UC Riverside
UC Riverside Electronic Theses and Dissertations

Title
Optimal Variance Estimation for a Multivariate Markov Chain Central Limit Theorem

Permalink
https://escholarship.org/uc/item/0rw0k5mq

Author
Liu, Ying

Publication Date
2017

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Optimal Variance Estimation for a Multivariate Markov Chain Central Limit Theorem

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Applied Statistics

by

Ying Liu

March 2017

Dissertation Committee:

Dr. James M. Flegal, Chairperson
Dr. Subir Ghosh
Dr. Kurt Schwabe

Copyright by
Ying Liu
2017

The Dissertation of Ying Liu is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

I am grateful to my supervisor, Dr. James Flegal for his friendship and the priceless Ph.D training he provided. When I almost gave up, his trust and encouragement pushed me through after which a long road was yet to come. When I couldn't see a way out, talking to him lights up the darkness in more ways than one, and I learned from him being flexible and creative when solving problems. When I am willing to accept a good solution, his habit of pursuing for a better one has been fundamentally influenced me and shaped my working style, which I believe will benefit me profoundly. It is a blessing for me to experience such a challenging, fulfilling and exciting Ph.D experience. I will always owe a great deal of gratitude for his mentorship.

I am thankful to Dr. Xinping Cui, who considered me as a Ph.D candidate five years ago, thus opening the door to this wonderful statistics world. I also want to express my gratitude towards the committee members and all the faculty. The knowledge I learned from them are the foundation of this thesis. I would also like to thank all the staff in the department for their support.

Dedication

My parents have been supportive since day one on my pursuit of knowledge. Their love is nourishing and teaches me to love myself, love other people, love what I do. As a result, I enjoy my daily study of statistics.

I am extremely thankful to Supawadee Wichitchan, Arnab Chowdhury and Hua Peng who helped me in many aspects. I am grateful to Chen Lin, Sakar Sigdel, and Matthew Arvanitis for their friendship. I want to thank Shangjie Xu who provided me the computational resources to finish the last piece of this thesis. I also want to thank all my other friends for their help and companionship.

ABSTRACT OF THE DISSERTATION

Optimal Variance Estimation for a Multivariate Markov Chain Central Limit Theorem

by

Ying Liu

Doctor of Philosophy, Graduate Program in Applied Statistics
University of California, Riverside, March 2017
Dr. James M. Flegal, Chairperson

Markov chain Monte Carlo (MCMC) methods are often used in Bayesian analysis to approximate expectations with respect to a target distribution. Monte Carlo standard errors (MCSEs) can be used to determine the desired number of dependent samples, as well as to construct confidence intervals of MCMC estimates. Various techniques have been suggested to estimate MCSE, but a fundamental problem is to choose an appropriate bandwidth. Previous research shows that a bandwidth proportional to $n^{1/3}$ is optimal for certain estimators, however the proportional constant is unknown. As a result, $n^{1/3}$ is suggested although sub-optimal due to the missing proportional constant. In practice, $n^{1/2}$ was also considered to account for the constant but its asymptotic performance is worrisome.

Existing literature almost always considers the above issues under univariate settings but Bayesian analysis normally involves multiple parameters. Computation time is a major challenge to estimate multivariate MCSE, where large amount of dependent samples are involved. Therefore multivariate estimators of MCSE that delivers fast and accurate estimation is desirable.

This dissertation addresses the above two problems. I consider a family of estimators and established conditions under which their mean squared consistency exist. The results have a direct

application in bandwidth selection and also suggests a bandwidth proportional to $n^{1/3}$. The proportional constant can be obtained based on the proof of mean squared consistency. I further suggest to approach the proportional constant with a pilot estimate. The suggested bandwidth shows superior performances compared with the commonly used bandwidth $n^{1/3}$ or $n^{1/2}$. The above results are established under multivariate setting which not only covers the long-standing univariate bandwidth selection problem, but also brings up the multivariate question with a solution.

To tackle the computational problem in multivariate setting, I propose a family of new estimators and prove strong consistency of these estimators. The new estimators are fast to compute and have comparable performances to spectral variance estimators with a slightly inflated variance.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Bayesian analysis provides a framework that combines data with available information when making statistical inferences. Under such a framework, parameters are treated as random variables and analysis is made based on the posterior distribution of interested parameters. Posterior expectation is commonly used as a point estimate of the parameter. Obtaining such an expectation requires complex and often high-dimensional integrals. If independent samples from posterior distribution are available, a sample mean can be used to approximate the integral via Monte Carlo methods. However in most of the cases, obtaining independent samples from a posterior distribution is challenging. Markov chain Monte Carlo (MCMC) methods have been extremely popular under these settings and the integral can be approximated by dependant samples from a Markov chain with an invariant distribution equal to the posterior distribution.

Let F be the target posterior distribution with support $\mathsf{X} \in \mathbb{R}^d$ and $g : \mathsf{X} \to \mathbb{R}^p$ be a F-integrable function. Suppose we are interested in estimating the p-dimensional vector

$$\theta := E_F g = \int_{\mathsf{X}} g(x) dF.$$

Consider the case where independent sampling from posterior distribution F is challenging, but one can obtain a Markov chain with invariant distribution F. Hence, θ can be approximated using dependent samples from the Markov chain. To this end, Let $X = \{X_t, t \geq 1\}$ be a Harris ergodic Markov chain with invariant distribution F, then with probability 1,

$$\theta_n := \frac{1}{n}\sum_{t=1}^{n} g(X_t) \to \theta \text{ as } n \to \infty.$$

The Monte Carlo error, $\theta_n - \theta$, reflects the accuracy of the estimator and the sampling distribution is available via a Markov chain central limit theorem (CLT). That is, if there exists a positive definite symmetric matrix Σ such that

$$\sqrt{n}(\theta_n - \theta) \xrightarrow{d} N_p(0,\Sigma) \text{ as } n \to \infty. \tag{1.0.1}$$

Using the CLT (1.0.1) requires estimation of Σ, which is difficult since it captures the covariance structure in F and the covariance due to the dependent sample, that is,

$$\Sigma = \text{Var}_F(g(X_t)) + \sum_{s=1}^{\infty}[\text{Cov}_F(g(X_t), g(X_{t+s})) + \text{Cov}_F(g(X_t), g(X_{t+s}))^T].$$

Σ is usually unknown and a good estimate $\hat{\Sigma}$ is essential to construct confidence regions for θ and further to terminate a simulation. This dissertation focuses on problems related to $\hat{\Sigma}$ under multivariate settings. As I will show in chapter 2 how to obtain reliable estimates of Σ. I will also introduce a family of $\hat{\Sigma}$ that are fast to compute in chapter 3. Given $\hat{\Sigma}$, various stopping rules have been established. Jones et al. (2006) suggest a fixed-width stopping rule (FWSR) under univariate setting, which terminates the chain the first time a confidence interval of the estimate is sufficiently small. Jones et al. (2006) and Flegal et al. (2008) show that FWSR is superior than visual inspections and convergence diagnostics. FWSR requires a pre-specified value ε as the threshold when comparing confidence interval width. Flegal and Gong (2013) advocate relative standard deviation FWSR that terminates the chain when uncertainty caused by computation is relatively small compared with posterior uncertainty of θ, hence eliminates the specification of ε. Gong and Flegal (2015) further establish relative standard deviation FWSR based on the diagonal element of Σ in high-dimensional settings. Vats et al. (2015a) also addressed the multivariate termination rule utilizing all entries of Σ. We first introduce univariate FWSR and relative standard deviation stopping rules, which apply to the cases where θ is univariate. It is also related to the diagonal terms of Σ, hence can be used for multivariate terminations.

