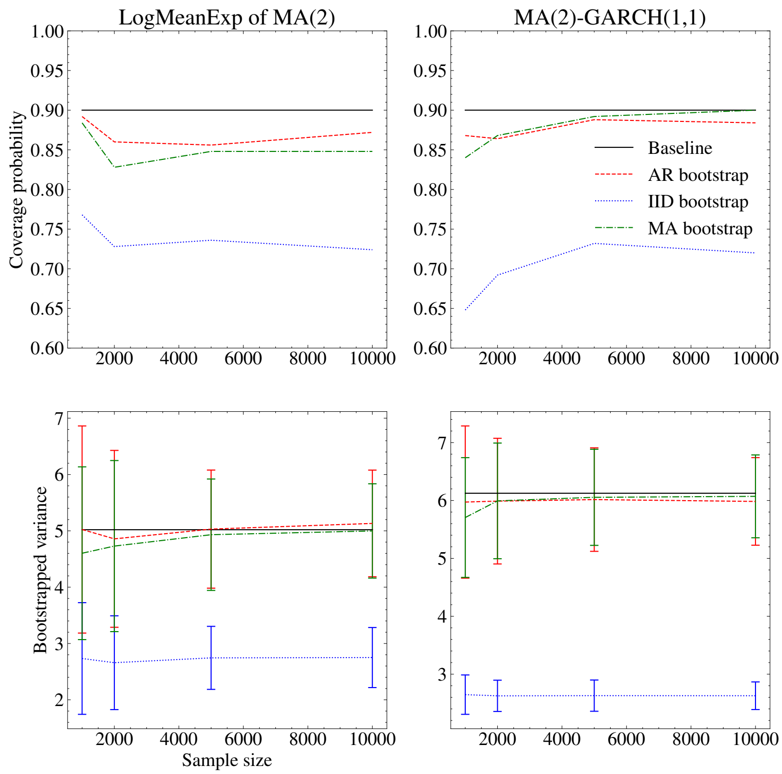

Figure 3: Estimated coverage probability of the bootstrap procedures with target level of 90% shown as solid line (top) and average plus/minus standard deviation of the estimated variances with target level σ_∞ shown as solid line (bottom).

We set $\theta_j = 2^{-j}$ and consider five scenarios:

- $MA(0)$ corresponds to the *iid* setting,
- $MA(2)$ corresponds to short-term serial dependence,
- $MA(20)$ corresponds to medium-term serial dependence.
- *LogMeanExp* corresponds to a transformation of a nonlinear process with short-term serial dependence.
- $MA(2) - GARCH(1, 1)$ corresponds to a nonlinear stochastic volatility process.

Bootstrap methods. We apply three bootstrap procedures of the form

$$X_i^* = \frac{V_i}{\overline{V}_n} X_i$$

with

(i) $V_i \sim \mathcal{N}(1, 1)$ corresponding to Example 2.2 (`IID bootstrap`);

(ii) V_i according to the moving average block bootstrap of Bühlmann (1993), see Example E.1 in the supplementary material, with $m_n = \lfloor n^{1/3} \rfloor$ (`MA bootstrap`);

(iii) V_i according to our new Construction 3.1 with parameter $\beta = \sqrt{2} - 1$ (`AR bootstrap`).

Evaluation. For each simulated time series and method, we generate 250 bootstrap samples and compute the sample variance and a 90%-confidence interval. We repeat this procedure $M = 250$ times. We assess the performance by a) mean and standard deviation of the estimated asymptotic variance σ_∞, b) coverage probability of the resulting confidence interval, c) computation time. See Appendix F for computational details of the evaluation metrics. The corresponding source code is provided at `https://github.com/nicolaipalm/online-bootstrap-implementation`.

4.2 Results

Validity. We start by checking the validity of confidence intervals constructed from the various bootstrap methods. Figure 1 and the top panels of Figure 3 plot

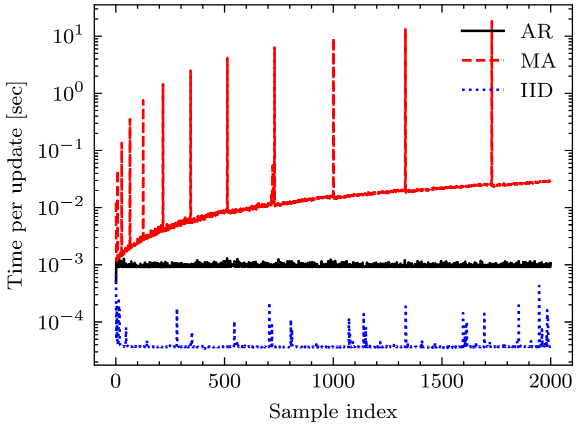

Figure 4: Computation time per online update of 200 bootstrap samples as the algorithms progress through a stream of samples.

the respective coverage probabilities against the sample size. The target level of 90% is indicated by the solid line. When the data is generated as an *iid* sequence (left panel of Figure 1), all three bootstrap variants have approximately correct coverage, especially for large samples. When there is dependence in the data (remaining panels), the `IID bootstrap` fails catastrophically, however. This is even the case for the linear MA(2) scenario, where dependence is weak and short-term (middle panel of Figure 1). The two dedicated time series bootstraps (`MA` and `AR`) achieve approximately correct coverage in all five scenarios, even in the presence of nonlinear dependencies and transformations of the sample average (top panels of Figure 3).

Accuracy. We now dive deeper into how well the bootstrap methods estimate the true variance σ_∞. Figure 2 and the bottom panels of Figure 3 show average estimates plus/minus their standard deviation and the target level. Unsurprisingly, the `IID` bootstrap works best when the data is actually an independent sequence (left panel of Figure 2). In particular, it has virtually no bias, and the smallest variance among all methods. In the time series settings it fails. On the other hand, the two time series bootstraps approach the target level, with bias and variance decreasing with the sample size in all five scenarios. The MA bootstrap appears to have a slightly smaller variance compared to our new AR method. This reflects the small statistical cost we pay for its computational advantage, see our comments after Theorem 3.6.

Computation time. The true benefit of the newly proposed scheme is the ability to compute it with cheap online updates. Figure 4 shows the computation time of an update step when the three bootstrap methods are used to generate 250 bootstrap samples in an online setting. We see that `AR` and `IID` require a small, constant amount of time for every update as the algorithms progress. The `IID` bootstrap is fastest, but invalid for time series data. The blockwise bootstrap `MA` allows for cheap online updates as long as the block size remains constant. It occurs a huge cost whenever the block size needs to be increased: one must regenerate all past and current bootstrap weights and recalculate the bootstrap averages with new weights. This shows as large spikes in Figure 4. Additionally, the time it requires quickly increases with time. The last block update at around just 1700 samples already takes 20 seconds, where the other two methods remain in the milliseconds. It is not reasonable to compute `MA` on much longer data streams. Our new `AR` bootstrap on the other hand remains fast and valid.

5 DISCUSSION

We close with a discussion of potential applications and current limitations of our method.

