

Estimating autocorrelation from multiple Markov chains

Medha Agarwal

Department of Mathematics and Statistics

IIT Kanpur

`medhaaga@iitk.ac.in`

Dootika Vats

Department of Mathematics and Statistics

IIT Kanpur

`dootika@iitk.ac.in`

June 12, 2020

Abstract

This is best abstract in the world.

.
. .
. .
. .
. .

1 Introduction

Markov chain Monte Carlo (MCMC) is a sampling method used to obtain correlated realizations from a stochastic process. Due to this serial correlation in observations, a large number of MCMC samples are required to estimate features of the target distribution as compared to the independent ones. Therefore, it becomes imperative for an MCMC practitioner to have an idea about underlying dependence structure in the Markov chains. Autocovariance function (ACF) characterizes this dependence for a stationary Markov chain. A famous sample autocovariance function is the empirical estimator, which unfortunately comes with its drawbacks. Firstly, it is not immune to outliers (see Ma and Genton (2000)) and secondly, it is usually non-informative in case of slowly mixing Markov chains for decently finite samples. We address the second issue in this paper by using multiple representative samples of the target distribution via parallel Markov chains.

Calculation of exact theoretical ACF has been done for simple models like autoregressive (AR), moving average (MA) (Quenouille (1947)), and autoregressive-moving-average (ARMA) (Box et al. (2015)) models. However, for most of the complex processes, we rely on sample ACF to understand the correlation in the chain at a particular lag. For a slowly mixing Markov chain or a multimodal target distribution, Markov chains explore the sample space only partially for even high sample size. In such situations, single chain empirical ACF will underestimate the autocovariance giving misleading inferences. With convenient multi-core computation resources, it is possible to get multiple Markov chains for a target distribution. We propose a new autocovariance estimator called replicated autocovariance function (R-ACF) with a simple fix in centering a chain about the overall mean. An important application of ACF is in calculating the asymptotic variance in the CLT of a Markov chain process.

Suppose F is the target distribution on support $\chi \subseteq \mathbb{R}^d$. For $g : \chi \rightarrow \mathbb{R}^p$ be an F -integrable function. We are interested in

$$\mu_g = \mathbb{E}_F[g(X)] = \int_{\chi} g(x)F(dx)$$

Suppose we have m replications of Markov chains with n samples in each chain. Let $\{X_{st}\}_{t \geq 1}$ denote the s^{th} Harris ergodic (aperiodic, F -irreducible, and positive Harris recurrent) Markov chain having invariant distribution F . We will use the following Monte Carlo estimator in order to estimate μ for the s^{th} chain

$$\hat{\mu}_s = \frac{1}{n} \sum_{i=1}^n g(X_{si}) \xrightarrow{a.s.} \mu \text{ as } n \rightarrow \infty$$

Throughout this paper, we will deal with the function $g(X) = X$. We denote the sample mean of s^{th} Markov chain by \bar{X}_s and the overall mean by $\bar{\bar{X}} = \sum_{s=1}^m \bar{X}_s / m$. We are interested in Monte Carlo error, i.e. $\bar{\bar{X}} - \mu$ to assess the reliability of simulations. The sampling distribution for Monte Carlo error of s^{th} chain is available via Markov chain CLT (Jones et al. (2004)) if there exists a positive definite matrix Σ such that

$$\sqrt{n}(\bar{X}_s - \mu) \xrightarrow{d} N(0, \Sigma)$$

With m i.i.d. samples of \bar{X}_s , $s \in \{1, \dots, m\}$, the MCSE is given by Σ/mn . We are interested in consistent estimates of Σ for two reasons. Firstly, to construct asymptotically valid confidence intervals and secondly, to determine when to stop the simulations. For this purpose, we use the class of Multivariate Spectral Variance Estimators (MSVE) wherein we use R-ACF to estimate the autocovariance. We call this replicated spectral variance (RSV) estimator (see Argon and Andradóttir (2006) for replicated batch means).

MCMC ensures asymptotic convergence, however this is not practically very useful where the number of samples is finite. There are many criteria for terminating simulation like fixed-time

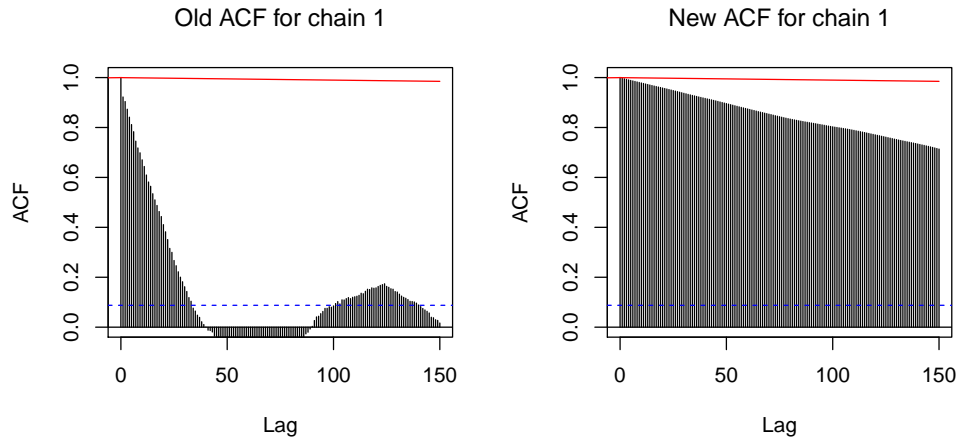
rule, sequential fixed-volume rule (see Glynn et al. (1992)), and using Gelman Rubin diagnostics (Gelman et al. (1992)). The relative fixed-volume rule has been thoroughly discussed for MCMC by Flegal and Gong (2015) and Gong and Flegal (2016). We will use method of effective sample size (ESS) defined by Vats et al. (2019) for terminating the simulations. The quality of estimation of Σ is critical in determining that the simulations are not stopped prematurely.

Markov chains are serially correlated and therefore $\text{Var}_F(X)$ is non-informative about the correlations in Markov chain. It is important to construct an estimator for Σ which is very close to underlying reality. There are many techniques for estimating Σ like batch means (BM), overlapping batch means (OBM), and spectral variance estimator (SVE). Due to their highly accurate results, we will only focus on SV estimates in this paper. Strong consistency for MSVE has been addressed by Vats et al. (2018). Bias and variance calculations for a variant of SV estimator with known mean is done by Hannan (2009). In this paper, we will address the asymptotic properties (basically strong consistency), bias and variance calculations for RSV estimator.

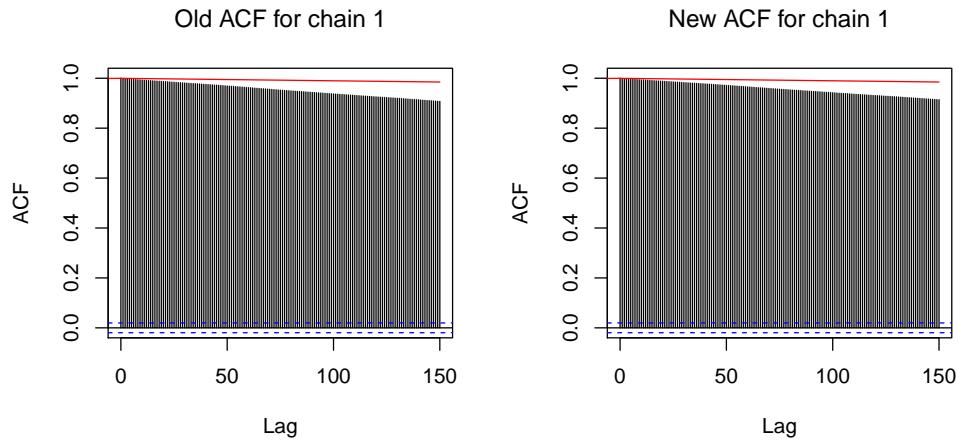
The most common practice would be to combine the SVE from m chains by averaging over them. We call this technique average spectral variance (ASV) estimator. We will address the issue of slowly mixing Markov chains and show that RSV performs better than ASV. Whereas on one side RSV gives better estimates than ASV for a poor Markov chain, it is as good as ASV for a good Markov chain. We will exemplify this property of RSV for a bi-modal target distribution in subsection 5.2. This appeals intuitively because for a fast mixing Markov chain, all the chains will give almost same estimates for mean.

In the next section we introduce the new sample autocovariance estimator and calculate its bias for finite samples. We observe that the bias term for R-ACF is similar to the bias of empirical ACF except for a few small order end-effect terms that vanish as $n \rightarrow \infty$. In section 3, we introduce three estimators for Σ and study the properties of RSV estimator through proof of strong consistency, bias, and variance calculations. We will elaborate on the estimator for ESS using RSV in section 4. Section 5 gives a simulation study on three different examples for the proposed estimators. There are only a handful of stochastic processes where the true autocovariance and Σ are known. To compare the estimators when the truth is known, we use a slowly mixing vector autoregressive process of order 1 (VAR(1)). We successfully show in section 5.1 that our estimators yield better results as compared to the classical sample ACF and SVE. A more promising advantage of our estimators is observed when the target distribution is multi-modal. For this purpose we use a bivariate bi-modal probability distribution introduced by Gelman and Meng (1991) in section 5.2. All the important proofs are given in the Appendix.

Example 1 (Motivating example). gj



(a) $n = 500$



(b) $n = 10^4$

Figure 1: Autocorrelation plot for ACF and R-ACF for first chain out of two parallel Markov chains. First column corresponds to ACF calculated using arithmetic mean of first Markov chain and the second column corresponds to the one calculated using global mean of two Markov chains.(a): $n = 500$, not converged yet; (b): $n = 10^4$, converged

2 Replicated Autocovariance Function

For a stationary stochastic process $\{X_t\}_{t \geq 1}$, let $\Gamma(k)$ denote the lag $-k$ autocovariance. In time series literature, we generally encounter the following estimator for autocovariance.

$$\hat{\Gamma}(k) = \frac{1}{n} \sum_{t=1}^{n-|k|} (X_t - \bar{X})(X_{t+|k|} - \bar{X})^T$$

Priestley (1981) shows that this estimator is biased with the main bias term equal to $|k|\Gamma(k)/n$, ignoring the small order terms arising due to estimation of μ . An unbiased estimator with a divisor of $n - |k|$ instead of n exists, say $\Gamma^*(k)$. Nevertheless, the biased estimator is preferred for its lower mean square error (Priestley (1981)). For a small k relative to n , the two estimators are almost the same. However, for larger relative k the variance of $\hat{\Gamma}(k)$ compensates for larger bias rendering a smaller mean square error than $\Gamma^*(k)$.

A commonly faced problem with the MCMC algorithms is the slow mixing property. In case of multi modal target distributions, it has been observed that even decently mixing Markov chains fail to explore the sample space for the first few thousand observations. This leads to underestimation of autocovariance for a single Markov chain. We propose a simple and neat fix for estimating ACF while dealing with slow convergence. Suppose we have m replications of Markov chains for same target distribution, the overall mean $\bar{\bar{X}} = \sum_{i=1}^m \bar{X}_i/m$ provides a better estimate for the expectation value when the number of simulation per chain are not sufficient for the convergence to kick in. The replicated autocovariance estimator or s^{th} Markov chain is given by:

$$\hat{\Gamma}_{RAC,s}(k) = \frac{1}{n} \sum_{t=1}^{n-k} (X_{st} - \bar{\bar{X}})(X_{s(t+k)} - \bar{\bar{X}})^T$$

If the starting points of these parallel Markov chains are sufficiently dispersed, R-ACF provides more realistic estimation of lag covariances. It is a known problem that getting true autocovariance in closed form for non-simple problems is difficult. In Figure 1, we will illustrate the comparison between ACF and R-ACF for a slowly mixing VAR-1 example, where the truth is known. We wish to make two points here: firstly, for small sample size, we observe that ACF gives extremely misleading estimates of autocovariance whereas R-ACF is very close to reality. Figure 1a depicts the scenario where the convergence is not achieved yet. Secondly, for a large sample size, the estimates from R-ACF as well as ACF are equivalent. Figure 1b depicts the scenario when convergence is achieved. This shows that R-ACF is at least as good as ACF; hence safe to use in any setting.