If conditions to guarantee a univariate Markov chain central limit theorem is satisfied so that

$$\sqrt{n}(\theta_n - \theta) \xrightarrow{d} N(0, \sigma^2) \text{ as } n \to \infty \tag{1.0.2}$$

where $\sigma^2 = \text{Var}_F(g(X_t)) + 2\sum_{s=1}^{\infty} \text{Cov}_F(g(X_t), g(X_{t+s}))$, which are the diagonal terms of Σ. Given an estimate of σ^2 say $\hat{\sigma}^2$, FWSR Glynn and Whitt (1992); Jones et al. (2006) terminates simulation at

$$t_1(\varepsilon) = \inf\{n \geq 0 : 2z_{\delta/2}\hat{\sigma}/\sqrt{n} + p(n) \leq \varepsilon\}$$

where $\varepsilon > 0$ is the desired interval half-width and $p(n)$ is a positive function that decreases monotonically to ensure the simulation is not terminated prematurely in case $\hat{\sigma}^2$ is poorly estimated. A reasonable $p(n)$ is to fix the desired minimum simulation effort $n^* > 0$ and take $p(n) = \varepsilon I(n \leq n^*)$.

Covariance due to sample dependence and posterior variance both contribute to σ^2. Denote λ^2 the posterior variance of θ, then seperating λ^2 from σ^2 enables one to terminate the chain according to uncertainty caused by computation. Suppose $\hat{\lambda}^2$ is an estimator of λ^2, Flegal et al. (2008) consider the relative FWSR that terminates the simulation when the length of a confidence interval of θ_n is less than a εth fraction of λ^2

$$t_2(\varepsilon) = \inf\{n \geq 0 : 2z_{\delta/2}\hat{\sigma}/\sqrt{n} + p(n) \leq \varepsilon\hat{\lambda}.\}$$

The benefit of using t_2 is that only a fraction is selected hence it can be used in all settings regardless of various magnitude of the parameter.

The above two stopping rules apply to univariate parameter θ. The following two stopping rules consider multivariate θ. Let θ_i be the ith element of θ and $\theta_{n,i}$ be the corresponding sample mean. Denote λ_i^2 the posterior variance associated with θ_i and its estimate is denoted by $\hat{\lambda}_i^2$. If for each θ_i, univariate CLT (1.0.2) exists, that it there exists a finite constant σ_i such that

$$\sqrt{n}(\theta_{n,i} - \theta_i) \xrightarrow{d} N(0, \sigma_i^2) \text{ as } n \to \infty,$$

then the multivariate Markov chain can be terminated if t_2 is satisfied for all θ_i. Gong and Flegal (2015) advocate the standard deviation FWSR which terminates the simulation at

$$T_1(\varepsilon) = \sup_{\{i=1,\ldots,p\}} \inf\{n \geq 0 : 2z_{\delta/2}\hat{\sigma}_i/\sqrt{n} + p(n) \leq \varepsilon\hat{\lambda}_i.\}$$

Due to the cross-correlation among θ_i, incorporating the off-diagonal elements for termination is reasonable. Let Λ be the posterior covariance matrix associated with θ. Suppose $\hat{\Sigma}$ and $\hat{\Lambda}$ are estimates of Σ and Λ, repectively. $\hat{\Sigma}$ enables one to construct confidence region for θ and denote its volume as V. When (1.0.1) is satisfied, the relative standard deviation fixed-volume stopping rule was suggested by Vats et al. (2015a), terminates a simulation at

$$T_2(\varepsilon) = \inf\{n \geq 0 : V^{1/p} + n^{-1} \leq \varepsilon|\hat{\Lambda}|^{1/2p}\}.$$

The rest of this chapter is organized as follows. Batch means, overlapping batch means, spectral variance estimation of $\hat{\Sigma}$ are first introduced, together with a discussion of their relationship and strong consistency. Existing bias and variance results of these estimators are then given, as they are crucial to the optimal bandwidth selection of the estimators, which is addressed afterwards. Then the computational challenge of high-dimensional problem is explained. Three examples used throughout the dissertation are illustrated at the end of this chapter.

1.1 Commonly used estimators

Denote $\gamma(s) = \mathrm{E}_F[(Y_t - \theta)(Y_{t+s} - \theta)]$, where $Y_t = g(X_t)$ and $\theta = E_F g$. Consider estimating

$$\sigma^2 = \mathrm{Var}_F(Y_1) + 2\sum_{s=1}^{\infty} \gamma(s).$$

in univariate setting, which is equivalent to estimating of diagonal entries in Σ.

Batch means (BM) estimator assumes the independence of a non-overlapping batch means. Sample variance of these batch means are used to estimate σ^2, adjusted by batch length of b. Suppose the number of iteration equals to $n = ab$. For $l = 0, ..., (a-1)$, define $\bar{Y} = b^{-1}\sum_{t=1}^{b} Y_{lb+t}$. BM estimator is defined as

$$\hat{\sigma}^2_{BM} = \frac{b}{a-1}\sum_{l=0}^{a-1}(\bar{Y}_l - \bar{Y})^2.$$

Non-overlapping batch means is a simplified version of overlapping batch means (OBM) based on $(n-b+1)$ batches with equal length. Define $\bar{Y}_l(b) = b^{-1}\sum_{t=1}^{b} Y_{l+t}$ for $l = 0, ..., (n-b)$, The estimator is defined as

$$\hat{\sigma}^2_{OBM} = \frac{nb}{(n-b)(n-b+1)}\sum_{l=0}^{n-b}(\bar{Y}_l(b) - \bar{Y})^2.$$

Meketon and Schmeiser (1984) proposition 1 and 2 show that $\hat{\sigma}^2_{OBM}$ is weighted average of $\hat{\sigma}^2_{BM}$ and their bias are equivalent. Overlapping of bacths in OBM reduces variance the BM estimator. Variance ratio of BM and OBM is 3/2, see Meketon and Schmeiser (1984) proposition 3 and Flegal and Jones (2010). As a result, $\hat{\sigma}^2_{OBM}$ generates confidence intervals with less variability. Welch (1987) analysed relationship between variance and the amount of overlapping batches. It showed that estimator variance can be largely reduced with modest overlapping of batches, pointing a middle-ground between $\hat{\sigma}^2_{BM}$ and $\hat{\sigma}^2_{OBM}$. Despite the variance reduction of OBM estimator, it requires significantly more computation compared with $\hat{\sigma}^2_{BM}$. This computation is especially challenging under high-dimensional MCMC settings. Chapter 3 of this thesis introduces a new estimator which is a batch means version of spectral variance estimator as a solution of computational burden for high-dimensional problems.

Spectral variance estimator was applied in non-parametric spectral density estimation. Since estimating σ^2 is tantamount to spectral density estimate at zero frequency Fishman (1978); Welch (1987), they are also used to approximate σ^2 under MCMC settings. These estimators approximate an infinite summation of $\gamma(s)$ by truncated summation of sample auto-covariance $\hat{\gamma}(s)$. Spectral variance (SV) estimator is defined as

$$\hat{\sigma}^2_{sv} = \hat{\gamma}(0) + 2\sum_{s=1}^{b-1} w_n(s)\hat{\gamma}(s),$$

where

$$\hat{\gamma}(s) = \frac{1}{n}\sum_{t=1}^{n-s}(Y_t - \bar{Y})(Y_{t+s} - \bar{Y}),$$

$w_n(s)$ is the *lag window* and b is the *truncation point*.