5.1 Applications in machine learning

Empirical risk minimizers. Consider a parametrized prediction model f_θ, a loss function L and the empirical risk minimizer

$$\widehat{\theta} = \arg\min_\theta \sum_{i=1}^n L(f_\theta, Z_i). \tag{1}$$

Under some regularity conditions, one usually has (see, e.g., Giordano et al., 2019)

$$\widehat{\theta} - \theta_0 \approx -H^{-1}\frac{1}{n}\sum_{i=1}^n \nabla_\theta L(f_{\theta_0}, Z_i), \tag{2}$$

where $\theta_0 = \arg\min_\theta \mathbb{E}[L(f_\theta, Z_i)]$ minimizes the true risk and $H = \mathbb{E}[\nabla_{\theta\theta} L(f_{\theta_0}, Z_i)]$ is the expected loss Hessian. In machine learning, the uncertainty of predictions $f_{\widehat{\theta}}(x)$ is more interesting than the parameter θ. A first-order Taylor approximation and (2) give

$$f_{\widehat{\theta}} - f_{\theta_0} \approx -\nabla_\theta f_{\theta_0} H^{-1}\frac{1}{n}\sum_{i=1}^n \nabla_\theta L(f_{\theta_0}, Z_i).$$

Now define the bootstrapped parameter

$$\widehat{\theta}^* = \arg\min_\theta \sum_{i=1}^n \frac{V_i}{\overline{V}_n} L(f_\theta, Z_i),$$

for which similar arguments yield

$$f_{\widehat{\theta}^*} - f_{\widehat{\theta}} \approx -\nabla_\theta f_{\theta_0}(x) H^{-1}\frac{1}{n}\sum_{i=1}^n \frac{V_i}{\overline{V}_n}\nabla_\theta L(f_{\theta_0}, Z_i).$$

Applying Corollary 3.7 to the average on the far right of the last display now shows that the distribution of $f_{\widehat{\theta}^*} - f_{\widehat{\theta}}$ appropriately reflects the uncertainty of $f_{\widehat{\theta}} - f_{\theta_0}$.

SGD and online convex optimization. In practice, optimization problems like (1) are often solved using algorithms from online convex optimization, e.g., stochastic gradient descent. In the *iid* setting, it is well established that approximation (2) also holds for averaged SGD updates (Ruppert, 1988; Polyak and Juditsky, 1992; Fang et al., 2018; Zhong et al., 2023). Similar results for SGD on time series data were established recently by, e.g., Godichon-Baggioni (2019); Godichon-Baggioni et al. (2023); Liu et al. (2023).

Bandit algorithms. Our new bootstrap scheme can also be incorporated into bandit algorithms, similar to Wan et al. (2023) in the *iid* case. In a simple multi-armed bandit, an agent picks some arm $A_t \in \{1, \ldots, K\}$ and receives reward R_{t,A_t} in return, at every time t. If $S_a = \{t\colon A_t = a\}$ is the set of times action a was played, the expected reward of arm k can be estimated as a simple sample average

$$\widehat{r}_a = \frac{1}{|S_a|} \sum_{t \in S_a} R_{t,a}.$$

Wan et al. (2023) proposed to use an independent multiplier bootstrap to quantify the uncertainty about $\widehat{r}_a$, and use this to guide the exploration/exploitation-trade-off of the algorithm. With our new method, the assumption that rewards $R_{t,a}$ are independent can be relaxed.

5.2 Limitations

Stationarity assumption. As mentioned in Section 3.1, stationarity of the series $(X_i)_{i\in\mathbb{N}}$ is a common assumption in the time series literature. In applications, stationarity is often ensured by pre-processing steps. Technically, these steps should also be accounted for in uncertainty quantification, but this is difficult to do with generality. Another way to alleviate this issue would be to extend our results to averages of nonstationary series. Some recent developments in this field (Merlevède and Peligrad, 2020) can likely be adapted in future work.

Negative weights. In Construction 3.1, we explicitly defined the weights V_i to follow a Gaussian AR-process. This choice is motivated by mathematical convenience: by construction, the distribution of the bootstrapped average (conditional on the data) is normal — no central limit theorem is required. A downside is that the bootstrap weights V_i can be negative. When the X_i's are positive variables (counts, lengths, prices, etc.) and the sample size is small, this may be problematic. The bootstrapped average $n^{-1}\sum_{i=1}^n \frac{V_i}{\overline{V}_n} X_i$ could become negative, which results in meaningless estimates. While this is not an issue asymptotically, it might be preferable to work with a strictly positive sequence $(V_i)_{i=1}^n$. To prove the validity of such schemes would then require new central limit theorems for time series with increasing dependence, which is an interesting problem for future work.

References

Bosq, D. (2012). *Nonparametric statistics for stochastic processes: estimation and prediction*, volume 110. Springer Science & Business Media.

Bühlmann, P. L. (1993). *The blockwise bootstrap in time series and empirical processes.* PhD thesis, ETH Zurich.

Burke, M. D. (1998). A gaussian bootstrap approach to estimation and tests. In Szyszkowicz, B., editor, *Asymptotic Methods in Probability and Statistics*, pages 697–706. North-Holland, Amsterdam.

Cesa-Bianchi, N. and Lugosi, G. (2006). *Prediction, learning, and games.* Cambridge university press.

Dimakopoulou, M., Ren, Z., and Zhou, Z. (2021). Online multi-armed bandits with adaptive inference. *Advances in Neural Information Processing Systems*, 34:1939–1951.

Efron, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 7(1):1 – 26.

Fang, Y., Xu, J., and Yang, L. (2018). Online bootstrap confidence intervals for the stochastic gradient descent estimator. *Journal of Machine Learning Research*, 19:1–21.

Giordano, R., Stephenson, W., Liu, R., Jordan, M., and Broderick, T. (2019). A swiss army infinitesimal jackknife. *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, 89:1139–1147.

Godichon-Baggioni, A. (2019). Online estimation of the asymptotic variance for averaged stochastic gradient algorithms. *Journal of Statistical Planning and Inference*, 203:1–19.

Godichon-Baggioni, A., Werge, N., and Wintenberger, O. (2023). Learning from time-dependent streaming data with online stochastic algorithms. *Transactions on Machine Learning Research.*

Hall, P., Horowitz, J. L., and Jing, B.-Y. (1995). On blocking rules for the bootstrap with dependent data. *Biometrika*, 82(3):561–574.

Hamilton, J. D. (2020). *Time series analysis*. Princeton university press.

Kotz, S. and Johnson, N. L. (1992). Breakthroughs in statistics: Volumes i and ii.

Künsch, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *The Annals of Statistics*, pages 1217–1241.

Lindner, A. M. (2009). Stationarity, mixing, distributional properties and moments of garch (p, q)–processes. In *Handbook of financial time series*, pages 43–69. Springer.

Liu, R., Chen, X., and Shang, Z. (2023). Statistical inference with stochastic gradient methods under ϕ-mixing data.

Merlevède, F. and Peligrad, M. (2020). Functional clt for nonstationary strongly mixing processes. *Statistics & Probability Letters*, 156:108581.

Politis, D. N. and Romano, J. P. (1993). Nonparametric resampling for homogeneous strong mixing random fields. *Journal of Multivariate Analysis*, 47(2):301–328.

Polyak, B. T. and Juditsky, A. B. (1992). Acceleration of stochastic approximation by averaging. *SIAM journal on control and optimization*, 30(4):838–855.

Ramprasad, P., Li, Y., Yang, Z., Wang, Z., Sun, W. W., and Cheng, G. (2022). Online bootstrap inference for policy evaluation in reinforcement learning. *Journal of the American Statistical Association*, pages 1–14.

Rubin, D. B. (1981). The Bayesian Bootstrap. *The Annals of Statistics*, 9(1):130 – 134.

Ruppert, D. (1988). Efficient estimations from a slowly convergent robbins-monro process. Technical report, Cornell University Operations Research and Industrial Engineering.

Shalev-Shwartz, S. et al. (2012). Online learning and online convex optimization. *Foundations and Trends® in Machine Learning*, 4(2):107–194.

Van der Vaart, A. W. (2000). *Asymptotic statistics*, volume 3. Cambridge university press.

Van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer New York, New York, NY.

Wan, R., Wei, H., Kveton, B., and Song, R. (2023). Multiplier bootstrap-based exploration. *Proceedings of the 40th International Conference on Machine Learning*, 202:35444–35490.