The bias for R-ACF is similar to the bias of sample autocovariance estimator in case of single chain except for a few small order terms that shall be later addressed in theorem 1. Before getting to the

bias results for $\hat{\Gamma}_{RAV}(k)$ for any lag k , we introduce an additional notation. For $q \geq 1$, we define

$$\Phi^{(q)} = \sum_{-\infty}^{\infty} |k|^q \|\Gamma(k)\|$$

We denote $\Phi^{(1)}$ by Φ .

Theorem 1.

$$\mathbb{E}[\hat{\Gamma}_{RAV}(k)] = \left(1 - \frac{|k|}{n}\right) \left(\Gamma(k) - \frac{\Sigma}{mn} - \frac{\Phi}{mn^2}\right) + \frac{2(m-1)|k|}{m} \frac{1}{n^2} \left(\Sigma - \sum_{h=0}^{n-1} \Gamma(h)\right) + o(n^{-2})$$

Proof. The estimator for autocovariance with divisor n is suggested by Parzen et al. (1961) for its lesser mean square error despite having greater bias than the unbiased estimator with an $n - |k|$ divisor. We know that $\hat{\Gamma}_s(k)$ is asymptotically unbiased with a $o(1)$ bias term that remains fixed for any k . The bias term is given by Priestley (1981) as

$$\mathbb{E}[\hat{\Gamma}(k)] = \left(1 - \frac{|k|}{n}\right) (\Gamma(k) - \text{Var}(\bar{X})) \quad (1)$$

By proposition 1 in Song and Schmeiser (1995)

$$\text{Var}(\bar{X}) = \frac{\Sigma}{n} + \frac{\Phi}{n^2} + o(n^{-2}) \quad (2)$$

As a consequence, if $\text{Var}(\bar{X})$ is finite, then $\text{Var}(\bar{X}) \rightarrow 0$ as $n \rightarrow \infty$. We can break $\hat{\Gamma}_{RAV}$ into four parts for all $k \geq 1$ as:

$$\hat{\Gamma}_{RAV} = \frac{1}{m} \sum_{s=1}^m \hat{\Gamma}_s(k) - \frac{1}{mn} \sum_{s=1}^m \sum_{t=1}^{|k|} A_{st}^T - \frac{1}{mn} \sum_{s=1}^m \sum_{t=n-|k|+1}^n A_{st} + \frac{n-|k|}{mn} \sum_{s=1}^m (\bar{X}_s - \bar{\bar{X}})(\bar{X}_s - \bar{\bar{X}})^T,$$

where $A_{st} = (X_{st} - \bar{X}_s)(\bar{X}_s - \bar{\bar{X}})^T$

The expectation value of the first term is given by equation 1 and 2 while the other three terms contribute $o(1)$ terms. The complete proof evaluating the expectation values of last three terms can be found in Appendix subsection B.1. \square

Remark 1. Using theorem 1 and equation 1,

$$\mathbb{E}[\hat{\Gamma}_{RAV}(k) - \hat{\Gamma}(k)] = \frac{|k|}{n} \left(1 - \frac{1}{m}\right) \left(\frac{\Sigma}{n} + \frac{\Phi}{n^2}\right) + \frac{2(m-1)|k|}{m} \frac{1}{n^2} \left(\Sigma - \sum_{h=0}^{n-1} \Gamma(h)\right) + o(n^{-2})$$

We know that the sample autocovariance $\hat{\Gamma}(k)$ underestimates the truth due to a negative bias of

$|k|\Gamma(k)/n$ for finite n . However, in practice, it is preferred to overestimate the autocovariance for finite n than to underestimate it in order to have a more realistic idea about the correlation in a Markov chain at a particular lag. Note that the RHS in the above expression is positive. As a consequence, R-ACF will counter this negative bias and give more reliable estimates for ACF.

Corollary 1.

$$\mathbb{E}[\hat{\Gamma}_{RAC}(0)] = \Gamma(0) + \frac{\Sigma}{mn} + \frac{\Phi}{mn^2} + o(n^{-2})$$

Proof. The proof is in Appendix subsection B.1. □

3 Variance Estimators for multiple Markov chains

Let $\{X_t\}_{t \geq 1}$ be a stochastic process. Denote lag- k autocovariance matrix by $\Gamma(k)$, which is given by,

$$\Gamma(k) = \Gamma(-k) = \mathbb{E}[(X_t - \mathbb{E}[X])(X_{t+k} - \mathbb{E}[X])^T]$$

Recall that $\hat{\Gamma}(k)$ is the empirical autocovariance function given by

$$\hat{\Gamma}(k) = \frac{1}{n} \sum_{t=1}^{n-|k|} (X_t - \bar{X})(X_{t+|k|} - \bar{X})^T$$

The spectral variance estimator is the weighted and truncated sum of lag- k autocovariances, given by:

$$\hat{\Sigma}_{SV} = \sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right) \hat{\Gamma}_n(k)$$

where $w(\cdot)$ is the lag window and b_n is the truncation point.

Consider m Markov chains where \bar{X}_s , $\hat{\Gamma}_s(k)$ and $\hat{\Sigma}_{SV,s}$ are the sample mean, sample lag- k autocovariance and spectral variance estimators respectively for s^{th} Markov chain, $s \in 1, \dots, m$. We will consider two estimators for spectral variance calculations using multiple Markov chains.

3.1 Average Spectral Variance

Recall that for s^{th} Markov chain, the spectral variance estimator is given by,

$$\begin{aligned} \hat{\Sigma}_{SV,s} &= \sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right) \hat{\Gamma}_s(k) \\ \hat{\Gamma}_s(k) &= \frac{1}{n} \sum_{t=1}^{n-k} (X_{st} - \bar{X}_s)(X_{s(t+k)} - \bar{X}_s)^T \end{aligned}$$

The average spectral variance (Σ_{ASV}) is computed by averaging the SV estimates from all m Markov chains.

$$\hat{\Sigma}_{ASV} = \frac{1}{m} \sum_{s=1}^m \hat{\Sigma}_{SV,s}$$

3.2 Replicated Spectral Variance

Now we introduce the Replicated Spectral Variance Estimator (RSV). $\hat{\Gamma}_{RAC,s}(k)$ is the sample replicated lag- k autocovariance estimator for the s^{th} MC using overall mean ($\bar{\bar{X}}$) instead of \bar{X}_s for centering. We construct an RSV estimator $\hat{\Sigma}_{RSV}$ using this replicated version of autocovariance estimator. Consider,

$$\begin{aligned} \hat{\Sigma}_{RSV,s} &= \sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right) \hat{\Gamma}_{RAC,s}(k) \\ \hat{\Sigma}_{RSV} &= \frac{1}{m} \sum_{s=1}^m \hat{\Sigma}_{RSV,s} \end{aligned}$$

The RSV estimator centers the data around the global mean. We are interested in proving the strong consistency, finding the bias and variance of RSV.

Assumption 1 (Strong Invariance Principle (SIP)). We assume the SIP holds for the Markov chain. Here $\{X_t\}_{t \geq 1}$ is the stochastic process that follows SIP with mean $\mu = \mathbb{E}[X]$. Let $\|\cdot\|$ denote the Euclidean norm. Through SIP, there exists a $p \times p$ lower triangular matrix L , an increasing function ψ on the integers, a finite random variable D , and a sufficiently rich probability space such that, with probability 1,

$$\left\| \sum_{t=1}^n X_t - n\mu - LB(n) \right\| < D\psi(n)$$

Let $\{B(n)\}_{n \geq 0}$ denotes a standard p -dim Brownian motion and $B^{(i)}$ denotes its i^{th} component. As shown in the results Kuelbs and Philipp (1980), SIP will hold for $\psi(n) = n^{1/2-\lambda}$, $\lambda > 0$ for Markov chains commonly encountered in MCMC settings. We know that $\psi(n)$ is an inherent property of the stochastic process that satisfies the CLT. This implies that $\psi(n)$ is at least $o(\sqrt{n})$. Using Law of Iterative Logarithms (LIL), a tighter bound for $\psi(n)$ is given by Strassen (1964) as $\mathcal{O}(\sqrt{n \log \log n})$.

Assumption 2. We assume that all the conditions given by Vats et al. (2018) for strong consistency of the spectral variance estimator hold true. As a consequence,

$$\hat{\Sigma}_{SV,s} \xrightarrow{a.s.} \Sigma \text{ as } n \rightarrow \infty \quad \forall s \in \{1, \dots, m\}$$

Theorem 2. *Let the Assumptions 1,2 hold and $n^{-1}b_n \log \log n \rightarrow 0$ as $n \rightarrow \infty$, then $\hat{\Sigma}_{RSV} \rightarrow \Sigma$ with probability 1, as $n \rightarrow \infty$.*

Proof. We complete the proof by showing that $\hat{\Sigma}_{RSV}^{ij} - \tilde{\Sigma}^{ij}$ converges to 0 almost surely for all $i, j \in \{1, \dots, p\}$ and $\tilde{\Sigma}^{ij} \rightarrow \Sigma^{ij}$ with probability 1 as $n \rightarrow \infty$. We will show that

$$\hat{\Sigma}_{RSV}^{ij} - \tilde{\Sigma}^{ij} = g_1(n)D^2 + g_2(n)D + g_3(n)$$

where D is the finite random variable associated with SIP. We will show that

$$\begin{aligned} g_1(n) &= (4 + C_1) \frac{b_n \psi^2(n)}{n^2} - 4 \frac{\psi^2(n)}{n^2} \\ g_2(n) &= 2\sqrt{2} \|L\| p^{1/2} (1 + \epsilon) \left[(4 + C_1) \frac{b_n \psi(n) \sqrt{n \log \log n}}{n^2} - 4 \frac{\psi(n) \sqrt{n \log \log n}}{n^2} \right] \\ g_3(n) &= \|L\|^2 p (1 + \epsilon)^2 \left[(4 + C_1) \frac{b_n \log \log n}{n} - 4 \frac{\log \log n}{n} \right] \end{aligned}$$

Using $\psi(n) = \mathcal{O}(\sqrt{n \log \log n})$, $b_n = o(n)$, and $b_n \log \log n = o(n)$, we will finally prove that $g_1(n), g_2(n), g_3(n) \rightarrow 0$ with probability 1, as $n \rightarrow \infty$. The proof is in Appendix subsection B.2 \square

We are interested in finding the bias for RSV estimator as a function of n . The proof below applies to all the important range of lag windows $w(x)$; where $w(x)$ is a continuous, even function with $w(0) = 1, |w(x)| < c$, and $\int_{-\infty}^{\infty} w^2(x) dx < \infty$. Define

$$W_n = \frac{1}{2\pi} \sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right)$$

We have $W_n < \infty$ for all the important lag windows.

Assumption 3. Let $\Gamma_s(k)$ be the lag- k autocovariance for s^{th} Markov chain and $w(x)$ be the lag window. We assume that there exists a $q \geq 0$ such that

- a. $\sum_{-\infty}^{\infty} |k|^q \|\Gamma(k)\| < \infty$
- b. $\lim_{x \rightarrow 0} \frac{1 - w(x)}{|x|^q} = k_q < \infty$
- c. $\frac{b_n^q}{n} \rightarrow 0$ as $n \rightarrow \infty$

If k_q is finite for some q_0 , then it is zero for $q < q_0$. q is taken to be the largest number satisfying the first two conditions above.

Theorem 3. *The limiting bias of replicated spectral variance is given by:*

$$\lim_{n \rightarrow \infty} b_n^q \mathbb{E} \left[\hat{\Sigma}_{RSV} - \Sigma \right] = -k_q \Phi^{(q)}$$

Proof.

$$\mathbb{E} \left[\hat{\Sigma}_{RSV} - \Sigma \right] = \sum_{k=-b_n+1}^{b_n-1} w \left(\frac{|k|}{b_n} \right) \mathbb{E} \left[\hat{\Gamma}_{RAC}(k) \right] - \sum_{k=-\infty}^{\infty} \Gamma(k)$$

Using theorem 1, we can write the expectation of $\hat{\Gamma}_{RAC}$ as a sum of equation 1 and some small order terms. The proof is given in the Appendix subsection B.1. \square

Assumption 4. Let $\{X_{st}\}_{t \geq 1}$ be the s^{th} Markov chain for which SIP holds.

a. $\mathbb{E}[D^4] < \infty$ where D is a finite random variable associated with the SIP

b. $\mathbb{E}[X^4] < \infty$ where $\{X\}$ is the stochastic process

Theorem 4. If Assumption 4 hold, $b_n^{-1} n \text{Var} \left(\hat{\Sigma}_{RSV}^{ij} \right) = [\Sigma_{ii}\Sigma_{jj} + \Sigma_{ij}^2] \int_{-\infty}^{\infty} w(x)^2 dx + o(1)$.