A popular SV estimator is *Bartlett* estimator with window function

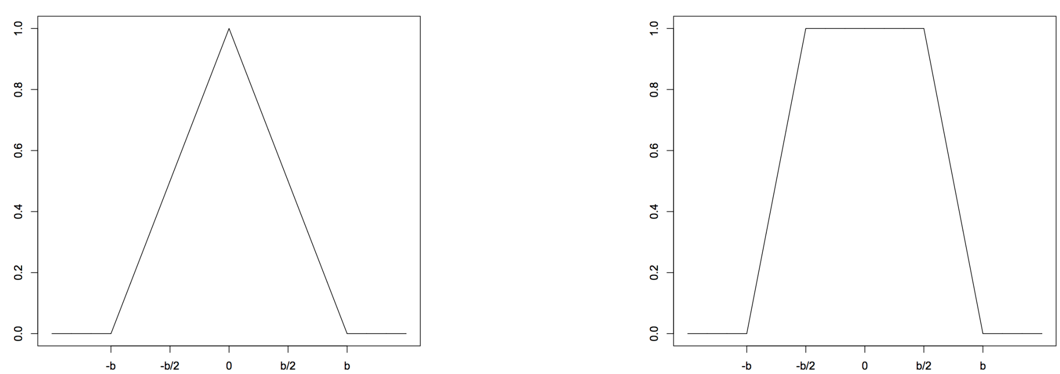
$$w_{bt}(s) = (1 - s/b)I(|s| < b).$$

Figure 1.1a shows the plot of $w_{bt}(s)$. The resulting estimator is

$$\hat{\sigma}^2_{bt} = \hat{\gamma}(0) + 2\sum_{s=1}^{b-1}(1 - s/b)\hat{\gamma}(s).$$

Politis and Romano (1995) introduced a family of spectral variance estimators based flat top window functions. These window functions are constructed from windows such as *Bartlett* window by letting the function equal to 1 near zero. It was demonstrated by Politis and Romano (1995) flat top window results in significant reduction of bias compared with SV estimator with original windows. Flat top window constructed from *Bartlett* window w_{ft} shown in figure 1.1b. It has the following expression

$$w_{ft}(s) = \begin{cases} 1 & \text{for } |s| \leqslant \frac{b}{2} \\ 2(1-|s|/b) & \text{for } \frac{b}{2} < |s| \leqslant b \\ 0 & \text{for } |s| > b \end{cases}$$



(a) *Bartlett* lag window

(b) *Bartlett* flat-top lag window

Figure 1.1 *Bartlett* lag window and *Bartlett* flat-top lag window with batch size b.

SV estimator with *Bartlett* window is known to be equivalent, except for some end-effect terms, to overlapping batch means (OBM), see Meketon and Schmeiser (1984) proposition 5. Song and Schmeiser (1993) establish properties such as non-negativity, location invariance, bias and variance of quadratic-form estimators including BM, OBM and *Bartlett* SV estimators. The quadratic forms of OBM and *Bartlett* estimator also show that the two are equivalent asymptotically. Although BM is a simpler version of OBM, it is not a SV estimator Song and Schmeiser (1993).

Estimation of Σ via multivariate batch means (mBM) has been discussed in Chen and Seila (1987) and Charnes (1995). More recently, Vats et al. (2015a) provide necessary conditions for strong consistency. Let $\bar{Y} = n^{-1}\sum_{t=1}^{n} Y_t$ and $n = ab$. For $l = 0, 1, ..., a-1$, mean vector for batch l is denoted by $\bar{Y}_l(b) = b^{-1}\sum_{t=1}^{b} Y_{lb+t}$, where b is the batch size, then

$$\hat{\Sigma}_{bm} = \frac{b}{a-1}\sum_{l=0}^{a-1}(\bar{Y}_l(b) - \bar{Y})(\bar{Y}_l(b) - \bar{Y})^T.$$

We also consider the following multivariate generalization of OBM. Let $\bar{Y}_l(b) = b^{-1}\sum_{t=1}^{b} Y_{l+t}$, there are $n-b$ batches of length b in multivariate OBM estimator with the following expression

$$\hat{\Sigma}_{obm} = \frac{nb}{(n-b)(n-b+1)}\sum_{l=0}^{n-b+1}(\bar{Y}_l(b) - \bar{Y})(\bar{Y}_l(b) - \bar{Y})^T.$$

Multivariate spectral variance (mSV) estimator is introduced by Vats et al. (2015b) along with conditions on Markov chain and window functions to guarantee strong consistency. mSV is defined as

$$\hat{\Sigma}_{sv} = \hat{\gamma}(0) + \sum_{s=1}^{b-1} w_n(s)[\hat{\gamma}(s) + \hat{\gamma}(s)^T],$$

where

$$\hat{\gamma}(s) = \frac{1}{n}\sum_{t=1}^{n-s}(Y_t - \bar{Y})(Y_{t+s} - \bar{Y})^T.$$

Strong consistency of the above estimators have been established by various author, indicating that they are asymptotically correct estimate of Σ. In paticular, Damerdji (1991) previously established the strong consistency of $\hat{\sigma}^2_{sv}$ estimators by showing an equivalent expression of $\hat{\sigma}^2_{sv}$ converges to σ^2, and that the difference between them goes to 0. *Strong invariance principle* is used to build connection between Brownian motion and $\{X_t\}$, thus convergence was established based on Brownian motion results. Flegal and Jones (2010) weakened the uniformly ergodicity condition required by that of Damerdji (1991, 1994) and proved strong consistency of $\hat{\sigma}^2_{sv}$ under geometrically ergodicity and other conditions that are easy to verify. Strong consistency of $\hat{\sigma}^2_{BM}$ and $\hat{\sigma}^2_{OBM}$ were addressed by Damerdji (1994) under uniformly ergodicity. Jones et al. (2006) lemma 2 povides uniformly and geometrically ergodicity conditions required by *strong invariant principle*, hence strong consistency of $\hat{\sigma}^2_{BM}$ can be relaxed to geometric ergodicity. Flegal and Jones (2010) further estabished strong consistency of $\hat{\sigma}^2_{BM}$ and $\hat{\sigma}^2_{OBM}$ under geometrically ergodicity. Strong consistency of multivariate SV estimator $\hat{\Sigma}_{sv}$ was shown by Vats et al. (2015b).

1.1.1 Bias and Variance

Bias of $\hat{\sigma}^2_{BM}$ and $\hat{\sigma}^2_{OBM}$ were considered by Chien et al. (1997); Goldsman and Meketon (1986); Song and Schmeiser (1993, 1995) based on the following theorem.

Theorem 1. (Chien et al., 1997) Suppose process $\{X_t\}$ is stationary and $\phi-$mixing with $E[X_1^{12}] < \infty$ and $\phi_k = O(k^{-9})$. Consider the case where $E(X_1) = 0$.

$$
\begin{aligned}
E[\hat{\sigma}^2_{BM}] &= E\left(\frac{b}{a-1}\sum_{l=0}^{a-1}(\bar{X}_l - \bar{X})^2\right) \\
&= \frac{b}{a-1}E\left(\sum_{l=0}^{a-1}\bar{X}_l^2 - a\bar{X}^2\right) \\
&= \frac{ab}{a-1}\left(\mathrm{Var}(\bar{X}_l^2) - \mathrm{Var}(\bar{X}^2)\right) \\
&= \sigma^2 + \frac{(a+1)\Gamma}{ab} + o\left(\frac{1}{b}\right),
\end{aligned}
$$

where $\Gamma = -2\sum_{s=1}^{\infty} s\gamma(s)$.

Definition of $\phi-$mixing is in appendix 1.B. The mixing condition requires a chain to be uniformly ergodic Flegal and Jones (2010). From theorem 1,

$$\text{Bias}(\hat{\sigma}^2_{BM}) = \frac{\Gamma}{b} + o\left(\frac{1}{b}\right).$$

Song and Schmeiser (1995) proposition 2 and Song and Schmeiser (1993) equation (14)-(17) show the same bias results for $\hat{\sigma}^2_{BM}$ and $\hat{\sigma}^2_{OBM}$ under certain conditions, that is

$$\lim_{b\to\infty,\ n/b\to\infty} bn \cdot \text{Bias}(\hat{\sigma}^2_{BM}/n) = \Gamma,$$

and

$$\lim_{b\to\infty,\ n/b\to\infty} bn \cdot \text{Bias}(\hat{\sigma}^2_{OBM}/n) = \Gamma.$$

Univariate SV estimators are previously addressed by Damerdji (1991, 1994); Flegal and Jones (2010).