Zhong, Y., Kuffner, T., and Lahiri, S. (2023). Online bootstrap inference with nonconvex stochastic gradient descent estimator.

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. Yes - see Section 3, i.p. Algorithm 1
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. Yes - see end of Section 3.1 and 4.2/Computation time
   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. Yes - `https://github.com/nicolaipalm/online-bootstrap-implementation`

2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. Yes - see (A1)–(A3)
   (b) Complete proofs of all theoretical results. Yes - see supplementary material
   (c) Clear explanations of any assumptions. Yes - see Section 3.2 Theory

3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). Yes - see Section 4.1, Appendix F and `https://github.com/nicolaipalm/online-bootstrap-implementation`.
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). Not Applicable
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). Yes - see Section 4.1 Evaluation
   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). Not Applicable

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
   (a) Citations of the creator if your work uses existing assets. Not Applicable
   (b) The license information of the assets, if applicable. Not Applicable

(c) New assets either in the supplemental material or as a URL, if applicable. Yes - see `https://github.com/nicolaipalm/online-bootstrap-implementation`.

(d) Information about consent from data providers/curators. Not Applicable

(e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. Not Applicable

5. If you used crowdsourcing or conducted research with human subjects, check if you include:

   (a) The full text of instructions given to participants and screenshots. Not Applicable

   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. Not Applicable

   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. Not Applicable

Supplementary Materials

A Proving consistency of bootstrapping schemes

Proving a bootstrapping scheme to be consistent mostly proceeds in the following two steps

S1 Prove that the random variables X_i satisfy some central limit theorem and

S2 prove that the bootstrapped random variables X_i^* satisfy some central limit theorem with the same limit distribution.

In practice, a lot of central limit theorems are already well established each covering different assumptions on the random variables X_i, i.e. S1 is given. For the sake of clarity, we assume $\mathbb{E}(X_i) = 0$. Lemma 3.4 treats the non-centered case.

Observe that

$$\mathbb{V}\mathrm{ar}^V \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} V_i X_i) \right)$$

is a random variable induced by X_i where we write $\mathbb{V}\mathrm{ar}^V = \mathbb{V}\mathrm{ar}^*$ with respect to the explicit construct $X^* = VX$. S2 implies

$$\mathbb{V}\mathrm{ar}^V \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} V_i X_i \right) \overset{n\to\infty}{\to} \sigma_\infty^2$$

in probability on the sequence $(X_i)_{i\in\mathbb{N}}$ where σ_∞^2 denotes the asymptotic (finite) variance

$$\lim_{n\to\infty} \mathbb{V}\mathrm{ar} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \right) = \sigma_\infty^2.$$

Classically, such results are derived by proving

$$\mathbb{E}^X \left[\mathbb{V}\mathrm{ar}^V \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} V_i X_i \right) \right] \overset{n\to\infty}{\to} \sigma_\infty^2 \qquad \text{and} \qquad \mathbb{V}\mathrm{ar}^X \left[\mathbb{V}\mathrm{ar}^V \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} V_i X_i \right) \right] \overset{n\to\infty}{\to} 0,$$

and then using Chebyshev's inequality. Without loss of gernality assuming $\mathbb{E}(X_i) = 0$ (see Lemma 3.4), a straightforward calculation exhibits

$$\mathbb{V}\mathrm{ar} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \right) = \frac{1}{n} \sum_{i,j=1}^{n} \mathbb{C}\mathrm{ov}(X_i, X_j),$$

$$\mathbb{E}^X \left[\mathbb{V}\mathrm{ar}^V \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} V_i X_i \right) \right] = \frac{1}{n} \sum_{i,j=1}^{n} \mathbb{C}\mathrm{ov}(V_i, V_j)\mathbb{C}\mathrm{ov}(X_i, X_j).$$

This suggests that we need $\mathbb{C}\mathrm{ov}(V_i, V_j) \approx 1$ for i, j great enough and $\mathbb{C}\mathrm{ov}(X_i, X_j) \not\approx 0$. However, calculating

$$\mathbb{V}\mathrm{ar}^X \left[\mathbb{V}\mathrm{ar}^V \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} V_i X_i \right) \right] = \frac{1}{n^2} \sum_{i_1,\dots,i_4=1}^{n} \mathbb{C}\mathrm{ov}(V_{i_1}, V_{i_2})\mathbb{C}\mathrm{ov}(V_{i_3}, V_{i_4})\mathbb{C}\mathrm{ov}(X_{i_1}X_{i_2}, X_{i_3}X_{i_4})$$

requires $\mathbb{C}\mathrm{ov}^V(V_i, V_j) \approx 0$ for sufficiently many i, j to vanish. In the next section we provide corresponding formal results about the asymptotic variance. These steps will be worked out in detail in the following sections.

B Proof of Lemma 3.4

It holds

$$\frac{1}{n}\sum_{i=1}^{n}\frac{V_i}{\overline{V}_n}X_i - \frac{1}{n}\sum_{i=1}^{n}X_i = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{V_i}{\overline{V}_n}-1\right)X_i = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{V_i}{\overline{V}_n}-1\right)(X_i-\mathbb{E}[X_i]) + \mathbb{E}[X_1]\underbrace{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{V_i}{\overline{V}_n}-1\right)}_{=0}.$$

Now

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\left(\frac{V_i}{\overline{V}_n}-1\right)(X_i-\mathbb{E}[X_i]) &= \frac{1}{n}\sum_{i=1}^{n}(V_i-1)(X_i-\mathbb{E}[X_i]) + \frac{1}{n}\sum_{i=1}^{n}\left(\frac{V_i}{\overline{V}_n}-V_i\right)(X_i-\mathbb{E}[X_i]) \\ &= \frac{1}{n}\sum_{i=1}^{n}(V_i-1)(X_i-\mathbb{E}[X_i]) + \left(\frac{1}{\overline{V}_n}-1\right)\frac{1}{n}\sum_{i=1}^{n}V_i(X_i-\mathbb{E}[X_i]).\end{aligned}$$

We observe

$$\mathbb{V}\mathrm{ar}(\overline{V}_n) = \frac{1}{n^2}\sum_{i,j=1}^{n}\mathbb{C}\mathrm{ov}(V_i,V_j) \le \frac{1}{n^2}\sum_{i=1}^{n}\sum_{h\in\mathbb{Z}}v(i,i+|h|) \overset{n\to\infty}{\to} 0,$$

by (iv) of Lemma C.2. Chebychev's inequality then yields $\overline{V}_n \overset{P}{\to} \mathbb{E}(V_i)=1$ and, consequently, $1/\overline{V}_n - 1 \overset{P}{\to} 0$. Now note that

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n}V_i(X_i-\mathbb{E}[X_i])\right] = \mathbb{E}^V\left[\frac{1}{n}\sum_{i=1}^{n}V_i\mathbb{E}^X[(X_i-\mathbb{E}[X_i])]\right] = 0.$$