Proof. Recall that $\tilde{\Sigma}$ is the averaged pseudo spectral variance estimator which uses the unobserved mean. Further, let $\tilde{\Sigma}^{ij}$ be the $(i, j)^{th}$ element of this matrix. We will prove that the variance of $\hat{\Sigma}_{RSV}$ and $\tilde{\Sigma}$ are equivalent for large sample sizes. The proof is in Appendix subsection B.4 \square

4 Effective sample size

Estimating the MCSE is crucial for determining when to terminate the simulations. The existing stopping rules rely on the strong consistency of the estimator of Σ . We will use the *relative standard deviation fixed-volume sequential stopping rule* proposed in Vats et al. (2019) when multivariate Markov chain central limit theorem holds. This stopping rule terminates the simulations on the basis of size of confidence interval relative to inherent variability in the target distribution. It terminates after the total number of samples, i.e. mn exceed some pre-specified n^* iterations to prevent pre-mature termination due to bad estimates of Σ and Λ . The rule is given by

$$\text{Volume of confidence region}^{1/p} + n^{-1} < \epsilon |\Lambda_n|^{1/2p}$$

where ϵ is the tolerance level, $|\cdot|$ denotes the determinant and Λ_n is the sample covariance matrix given by

$$\frac{1}{mn-1} \sum_{s=1}^m \sum_{t=1}^n (X_{st} - \bar{X})(X_{st} - \bar{X})^T$$

The determinant of Monte Carlo standard error is called generalized variance in Wilks (1932) and gives a metric for volume of confidence interval. Vats et al. (2019) showed that if the estimator for Σ is strongly consistent, the stopping rule is asymptotically valid, in that the confidence regions created at termination have the right coverage probability as $\epsilon \rightarrow 0$. Since $\hat{\Sigma}_{RSV}$ is a strongly consistent estimator of Σ , w

$$\left| \hat{\Sigma}_{RSV} \right|^{1/p} + (mn)^{-1} < \epsilon |\Lambda_{mn}|^{1/2p}$$

Another way to get more accurate stopping rule is to use RSV estimator for estimating Effective Sample Size (ESS) in multivariate settings. RSV estimator better captures the standard error of target distribution by considering the global sample mean across the Markov chains; which might otherwise get lost when considering a single localized slowly mixing Markov chain. We therefore define our ESS rule as

$$\text{ESS} = mn \left(\frac{|\Lambda_{mn}|}{\left| \hat{\Sigma}_{RSV} \right|} \right)^{1/p}$$

When there is no correlation in Markov chain $\hat{\Sigma}_{RSV} = \Lambda_{mn}$ and therefore, $\text{ESS} = n$. Both the rules involve the ratio of generalized variances and therefore the relative standard deviation fixed-volume sequential stopping rule is asymptotically equivalent to stopping when ESS is greater than $W_{p,\alpha,\epsilon}$ which depends on dimension of state-space, level of confidence of the confidence region, and the relative precision required.

5 Examples

In the following examples we compare the performance of asymptotic variance estimators (ASVE and RSVE) for Markov chain CLT. We will successfully illustrate that the replicated spectral variance estimator performs better than the average spectral variance estimator in terms of ESS calculations, coverage probabilities, and sample distribution of Σ . We will also show that in case of nicely mixing Markov chains, these estimators give almost equivalent results. As a consequence, while our replicated version of variance estimator is better in cases where the Markov chains do not explore the entire sample space in finite time, it is harmless to be used in cases where they do. [about this new FFT based method for faster RSVE calculations](#)

5.1 Vector Autoregressive Process

We examine the vector autoregressive process of order 1 (VAR(1)) where the true autocovariance is known in closed form. We have already illustrated the better performance of R-ACF over traditional ACF on a slowly mixing VAR(1) process in Figure 1. Here we will examine the performance of RSV estimator on the same process. Consider a p -dimensional VAR(1) process $\{X_t\}_{t \geq 1}$ such that

$$X_t = \Phi X_{t-1} + \epsilon_t$$

where $X_t \in \mathbb{R}^p$, Φ is a $p \times p$ matrix, $\epsilon_t \stackrel{i.i.d.}{\sim} N(0, \Omega)$, and Ω is a positive definite $p \times p$ matrix.

The invariant distribution for this Markov chain is $N(0, \Psi)$ where $\vec{\Psi} = (I_{p^2} - \Phi \otimes \Phi)\vec{\Omega}$. The lag- k autocovariance can be calculated easily for $k > 0$ as

$$\begin{aligned}\Gamma(k) &= \Phi^k \Psi \\ \Gamma(-k) &= \Psi(\Phi^T)^k\end{aligned}$$

The Markov chain is geometrically ergodic when the spectral norm of Φ is less than 1 (Tjstheim (1990)). The CLT holds for the invariant distribution, therefore, Σ exists and is given by

$$\Sigma = (1 - \Phi)^{-1}\Psi + \Psi(1 - \Phi^T)^{-1} - \Psi$$

Let ϕ_{max} be the largest absolute eigenvalue of Φ such that $|\phi_{max}| < 1$. The larger it is, the slower the Markov chain mixes. For our case, we require a slowly mixing VAR(1) process. We consider a bivariate example with $\phi_{max} = 0.9999$. We run five parallel Markov chains with their starting points evenly distributed about the center of invariant distribution. Figure 2a shows that for the first 500 samples, the chains have not explored the sample space well unlike Figure 2b with $5 \cdot 10^4$ samples.

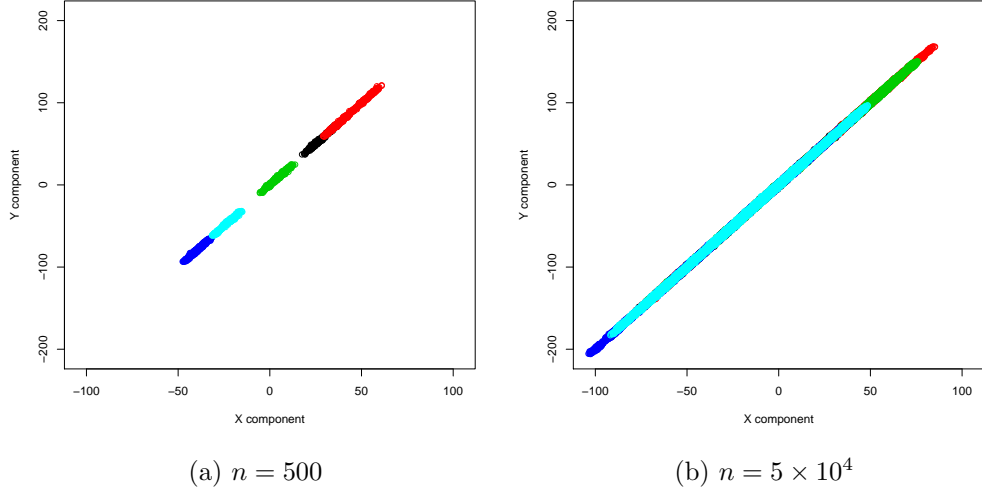


Figure 2: Scatter plot of five Markov chains for different sample sizes.

5.2 Boomerang Distribution

We will use a bivariate bi-modal distribution introduced by Gelman and Meng (1991) which has Gaussian conditional distributions in both directions. This allows us to sample parallel Markov chains using the Gibbs sampler. Let x and y be two random variable that are jointly distributed

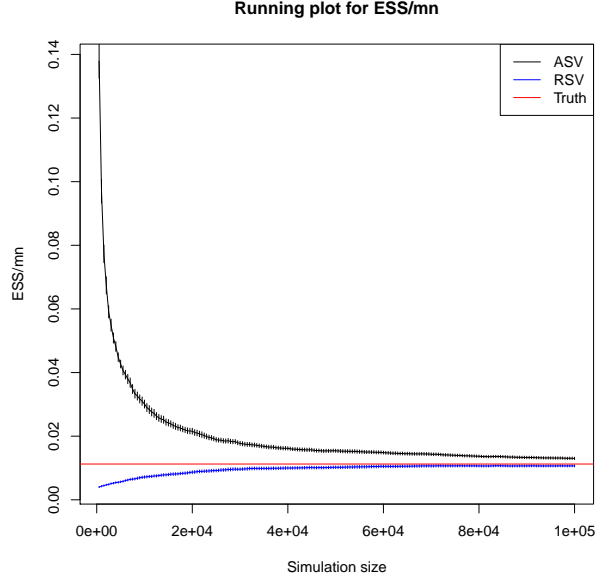


Figure 3: Running plots for \hat{ESS}/mn calculated using ASV (black) and RSV (blue). The true ESS/mn (red) is known.

as

$$f(x, y) \propto \exp \left(-\frac{1}{2} [Ax^2y^2 + x^2 + y^2 - 2Bxy - 2C_1x - 2C_2y] \right)$$

The conditional distribution of x given y and vice versa is a normal distribution given by:

$$\begin{aligned} x_1 \mid x_2 &\sim N \left(\frac{Bx_2 + C_1}{Ax_2^2 + 1}, \frac{1}{Ax_2^2 + 1} \right) \\ x_2 \mid x_1 &\sim N \left(\frac{Bx_1 + C_2}{Ax_1^2 + 1}, \frac{1}{Ax_1^2 + 1} \right) \end{aligned}$$

We use a carefully chosen parameterization of $A = 1$, $B = 3$, $C = 8$ which ensures bimodality for our purpose. Let n be the number of samples in each chain and m be the number of Markov chain replications. Finding the actual mean of this distribution in closed form is difficult. Therefore, we use numerical integration with fine tuning to calculate it. We sample two parallel Markov chains with well-separated starting values.

In Figure 5, we demonstrate the “sticky” nature of the Markov chains. For the first thousand samples, both the chains are oblivious of the existence of another mode. By ten thousand samples, both the Markov chains have explored the two modes. This property will affect the single chain estimators like ACF. In Figure 6a, the ACF is severely underestimated because the chain-1 has not jumped to the other mode. Whereas, in Figure 6b, both ACF and R-ACF give almost similar results indicating that both modes have been explored by chain-1.

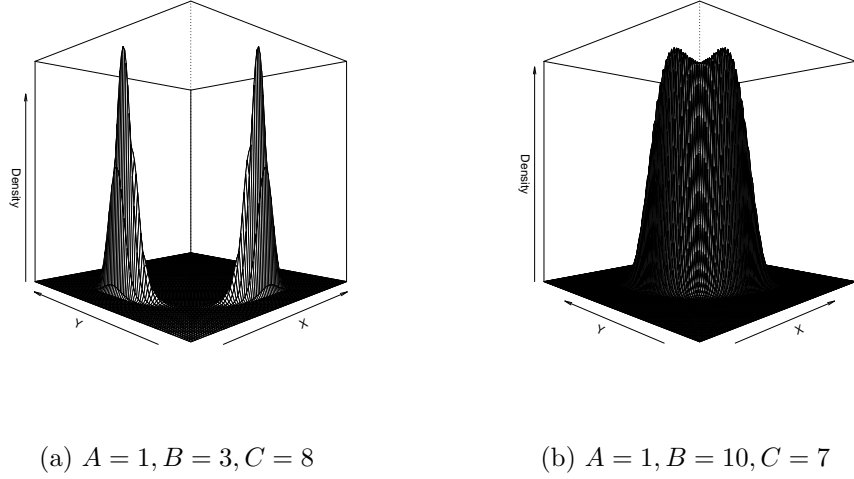


Figure 4: Perspective plots of the target distribution for two different parameterizations.