Damerdji (1994) derived variance expression of $\hat{\sigma}^2_{BM}$ and $\hat{\sigma}^2_{OBM}$ under uniformly ergodicity by combining Brownian motion results and *strong invariance principle*. Flegal and Jones (2010) obtained the same expressions under geometrically ergodicity. Variance of $\hat{\sigma}^2_{BM}$ was also discussed previously by Chien et al. (1997) under same conditions as theorem 1. These authors show that under certain conditions,

$$\text{Var}(\hat{\sigma}^2_{BM}) = \frac{2b}{n}\sigma^4 + o\left(\frac{b}{n}\right),$$

and

$$\text{Var}(\hat{\sigma}^2_{OBM}) = \text{Var}(\hat{\sigma}^2_{bt}) = \frac{4}{3}\frac{b}{n}\sigma^4 + o\left(\frac{b}{n}\right).$$

Asymptotic mean-squared error can be obtained based on the bias and variance results. In particular, Damerdji (1995) proved mean square consistency of $\hat{\sigma}^2_{BM}$ and $\hat{\sigma}^2_{OBM}$ under uniformly ergodicity given their bias and variance expressions. Flegal and Jones (2010) showed the mean squared consistency of $\hat{\sigma}^2_{BM}$ and $\hat{\sigma}^2_{OBM}$ under geometrically ergodicity. The bias and variance expression were derived based on existing literature that requires uniformly ergodicity.

Politis and Romano (1995) discussed bias and varianc of flat top estimator for nonparametric spectral density estimation. Bias of the estimator were derived according to different auto-covariance decreasing rate of $\{X_t\}$. The new estimator has significant bias reduction effect especially for those $\{X_t\}$ with faster die down rate. Variance of $\hat{\sigma}^2_{ft}$ estimator is slightly inflated but at the same rate of b/n as that of $\hat{\sigma}^2_{bt}$. Multivariate flat top SV estimator for nonparametric spectral density estimation was advocated by Politis and Romano (1996). Bias and mean square error were also discussed for various auto-covariance decreasing rates. The flat top estimators were further applied to nonparametric estimation of multivariate density function by Politis and Romano (1999). Some of the above results are listed as follows in the setting of time series and spectral density estimation.

Suppose notation $A_n \sim B_n$ means $A_n/B_n \to 1$ as $n \to \infty$. Under certain conditions, the asymptotic variance of $\hat{\sigma}^2_{bt}$ is given by

$$\mathrm{Var}(\hat{\sigma}^2_{bt}) \sim \frac{4}{3}\frac{b}{n}\sigma^4 \tag{1.1.1}$$

Romano (1994). The following theorem provides bias of $\hat{\sigma}^2_{ft}$ estimate with an optimal choice of truncation points in terms of MSE.

Theorem 2.(Romano (1994)) Let $\{X_t\}$ be a stationary time series. Assume $\sum_{s=\infty}^{s=\infty} |s|^r|\gamma(s)| < \infty$ for some positive integer r. Suppose we take $b = 2m$ and that $b \to \infty$ as $n \to \infty$, but with $b^r/n \to 0$ (or $b = \lfloor n^\nu \rfloor$ for some $0 < \nu < 1/r$). Then

$$\mathrm{Bias}(\hat{\sigma}^2_{ft}) = o(1/b^r).$$

If in addition X_t is such that the condition to guarantee equation (1.1.1) holds are satisfied, then

$$\text{Var}(\hat{\sigma}^2_{ft}) \sim \frac{8}{3}\frac{b}{n}\sigma^4.$$

By choosing $b = C_1 n^{1/(2r+1)}$ for some constant C_1, $MSE(\hat{\sigma}^2_{ft})$ is minimized and is of order $O(n^{-2r/(2r+1)})$.

Theorem 3. (Romano (1994)) Let $\{X_t\}$ be a stationary time series. Assume that the autocovariance $\gamma(s)$ decreases geometrically fast, i.e. $\gamma(s) \leq De^{-d|s|}$. Suppose we take $b \sim Alogn$ for a constant $A > 0$. Then

$$\text{Bias}(\hat{\sigma}^2_{ft}) = O(e^{-\theta b/2}).$$

If in addition X_t is such that the condition to guarantee equation (1.1.1) holds are satisfied, then

$$\text{Var}(\hat{\sigma}^2_{ft}) \sim \frac{8}{3}\frac{b}{n}\sigma^4.$$

By choosing $b_n \sim Alogn$, $MSE(\hat{\sigma}^2_{ft})$ is minimized and is of order $O(logn/n)$.

Proof of the above bias results see appendix 1.A. There are many conditions under which (1.1.1) holds. Lahiri (1999) provide the conditions and detailed approximation of bias and variance of *Bartlett* estimator and we will prove the MSE consistency of flat top estimate under the same conditions.

Theorem 4. Lahiri (1999) Assume $E_\pi|X_t|^{6+\delta} < \infty$, and $\sum_{k=1}^{\infty} k^2(\alpha_X(k))^{\frac{\delta}{\delta+6}} < \infty$ for some $\delta > 0$. If $b \to \infty$ as $n \to \infty$ but with $b = o(n^{1/2})$ (or $b = \lfloor n^\nu \rfloor$ for some $0 < \nu < 1/2$), then

$$\text{Bias}(\hat{\sigma}^2_{bt}) = \frac{1}{b}\Gamma + o(1/b),$$

$$\text{Var}(\hat{\sigma}^2_{bt}) = \frac{4}{3}\frac{b}{n}\sigma^4 + o(b/n).$$

Bias and variance of $\hat{\sigma}^2_{bt}$ in theorem 4 coincide with results in that previous section.

Existing results consider univariate bias and variance of $\hat{\sigma}^2_{bt}$ and $\hat{\sigma}_{ft}$ in non-MCMC contexts. Multivariate bias and variance results of $\hat{\Sigma}_{sv}$ with other window functions under MCMC settings have not been addressed. In this thesis, mean square error expression of a family of $\hat{\Sigma}_{sv}$ is derived and their mean square error consistency is established.

1.2 Optimal bandwidth

The choice of batch size for mBM and bandwidth for mSV estimators largely influences the performance of the estimators. Note that batch size and bandwidth selection are similar problems with different nomenclature since the overlapping BM estimator is equivalent to a SV estimator using a modified Bartlett window, see e.g. Welch (1987), Song and Schmeiser (1993), or Meketon and Schmeiser (1984). For more general dependent processes, (Song and Schmeiser (1995)) and Damerdji (1995) consider univariate BM estimators and obtain optimal bandwidths that minimize the asymptotic means squared error based on bias and variance results in previous section. Flegal and Jones (2010) also consider BM and OBM estimators for MCMC simulations under weaker mixing and moment conditions. These papers show the asymptotic mean-squared error for $\hat{\sigma}^2_{BM}$ and $\hat{\sigma}^2_{OBM}$ are

$$\text{MSE}(\hat{\sigma}^2_{BM}) = \frac{\Gamma^2}{b^2} + \frac{2b\sigma^4}{n} + o\left(\frac{1}{b^2}\right) + o\left(\frac{b}{n}\right),$$

and

$$\text{MSE}(\hat{\sigma}^2_{OBM}) = \frac{\Gamma^2}{b^2} + \frac{4b\sigma^4}{3n} + o\left(\frac{1}{b^2}\right) + o\left(\frac{b}{n}\right).$$

The resulting bandwidth that minimized mean-squared errors are

$$\hat{b}_{BM} = \left(\frac{\Gamma^2 n}{\sigma^4}\right)^{1/3} \quad \text{and} \quad \hat{b}_{OBM} = \left(\frac{3\Gamma^2 n}{2\sigma^4}\right)^{1/3}.$$

The optimal bandwidths are proportional to $n^{1/3}$ with Γ and σ^2 being unknown quantities determined by Markov chain. As a result, $n^{1/3}$ was widely used as bandwidth to achieve an asymptotically correct increasing rate. Practically speaking, using a bandwidth of $n^{1/3}$ has poor performance for finite sample simulations Flegal and Jones (2010). Hence, $n^{1/2}$ is routinely used avoiding estimation of the unknown proportionality constants. However, the long-run performance of such an approach is a concern.