The law of total variance gives

$$\begin{aligned}\mathbb{V}\mathrm{ar}\left[\frac{1}{n}\sum_{i=1}^{n}V_i(X_i-\mathbb{E}[X_i])\right] &= \mathbb{E}^V\left[\mathbb{V}\mathrm{ar}^X\left[\frac{1}{n}\sum_{i=1}^{n}V_i(X_i-\mathbb{E}[X_i])\right]\right] + \mathbb{V}\mathrm{ar}^V\Bigg[\underbrace{\mathbb{E}^X\left[\frac{1}{n}\sum_{i=1}^{n}V_i(X_i-\mathbb{E}[X_i])\right]}_{=0}\Bigg] \\ &= \mathbb{E}^V\left[\mathbb{V}\mathrm{ar}^X\left[\frac{1}{n}\sum_{i=1}^{n}V_i(X_i-\mathbb{E}[X_i])\right]\right] \\ &= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}[V_iV_j]\mathbb{C}\mathrm{ov}(X_i,X_j) \\ &\le \frac{2}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}|\mathbb{C}\mathrm{ov}(X_i,X_j)| && \text{by Cauchy-Schwarz and } \mathbb{E}[V_i^2]=2 \\ &\le \frac{1}{n}\sum_{h\in\mathbb{Z}}|C^X(h)| && \text{using that } (X_i)_{i\in\mathbb{N}} \text{ is stationary} \\ &= \mathcal{O}\left(\frac{1}{n}\right). && \text{by assumption (A3)}\end{aligned}$$

Applying Chebyshev's inequality again, we have shown

$$\left(\frac{1}{\overline{V}_n}-1\right)\frac{1}{n}\sum_{i=1}^{n}V_i(X_i-\mathbb{E}[X_i]) = o_P(1)\times\mathcal{O}_P(n^{-1/2}),$$

and the claim follows. □

C Asymptotic variance

In the following, we consider $(X_i)_{i\in\mathbb{N}}$ to be a real-valued strictly stationary stochastic process and abbreviate $v(i,j) = \mathbb{C}\mathrm{ov}(V_i, V_j)$.

We frequently sum over two indices, i.e. we consider a sum of the form $\sum_{i,j=1}^n a_{ij}$. The reader convinces himself that

$$\{(i,j)|i,j = 1,\dots,n\}$$

is the disjoint union of

$$\{(i,i+j)|j = 0,\dots,n, i = 1,\dots,n-j\} \text{ and } \{(i+j,i)|j = 1,\dots,n, i = 1,\dots,n-j\}.$$

If we assume $a_{ij} = a_{ji}$, then, we identify

$$\sum_{i,j=1}^n a_{ij} = \sum_{j=-n}^n \sum_{i=1}^{n-|j|} a_{i,i+|j|}. \tag{3}$$

Lemma C.1. *Assume the following:*

(i) $\mathbb{E}(X_i) = 0$ *for all* i.

(ii) $|v(i,j)| \le C$ *for some* $C \in \mathbb{R}$ *and all* i, j.

(iii) $\sum_{h=-\infty}^{\infty} |C^X(|h|)| < \infty$.

(iv) for all $\epsilon > 0$ *and* h *such that* $C^X(|h|) \neq 0$ *there exists* $n_{\epsilon,|h|}$ *such that*

$$|1 - v(i, i+|h|)| \le \epsilon$$

for all $i \ge n_{\epsilon,|h|}$ *and*

(v) $\lim_{n\to\infty} \frac{1}{n} \sum_{h=-n}^n n_{\epsilon,|h|} |C^X(|h|)| = 0$

Then, the asymptotic variance

$$\mathbb{V}\mathrm{ar}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i\right) \overset{n\to\infty}{\to} \sigma_\infty^2 = \frac{1}{n}\sum_{i,j=1}^{\infty} \mathbb{C}\mathrm{ov}(X_i, X_j)$$

exists,

$$\mathbb{E}^X\left[\mathbb{V}\mathrm{ar}^V\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n V_i X_i\right) - \sigma_\infty^2\right] = \mathcal{O}\left(\frac{1}{n}\sum_{h=-n}^n n_{\epsilon,|h|}|C^X(|h|)| + \epsilon \sum_{h=-n}^n |C^X(|h|)|\right)$$

and, in particular,

$$\mathbb{E}^X\left[\mathbb{V}\mathrm{ar}^V\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n V_i X_i\right)\right] \overset{n\to\infty}{\to} \sigma_\infty^2.$$

Proof. We observe

$$\mathbb{V}\mathrm{ar}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i\right) = \sum_{i,j=1}^n \mathbb{C}\mathrm{ov}(X_i, X_j).$$

Therefore, its asymptotic variance (if it exists) is given by

$$\mathbb{V}\mathrm{ar}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i\right) \overset{n\to\infty}{\to} \sigma_\infty^2 = \frac{1}{n}\sum_{i,j=1}^{\infty} \mathbb{C}\mathrm{ov}(X_i, X_j).$$

Furthermore, we calculate

$$\begin{aligned}
\mathbb{V}\mathrm{ar}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i\right) &= \frac{1}{n}\sum_{i,j=1}^{n} \mathbb{C}\mathrm{ov}(X_i, X_j) \\
&= \frac{1}{n}\sum_{h=-n}^{n}\sum_{i=1}^{n-|h|} \mathbb{C}\mathrm{ov}(X_i, X_{i+|h|}) && \text{by (3)} \\
&= \frac{1}{n}\sum_{h=-n}^{n}\sum_{i=1}^{n-|h|} C^X(|h|) && \text{by stationarity of } (X_i)_{i\in\mathbb{N}} \\
&= \sum_{h=-n}^{n} \frac{n-|h|}{n} C^X(|h|). \\
&\leq \sum_{h=-\infty}^{\infty} |C^X(|h|)|.
\end{aligned}$$

Therefore, the asymptotic variance exists by (iii).

We calculate

$$\begin{aligned}
\mathbb{E}^X\left[\mathbb{V}\mathrm{ar}^V\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} V_i X_i\right)\right] &= \mathbb{E}^X\left(\frac{1}{n}\sum_{i,j=1}^{n} v(i,j) X_i X_j\right) \\
&= \frac{1}{n}\sum_{i,j=1}^{n} v(i,j)\mathbb{E}(X_i X_j) \\
&= \frac{1}{n}\sum_{i,j=1}^{n} v(i,j)\mathbb{C}\mathrm{ov}(X_i, X_j). && \text{by assumption (i)}
\end{aligned}$$

Hence,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i,j=1}^{n} (v(i,j)-1)\mathbb{C}\mathrm{ov}(X_i, X_j) = \lim_{n\to\infty} \mathbb{E}^X\left[\mathbb{V}\mathrm{ar}^V\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} V_i X_i\right)\right] - \sigma_\infty^2.$$

By (3) and stationarity of $(X_i)_{i\in\mathbb{N}}$ we obtain

$$\frac{1}{n}\sum_{h=-n}^{n}\sum_{i=1}^{n-|h|} (v(i,i+|h|)-1)C^X(|h|) = \frac{1}{n}\sum_{i,j=1}^{n} (v(i,j)-1)\mathbb{C}\mathrm{ov}(X_i, X_j).$$

Therefore, we consider

$$\frac{1}{n}\sum_{h=-n}^{n}\sum_{i=1}^{n-|h|} (v(i,i+|h|)-1)C^X(|h|)$$

and prove that it converges to zero for $n \to \infty$.