We also run five parallel Markov chains with well-separated starting points. In Table 1, we give the coverage probabilities for 95% confidence interval for both the estimator. We can see that the RSVE gives better coverage probabilities for all the n . For smaller sample size, the coverage probability of RSVE is significantly higher than ASVE. As the number of samples per chain increases, they start coming closer due to the strong consistency of both the estimators.

n	$m = 2$		$m = 5$	
	ASV	RSV	ASV	RSV
10^3	0.612	0.689	0.602	0.753
$2 \cdot 10^3$	0.693	0.751	0.735	0.827
$5 \cdot 10^3$	0.826	0.854	0.847	0.880
10^4	0.862	0.868	0.884	0.907
$2 \cdot 10^4$	0.899	0.906	0.922	0.934

Table 1: Coverage probabilities for parameter values $A = 1, B = 3, C = 8$.

A good estimate of ESS is crucial to determine when to stop the simulations. We can see in Figure 7 that for the first few thousand samples, ASVE gives misleadingly higher $E\hat{SS}/mn$ than RSVE. This can cause us to stop the sampling before the sample space has been explored by the chains. The inferences derived from this chain would then be severely non-informative.

To examine the performance of RSV for a nicely mixing Markov chain, we use the parametrization of $A = 1, B = 10, C = 7$. This is also a bimodal distribution, however, the two modes interact

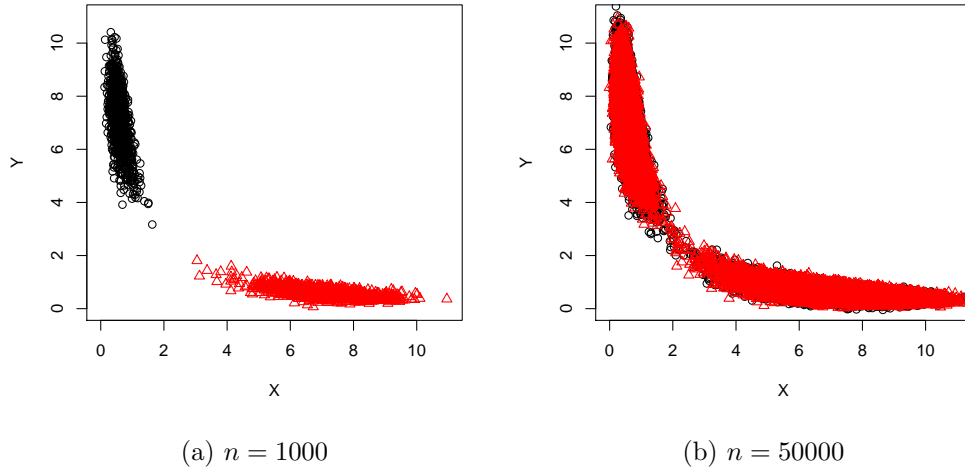


Figure 5: The scatter plots for two parallel Markov chains. (a): Both the chains are stuck in one of the mode (b): Both the chains have travelled and explored both the modes.

well with each other (Figure 4b). In Figure 8, we observe that both RSV and ASV give almost the same estimates for ESS. Similar is the case with coverage probabilities in Table 2 for a 95% confidence region.

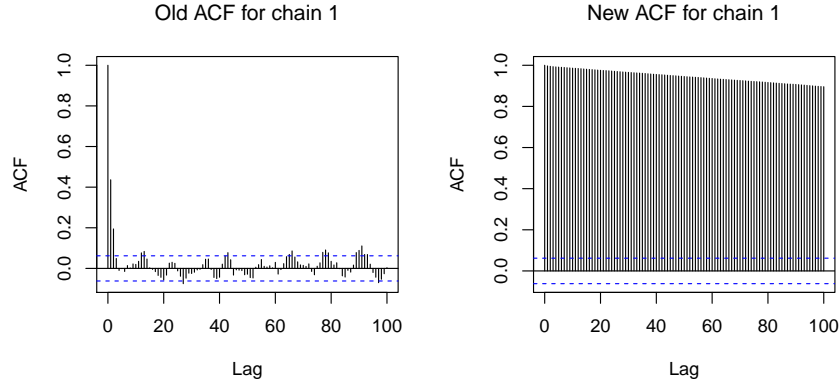
n	$m = 2$		$m = 5$	
	ASV	RSV	ASV	RSV
10^3	0.873	0.884	0.875	0.898
$2 \cdot 10^3$	0.881	0.888	0.897	0.914
$5 \cdot 10^3$	0.902	0.908	0.919	0.929
10^4	0.923	0.927	0.925	0.926
$5 \cdot 10^4$	0.951	0.952	0.939	0.939

Table 2: Coverage probabilities for parameter values $A = 1$, $B = 10$, $C = 7$.

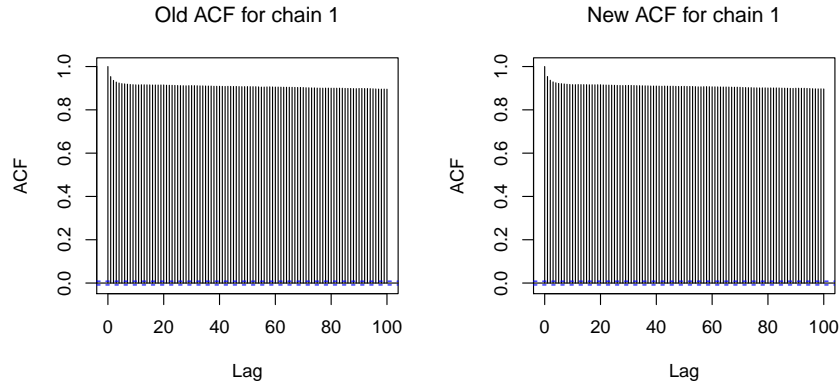
6 Discussion

A Preliminaries

Lemma 1. (*Csörgo and Révész (2014)*). Suppose Assumption 1 holds, then for all $\epsilon > 0$ and for almost all sample paths, there exists $n_0(\epsilon)$ such that $\forall n \geq n_0$ and $\forall i = 1, \dots, p$



(a) $n = 1000$



(b) $n = 50000$

Figure 6: ACF and R-ACF for component-1 of chain-1 at two different sample sizes.

$$\sup_{0 \leq t \leq n-b_n} \sup_{0 \leq s \leq b_n} \left| B^{(i)}(t+s) - B^{(i)}(t) \right| < (1+\epsilon) \left(2b_n \left(\log \frac{n}{b_n} + \log \log n \right) \right)^{1/2},$$

$$\sup_{0 \leq s \leq b_n} \left| B^{(i)}(n) - B^{(i)}(n-s) \right| < (1+\epsilon) \left(2b_n \left(\log \frac{n}{b_n} + \log \log n \right) \right)^{1/2}, \text{ and}$$

$$\left| B^{(i)}(n) \right| < (1+\epsilon) \sqrt{2n \log \log n}.$$

Lemma 2. *If Assumption 1 holds, then $\|\bar{X} - \mu\|_\infty, \|\bar{X}_s - \mu\|_\infty \xrightarrow{a.s.} 0$ as $n \rightarrow \infty$ for all $s \in \{1, \dots, m\}$*

Proof. Let $\|\cdot\|$ denote the Euclidean norm. Assumption 1 allows us to upper bound $\|\bar{X}_s - \mu\|_\infty$ using SIP. Following that, we show that the upper bound term converges to 0 as $n \rightarrow \infty$ using

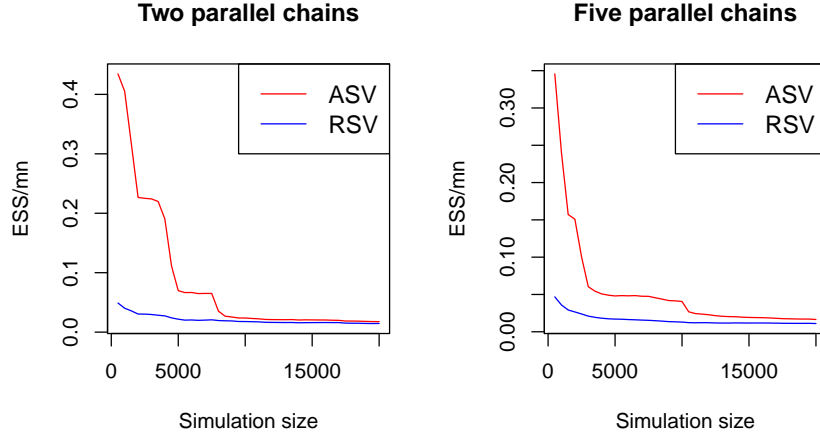


Figure 7: Running plot for $E\hat{S}S/mn$ using ASV and RSV calculated for two chains (left) and five chains (right). The parameters for target distribution are $A = 1$, $B = 3$, $C = 8$.

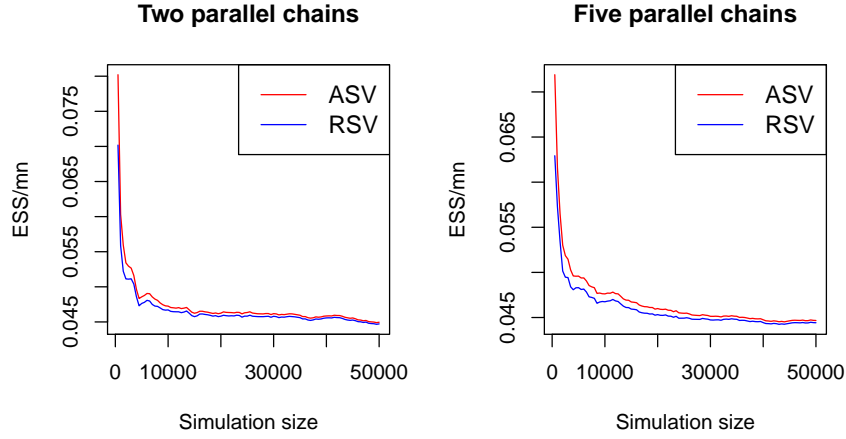


Figure 8: Running plot for $E\hat{S}S/mn$ using ASV and RSV calculated for two chains (left) and five chains (right). The parameters of target distribution are $A = 1$, $B = 10$, $C = 7$.

lemma 1.

$$\begin{aligned}
\|\bar{X}_s - \mu\|_\infty &\leq \|\bar{X}_s - \mu\| \\
&= \frac{1}{n} \left\| \sum_{t=1}^n X_{st} - n\mu \right\| = \frac{1}{n} \left\| \sum_{t=1}^n X_{st} - n\mu \pm LB(n) \right\| \\
&\leq \frac{1}{n} \left\| \sum_{t=1}^n X_{st} - \mu - LB(n) \right\| + \frac{\|LB(n)\|}{n} \\
&< \frac{D\psi(n)}{n} + \frac{\|LB(n)\|}{n} \\
&< \frac{D\psi(n)}{n} + \frac{1}{n} \|L\| \left(\sum_{i=1}^p |B^{(i)}(k)|^2 \right)^{1/2} \\
&\leq \frac{D\psi(n)}{n} + \frac{1}{n} \|L\| p^{1/2} (1 + \epsilon) \sqrt{2n \log \log n} \\
&\xrightarrow{a.s.} 0 \text{ as } n \rightarrow \infty
\end{aligned}$$

Similarly one can easily show that $\|\bar{\bar{X}} - \mu\|_\infty \rightarrow 0$ as $n \rightarrow \infty$ using the same upper-bound as:

$$\|\bar{\bar{X}} - \mu\|_\infty \leq \|\bar{\bar{X}} - \mu\| = \frac{1}{m} \left\| \sum_{j=1}^m (\bar{X}_j - \mu) \right\| \leq \frac{1}{m} \sum_{j=1}^m \|\bar{X}_j - \mu\| \xrightarrow{a.s.} 0 \text{ as } n \rightarrow \infty$$

□

B Proof of Theorems

B.1 Bias of autocovariance

Proof of Theorem 1.