Optimal bandwidth forOBM is also optimal for $\hat{\sigma}^2_{bt}$ due to their asymptotic equivalence. This can also be obtained from theorem 4. To our best knowledge, optimal bandwidth that minimized mean squared error for other SV estimators received no attention under MCMC settings. A main challenge is to obtain asymptotic variance and bias for the estimator. Given bias and variance hence optimal bandwidth expression, the unknown proportional constant is again frustrating. Chapter 2 of the thesis aims to solve these two problems. First of all, I consider asymptotic bias and variance for a family of estimators that resemble mSV estimators, which provides a natural connection to calculating the variance and bias for a larger class of mSV estimators. As before, the optimal bandwidth is proportional to $n^{1/3}$. This is an important extension since prior results focus on univariate BM estimators. It also justifies recommended bandwidths in Vats et al. (2015b) inherited from univariate results. Secondly, we advocate a bandwidth proportional to $n^{1/3}$ and provide an estimate of the proportionality constant. In particular, pilot estimates from flat top estimator and iterative plug-in methods are considered for the proportional constant. Simulation results show that pilot estimates with iterative plug-in method have smaller variance compared with flat top pilot, but it requires more computational efforts. We also provide guidance on optimal bandwidth selection in multivariate settings given the pilot estimates. The suggested optimal bandwidth yields $\hat{\Sigma}$ with a significant better performance than bandwidth of $n^{1/3}$ or $n^{1/2}$.

Bandwidth estimation for BM and SV estimators is related to the bandwidth selection in nonparametric kernel density estimation where many data based methods have been proposed. An interested reader is directed to Bowman (1984), Jones et al. (1996), Silverman (1986), Woodroofe

(1970), Hall (1980), Sheather (1983), Sheather and Jones (1991). In the context of kernel density estimation, the bandwidth is often chosen so that the asymptotic mean integrated squared error of the kernel density estimator is minimized. The proportionality constant in bandwidth expression can be approached by plugging in a pilot estimate, which again requires a proper bandwidth. Our work is motivated by flat top pilot estimates Politis (2003, 2009) and iterative plug-in method Brockmann et al. (1993); Bühlmann (1996) because their bandwidth selection of pilots are well-established. These methods were addressed in the context of spectral estimation that is closely related to the estimation of Σ. Hence we consider similar approaches when estimating the proportionality constant of $n^{1/3}$ for mBM and mSV estimator in MCMC. Flat top pilot estimates in quantile estimation have been considered by Liu et al. (2016) to determine the number of draws needed for Bayesian credible interval.

1.3 Computational challenge

Normally a large sample size n is required to calculate θ_n. Moreover, Bayesian analysis usually involves multiple parameters as shown by the toy example in 1.1.1, where there is a total of 20 parameters. Therefore expensive computation is a major challenge to estimate Σ for high-dimensional problems.

The mSV estimator has been widely used to estimate Σ in other fields, such as nonparametric density function estimation and spectral density estimation, see (Politis and Romano (1996), Politis and Romano (1999)). One advantage of these estimators is the flexibility of choosing various window functions, making it possible to improve the estimates by choosing better performed window function. Vats et al. (2015b) recently introduced multivariate spectral variance (mSV) estimator under MCMC setting and provided theoretical justification of the methods. Despite the popularity in other fields such as time series analysis where sample size are moderate, expensive computing restricts the application of mSV in MCMC problems. As a result, only mBM is considered in high dimensional problem. $\hat{\Sigma}_{BM}$ is fast to compute but does not allow one to improve the performance with windows other than *Bartlett* window. The lack of study in this area discourages one to monitor $\hat{\Sigma}$ and terminate Markov chain in a sensible way.

Chapter 3 of the thesis advocate a new family of estimators that are fast to compute yet allowing use of better-performed window function to improve performance of mBM. Simulation shows that new estimators with flat top window are superior to mBM. In the meantime, the new estimators have significant reduction of computation time compared with mSV, hence provides an applicable solution to the problem faced by multivariate MCMC methods. I prove the strong consistency of the new estimators, followed with the discussion of their minor sacrifice on the convergence rate compared with mSV. The performance of the new estimators are illustrated by univariate and multivariate auto-regressive models. These simulations coincide with the theoretical results, showing that the new estimators converge to the correct value, and as dimension or chain length increases, the new estimators save significant amount of time compared with mSV. The variance of the new estimators are slightly larger than mSV, but the ratio between the variance of a new estimator and the corresponding mSV estimator with the same window function are usually less than two, which seems to be negligible since the actual variance of these estimators are already small given a chain with a reasonable length.

1.4 Examples

Three examples are used for simulation study. Markov chain generated from auto-regressive (AR(1)) process is used to illustrate univariate results. Vector auto-regressive (VAR(1)) model and a Bayesian dynamic space-time model are considered for multivariate simulations. In this section, some basic results of these examples are introduced.

1.4.1 AR(1)

Suppose ε_i are i.i.d N(0,1). Consider the following autoregressive process of order 1 (AR(1)):

$$X_i = \phi X_{i-1} + \varepsilon_i \quad \text{for } i = 1, 2, ...$$

Consider approximating $\theta = E[X_i]$ by $\theta_n = \bar{X}_n$. The Markov chain satisfies

$$\mathrm{Var}(X_1) = \frac{1}{1-\phi^2},$$
$$\mathrm{Cov}(X_1, X_s) = \frac{\phi^{s-1}}{1-\phi^2}.$$

For $|\phi| < 1$, the Markov chain is geometrically ergodic.

AR(1) model is considered because the true value of σ^2 in Markov chain CLT is available. The usually unknown quantity Γ in the optimal bandwidth expression is also available for this example. Performances of new estimator and suggested optimal bandwidth can be evaluated by comparing estimate of σ^2 with true value. True proportional constants are known given Γ and σ^2, hence pilot estimates from flat top and iterative plug-in method can be evaluated. Related calculations are as follows.

$$\sigma^2 = \mathrm{Var}[X_1] + 2\sum_{s=1}^{\infty} \mathrm{Cov}(X_t, X_{t+s})$$

$$
\begin{aligned}
&= \frac{1}{1-\phi^2} + 2\sum_{s=1}^{\infty} \frac{\phi^s}{1-\phi^2} \\
&= \frac{1}{1-\phi^2} + 2\lim_{n\to\infty}\sum_{s=1}^{n} \frac{\phi^s}{1-\phi^2} \\
&= \frac{1}{1-\phi^2} + 2\lim_{n\to\infty}\left[\frac{\phi(1-\phi^{n-1})}{1-\phi}\right] \\
&= \frac{1}{1-\phi^2} + \frac{2}{1-\phi^2}\cdot\frac{\phi}{1-\phi} \\
&= \frac{1+\phi}{(1-\phi^2)(1-\phi)} \\
&= \frac{1}{(1-\phi)^2}
\end{aligned}
$$