We split the sum

$$\sum_{i=1}^{n-|h|} (v(i,i+|h|)-1)C^X(|h|) = \sum_{i=1}^{n_{\epsilon,|h|}} (v(i,i+|h|)-1)C^X(|h|) + \sum_{i=n_{\epsilon,|h|}+1}^{n-|h|} (v(i,i+|h|)-1)C^X(|h|)$$

and prove that the absolute values of both summands converge to zero. For fixed h we calculate

$$\begin{aligned}
\left|\sum_{i=n_{\epsilon,|h|}+1}^{n-|h|} (v(i,i+|h|)-1)C^X(|h|)\right| &\leq \sum_{i=n_{\epsilon,|h|}+1}^{n-|h|} \epsilon|C^X(|h|)| && \text{by assumption (iv).} \\
&= \epsilon(n-|h|-n_{\epsilon,|h|}-1)|C^X(|h|)|
\end{aligned}$$

for n great enough. Thus,

$$\begin{aligned}\left|\frac{1}{n}\sum_{h=-n}^{n}\sum_{i=n_{\epsilon,|h|}+1}^{n-|h|}(v(i,i+|h|)-1)C^X(|h|)\right| &\leq \frac{\epsilon}{n}\sum_{h=-n}^{n}(n-|h|-n_{\epsilon,|h|-1}-1)|C^X(|h|)| \\ &\leq \frac{\epsilon}{n}\sum_{h=-n}^{n} n|C^X(|h|)| \\ &= \epsilon\sum_{h=-n}^{n}|C^X(|h|)|. \end{aligned} \tag{4}$$

We further observe

$$\epsilon\sum_{h=-n}^{n}|C^X(|h|)| \overset{n\to\infty}{\to} \epsilon\sum_{h\in\mathbb{Z}}|C^X(|h|)| < \infty$$

by assumption (iii). Since ϵ might be chosen arbitrarily small for great enough n, we obtain

$$\left|\frac{1}{n}\sum_{h=-n}^{n}\sum_{i=n_{\epsilon,|h|}+1}^{n-|h|}(v(i,i+|h|)-1)C^X(|h|)\right| \overset{n\to\infty}{\to} 0.$$

At last, using assumption (ii) and (iv) we calculate

$$\begin{aligned}\left|\frac{1}{n}\sum_{h=-n}^{n}\sum_{i=1}^{n_{\epsilon,|h|}}(v(i,i+|h|)-1)C^X(|h|)\right| &\leq \frac{1}{n}\sum_{h=-n}^{n}\sum_{i=1}^{n_{\epsilon,|h|}}|(v(i,i+|h|)-1)||C^X(|h|)| \\ &\leq \frac{(C+1)}{n}\sum_{h=-n}^{n}\sum_{i=1}^{n_{\epsilon,|h|}}|C^X(|h|)| \\ &= \frac{(C+1)}{n}\sum_{h=-n}^{n} n_{\epsilon,|h|}|C^X(|h|)| \qquad (5) \\ &\overset{n\to\infty}{\to} 0 \end{aligned}$$

□

Lemma C.2. *Assume the following:*

(i) $\mathbb{E}(X_i) = 0$ *for all* i.

(ii) $|v(i,j)| \leq C$ *for some* $C \in \mathbb{R}$ *and all* i, j.

(iii) for all $i \in \mathbb{N}$ *the sum* $\sum_{h\in\mathbb{Z}} v(i, i+|h|) = C_i \in \mathbb{R}$ *is finite and*

(iv) $\lim_{n\to\infty} \frac{1}{n^2}\sum_{i=1}^{n} C_i = 0$

(v) $\mathbb{E}(X_i^8) < \infty$ *and* $\sum_{i=1}^{\infty} \alpha(i)^{\frac{1}{2}} < \infty$

where α *is the dependence coefficient of the* α*-mixing sequence* $(X_i)_{i\in\mathbb{N}}$.

Then,

$$\mathbb{V}\mathrm{ar}^X\left[\mathbb{V}\mathrm{ar}^V\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} V_i X_i\right)\right] \leq \mathcal{O}\left(\frac{1}{n^2}\sum_{i=1}^{n} C_i\right) \overset{n\to\infty}{\to} 0.$$

Proof. By Lemma G.1, (v) implies

$$\sum_{h=-\infty}^{\infty} C^X(|h|) = K \in \mathbb{R}. \tag{6}$$

Using (i), we calculate

$$\mathbb{V}\mathrm{ar}^X\left[\mathbb{V}\mathrm{ar}^V\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n V_iX_i\right)\right] = \frac{1}{n^2}\sum_{i_1,\ldots,i_4=1}^n v(i_1,i_2)v(i_3,i_4)\mathbb{C}\mathrm{ov}(X_{i_1}X_{i_2}, X_{i_3}X_{i_4}). \tag{7}$$

We consider the following term

$$\frac{1}{n^2}\sum_{i_1,\ldots,i_4=1}^n v(i_1,i_2)v(i_3,i_4)(C^X(|i_1-i_3|)C^X(|i_2-i_4|) + C^X(|i_1-i_4|)C^X(|i_2-i_3|)).$$

Substituting $i_2 = i_1+s_1, i_3 = i_1+s_2, i_4 = i_1+s_3$ and using Lemma G.2, (7) is asymptotically equivalent to

$$\begin{aligned}\frac{1}{n^2}\sum_{i_1=1}^n\sum_{s_1,s_2,s_3\in\mathbb{Z}} &(v(i_1,i_1+|s_1|)v(i_1+|s_2|,i_1+|s_3|)\\ &\cdot (C^X(|s_2|)C^X(|s_1-s_3|) + C^X(|s_3|)C^X(|s_1-s_2|))).\end{aligned} \tag{8}$$

We observe further $\sum_{h\in\mathbb{Z}} C^X(|h-k|) = \sum_{h\in\mathbb{Z}} C^X(|h|)$ for all $k\in\mathbb{Z}$ and, hence,

$$\begin{aligned}\sum_{s_1,s_3\in\mathbb{Z}} v(i_1,i_1+|s_1|)C^X(|s_1-s_3|) &= \sum_{s_1\in\mathbb{Z}} v(i_1,i_1+|s_1|)\sum_{s_3\in\mathbb{Z}} C^X(|s_1-s_3|)\\ &= \sum_{s_1\in\mathbb{Z}} v(i_1,i_1+|s_1|)\sum_{s_3\in\mathbb{Z}} C^X(|s_3|)\\ &= C_{i_1}K,\end{aligned}$$

by (iii) and (vi) and similarly

$$\sum_{s_1,s_2\in\mathbb{Z}} v(i_1,i_1+|s_1|)C^X(|s_1-s_2|) = C_{i_1}K.$$

Combined with (ii), we obtain

$$(8) \leq \frac{1}{n^2}\sum_{i_1=1}^n 2C_{i_1}CK^2 \overset{n\to\infty}{\to} 0 \qquad \text{by (iv).} \tag{9}$$

This establishes the claim. □

Lemma C.3. *Suppose the following conditions hold:*

(i) $\mathbb{E}(X_i) = 0$ *and* $\mathbb{E}(X_i^8) < \infty$ *for all* i *and* $\sum_{h=-\infty}^{\infty} |C^X(|h|)| < \infty$.

(ii) $\sum_{i=1}^{\infty}\alpha(i)^{\frac{1}{2}} < \infty$.

(iii) $|v(i,j)| \leq C$ *for some* $C\in\mathbb{R}$ *and all* i,j.

(iv) for all $\epsilon > 0$ *and* h *such that* $C^X(|h|) \neq 0$ *there exists* $n_{\epsilon,|h|}$ *such that*

$$|1 - v(i, i+|h|)| \leq \epsilon$$

for all $i \geq n_{\epsilon,|h|}$ *and* $\lim_{n\to\infty}\frac{1}{n}\sum_{h=-n}^n n_{\epsilon,|h|}|C^X(|h|)| = 0$

(v) for all $i\in\mathbb{N}$ *the sum* $\sum_{h\in\mathbb{Z}} v(i,i+|h|) = C_i \in \mathbb{R}$ *is finite and* $\lim_{n\to\infty}\frac{1}{n^2}\sum_{i=1}^n C_i = 0$.