$$\begin{aligned}
\hat{\Gamma}_{RAC,s}(k) &= \frac{1}{n} \sum_{t=1}^{n-|k|} \left(X_{st} - \bar{\bar{X}} \right) \left(X_{s(t+k)} - \bar{\bar{X}} \right)^T \\
&= \left[\frac{1}{n} \sum_{t=1}^{n-|k|} \left(X_{st} - \bar{X}_s \right) \left(X_{s(t+k)} - \bar{X}_s \right)^T \right] + \left[\frac{1}{n} \sum_{t=1}^{|k|} \left(\bar{X}_s - \bar{\bar{X}} \right) \left(\bar{X}_s - X_{st} \right)^T \right] \\
&\quad + \left[\frac{1}{n} \sum_{t=n-|k|+1}^n \left(\bar{X}_s - X_{st} \right) \left(\bar{X}_s - \bar{\bar{X}} \right)^T \right] + \left[\frac{n-|k|}{n} \left(\bar{X}_s - \bar{\bar{X}} \right) \left(\bar{X}_s - \bar{\bar{X}} \right)^T \right]
\end{aligned}$$

$$= \hat{\Gamma}_s(k) - \frac{1}{n} \sum_{t=1}^{|k|} A_{st}^T - \frac{1}{n} \sum_{t=n-|k|+1}^n A_{st} + \frac{n-|k|}{n} (\bar{X}_s - \bar{\bar{X}}) (\bar{X}_s - \bar{\bar{X}})^T, \quad (3)$$

where $A_{st} = (X_{st} - \bar{X}_s)(\bar{X}_s - \bar{\bar{X}})^T$. We will study the expectations of the each of the above terms. Without loss of generality, consider A_{11} ,

$$\begin{aligned} \mathbb{E}[A_{11}] &= \mathbb{E} \left[(X_{11} - \bar{X}_1) (\bar{X}_1 - \bar{\bar{X}})^T \right] \\ &= \mathbb{E} [X_{11} \bar{X}_1^T] - \mathbb{E} [X_{11} \bar{\bar{X}}^T] + \mathbb{E} [\bar{X}_1 \bar{\bar{X}}^T] - \mathbb{E} [\bar{X}_1 \bar{X}_1^T] \\ &= \mathbb{E} [X_{11} \bar{X}_1^T] - \frac{1}{m} \mathbb{E} [X_{11} \bar{X}_1^T] - \frac{m-1}{m} \mathbb{E} [X_{11} \bar{X}_2^T] + \frac{1}{m} \mathbb{E} [\bar{X}_1 \bar{X}_1^T] + \frac{m-1}{m} \mathbb{E} [\bar{X}_1 \bar{X}_2^T] - \mathbb{E} [\bar{X}_1 \bar{X}_1^T] \\ &= \frac{m-1}{m} \left(\mathbb{E} [X_{11} \bar{X}_1^T] - \mathbb{E} [X_{11} \bar{X}_2^T] + \mathbb{E} [\bar{X}_1 \bar{X}_2^T] - \mathbb{E} [\bar{X}_1 \bar{X}_1^T] \right) \\ &= \frac{m-1}{m} \left(\frac{1}{n} \sum_{t=1}^n \mathbb{E} [X_{11} X_{1t}^T] - \mathbb{E} [X_{11}] \mathbb{E} [\bar{X}_2^T] + \mathbb{E} [\bar{X}_1] \mathbb{E} [\bar{X}_2^T] - \text{Var} [\bar{X}_1] - \mathbb{E} [\bar{X}_1] \mathbb{E} [\bar{X}_1^T] \right) \\ &= \frac{m-1}{m} \left(\frac{1}{n} \sum_{k=0}^{n-1} \Gamma(k) + \mu \mu^T - \mu \mu^T + \mu \mu^T - \text{Var} [\bar{X}_1] - \mu \mu^T \right) \\ &= \frac{m-1}{mn} \left(\sum_{k=0}^{n-1} \Gamma(k) - n \text{Var} [\bar{X}_1] \right). \end{aligned} \quad (4)$$

Similarly,

$$\mathbb{E} [A_{11}^T] = \mathbb{E} [A_{11}]^T = \frac{m-1}{mn} \left(\sum_{k=0}^{n-1} \Gamma(k) - n \text{Var} [\bar{X}_1] \right)^T. \quad (5)$$

Further,

$$\begin{aligned} \mathbb{E} \left[(\bar{X}_s - \bar{\bar{X}}) (\bar{X}_s - \bar{\bar{X}})^T \right] &= \mathbb{E} [\bar{X}_s \bar{X}_s^T - \bar{X}_s \bar{\bar{X}}^T - \bar{\bar{X}} \bar{X}_s^T + \bar{\bar{X}} \bar{\bar{X}}^T] \\ &= \mathbb{E} [\bar{X}_s \bar{X}_s^T - \bar{\bar{X}} \bar{\bar{X}}^T] \\ &= \left(\mathbb{E} [\bar{X}_1 \bar{X}_1^T] - \mathbb{E} [\bar{\bar{X}} \bar{\bar{X}}^T] \right) \\ &= \left(\text{Var}(\bar{X}_1) + \mu \mu^T - \text{Var}(\bar{\bar{X}}) - \mu \mu^T \right) \\ &= \frac{m-1}{m} \text{Var}(\bar{X}_1). \end{aligned} \quad (6)$$

Using (1), (4), (5), and (6) in (3),

$$\begin{aligned}
& \mathbb{E} \left[\hat{\Gamma}_{RAC,s}(k) \right] \\
&= \mathbb{E} \left[\hat{\Gamma}_s(k) \right] - \frac{1}{n} \left(\sum_{t=1}^{|k|} \mathbb{E}[A_{1t}^T] + \sum_{t=n-|k|+1}^n \mathbb{E}[A_{1t}] \right) + \left(1 - \frac{|k|}{n} \right) \left(1 - \frac{1}{m} \right) \text{Var}(\bar{X}_1) \\
&= \mathbb{E} \left[\hat{\Gamma}_s(k) \right] - \frac{2|k|}{n} \frac{m-1}{mn} \left(\sum_{h=0}^{n-1} \Gamma(h) - n \text{Var}(\bar{X}_1) \right) + \left(1 - \frac{|k|}{n} \right) \left(1 - \frac{1}{m} \right) \text{Var}(\bar{X}_1).
\end{aligned}$$

DV: Instead of the “2”, should it be something that has $\Gamma(h) + \Gamma(h)^T$?

For a diagonal element of Γ ,

$$\begin{aligned}
& \mathbb{E} \left[\hat{\Gamma}_{RAC,s}^{ii} \right] \\
&= \mathbb{E} \left[\hat{\Gamma}_s^{ii}(k) \right] - \frac{2|k|}{n} \frac{m-1}{mn} \left(\sum_{h=0}^{n-1} \Gamma^{ii}(h) - n \text{Var}(\bar{X}_1)^{ii} \right) + \left(1 - \frac{|k|}{n} \right) \left(1 - \frac{1}{m} \right) \text{Var}(\bar{X}_1)^{ii}.
\end{aligned}$$

In the presence of positive correlation, the leftover term is positive. □

Proof of corollary 1.

$$\begin{aligned}
\hat{\Gamma}_{RAC}(0) &= \frac{1}{m} \sum_{s=1}^m \hat{\Gamma}_{RAC,s}(0) = \frac{1}{mn} \sum_{s=1}^m \sum_{t=1}^n (X_{st} - \bar{X}) (X_{st} - \bar{X})^T \\
&= \left[\frac{1}{mn} \sum_{s=1}^m \sum_{t=1}^n (X_{st} - \bar{X}_s) (X_{st} - \bar{X}_s)^T \right] + \left[\frac{1}{m} \sum_{s=1}^m (\bar{X}_s - \bar{X}) (\bar{X}_s - \bar{X})^T \right] \\
&= \frac{1}{m} \sum_{s=1}^m \hat{\Gamma}_s(0) + \frac{1}{m} \sum_{s=1}^m (\bar{X}_s - \bar{X}) (\bar{X}_s - \bar{X})^T.
\end{aligned}$$

Using (1) and (6) the expectation of $\hat{\Gamma}_{RAC}(0)$ can be written as

$$\begin{aligned}
\mathbb{E} \left[\hat{\Gamma}_{RAC}(0) \right] &= \mathbb{E} \left[\hat{\Gamma}_1(0) \right] + \left(1 - \frac{1}{m} \right) \text{Var}(\bar{X}_1) \\
&= \Gamma(k) - \frac{1}{m} \text{Var}(\bar{X}_1).
\end{aligned}$$
□

B.2 Strong consistency argument

We are going to consider a class of pseudo spectral variance estimator denoted by $\tilde{\Sigma}$ which uses data centered around the unobserved actual mean μ . Such pseudo covariance estimator and auto-

covariance estimator for the j^{th} chain is denoted by $\tilde{\Sigma}_s$ and $\tilde{\Gamma}_s$ respectively.

$$\begin{aligned}\tilde{\Gamma}_s(k) &= \frac{1}{n} \sum_{t=1}^{n-|k|} (X_{st} - \mu)(X_{s(t+k)} - \mu)^T \\ \tilde{\Sigma}_s &= \sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right) \tilde{\Gamma}_s(k).\end{aligned}$$

The average pseudo spectral variance estimator (APSV) is given by

$$\tilde{\Sigma} = \frac{1}{m} \sum_{s=1}^m \tilde{\Sigma}_s$$

Further, let

$$\begin{aligned}M_1 &= \frac{1}{m} \sum_{s=1}^m \left\{ \sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right) \sum_{t=1}^{n-|k|} \frac{1}{n} \left[(X_{st} - \mu)_i (\mu - \bar{X})_j + (\mu - \bar{X})_i (X_{s(t+k)} - \mu)_j \right] \right\}, \\ M_2 &= (\mu - \bar{X})_i (\mu - \bar{X})_j \sum_{k=-b_n+1}^{b_n-1} \left(1 - \frac{|k|}{n}\right) w\left(\frac{k}{b_n}\right).\end{aligned}$$

Lemma 3. For the RSV estimator $\hat{\Sigma}_{RSV}^{ij} = \tilde{\Sigma}^{ij} + M_1 + M_2$ and

$$|M_1 + M_2| \leq D^2 g_1(n) + D g_2(n) + g_3(n)$$

where

$$\begin{aligned}g_1(n) &= (4 + C_1) \frac{b_n \psi^2(n)}{n^2} - 4 \frac{\psi^2(n)}{n^2} \rightarrow 0 \\ g_2(n) &= 2\sqrt{2} \|L\| p^{1/2} (1 + \epsilon) \left[(4 + C_1) \frac{b_n \psi(n) \sqrt{n \log \log n}}{n^2} - 4 \frac{\psi(n) \sqrt{n \log \log n}}{n^2} \right] \rightarrow 0 \\ g_3(n) &= \|L\|^2 p (1 + \epsilon)^2 \left[(4 + C_1) \frac{b_n \log \log n}{n} - 4 \frac{\log \log n}{n} \right] \rightarrow 0.\end{aligned}$$

Proof. The proof follows from standard algebraic calculations and is presented here for completeness. Consider,

$$\begin{aligned}\hat{\Sigma}_{RSV}^{ij} &= \frac{1}{m} \sum_{s=1}^m \sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right) \frac{1}{n} \sum_{t=1}^{n-|k|} (X_{st} - \bar{X})_i (X_{s(t+k)} - \bar{X})_j\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{m} \sum_{s=1}^m \sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right) \frac{1}{n} \sum_{t=1}^{n-|k|} \left[\left\{ (X_{st} - \mu)_i + (\mu - \bar{X})_i \right\} \left\{ (X_{s(t+k)} - \mu)_j + (\mu - \bar{X})_j \right\} \right] \\
&= \frac{1}{m} \sum_{s=1}^m \sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right) \frac{1}{n} \sum_{t=1}^{n-|k|} \left[(X_{st} - \mu)_i (X_{s(t+k)} - \mu)_j + (X_{st} - \mu)_i (\mu - \bar{X})_j \right. \\
&\quad \left. + (\mu - \bar{X})_i (X_{s(t+k)} - \mu)_j + (\mu - \bar{X})_i (\mu - \bar{X})_j \right] \\
&= \tilde{\Sigma}^{ij} + \left[(\mu - \bar{X})_i (\mu - \bar{X})_j \sum_{k=-b_n+1}^{b_n-1} \left(1 - \frac{|k|}{n}\right) w\left(\frac{k}{b_n}\right) \right] \\
&\quad + \frac{1}{m} \sum_{s=1}^m \sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right) \sum_{t=1}^{n-|k|} \left[\frac{1}{n} (X_{st} - \mu)_i (\mu - \bar{X})_j + \frac{1}{n} (\mu - \bar{X})_i (X_{s(t+k)} - \mu)_j \right] \\
&= \tilde{\Sigma}^{ij} + M_1 + M_2.
\end{aligned}$$