To calculate Γ, first let $S = \sum_{s=1}^{n} s\phi^s$, then

$$S = \phi + 2\phi^2 + 3\phi^3 ... + n\phi^n,$$

$$\phi S = \phi^2 + 2\phi^3 + 3\phi^4 + ... + (n-1)\phi^n + n\phi^{n+1}.$$

Hence

$$(1-\phi)S = \phi + \phi^2 + \phi^3 ... + \phi^n - n\phi^{n+1} = \frac{\phi(1-\phi^{n-1})}{1-\phi} - n\phi^{n+1},$$

resulting

$$S = \frac{\phi(1-\phi^{n-1})}{(1-\phi)^2} - \frac{n\phi^{n+1}}{1-\phi}.$$

Therefore

$$
\begin{aligned}
\Gamma &= 2\sum_{s=1}^{\infty} s\cdot \mathrm{Cov}(X_t, X_{t+s}) \\
&= 2\lim_{n\to\infty}\sum_{s=1}^{n} s\cdot\frac{\phi^s}{1-\phi^2} \\
&= \frac{2}{1-\phi^2}\lim_{n\to\infty}\sum_{s=1}^{n} s\phi^s \\
&= \frac{2}{1-\phi^2}\lim_{n\to\infty}\sum_{s=1}^{n}\left[\frac{\phi(1-\phi^{n-1})}{(1-\phi)^2} - \frac{n\phi^{n+1}}{1-\phi}\right]
\end{aligned}
$$

$$= \frac{2}{1-\phi^2}\left[\frac{\phi}{(1-\phi)^2}\right]$$
$$= \frac{2\phi}{(1-\phi^2)(1-\phi)^2}.$$

1.4.2 VAR(1)

As in the AR(1) model, Σ for vector auto-regressive (VAR(1)) model is known hence can be used to evaluate different estimates. For $i = 1, 2, ...$, consider p-dimensional vector autoregressive process of order 1 (VAR(1))

$$X_i = \Phi X_{i-1} + \varepsilon_i,$$

where $X_i \in \mathbb{R}^p$, ε_i are i.i.d $N_p(0, I_p)$ and Φ is a $p \times p$ matrix. Let $\otimes$ be the Kronecker product. When the largest eigenvalue of Φ in absolute value is less than 1, the Markov chain is geometrically ergodic Tjøstheim (1990) with invariant distribution $N_p(0, V)$, where $vec(V) = (I_{p^2} - \Phi \otimes \Phi)^{-1} vec(I_p)$. Consider approximating $\theta = EX_i$ by $\theta_n = \bar{X}_n$, we would like to estimate

$$\Sigma = \text{Var}[X_1] + 2\sum_{s=1}^{\infty} \text{Cov}(X_1, X_1 + s)$$
$$= (I_p - \Phi)^{-1} V + V(I_p - \Phi)^{-1} - V.$$

To construct a geometrically ergodic VAR(1) Markov chain, the largest eigenvalue of Φ should be less than 1. First of all, a positive semi-definite matrix is constructed from product of a matrix and its transpose, say $B = AA^T$ where A is a $p \times p$ matrix with each entry generated from standard normal distribution. The eigen decomposition of positive semi-definite matrix B always exists. Suppose $B = U\Lambda U^T$ where $\Lambda = diag\{\lambda_1, \lambda_2, ..., \lambda_p\}$ is a diagonal matrix with eigen-values λ_i, $i = 1, 2, ..., p$. The ith column of U is the corresponding eigen-vector. Let $m = max_{i=1}^{p} |\lambda_i|$, then $\Phi_0 = B/(m + 0.001) = U\Lambda/(m + 0.001)U^T$ has eigen values $|\lambda_i/(m + 0.001)| < 1$ for $i = 1, ..., p$. We would like to consider various strength of auto-covariance and cross auto-covariance in the simulation, therefore consider a series of $\Phi = k \cdot \Phi_0$, where $k \in (0,1)$ so that the eigen-values of Φ remains less than 1, meanwhile larger k implies stronger auto-covariance and cross auto-covariance.

1.4.3 Bayesian dynamic spatial-temporal model

Spatial-temporal models are often applied to model data at several locations in a certain region over a period of time. Usually space is treated as continuous and time is viewed as discrete in these models. Without the time effect, spatial-temporal model becomes a spatial models.

This example is applied to `NETemp` data described in `R` package `spBayes` Finley et al. (2013). Choose a subset of locations by letting variable UTMX$>$ 6000000 and UTMY$>$ 3250000, we consider 10 nearby weather stations. For each weather station, elevation and temperature from 126 consecutive months are available. We consider the first 12 month (year 2000) for these 10 stations.

Spatial model. Consider building regression model to explain response variables over a region. Suppose $y(s)$ denotes reponse at location s for $s = 1,2,...,N_s$. Let $x(s)$ be a $k \times 1$ vector of predictors. The following spatial model Gelfand et al. (2003) allows responses at closer locations to have stronger dependence.

$$y(s) = \boldsymbol{x}(s)^T \boldsymbol{\beta} + u(s) + \varepsilon(s), \;\; \varepsilon_t \sim N(0, \tau^2),$$

where $u(s) \sim GP(0, C(\cdot, \sigma^2, \phi))$ is a zero mean stationary Gaussian process. Gaussian process is one of the most commonly used stochastic process when modelling dependent data over time or space. It is determined by its mean and variance functions as follows. $\mathrm{E}[u(s)] = 0$, $\mathrm{Var}[u(s)] = \sigma^2$, $\mathrm{Cov}[u(s_1), u(s_2)] = \sigma^2 \rho(s_1, s_2; \phi)$, where ϕ is a valid two-dimensional correlation function. By specifying $u(s) \sim GP(0, C(\cdot, \sigma^2, \phi))$, it conveniently describes variance and covariance structure for all locations. A possible distribution of $\boldsymbol{y} = (y(1), y(2), ..., y(N_s))^T$ is

$$\boldsymbol{y} \sim N(\boldsymbol{x}(s)^T \boldsymbol{\beta},\ \Sigma(\Theta)).$$

In the above distribution, $\Theta = \{\sigma, \tau, \phi\}$ and $\Sigma(\Theta) = (\sigma^2 + \tau^2)I + H(\phi)$ where I is identity matrix and $H(\phi)$ is a $k \times k$ matrix. The ijth element of $H(\phi)$ is $H_{ij}(\phi) = exp(-\|s_i - s_j\|/\phi)$, where $\|s_i - s_j\|$ is the Euclidean distance between locations $s_i,\ s_j$. Therefore, dependence between two locations decreases as their distance increases. This model is attractive when describing the correlation structure of location. To further model time series observed at these locations, consider spatial-temporal model incorporates time effect.

Spatial-temporal model. We apply a spatial-temporal model to the temperature data from 12 month and 10 locations. Suppose $y_t(s)$ denote the temperature observed at location s and time t for $s = 1, 2, ..., N_s$ and $t = 1, 2, ..., N_t$. Let $x_t(s)$ be a $k \times 1$ vector of predictors and β_t be a $k \times 1$ coefficient vector, which is a purely time component. $u_t(s)$ denotes a space-time component. The model is

$$y_t(s) = \boldsymbol{x}_t(s)^T \boldsymbol{\beta}_t + u_t(s) + \varepsilon_t(s),\ \ \varepsilon_t \sim N(0, \tau_t^2),$$

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \boldsymbol{\eta}_t;\ \ \boldsymbol{\eta}_t \sim N_p(0, \Sigma_\eta),$$

$$u_t(s) = u_{t-1}(s) + w_t(s);\ \ w_t(s) \sim GP(0, C_t(\cdot, \sigma_t^2, \phi_t)).$$

$GP(0, C_t(\cdot, \sigma_t^2, \phi_t))$ is a spatial Gaussian process where $C_t(s_1, s_2; \sigma_t^2, \phi_t) = \sigma_t^2 \rho(s_1, s_2; \phi_t)$. $\rho(\cdot; \phi)$ is an exponential correlation function with ϕ controlling the correlation decay, and σ_t^2 represents the spatial variance components. We are interested in estimating posterior expectation of 185 parameters $\theta = (\boldsymbol{\beta}_t,\ u_t(s),\ \sigma_t^2,\ \Sigma_\eta,\ \tau_t^2,\ \phi_t)$, their prior follows `spDynlM` function in `spBayes` package.