Then

$$\mathbb{V}\mathrm{ar}^V\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n V_i(X_i - \mathbb{E}(X_1))\right) \overset{P}{\to} \sigma_\infty^2.$$

Proof. Follows immediately from Lemma C.1, Lemma C.2, and Chebyshev's inequality. ∎

Lemma C.4. *Let V_i be as in Construction 3.1. Then,*

1. $V_i \sim \mathcal{N}(1,1)$ *for all* i
2. *for all* $h \in \mathbb{Z}$ *and* $i \in \mathbb{N}$ *it holds*
$$0 \leq \mathbb{C}\mathrm{ov}(V_i, V_{i+|h|}) = \prod_{k=1}^{|h|}(1-(i+k)^{-\beta}) \leq 1$$
3. *for all* $\epsilon < 1$, $h \in \mathbb{Z}$ *and* $n_{\epsilon,|h|} = (1 - \sqrt[|h|]{1-\epsilon})^{-\frac{1}{\beta}}$ *it holds*
$$|1 - \mathbb{C}\mathrm{ov}(V_i, V_{i+|h|})| \leq \epsilon$$
for all $i \geq n_{\epsilon,|h|}$
4. $C_i = \sum_{h\in\mathbb{Z}} \mathbb{C}\mathrm{ov}(V_i, V_{i+|h|}) \leq 4i^\beta + B$ *for some constant* $B \in \mathbb{R}$ *independent of* i *and, in particular,*
$$\lim_{n\to\infty} \frac{1}{n^2}\sum_{i=1}^{n} C_i = 0.$$

Proof. Observe all V_i to be normally distributed as sums of normally distributed random variables. Inductively, we obtain $\mathbb{V}\mathrm{ar}(V_i) = 1 = \mathbb{E}(V_i)$. This proves 1.

We calculate

$$\begin{aligned}
0 &\leq \mathbb{C}\mathrm{ov}(V_i, V_{i+|h|}) \\
&= \rho_{i+|h|}\mathbb{C}\mathrm{ov}(V_i, V_{i+|h|-1}) \\
&= \prod_{k=1}^{|h|}\rho_{i+k}\mathbb{V}\mathrm{ar}(V_i) && \text{inductively} \\
&= \prod_{k=1}^{|h|}\rho_{i+k} \\
&= \prod_{k=1}^{|h|}(1-(i+k)^{-\beta}) \\
&\leq (1-(i+|h|)^{-\beta})^{|h|}
\end{aligned}$$

for $h \in \mathbb{Z}$ and, in particular,

$$|\mathbb{C}\mathrm{ov}(V_i, V_{i+h})| \leq 1$$

for all i, h which proves 2.

Furthermore,

$$|1 - \mathbb{C}\mathrm{ov}(V_i, V_{i+|h|})| = 1 - \prod_{k=1}^{|h|}(1-(i+k)^{-\beta}) \leq 1 - (1-i^{-\beta})^{|h|}.$$

For any $\epsilon \in (0,1)$, we calculate

$$(1-(1-i^{-\beta})^{|h|}) \leq \epsilon \qquad \Leftrightarrow \qquad 1 - i^{-\beta} \geq \sqrt[|h|]{1-\epsilon} \qquad \Leftrightarrow \qquad i \geq (1-\sqrt[|h|]{1-\epsilon})^{-\frac{1}{\beta}}.$$

Hence, we obtain 3.

Furthermore,

$$\begin{aligned}
\sum_{h\in\mathbb{Z}} \mathbb{C}\mathrm{ov}(V_i, V_{i+|h|}) &\leq 2\sum_{h=0}^{\infty}(1-(i+h)^{-\beta})^h \\
&= 2\left[\sum_{h=0}^{i}(1-(i+h)^{-\beta})^h + \sum_{h=i+1}^{\infty}(1-(i+h)^{-\beta})^h\right] \\
&\leq 2\left[\sum_{h=0}^{\infty}(1-(2i)^{-\beta})^h + \sum_{h=0}^{\infty}(1-(2h)^{-\beta})^h\right] \\
&= 2\left[(2i)^{\beta} + \sum_{h=0}^{\infty}(1-(2h)^{-\beta})^h\right] \qquad \text{by the gemoetric series} \\
&= C_i < \infty,
\end{aligned}$$

where the last step follows from

$$0 \leq \sum_{h=0}^{\infty}(1-(2h)^{-\beta})^h < \infty$$

for all $\beta \in (0, 1)$. Thus, we calculate

$$0 \leq \lim_{n\to\infty} \frac{1}{n^2}\sum_{i=1}^{n} C_i = \lim_{n\to\infty}\frac{4}{n^2}\sum_{i=1}^{n} i^{\beta} \leq \lim_{n\to\infty}\frac{4}{n}n^{\beta} = 0 \tag{10}$$

and, hence, we obtain 4. □

Proof of Theorem 3.6. To prove the parts (a)–(c) of the theorem it suffices to check the conditions of Lemma C.3, which also imply the conditions of Lemmas C.1 to C.3. Conditions (i) and (ii) are given by assumption of the theorem. Conditions (iii)–(v) were proven in Lemma C.4.

For (iv) we calculated $n_{\epsilon,|h|} = (1 - \sqrt[|h|]{1-\epsilon})^{-\frac{1}{\beta}}$ in Lemma C.4. We approximate

$$f(\epsilon) = 1 - \sqrt[|h|]{1-\epsilon} \approx \frac{\epsilon}{|h|}$$

through a first order Taylor series around $\epsilon = 0$. Thus, we calculate

$$n_{\epsilon,|h|} = (1 - \sqrt[|h|]{1-\epsilon})^{-\frac{1}{\beta}} \approx \left(\frac{|h|}{\epsilon}\right)^{\frac{1}{\beta}} = \mathcal{O}(|h|^{\frac{1}{\beta}}).$$

In order to determine an optimal β we minimize the MSE

$$\begin{aligned}
&\mathbb{E}^X\left[\left(\mathbb{V}\mathrm{ar}^V\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} V_i X_i\right) - \sigma_\infty^2\right)^2\right] \\
&= \mathbb{E}^X\left[\mathbb{V}\mathrm{ar}^V\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} V_i X_i\right) - \sigma_\infty^2\right]^2 + \mathbb{V}\mathrm{ar}^X\left[\mathbb{V}\mathrm{ar}^V\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} V_i X_i\right)\right]
\end{aligned}$$

for $n \to \infty$. From (9) and (10), we get

$$\mathbb{V}\mathrm{ar}^X\left[\mathbb{V}\mathrm{ar}^V\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} V_i X_i\right)\right] = \mathcal{O}\left(\frac{1}{n^2}\sum_{i=1}^{n} C_i\right) = \mathcal{O}(n^{\beta-1}).$$

From (4) and (5), we get

$$\mathbb{E}^X\left[\mathbb{V}\mathrm{ar}^V\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} V_i X_i\right) - \sigma_\infty^2\right] = \mathcal{O}\left(\frac{1}{n}\sum_{h=-n}^{n} n_{\epsilon,|h|}|C^X(|h|)| + \epsilon\sum_{h=-n}^{n}|C^X(|h|)|\right).$$

We determine a sequence of $\epsilon_n \to 0$ in order to balance both.