Consequently

$$|\hat{\Sigma}_{RSV}^{ij} - \tilde{\Sigma}^{ij}| = |M_1 + M_2| \leq |M_1| + |M_2|.$$

Consider,

$$\begin{aligned}
&|M_1| \\
&= \left| \frac{1}{m} \sum_{s=1}^m \left\{ \sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right) \sum_{t=1}^{n-|k|} \frac{1}{n} \left[(X_{st} - \mu)_i (\mu - \bar{X})_j + (\mu - \bar{X})_i (X_{s(t+k)} - \mu)_j \right] \right\} \right| \\
&\leq \frac{1}{m} \sum_{s=1}^m \left\{ \sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right) \left[\frac{1}{n} \left| \sum_{t=1}^{n-|k|} (X_{st} - \mu)_i (\mu - \bar{X})_j \right| + \frac{1}{n} \left| \sum_{t=1}^{n-|k|} (\mu - \bar{X})_i (X_{s(t+k)} - \mu)_j \right| \right] \right\} \\
&\leq \frac{1}{m} \sum_{s=1}^m \left\{ \sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right) \left[\frac{1}{n} \left| \sum_{t=1}^{n-|k|} (X_{st} - \mu)_i \right| |(\mu - \bar{X})_j| + \frac{1}{n} |(\mu - \bar{X})_i| \left| \sum_{t=1}^{n-|k|} (X_{s(t+k)} - \mu)_j \right| \right] \right\} \\
&\leq \frac{\|(\bar{X} - \mu)\|_\infty}{m} \sum_{s=1}^m \sum_{k=-b_n+1}^{b_n-1} \left[\frac{1}{n} \left\| \sum_{t=1}^{n-|k|} (X_{st} - \mu) \right\|_\infty + \frac{1}{n} \left\| \sum_{t=1}^{n-|k|} (X_{s(t+k)} - \mu) \right\|_\infty \right] \\
&\leq \frac{\|(\bar{X} - \mu)\|_\infty}{m} \\
&\quad \times \sum_{s=1}^m \sum_{k=-b_n+1}^{b_n-1} \left[\frac{1}{n} \left\| \sum_{t=n-|k|+1}^n (X_{st} - \mu) - n(\bar{X}_s - \mu) \right\|_\infty + \frac{1}{n} \left\| \sum_{t=1}^{|k|} (X_{st} - \mu) - n(\bar{X}_s - \mu) \right\|_\infty \right]
\end{aligned}$$

$$\begin{aligned}
&\leq \frac{\|(\bar{\bar{X}} - \mu)\|_\infty}{m} \sum_{s=1}^m \sum_{k=-b_n+1}^{b_n-1} \left[\frac{1}{n} \left\| \sum_{t=n-|k|+1}^n (X_{st} - \mu) \right\|_\infty + \frac{1}{n} \left\| \sum_{t=1}^{|k|} (X_{st} - \mu) \right\|_\infty + 2\|\bar{X}_s - \mu\|_\infty \right] \\
&\leq \frac{\|(\bar{\bar{X}} - \mu)\|_\infty}{m} \sum_{s=1}^m \sum_{k=-b_n+1}^{b_n-1} \frac{1}{n} \left[\left\| \sum_{t=n-|k|+1}^n (X_{st} - \mu) \right\|_\infty + \left\| \sum_{t=1}^{|k|} (X_{st} - \mu) \right\|_\infty \right] \\
&\quad + 2(2b_n - 1)\|\bar{\bar{X}} - \mu\|_\infty \|\bar{X}_1 - \mu\|_\infty
\end{aligned}$$

Using SIP on summation of k terms, we obtain the following upper bound for $|M_1|$

$$\begin{aligned}
|M_1| &< 2\|(\bar{\bar{X}} - \mu)\|_\infty \sum_{k=-b_n+1}^{b_n-1} \left[\frac{D\psi(k)}{n} + \frac{\|L\|p^{1/2}(1+\epsilon)\sqrt{2k \log \log k}}{n} \right] \\
&\quad + 2(2b_n - 1)\|\bar{\bar{X}} - \mu\|_\infty \|\bar{X}_1 - \mu\|_\infty \\
&\leq 2(2b_n - 1)\|(\bar{\bar{X}} - \mu)\|_\infty \left[\frac{D\psi(n)}{n} + \frac{\|L\|p^{1/2}(1+\epsilon)\sqrt{n \log \log n}}{n} \right] \\
&\quad + 2(2b_n - 1)\|\bar{\bar{X}} - \mu\|_\infty \|\bar{X}_1 - \mu\|_\infty.
\end{aligned}$$

We use Lemma 2 to upper bound the above term as

$$4(2b_n - 1) \left[\frac{D\psi(n)}{n} + \frac{\|L\|p^{1/2}(1+\epsilon)\sqrt{n \log \log n}}{n} \right]^2 \quad (7)$$

Using the bound on $\psi(n)$ given by Strassen (1964), equation (7) becomes $\mathcal{O}(b_n \log \log n/n)$ which converges to 0 as $n \rightarrow \infty$. Now we will show that $M_2 \xrightarrow{a.s.} 0$ as $n \rightarrow \infty$

$$\begin{aligned}
|M_2| &= \left| \frac{1}{m} \sum_{s=1}^m \left\{ \left(\mu - \bar{\bar{X}} \right)_i \left(\mu - \bar{\bar{X}} \right)_j \sum_{k=-b_n+1}^{b_n-1} \left(1 - \frac{|k|}{n} \right) w \left(\frac{k}{b_n} \right) \right\} \right| \\
&\leq \|\bar{\bar{X}} - \mu\|_\infty^2 \left[\sum_{k=-b_n+1}^{b_n-1} \left(1 - \frac{|k|}{n} \right) w \left(\frac{k}{b_n} \right) \right] \\
&< \|\bar{\bar{X}} - \mu\|_\infty^2 \left[\sum_{k=-b_n+1}^{b_n-1} \left| w \left(\frac{k}{b_n} \right) \right| \right] \\
&= b_n \|\bar{\bar{X}} - \mu\|_\infty^2 \left[\frac{1}{b_n} \sum_{k=-b_n+1}^{b_n-1} \left| w \left(\frac{k}{b_n} \right) \right| \right]
\end{aligned}$$

$$\begin{aligned}
&\leq b_n \|\bar{\bar{X}} - \mu\|_\infty^2 \int_{-\infty}^{\infty} |w(x)| dx \\
&\leq C b_n \|\bar{\bar{X}} - \mu\|^2
\end{aligned}$$

Using Lemma 2 to upper bound the above term as

$$C b_n \left[\frac{D\psi(n)}{n} + \frac{\|L\| p^{1/2} (1+\epsilon) \sqrt{n \log \log n}}{n} \right]^2 \quad (8)$$

Complete the algebra here

$$\begin{aligned}
g_1(n) &= (4 + C_1) \frac{b_n \psi^2(n)}{n^2} - 4 \frac{\psi^2(n)}{n^2} \rightarrow 0 \\
g_2(n) &= 2\sqrt{2} \|L\| p^{1/2} (1+\epsilon) \left[(4 + C_1) \frac{b_n \psi(n) \sqrt{n \log \log n}}{n^2} - 4 \frac{\psi(n) \sqrt{n \log \log n}}{n^2} \right] \rightarrow 0 \\
g_3(n) &= \|L\|^2 p (1+\epsilon)^2 \left[(4 + C_1) \frac{b_n \log \log n}{n} - 4 \frac{\log \log n}{n} \right] \rightarrow 0.
\end{aligned}$$

□

Proof of theorem 2. We have the following decomposition,

$$\begin{aligned}
&\tilde{\Sigma}^{ij} \\
&= \frac{1}{m} \sum_{s=1}^m \sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right) \frac{1}{n} \sum_{t=1}^{n-|k|} (X_{st} - \mu)_i (X_{s(t+k)} - \mu)_j \\
&= \frac{1}{m} \sum_{s=1}^m \sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right) \frac{1}{n} \sum_{t=1}^{n-|k|} (X_{st} \pm \bar{X}_s - \mu)_i (X_{s(t+k)} \pm \bar{X}_s - \mu)_j \\
&= \hat{\Sigma}_{SV}^{ij} + \frac{1}{m} \sum_{s=1}^m \sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right) \frac{1}{n} \sum_{t=1}^{n-|k|} \left[(X_{st} - \bar{X}_s)_i (\bar{X}_s - \mu)_j + (\bar{X}_s - \mu)_i (X_{s(t+k)} - \bar{X}_s)_j \right] \\
&\quad + \left[\frac{1}{m} \sum_{s=1}^m (\bar{X}_s - \mu)_i (\bar{X}_s - \mu)_j \right] \left[\sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right) \left(1 - \frac{|k|}{b_n}\right) \right] \\
&= \hat{\Sigma}_{SV}^{ij} + N_1 + N_2,
\end{aligned}$$

where

$$N_1 = \frac{1}{m} \sum_{s=1}^m \sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right) \frac{1}{n} \sum_{t=1}^{n-|k|} \left[(X_{st} - \bar{X}_s)_i (\bar{X}_s - \mu)_j + (\bar{X}_s - \mu)_i (X_{s(t+k)} - \bar{X}_s)_j \right]$$

$$N_2 = \left[\frac{1}{m} \sum_{s=1}^m (\bar{X}_s - \mu)_i (\bar{X}_s - \mu)_j \right] \left[\sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right) \left(1 - \frac{|k|}{b_n}\right) \right].$$

Using the above and Lemma 3,

$$\begin{aligned} \left| \hat{\Sigma}_{RSV}^{ij} - \Sigma^{ij} \right| &= \left| \hat{\Sigma}_{SV}^{ij} - \Sigma^{ij} + N_1 + N_2 + M_1 + M_2 \right| \\ &\leq \left| \hat{\Sigma}_{SV}^{ij} - \Sigma^{ij} \right| + |N_1| + |N_2| + |M_1 + M_2| \end{aligned}$$

By Assumption 2, the first term goes to with probability 1 and by Lemma 3, the third term goes to 0 with probability 1 as $n \rightarrow \infty$. What is left to show is that $|N_1| \rightarrow 0$ and $|N_2| \rightarrow 0$ with probability 1

$$\begin{aligned} |N_1| &= \left| \frac{1}{m} \sum_{s=1}^m \sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right) \frac{1}{n} \sum_{t=1}^{n-|k|} \left[(X_{st} - \bar{X}_s)_i (\bar{X}_s - \mu)_j + (\bar{X}_s - \mu)_i (X_{s(t+k)} - \bar{X}_s)_j \right] \right| \\ &\leq \left| \frac{1}{m} \sum_{s=1}^m \sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right) \frac{1}{n} \sum_{t=1}^{n-|k|} \left[(X_{st} - \bar{X}_s)_i (\bar{X}_s - \mu)_j \right] \right| \\ &\quad + \left| \frac{1}{m} \sum_{s=1}^m \sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right) \frac{1}{n} \sum_{t=1}^{n-|k|} \left[(\bar{X}_s - \mu)_i (X_{s(t+k)} - \bar{X}_s)_j \right] \right| \end{aligned}$$