Elevation is the only predictor in this example, hence $\boldsymbol{\beta}_t = (\beta_t^{(0)},\ \beta_t^{(1)})^T$ for $t = 1, 2, ..., 12$, where $\beta_t^{(0)}$ is intercept and $\beta_t^{(1)}$ is coefficient of covariate elevation. In this example, covariate elevation stays the same over time, therefore it is a simpler version of the spatial-temporal model. Compared with spatial model, time effects of both $\boldsymbol{\beta}_t$ and $u_t(s)$ are introduced by the transition model from $t-1$ to t. $\boldsymbol{\beta}_t$ is centered at $\boldsymbol{\beta}_{t-1}$ with a covariance matrix determined by Σ_η. $\boldsymbol{u}_t = (u_t(1), u_t(2), ..., u_t(N_s))^T$ is centered at $\boldsymbol{u}_{t-1}^T = (u_{t-1}(1), u_{t-1}(2), ..., u_{t-1}(N_s))$ with a covariance matrix determined $\{\sigma_t^2, \phi_t\}$ in the Gaussian process so that $u_t(s_1),\ u_t(s_2)$ with closer $s_1,\ s_2$ have higher correlation.

We will consider the intercept parameter for two consecutive month $\beta_1^{(0)}$ and $\beta_2^{(0)}$. Let $\{X_i = (\beta_{1,i}^{(0)}, \beta_{2,i}^{(0)})^T\}_{i=1}^{\infty}$ be posterior sample of $(\beta_1^{(0)},\ \beta_2^{(0)})$ generated from `spDynlM` function. We are interested in estimating

$$\Sigma = \text{Var}(X_i) + \sum_{s=1}^{\infty}[\gamma(s) + \gamma(s)^T].$$

1.A Appendix of Chapter 1

A Proof of bias of theorem 3

Let $B = b$ and $b = b/2$ for simplicity of the proof. First we prove that $\text{Bias}(\hat{\sigma}_{ft}^2) = O(e^{-\theta b})$. Since $\gamma(s) = O(e^{-\theta|s|})$, $\gamma(s) \leq C \cdot e^{-\theta|s|}$ for some constant C.

$$\text{Bias}(\hat{\sigma}_{ft}^2) = E[\hat{\sigma}_{ft}^2] - \sigma^2 = A_1 + A_2 + A_3$$

where

$$A_1 = \frac{1}{2\pi}\sum_{s=-n+1}^{n-1}(w(s)-1)\gamma(s)$$

$$A_2 = -\frac{1}{2n\pi}\sum_{s=-n+1}^{n-1}|s|w(s)\gamma(s)$$

$$A_3 = -\frac{1}{2\pi}\sum_{|s|>n}\gamma(s).$$

$$|A_3| \leq \frac{1}{2\pi}\frac{1}{e^{\theta n}}\sum_{|s|>n}\gamma(s)e^{\theta n} \leq \frac{C}{2\pi}\frac{1}{e^{\theta n}}\sum_{|s|>n}e^{-(|s|-n)\theta} = O(e^{-\theta n}).$$

$$|A_2| \leq \frac{1}{2\pi n}\sum_{s=-n+1}^{n-1}|s|w(s)\gamma(s) \leq \frac{1}{2\pi n}\sum_{s=-n+1}^{n-1}|s|\gamma(s) = O(\frac{1}{n}).$$

Since if $\gamma(s) = O(e^{-\theta|s|})$, $\sum_{s=-n+1}^{n-1}|s|^r\gamma(s) < \infty$ for any $r > 0$. When $r = 1$, $\sum_{s=-n+1}^{n-1}|s|\gamma(s) < \infty$.

$$A_1 = a_1 + a_2 + a_3$$

where

$$a_1 = \frac{1}{2\pi}\sum_{|s|\leq b}(w(s)-1)\gamma(s)$$

$$a_2 = \frac{1}{2\pi}\sum_{b<|s|\leq B}(w(s)-1)\gamma(s)$$

$$a_3 = \frac{1}{2\pi}\sum_{B<|s|\leq n}(w(s)-1)\gamma(s)$$

$a_1 = 0$ since $w(s) = 1$ for $|s| \leq m$. Note $|\frac{s-b}{B-b}| \leq 1$ for $b < |s| \leq B$, we have

$$|a_2| \leq \frac{1}{\pi}\sum_{b<|s|\leq B}\frac{s-b}{B-b}\gamma(s) \leq \frac{1}{\pi}\frac{1}{e^{\theta b}}\sum_{b<|s|\leq B}e^{\theta b}\gamma(s)$$

$$\leq \frac{C}{\pi}\frac{1}{e^{\theta b}}\sum_{b<|s|\leq B}e^{-\theta(|s|-b)} = O(e^{-\theta b}).$$

When $B < |s| \leq n$, $w(s) = 0$, thus

$$|a_3| \leq \frac{1}{2\pi}\sum_{B<|s|\leq n}\gamma(s) \leq \frac{C}{2\pi}\frac{1}{e^{\theta B}}\sum_{B<|s|\leq n}e^{-\theta(|s|-B)} = O(e^{-\theta B}) = O(e^{-\theta b}).$$

We proved that $\text{Bias}(\hat{\sigma}^2_{ft}) = O(e^{-\theta b})$.

B Mixing conditions

Let $P(x,dy)$ be the Markov chain transition kernel of X on space $(\mathsf{X},\mathscr{B}(\mathsf{X}))$. Let $P^n(x,dy)$ denote the n-step transition kernel of X. For $x \in \mathsf{X}$ and $A \in \mathscr{B}(\mathsf{X})$,

$$P^n(x,A) = P(X_{t+n} \in A | X_t = x).$$

Suppose X is Harris ergodic with stationary distribution $\pi(\cdot)$, then for any initial probability distribution $\lambda(\cdot)$,

$$\|P(x,\cdot) - \pi(\cdot)\| \to 0 \quad (as)\ n \to \infty$$

where $\|\cdot\|$ denotes total variation norm.

Definition. A Markov chain with stationary distribution $\pi(\cdot)$ is geometric ergodic if

$$\|P(x,\cdot) - \pi(\cdot)\| \leq M(x)\rho^n \quad n = 1,2,3,...$$

for some $\rho < 1$.

Suppose X is defined on the probability space $(\Omega,\mathscr{F},P)$. Let $\mathscr{F}_k^m = \sigma(X_k,...,X_m)$ be the sigma field generated by $(X_k,...,X_m)$ and $L_2(\mathscr{F})$ be the family of all square integrable $\mathscr{F}-$measurable random variables. Jones (2004) discuss the connections between mixing conditions and convergence of Markov chain. We focus on the results of α-mixing and ρ-mixing in this section.

Definition. The sequence X is said to be strongly mixing (or α-mixing) if $\alpha(n) \to$ as $n \to \infty$ where

$$\alpha(n) := \sup_{k\geqslant 0} \sup_{\mathscr{A}\in\mathscr{F}_0^k,\ \mathscr{B}\in\mathscr{F}_{k+n}^\infty} |P(\mathscr{A}\cap\mathscr{B}) - P(\mathscr{A})P(\mathscr{B})|.$$

Definition. The sequence X is said to be asymptotically uncorrelated (or $\rho-$mixing) if $\rho(n) \to 0$ as $n \to \infty$ where

$$\rho(n) := \sup_{k\geqslant 0} \sup_{Z_1\in L_2(\mathscr{F}_0^k),\ Z_2\in L_2(\mathscr{F}_{k+n}^\infty)} |\mathrm{corr}(Z_1,Z_2)|.$$

It can be shown that for a Harris ergodic Markov chain X,

$$\alpha(n) = \sup_{\mathscr{A} \in \sigma(X_0),\ \mathscr{B} \in \sigma(X_n)} |P(\mathscr{A} \cap \mathscr{B}) - P(\mathscr{A})P(\mathscr{B})|$$

and

$$\rho(n) = \sup_{Z_1 \in L_2(\sigma(X_0)),\ Z_2 \in L_2(\sigma(X_n))} |\text{corr}(Z_1, Z_2)|$$

Bradley (1985). From now we say a non-negative sequence $\{a_n, n = 1, 2, 3, ...\}$ converges to 0 exponentially fast if $a_n = O(e^{-\theta n})$. The following facts about mixing conditions are true:

(i) Harris ergodic Markov chains are α-mixing. (ii) Geometric ergodic Markov chains have exponentially fast α-mixing. (iii) If a geometric ergodic Markov chain satisfies

$$\pi(dx)P(x, dy) = \pi(dy)P(y, dx), \quad x, y \in \mathsf{X}, \tag{1.A.1}$$

then it has exponentially fast $\rho-$mixing.