Assuming

$$A = \sum_{h \in \mathbb{Z}} |C^X(|h|)|, B = \sum_{h \in \mathbb{Z}} |h|^{\frac{1}{\beta}} C^X(|h|) < \infty.$$

Then, ϵ_n minimizing the sum of

$$\frac{1}{n} \sum_{h=-n}^{n} n_{\epsilon,|h|} |C^X(|h|)| \text{ and } \epsilon \sum_{h=-n}^{n} |C^X(|h|)|$$

is (asymptotically) given by

$$\epsilon_n = \min_{\epsilon > 0} \left(\frac{1}{n} B \epsilon^{-\frac{1}{\beta}} + A\epsilon \right).$$

Calculating the derivative with respect to ϵ yields

$$-\frac{1}{\beta n} B \epsilon^{-\frac{\beta+1}{\beta}} + A$$

whose only root is given by

$$\epsilon_n = \left(\frac{A}{B} \beta n \right)^{-\frac{\beta}{1+\beta}}.$$

A straightforward calculation shows

$$\frac{1}{n} \epsilon_n^{-\frac{1}{\beta}} + \epsilon_n = n^{-\frac{1+\beta}{1+\beta}} \left(\frac{A}{B} \beta n \right)^{\frac{1}{1+\beta}} + \epsilon_n = 2\epsilon_n.$$

Combined, we obtain the convergence rate

$$\mathbb{E}^X \left[\mathbb{V}\mathrm{ar}^V \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} V_i X_i \right) - \sigma_\infty^2 \right] = \mathcal{O} \left(\frac{1}{n} \epsilon_n^{-\frac{1}{\beta}} + \epsilon_n \right) = \mathcal{O}(2\epsilon_n) = \mathcal{O}(n^{-\frac{\beta}{1+\beta}}).$$

Finally, we pick β to minimize the MSE

$$E^X \left[\left(\mathbb{V}\mathrm{ar}^V \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} V_i X_i \right) - \sigma_\infty^2 \right)^2 \right] = \mathcal{O} \left(n^{-\frac{2\beta}{1+\beta}} \right) + \mathcal{O}(n^{\beta - 1}).$$

Minimizing

$$n^{-\frac{2\beta}{1+\beta}} + n^{\beta-1}$$

for $n \to \infty$ amounts in minimizing the sum of exponents

$$-\frac{2\beta}{1+\beta} + (\beta - 1) = \frac{\beta^2 - 2\beta - 1}{1+\beta}.$$

Its derivative is given by

$$\frac{(\beta+1)^2 - 2}{(\beta+1)^2}.$$

Thus the optimal $\beta \in (0, \frac{1}{2})$ is given by $\beta_{opt} = \sqrt{2} - 1$. □

D Central limit theorem

Proof of Theorem 3.5. Theorem 1.7 in Bosq (2012) with $\gamma = 4$ implies that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - \mathbb{E}(X_i)) \xrightarrow{d} \mathcal{N}(0, \sigma_\infty^2),$$

as $n \to \infty$, or, written differently,

$$\sup_{x\in\mathbb{R}}\left|\mathbb{P}\left\{\sqrt{n}(T_n - \mathbb{E}(T_n)) \le x\right\} - \Phi(x/\sigma_\infty)\right| \to 0, \tag{11}$$

where $T_n = n^{-1}\sum_{i=1}^n X_i$ and Φ the standard normal cumulative distribution function. It remains to show that the centered bootstrap average has the same limiting distribution. Recall from Lemma 3.4 that

$$\sqrt{n}(T_n^* - T_n) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{V_i}{\overline{V}_n}X_i - \frac{1}{\sqrt{n}}\sum_{i=1}^{n}X_i = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(V_i - 1)(X_i - \mathbb{E}[X_i]) + o_P(1).$$

The distribution of the sum on the right, conditional on $X_1, \dots, X_n$, is normal with mean 0 and variance

$$\sigma_n^2 := \mathbb{V}\mathrm{ar}^V\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(V_i - 1)(X_i - \mathbb{E}[X_i])\right) = \mathbb{V}\mathrm{ar}^V\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_i(X_i - \mathbb{E}[X_i])\right),$$

which converges in probability to σ_∞^2 by Lemma C.3. We therefore get

$$\begin{aligned}\sup_{x\in\mathbb{R}}|\mathbb{P}^*\{\sqrt{n}(T_n^* - T_n) \le x\} - \Phi(x/\sigma_\infty)| &\le \sup_{x\in\mathbb{R}}|\mathbb{P}^*\{\sqrt{n}(T_n^* - T_n) \le x\} - \Phi(x/\sigma_n)| + \sup_{x\in\mathbb{R}}|\Phi(x/\sigma_n) - \Phi(x/\sigma_\infty)| \\ &\xrightarrow{P} 0,\end{aligned}$$

where the rightmost supremum converges by the continuous mapping theorem. Together with (11), we obtain

$$\begin{aligned}&\sup_{x\in\mathbb{R}}|\mathbb{P}^*\{\sqrt{n}(T_n^* - T_n) \le x\} - \mathbb{P}\left\{\sqrt{n}(T_n - \mathbb{E}(T_n)) \le x\right\}| \\ &\le \sup_{x\in\mathbb{R}}|\mathbb{P}^*\{\sqrt{n}(T_n^* - T_n) \le x\} - \Phi(x/\sigma_\infty)| + \sup_{x\in\mathbb{R}}|\mathbb{P}\left\{\sqrt{n}(T_n - \mathbb{E}(T_n)) \le x\right\} - \Phi(x/\sigma_\infty)| \xrightarrow{P} 0.\end{aligned}$$

□

E Details on Moving average block bootstrap

Example E.1 (Multiplier block bootstrap with MA-weights)**.** *Consider the moving average (MA) process*

$$V_{j,n} = \sum_{j\in\mathbb{Z}} b_{j,n}\zeta_{i-j,n}$$

with

$$b_j = \begin{cases} m_n^{-1}(1 - |j|/m_n) & |j| \le m_n \\ 0 & \textit{else,}\end{cases}$$

where $\zeta_{i,n} \overset{iid}{\sim} Gamma(q_n, q_n)$ *with* $q_n = \frac{2}{3m_n} + \frac{1}{3m_n^2}$ *and* $m_n \sim Cn^{1/3}$ *asymptotically. Define the resampling scheme* $X_{i,n}^* = X_i V_{i,n}/\overline{V}_n$*.*

F Evaluation metrics

Recall that Theorem 3.6 yields a consistent variance estimator, i.e.

$$\mathbb{V}\mathrm{ar}^V\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{V_i}{\overline{V}_n}X_i\right) = \mathbb{V}\mathrm{ar}^V\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\frac{V_i}{\overline{V}_n} - 1\right)X_i\right) \to \sigma_\infty$$

in probability. Furthermore, Theorem 3.5 yields asymptotically consistent confidence intervals C_n at level α for $\theta = \mathbb{E}(X_1) = \mathbb{E}(X_i)$ by Lemma 23.3 of Van der Vaart (2000), i.e.

$$\lim \inf_{n\to\infty} P(\theta \in C_n(X_1, ..., X_n)) \geq 1 - \alpha.$$

Accordingly, given some realizations $x_1, ..., x_n$, we obtain

$$ev_1(x_1, ..., x_n) = \mathbb{V}\mathrm{ar}^V \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{V_i}{\overline{V}_n} x_i \right)$$

$$ev_2(x_1, ..., x_n) = \begin{cases} 1 & \theta \in C_n(x_1, ..., x_n) \\ 0 & \theta \notin C_n(x_1, ..., x_n) \end{cases}$$

which provide quantities of the accuracy of the bootstrap procedure. Observe

$$\mathbb{E}(ev_2(X_1, ..., X_n)) = P(\theta \in C_n(X_1, ..., X_n))$$

and

$$ev_1(X_1, ..., X_n) = \mathbb{V}\mathrm{ar}^V \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{V_i}{\overline{V}_n} X_i \right)$$

by the very construction.