Both terms are similar and we will show that the first term goes to 0. The proof for the second term is similar. Consider

$$\begin{aligned} &\left| \frac{1}{m} \sum_{s=1}^m \sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right) \frac{1}{n} \sum_{t=1}^{n-|k|} \left[(X_{st} - \bar{X}_s)_i (\bar{X}_s - \mu)_j \right] \right| \\ &\leq \frac{1}{m} \sum_{s=1}^m \sum_{k=-b_n+1}^{b_n-1} \left| w\left(\frac{k}{b_n}\right) \right| \frac{|(\bar{X}_s - \mu)_j|}{n} \left[\sum_{t=1}^{|k|} |(\bar{X}_s - X_{st})_i| \right] \\ &\leq \frac{1}{m} \sum_{s=1}^m \sum_{k=-b_n+1}^{b_n-1} \left| w\left(\frac{k}{b_n}\right) \right| \frac{|(\bar{X}_s - \mu)_j|}{n} \left[\sum_{t=1}^{|k|} |\mu - X_{st}|_i + |k| |(\bar{X}_s - \mu)_i| \right] \\ &\leq \frac{1}{m} \sum_{s=1}^m \sum_{k=-b_n+1}^{b_n-1} \left| w\left(\frac{k}{b_n}\right) \right| \frac{\|\bar{X}_s - \mu\|_\infty}{n} \left\| \sum_{t=1}^{|k|} (\mu - X_{st}) + |k| (\bar{X}_s - \mu) \right\|_\infty \\ &\leq \frac{1}{m} \sum_{s=1}^m \sum_{k=-b_n+1}^{b_n-1} \left| w\left(\frac{k}{b_n}\right) \right| \frac{\|\bar{X}_s - \mu\|_\infty}{n} \left(\left\| \sum_{t=1}^{|k|} (X_{st} - \mu) \right\|_\infty + |k| \|\bar{X}_s - \mu\|_\infty \right) \end{aligned}$$

$$\leq \frac{1}{m} \sum_{s=1}^m \sum_{k=-b_n+1}^{b_n-1} \frac{\|\bar{X}_s - \mu\|_\infty}{n} \left\| \sum_{t=1}^{|k|} (X_{st} - \mu) \right\| + \frac{1}{m} \sum_{s=1}^m \frac{b_n(b_n-1)}{n} \|\bar{X}_s - \mu\|_\infty^2.$$

Using SIP on the summation of k terms,

$$\begin{aligned} & \left| \frac{1}{m} \sum_{s=1}^m \sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right) \frac{1}{n} \sum_{t=1}^{n-|k|} \left[(X_{st} - \bar{X}_s)_i (\bar{X}_s - \mu)_j \right] \right| \\ & < \frac{1}{m} \sum_{s=1}^m \|\bar{X}_s - \mu\|_\infty \sum_{k=-b_n+1}^{b_n-1} \left[\frac{D\psi(k)}{n} + \frac{\|L\|p^{1/2}(1+\epsilon)\sqrt{2k\log\log k}}{n} \right] + \frac{1}{m} \sum_{s=1}^m \frac{b_n(b_n-1)}{n} \|\bar{X}_s - \mu\|_\infty^2 \\ & < \frac{(2b_n-1)}{m} \sum_{s=1}^m \|\bar{X}_s - \mu\|_\infty \left[\frac{D\psi(n)}{n} + \frac{\|L\|p^{1/2}(1+\epsilon)\sqrt{2n\log\log n}}{n} \right] + \frac{1}{m} \sum_{s=1}^m \frac{b_n(b_n-1)}{n} \|\bar{X}_s - \mu\|_\infty^2 \end{aligned}$$

Using Lemma 2, the above term can be upper bounded by,

$$\left(2b_n - 1 + \frac{b_n^2}{n} - \frac{b_n}{n} \right) \left[\frac{D\psi(n)}{n} + \frac{\|L\|p^{1/2}(1+\epsilon)\sqrt{2n\log\log n}}{n} \right]^2. \quad (9)$$

Using the bound on $\psi(n)$ given by Strassen (1964), (9) becomes $\mathcal{O}(b_n \log \log n / n)$ which converges to 0 as $n \rightarrow \infty$.

Similar to the above proof, second part of N_1 can also be written as a summation of k terms and the proof follows exactly as the first part. Now we prove that $N_2 \xrightarrow{a.s.} 0$ as $n \rightarrow \infty$. **Remove the following proof maybe because repetitive**

$$\begin{aligned} |N_2| &= \left| \frac{1}{m} \sum_{s=1}^m (\bar{X}_s - \mu)_i (\bar{X}_s - \mu)_j \sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right) \left(1 - \frac{|k|}{b_n}\right) \right| \\ &< \frac{1}{m} \sum_{s=1}^m \|\bar{X}_s - \mu\|_\infty^2 \sum_{k=-b_n+1}^{b_n-1} \left| w\left(\frac{k}{b_n}\right) \right| \\ &= \frac{b_n}{m} \sum_{s=1}^m \|\bar{X}_s - \mu\|_\infty^2 \frac{1}{b_n} \sum_{k=-b_n+1}^{b_n-1} \left| w\left(\frac{k}{b_n}\right) \right| \\ &\leq \frac{b_n}{m} \sum_{s=1}^m \|\bar{X}_s - \mu\|_\infty^2 \int_{-\infty}^{\infty} |w(x)| dx. \end{aligned}$$

Recall that the lag window is such that $\int_{-\infty}^{\infty} |w(x)| dx = C$ for some constant C . Using Lemma 2, N_2 can be upper bounded by

$$Cb_n \left[\frac{D\psi(n)}{n} + \frac{\|L\|p^{1/2}(1+\epsilon)\sqrt{2n\log\log n}}{n} \right]^2 \quad (10)$$

(10) is also $\mathcal{O}(b_n \log \log n/n)$, and hence converges to 0 as $n \rightarrow \infty$. \square

B.3 RSV Bias

Proof of theorem 3. The convoluted end effect terms in theorem 1 are all essentially $\mathcal{O}(1/n)$. We are interested in finding the asymptotic bias here. Therefore we can write the expectation of replicated autocovariance as

$$\mathbb{E} \left[\hat{\Gamma}_{RAC}(k) \right] = \left(1 - \frac{|k|}{n} \right) \Gamma(k) + \left(1 - \frac{|k|}{n} \right) \mathcal{O} \left(\frac{1}{n} \right) + \mathcal{O} \left(\frac{1}{n} \right). \quad (11)$$

Since $w \left(\frac{k}{b_n} \right) = 0$ for $|k| > b_n$. The last term in (11) is stated separately because it does not depend on k .

$$\begin{aligned} & \mathbb{E} \left[\hat{\Sigma}_{RSV} - \Sigma \right] \\ &= \sum_{k=-b_n+1}^{b_n-1} w \left(\frac{|k|}{b_n} \right) \mathbb{E} \left[\hat{\Gamma}_{RAC}(k) \right] - \sum_{k=-\infty}^{\infty} \Gamma(k) \\ &= \sum_{k=-b_n+1}^{b_n-1} w \left(\frac{|k|}{b_n} \right) \left(1 - \frac{|k|}{n} \right) \left[\left(\Gamma(k) + \mathcal{O} \left(\frac{1}{n} \right) \right) + \mathcal{O} \left(\frac{1}{n} \right) \right] - \sum_{k=-\infty}^{\infty} \Gamma(k) \\ &= \sum_{k=-b_n+1}^{b_n-1} w \left(\frac{|k|}{b_n} \right) \left(1 - \frac{|k|}{n} \right) \Gamma(k) - \sum_{k=-\infty}^{\infty} \Gamma(k) \\ &\quad + \sum_{k=-b_n+1}^{b_n-1} w \left(\frac{|k|}{b_n} \right) \left[\left(1 - \frac{|k|}{n} \right) \mathcal{O} \left(\frac{1}{n} \right) + \mathcal{O} \left(\frac{1}{n} \right) \right] \\ &= O_1 + O_2, \end{aligned}$$

where

$$\begin{aligned} O_1 &= \sum_{k=-n+1}^{n-1} \left\{ w \left(\frac{|k|}{b_n} \right) \left(1 - \frac{|k|}{n} \right) \Gamma(k) \right\} - \sum_{k=-\infty}^{\infty} \Gamma(k) \\ O_2 &= \sum_{k=-n+1}^{n-1} \left\{ w \left(\frac{|k|}{b_n} \right) \left[\left(1 - \frac{|k|}{n} \right) \mathcal{O} \left(\frac{1}{n} \right) + \mathcal{O} \left(\frac{1}{n} \right) \right] \right\}. \end{aligned}$$

1. We first solve for term O_1 by breaking it into three parts as in Hannan (2009).

$$- \sum_{|k| \geq n} \Gamma(k) + \sum_{k=-n+1}^{n-1} w \left(\frac{|k|}{b_n} \right) \frac{|k|}{n} \Gamma(k) - \sum_{k=-n+1}^{n-1} \left(1 - w \left(\frac{|k|}{b_n} \right) \right) \Gamma(k)$$

We deal with the three subterms of term **A** individually,

(a)

$$\left\| - \sum_{|k| \geq n} \Gamma(k) \right\| \leq \left| \frac{k}{n} \right|^q \sum_{|k| \geq n} \|\Gamma(k)\| = \frac{1}{b_n^q} \left| \frac{b_n}{n} \right|^q \underbrace{\sum_{|k| \geq n} |k|^q \|\Gamma(k)\|}_{< \infty} = o\left(\frac{1}{b_n^q}\right)$$

(b)

$$\left\| \sum_{k=-n+1}^{n-1} w\left(\frac{|k|}{n}\right) \frac{|k|}{n} \Gamma(k) \right\| \leq \frac{c}{n} \sum_{k=-n+1}^{n-1} |k| \|\Gamma(k)\|$$

For $q \geq 1$, we have, $\sum_{k=-n+1}^{n-1} |k|^q \|\Gamma(k)\| < \infty$

$$\frac{c}{n} \sum_{k=-n+1}^{n-1} |k| \|\Gamma(k)\| \leq \frac{c}{n} \sum_{k=-n+1}^{n-1} |k|^q \|\Gamma(k)\| = \frac{1}{b_n^q} \frac{b_n^q}{n} c \sum_{k=-n+1}^{n-1} |k|^q \|\Gamma(k)\| = o\left(\frac{1}{b_n^q}\right)$$

For $q < 1$:

$$c \sum_{k=-n+1}^{n-1} \frac{|k|}{n} \|\Gamma(k)\| \leq c \sum_{k=-n+1}^{n-1} \left| \frac{k}{n} \right|^q \|\Gamma(k)\| = \frac{1}{b_n^q} \frac{b_n^q}{n} c \sum_{k=-n+1}^{n-1} |k|^q \|\Gamma(k)\| = o\left(\frac{1}{b_n^q}\right)$$

(c) Next,

$$\sum_{k=-n+1}^{n-1} \left(1 - w\left(\frac{|k|}{b_n}\right)\right) \Gamma(k) = -\frac{1}{b_n^q} \sum_{k=-n+1}^{n-1} \left(\frac{|k|}{b_n}\right)^{-q} \left(1 - w\left(\frac{|k|}{b_n}\right)\right) |k|^q \Gamma(k).$$

The condition

$$\lim_{|x| \rightarrow 0} \frac{1 - w(x)}{|x|^q} = k_q < \infty$$

implies that $|k/b_n|^{-q} (1 - w(k/b_n))$ converges boundedly to k_q for each k ; where k_q is specific to the lag-window used. Taking the limiting value, we have

$$\lim_{n \rightarrow \infty} \sum_{k=-n+1}^{n-1} \left(1 - w\left(\frac{|k|}{b_n}\right)\right) \Gamma(k) = -\frac{k_q \Phi^{(q)}}{b_n^q}$$

2. Solving for O_2 . We will find an upper bound and prove that it is $\mathcal{O}(n^{-1})$:

$$\begin{aligned}
& \sum_{k=-n+1}^{n-1} \left\{ w\left(\frac{|k|}{b_n}\right) \left[\left(1 - \frac{|k|}{n}\right) \mathcal{O}\left(\frac{1}{n}\right) + \mathcal{O}\left(\frac{1}{n}\right) \right] \right\} \\
& \leq \sum_{k=-n+1}^{n-1} w\left(\frac{|k|}{b_n}\right) \mathcal{O}\left(\frac{1}{n}\right) \\
& = \mathcal{O}\left(\frac{1}{n}\right) \sum_{k=-n+1}^{n-1} w\left(\frac{|k|}{b_n}\right) \\
& = \mathcal{O}\left(\frac{1}{n}\right) = o\left(\frac{1}{b_n^q}\right).
\end{aligned}$$

We have shown that all, but one, terms are $o(1/b_n^q)$. Therefore, the limiting value of $b_n^q \mathbb{E}[\hat{\Sigma}_{RAC} - \Sigma]$ is proved to be same as that of $b_n^q \mathbb{E}[\tilde{\Sigma} - \Sigma]$ in Hannan (2009) with some extra $o(1)$ terms. This is quite intuitive as we are using a strongly consistent estimator $\bar{\bar{X}}$ for μ in our replicated autovariance estimator. Similar to theorem 1, the effect of estimating μ contributes only $o(1)$ terms. \square