Following standard practice, we estimate the mean E_1 and variance E_2 of $ev_1(X_1, ..., X_n)$ as well as the expected value E_3 of $ev_2(X_1, ..., X_n)$ by sampling $M = 250$ times from the underlying bootstrap distribution, i.e. realize $v_{1,j}, ...v_{n,j} \in \mathbb{R}$ according $v_{i,j} \sim V_i$ for $j = 1, ..., M$ and calculate

$$\hat{ev}_1(x_1, ..., x_n) = \frac{1}{M-1} \sum_{j=1}^{M} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{v_{i,j}}{\overline{v_n}} x_i \right)^2$$

$$\hat{ev}_2(x_1, ..., x_n) = \begin{cases} 1 & \theta \in \hat{C}_n(x_1, ..., x_n) \\ 0 & \theta \notin \hat{C}_n(x_1, ..., x_n) \end{cases}$$

where $\hat{C}_n(x_1, ..., x_n)$ denotes corresponding estimate of $C_n(x_1, ..., x_n)$.

G Auxilary results

Lemma G.1. *Assume*

$$\mathbb{E}(X_i^4) < \infty$$

and

$$\sum_{i=1}^{\infty} \alpha(i)^{\frac{1}{2}} < \infty.$$

Then,

$$\sum_{h=-\infty}^{\infty} |C^X(|h|)| < \infty.$$

Proof. We apply (corollary 1.1 Bosq) to $p = \frac{1}{2}$ and $r = q = \frac{1}{4}$. A straightforward calculation exhibits

$$\frac{1}{q} + \frac{1}{r} = 1 - \frac{1}{p}.$$

We obtain

$$|C^X(|h|)| \leq 2p\alpha(|h|)^{\frac{1}{2}} \mathbb{E}(X_1^4)^{\frac{1}{2}}$$

by stationarity of $(X_i)_{i\in\mathbb{N}}$. Hence,

$$\sum_{h=-\infty}^{\infty} |C^X(|h|)| < \infty$$

follows from

$$\sum_{i=1}^{\infty} \alpha(i)^{\frac{1}{2}} < \infty.$$

□

Lemma G.2. *Let $(X_i)_{i\in\mathbb{N}}$ be a zero mean, real valued, strictly stationary strong mixing stochastic process with eigth moments existing.*

Then, for all $a, b, c, d \in \mathbb{N}$

$$\begin{aligned} &|\mathbb{E}(X_aX_bX_cX_d) - \mathbb{E}(X_aX_b)\mathbb{E}(X_cX_d) - \mathbb{E}(X_aX_c)\mathbb{E}(X_bX_d) - \mathbb{E}(X_aX_d)\mathbb{E}(X_bX_c)| \\ &\leq C\alpha(\max\{i_2 - i_1, i_3 - i_2, i_4 - i_3\})^{\frac{1}{2}} \end{aligned}$$

with $\{a, b, c, d\} = \{i_1, i_2, i_3, i_4\}$ such that $i_1 \leq i_2 \leq i_3 \leq i_4$ and $C \in \mathbb{R}$ independent of $a, b, c, d \in \mathbb{N}$.

In particular, fixing a, we obtain

$$\mathbb{E}(X_aX_bX_cX_d) - \mathbb{E}(X_aX_b)\mathbb{E}(X_cX_d) = \mathbb{E}(X_aX_c)\mathbb{E}(X_bX_d) - \mathbb{E}(X_aX_d)\mathbb{E}(X_bX_c) + \epsilon_{b,c,d}$$

with $\epsilon_{b,c,d} \to 0$ if some b, c or d tends to infinity

Proof. In order to remain clarity, we only outline the proof rather than going into detaill.

Since

$$|\mathbb{E}(X_aX_bX_cX_d) - \mathbb{E}(X_aX_b)\mathbb{E}(X_cX_d) - \mathbb{E}(X_aX_c)\mathbb{E}(X_bX_d) - \mathbb{E}(X_aX_d)\mathbb{E}(X_bX_c)|$$

is invariant under permutation of the a, b, c, d, we may without loss of generality assume $a \leq b \leq c \leq d$, i.e. $a = i_1, ..., d = i_4$. Then, we obtain

$$\begin{aligned} |\mathbb{E}(X_aX_bX_cX_d) - \mathbb{E}(X_aX_b)\mathbb{E}(X_cX_d)| = |\mathbb{C}\mathrm{ov}(X_aX_b, X_cX_d)| &\leq K\alpha(c-b)^{1/2}\mathbb{E}(X_1^8)^{1/2} \\ |\mathbb{E}(X_aX_c)| = |\mathbb{C}\mathrm{ov}(X_a, X_c)| &\leq K\alpha(c-a)^{1/2}\mathbb{E}(X_1^4)^{1/2} \\ |\mathbb{E}(X_bX_c)| = |\mathbb{C}\mathrm{ov}(X_b, X_c)| &\leq K\alpha(c-b)^{1/2}\mathbb{E}(X_1^4)^{1/2} \end{aligned}$$

by applying Corollary 1.1 in Bosq (2012) with $q = r = 4, p = 2$ to the respective covariance and by using Cauchy-Schwarz inequality. Combined, triangle inequality yields

$$\begin{aligned} &|\mathbb{E}(X_aX_bX_cX_d) - \mathbb{E}(X_aX_b)\mathbb{E}(X_cX_d) - \mathbb{E}(X_aX_c)\mathbb{E}(X_bX_d) - \mathbb{E}(X_aX_d)\mathbb{E}(X_bX_c)| \\ &\leq C\alpha(c-b)^{\frac{1}{2}}. \end{aligned}$$

for some C independent of $a, ..., d$. By the very same argument we derive

$$\begin{aligned} |\mathbb{E}(X_aX_b)| = |\mathbb{C}\mathrm{ov}(X_a, X_c)| &\leq K\alpha(b-a)^{1/2}\mathbb{E}(X_1^4)^{1/2} \\ |\mathbb{E}(X_aX_c)| = |\mathbb{C}\mathrm{ov}(X_b, X_c)| &\leq K\alpha(b-a)^{1/2}\mathbb{E}(X_1^4)^{1/2} \\ |\mathbb{E}(X_aX_d)| = |\mathbb{C}\mathrm{ov}(X_b, X_c)| &\leq K\alpha(b-a)^{1/2}\mathbb{E}(X_1^4)^{1/2} \end{aligned}$$

Next, we apply Corollary 1.1 in Bosq (2012) with $q = 8/3, r = 8$ and $p = 2$ in order to obtain

$$\begin{aligned} |\mathbb{E}(X_aX_bX_cX_d)| = |\mathbb{C}\mathrm{ov}(X_a, X_bX_cX_d)| &\leq K\alpha(b-a)^{1/2}\mathbb{E}(X_1^8)^{1/8}\mathbb{E}((X_bX_cX_d)^{8/3})^{3/8} \\ &\leq K\alpha(b-a)^{1/2}\mathbb{E}(X_1^8)^{1/2} \end{aligned}$$

where we applied Hölder's inequality in the last step. Thus, triangle inequality yields

$$
\begin{aligned}
&|\mathbb{E}(X_aX_bX_cX_d) - \mathbb{E}(X_aX_b)\mathbb{E}(X_cX_d) - \mathbb{E}(X_aX_c)\mathbb{E}(X_bX_d) - \mathbb{E}(X_aX_d)\mathbb{E}(X_bX_c)| \\
&\leq C\alpha(b-a)^{\frac{1}{2}}
\end{aligned}
$$

and the same argument yields

$$
\begin{aligned}
&|\mathbb{E}(X_aX_bX_cX_d) - \mathbb{E}(X_aX_b)\mathbb{E}(X_cX_d) - \mathbb{E}(X_aX_c)\mathbb{E}(X_bX_d) - \mathbb{E}(X_aX_d)\mathbb{E}(X_bX_c)| \\
&\leq C\alpha(d-c)^{\frac{1}{2}}
\end{aligned}
$$

for some constant independent of $a, ..., d$. This proves the first claim.

The last claim follows by the strong mixing property of (X_i). This completes the proof. □