B.4 Variance Calculation

Theorem 4

Proof. Due to the strong consistency proof from theorem 2, as $n \rightarrow \infty$,

$$|\hat{\Sigma}_{RSV} - \tilde{\Sigma}| \rightarrow 0 \text{ with probability } 1. \quad (12)$$

Further, we have defined $g_1(n), g_2(n), g_3(n)$ such that as $n \rightarrow \infty$,

$$\begin{aligned}
g_1(n) &= (4 + C_1) \frac{b_n \psi^2(n)}{n^2} - 4 \frac{\psi^2(n)}{n^2} \rightarrow 0 \\
g_2(n) &= 2\sqrt{2} \|L\| p^{1/2} (1 + \epsilon) \left[(4 + C_1) \frac{b_n \psi(n) \sqrt{n \log \log n}}{n^2} - 4 \frac{\psi(n) \sqrt{n \log \log n}}{n^2} \right] \rightarrow 0 \\
g_3(n) &= \|L\|^2 p (1 + \epsilon)^2 \left[(4 + C_1) \frac{b_n \log \log n}{n} - 4 \frac{\log \log n}{n} \right] \rightarrow 0
\end{aligned}$$

We have shown from the proof of strong consistency that,

$$\begin{aligned}
& |\hat{\Sigma}_{RSV}^{ij} - \tilde{\Sigma}^{ij}| \\
& \leq \frac{1}{m} \sum_{s=1}^m \left| \sum_{k=-b_n+1}^{b_n-1} w\left(\frac{k}{b_n}\right) \sum_{t=1}^{n-|k|} \left[\left(\frac{(X_{st} - \mu)_i (\mu - \bar{\bar{X}})_j}{n} \right) + \left(\frac{(\mu - \bar{\bar{X}})_i (X_{s(t+k)} - \mu)_j}{n} \right) \right] \right|
\end{aligned}$$

$$\begin{aligned}
& +(\mu - \bar{X})(\mu - \bar{X})^T \sum_{k=-b_n+1}^{b_n-1} \left(\frac{n-|k|}{n} \right) w \left(\frac{k}{n} \right) \Big| \\
& < D^2 g_1(n) + D g_2(n) + g_3(n).
\end{aligned}$$

By (12), there exists an N_0 such that

$$\begin{aligned}
\left(\hat{\Sigma}_{RSV}^{ij} - \tilde{\Sigma}^{ij} \right)^2 &= \left(\hat{\Sigma}_{RSV}^{ij} - \tilde{\Sigma}^{ij} \right)^2 I(0 \leq n \leq N_0) + \left(\hat{\Sigma}_{RSV}^{ij} - \tilde{\Sigma}^{ij} \right)^2 I(n > N_0) \\
&\leq \left(\hat{\Sigma}_{RSV}^{ij} - \tilde{\Sigma}^{ij} \right)^2 I(0 \leq n \leq N_0) + \left(D^2 g_1(n) + D g_2(n) + g_3(n) \right)^2 I(n > N_0) \\
&:= g_n^*(X_{11}, \dots, X_{1n}, \dots, X_{m1}, \dots, X_{mn}).
\end{aligned}$$

But since by assumption $\mathbb{E}D^4 < \infty$ and the fourth moment is finite,

$$\mathbb{E} |g_n^*| \leq \mathbb{E} \left[\left(\hat{\Sigma}_{RSV}^{ij} - \tilde{\Sigma}^{ij} \right)^2 \right] + \mathbb{E} \left[\left(D^2 g_1(n) + D g_2(n) + g_3(n) \right)^2 \right] < \infty.$$

Thus, $\mathbb{E} |g_n^*| < \infty$ and further as $n \rightarrow \infty$, $g_n \rightarrow 0$ under the assumptions. Since $g_1, g_2, g_3 \rightarrow 0$, $\mathbb{E}g_n^* \rightarrow 0$. By the majorized convergence theorem (Zeidler, 2013), as $n \rightarrow \infty$,

$$\mathbb{E} \left[\left(\hat{\Sigma}_{RSV}^{ij} - \tilde{\Sigma}^{ij} \right)^2 \right] \rightarrow 0. \tag{13}$$

We will use (13) to show that the variances are equivalent. Define,

$$\xi \left(\hat{\Sigma}_{RSV}^{ij}, \tilde{\Sigma}^{ij} \right) = \text{Var} \left(\hat{\Sigma}_{RSV}^{ij} - \tilde{\Sigma}^{ij} \right) + 2\mathbb{E} \left[\left(\hat{\Sigma}_{RSV}^{ij} - \tilde{\Sigma}^{ij} \right) \left(\tilde{\Sigma}^{ij} - \mathbb{E} \left(\tilde{\Sigma}^{ij} \right) \right) \right]$$

We will show that the above is $o(1)$. Using Cauchy-Schwarz inequality followed by (13),

$$\begin{aligned}
\left| \xi \left(\hat{\Sigma}_{RSV}^{ij}, \tilde{\Sigma}^{ij} \right) \right| &\leq \left| \text{Var} \left(\hat{\Sigma}_{RSV}^{ij} - \tilde{\Sigma}^{ij} \right) \right| + \left| 2\mathbb{E} \left[\left(\hat{\Sigma}_{RSV}^{ij} - \tilde{\Sigma}^{ij} \right) \left(\tilde{\Sigma}^{ij} - \mathbb{E} \left(\tilde{\Sigma}^{ij} \right) \right) \right] \right| \\
&\leq \mathbb{E} \left[\left(\hat{\Sigma}_{RSV}^{ij} - \tilde{\Sigma}^{ij} \right)^2 \right] + 2 \left| \left(\mathbb{E} \left[\left(\hat{\Sigma}_{RSV}^{ij} - \tilde{\Sigma}^{ij} \right)^2 \right] \text{Var} \left(\tilde{\Sigma}^{ij} \right) \right)^{1/2} \right| \\
&= o(1) + 2 \left(o(1) \left(O \left(\frac{b_n}{n} \right) + o \left(\frac{b_n}{n} \right) \right) \right) \\
&= o(1).
\end{aligned}$$

Finally,

$$\begin{aligned}
& \text{Var} \left(\hat{\Sigma}_{RSV}^{ij} \right) \\
&= \mathbb{E} \left[\left(\hat{\Sigma}_{RSV}^{ij} - \mathbb{E} \left[\hat{\Sigma}_R^{ij} \right] \right)^2 \right] \\
&= \mathbb{E} \left[\left(\hat{\Sigma}_{RSV}^{ij} \pm \tilde{\Sigma}^{ij} \pm \mathbb{E} \left[\tilde{\Sigma}^{ij} \right] - \mathbb{E} \left[\hat{\Sigma}_{RSV}^{ij} \right] \right)^2 \right] \\
&= \mathbb{E} \left[\left(\left(\hat{\Sigma}_{RSV}^{ij} - \tilde{\Sigma}^{ij} \right) + \left(\tilde{\Sigma}^{ij} - \mathbb{E} \left[\tilde{\Sigma}^{ij} \right] \right) + \left(\mathbb{E} \left[\tilde{\Sigma}^{ij} \right] - \mathbb{E} \left[\hat{\Sigma}_{RSV}^{ij} \right] \right) \right)^2 \right] \\
&= \mathbb{E} \left[\left(\tilde{\Sigma}^{ij} - \mathbb{E} \left[\tilde{\Sigma}^{ij} \right] \right)^2 \right] + \mathbb{E} \left[\left(\left(\hat{\Sigma}_{RSV}^{ij} - \tilde{\Sigma}^{ij} \right) + \left(\mathbb{E} \left[\tilde{\Sigma}^{ij} \right] - \mathbb{E} \left[\hat{\Sigma}_{RSV}^{ij} \right] \right) \right)^2 \right] \\
&\quad + 2\mathbb{E} \left[\left(\tilde{\Sigma}^{ij} - \mathbb{E} \left[\tilde{\Sigma}^{ij} \right] \right) \left(\hat{\Sigma}_{RSV}^{ij} - \tilde{\Sigma}^{ij} \right) \right] + 2\mathbb{E} \left[\left(\tilde{\Sigma}^{ij} - \mathbb{E} \left[\tilde{\Sigma}^{ij} \right] \right) \left(\mathbb{E} \left[\tilde{\Sigma}^{ij} \right] - \mathbb{E} \left[\hat{\Sigma}_{RSV}^{ij} \right] \right) \right] \\
&= \text{Var} \left(\tilde{\Sigma}^{ij} \right) + \text{Var} \left(\hat{\Sigma}_{RSV}^{ij} - \tilde{\Sigma}^{ij} \right) + 2\mathbb{E} \left[\left(\hat{\Sigma}_{RSV}^{ij} - \tilde{\Sigma}^{ij} \right) \left(\tilde{\Sigma}^{ij} - \mathbb{E} \left(\tilde{\Sigma}^{ij} \right) \right) \right] + o(1) \\
&= \text{Var} \left(\tilde{\Sigma}^{ij} \right) + o(1)
\end{aligned}$$

Hannan (2009) has given the calculations for variance of $\tilde{\Sigma}$ as

$$\frac{n}{b_n} \text{Var}(\tilde{\Sigma}^{ij}) = [\Sigma^{ii} \Sigma^{jj} + (\Sigma^{ij})^2] \int_{-\infty}^{\infty} w^2(x) dx + o(1) \quad (14)$$

Plugging (14) into variance of $\hat{\Sigma}_{RSV}$ gives the result of the theorem. □

References

- Argon, N. T. and Andradóttir, S. (2006). Replicated batch means for steady-state simulations. *Naval Research Logistics (NRL)*, 53:508–524.
- Box, G. E., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.
- Csörgö, M. and Révész, P. (2014). *Strong approximations in probability and statistics*. Academic Press.
- Flegal, J. M. and Gong, L. (2015). Relative fixed-width stopping rules for markov chain monte carlo simulations. *Statistica Sinica*, pages 655–675.

- Gelman, A. and Meng, X.-L. (1991). A note on bivariate distributions that are conditionally normal. *The American Statistician*, 45(2):125–126.
- Gelman, A., Rubin, D. B., et al. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, 7(4):457–472.
- Glynn, P. W., Whitt, W., et al. (1992). The asymptotic validity of sequential stopping rules for stochastic simulations. *The Annals of Applied Probability*, 2(1):180–198.
- Gong, L. and Flegal, J. M. (2016). A practical sequential stopping rule for high-dimensional markov chain monte carlo. *Journal of Computational and Graphical Statistics*, 25(3):684–700.
- Hannan, E. J. (2009). *Multiple time series*, volume 38. John Wiley & Sons.
- Jones, G. L. et al. (2004). On the markov chain central limit theorem. *Probability surveys*, 1(299-320):5–1.
- Kuelbs, J. and Philipp, W. (1980). Almost sure invariance principles for partial sums of mixing b-valued random variables. *The Annals of Probability*, pages 1003–1036.
- Ma, Y. and Genton, M. G. (2000). Highly robust estimation of the autocovariance function. *Journal of time series analysis*, 21(6):663–684.
- Parzen, E. et al. (1961). An approach to time series analysis. *The Annals of Mathematical Statistics*, 32(4):951–989.
- Priestley, M. B. (1981). *Spectral analysis and time series: probability and mathematical statistics*. Number 04; QA280, P7.
- Quenouille, M. H. (1947). Notes on the calculation of autocorrelations of linear autoregressive schemes. *Biometrika*, 34(3/4):365–367.
- Song, W. T. and Schmeiser, B. W. (1995). Optimal mean-squared-error batch sizes. *Management Science*, 41(1):110–123.
- Strassen, V. (1964). An invariance principle for the law of the iterated logarithm. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 3:211–226.
- Tjstheim, D. (1990). Non-linear time series and markov chains. *Advances in Applied Probability*, 22(3):587–611.
- Vats, D., Flegal, J. M., and Jones, G. L. (2018). Strong consistency of multivariate spectral variance estimators in Markov chain Monte Carlo. *Bernoulli*, 24:1860–1909.
- Vats, D., Flegal, J. M., and Jones, G. L. (2019). Multivariate output analysis for markov chain monte carlo. *Biometrika*, 106(2):321–337.
- Wilks, S. S. (1932). Certain generalizations in the analysis of variance. *Biometrika*, pages 471–494.
- Zeidler, E. (2013). *Nonlinear functional analysis and its applications: III: variational methods and optimization*. Springer Science & Business Media.