To prove fact (i), consider coupling method Lindvall (2002) by constructing Markov chain $X' := \{X'_t\}$ which is also governed by $P(x, dy)$ but start with the invariant distribution $\pi(\cdot)$. Define $X'' := \{X''_t\}$ by

$$X''_t = \begin{cases} X_t & \text{if } t < T \\ X'_t & \text{if } t \geq T \end{cases}$$

where $T = \min\{t : X_t = X'_t\}$. X and X'' are equally distributed due to Markov property. Since

$$\|P^n(x, \cdot) - \pi(\cdot)\| = 2 \sup_{\mathscr{A} \in \mathscr{F}} (P(X_n \in \mathscr{A}) - P(X'_n \in \mathscr{A})),$$

and

$$\begin{aligned} P(X_n \in \mathscr{A}) - P(X'_n \in \mathscr{A}) &= P(X''_n \in \mathscr{A}) - P(X'_n \in \mathscr{A}) \\ &= [P(X''_n \in \mathscr{A}, n < T) + P(X''_n \in \mathscr{A}, n \geq T)] \\ &- [P(X'_n \in \mathscr{A}, n < T) + P(X'_n \in \mathscr{A}, n \geq T)] \\ &\leq P(n < T). \end{aligned}$$

Therefore

$$\|P^n(x,\cdot) - \pi(\cdot)\| \le 2P(n < T). \tag{1.A.2}$$

Let $A, B \in \mathscr{B}(\mathsf{X})$ be sets corresponding to $\mathscr{A}, \mathscr{B} \in \mathscr{F}$ respectively. Then

$$\begin{aligned}
\alpha(n) = & \sup_{\mathscr{A}\in\sigma(X_0),\ \mathscr{B}\in\sigma(X_n)} |P(\mathscr{A}\cap\mathscr{B}) - P(\mathscr{A})P(\mathscr{B})| \\
& \le |P(X_0 \in A \text{ and } X_n \in B) - \pi(A)\pi(B)| \\
& \le |\int_B [P^n(x,A) - \pi(A)]\pi(dx)| \\
& \le \int_B |P^n(x,A) - \pi(A)|\pi(dx) \\
& \le \int_B 2P(n < T)\pi(dx) = 2E_\pi[P(T > n)].
\end{aligned}$$

Since $E_\pi[P(T > n)] \to 0$ as $n \to 0$, fact (i) is proved.

For fact (ii), if $E_\pi M < \infty$, we can prove that

$$\begin{aligned}
\alpha(n) \le & \int_B |M(x)\rho^n|\pi(dx) \\
& \le E_\pi M \cdot \rho^n,
\end{aligned}$$

hence $\alpha(n) = O(e^{-\theta n})$ for an $\theta > 0$.

(Rosenblatt (1971) and Roberts and Rosenthal (1997)) show that if a Markov chain is geometric ergodic, it is ρ-mixing. By strong Markov property, it is exponentially fast ρ-mixing Bradley (1985) and (iii) is proved. From now on, assume $E[X_t] = 0$ without loss of generality. Define the lag s auto-covariance $\gamma(s) = \gamma(-s) := E_\pi[X_t X_{t+s}]$.

Proposition. If X is a geometric ergodic Markov chain and satisfies (1.A.1), then $\gamma(s) = O(e^{-\theta s})$ for a $\theta > 0$.

Proof. By fact (iii), there exits constant $N > 0$, C and $\theta > 0$, such that

$$\rho(n) \le C \cdot e^{-\theta n} \quad \text{for } n \ge N.$$

Therefore $|\gamma(s)| = |\text{corr}(X_t, X_{t+s})| \cdot \gamma(0) \leq \rho(s) \cdot \gamma(0) \leq C \cdot \gamma(0) \cdot e^{-\theta s}$ for $n \geq N$, which proves $\gamma(s) = O(e^{-\theta s})$.

Chapter 2

Optimal bandwidth selection

Choosing an appropriate bandwidth has been a long standing question for BM, OBM and SV estimators since it is crucial to the performances of the estimators. In this chapter, I consider optimal bandwidth selection for a family of mSV estimators. Recall that

$$\Sigma = \text{Var}_F(g(X_1)) + \sum_{s=1}^{\infty}[\text{Cov}_F(g(X_t), g(X_{t+s})) + \text{Cov}_F(g(X_t), g(X_{t+s}))^T].$$

Estimation of Σ via multivariate batch means (mBM) has been discussed in Chen and Seila (1987) and Charnes (1995). More recently, Vats et al. (2015a) provide necessary conditions for strong consistency. Univariate BM estimators for the diagonal terms of Σ have been studied previously by Chien et al. (1997), Chien et al. (1997), Damerdji (1994), Flegal and Jones (2010), and Jones et al. (2006). Multivariate spectral variance estimators (mSV) of Σ are also available along with necessary conditions for strong consistency Vats et al. (2015b). However, Vats et al. (2015a) suggest mSV are computationally expensive relative to mBM. Damerdji (1991), Damerdji (1994), and Flegal and Jones (2010) previously studied univariate SV estimators for the diagonal terms of Σ. One can also consider regenerative simulation (RS) estimators of Σ see e.g.Hobert et al. (2002); Mykland et al. (1995).

The choice of batch size for mBM and bandwidth for mSV estimators largely influences performance of the estimators. Note that batch size and bandwidth selection are similar problems with different nomenclature since the overlapping BM estimator is equivalent to a SV estimator using a modified Bartlett window, see e.g. Welch (1987), Song and Schmeiser (1993), or Meketon and Schmeiser (1984). The focus here is on optimal bandwidth selection for MCMC simulations, which has received limited attention in the literature. For more general dependent processes, Song and Schmeiser (1995) and Damerdji (1995) consider univariate BM estimators and obtain optimal bandwiths that minimize the asymptotic means squared error. Flegal and Jones (2010) also consider these estimators for MCMC simulations under weaker mixing and moment conditions. In short, these papers show the optimal bandwidth is proportional to $n^{1/3}$. However, there has been no work to our knowledge in MCMC settings with regard to estimating the proportionality constant. As a result, Flegal and Jones (2010) suggest using a bandwidth equal to $\lfloor n^{1/2} \rfloor$.

In this chapter, we derive an asymptotic expression of variance and bias for a larger class of mSV estimators resulting in an optimal bandwidth expression that minimizes the asymptotic means squared error. As before, the optimal bandwidth is proportional to $n^{1/3}$. This is an important extension since prior results focus on univariate BM estimators. This also justifies recommended bandwidths in Vats et al. (2015b) inherited from univariate results.

Although suggested by existing literature, using a bandwidth of $n^{1/3}$ has poor performance for finite sample simulations and $n^{1/2}$ is routinely used avoiding estimaiton of the unknown proportionality constants Flegal and Jones (2010). However, the long-run performance of such an approach is a concern. In this chapter, I advocate a bandwidth proportional to $n^{1/3}$ and provide an estimate of the proportionality constant, which has received no attenton in the MCMC literature. The proposed estimates resemble mSV estimators, which provides a natural connection to calculating the variance and bias for a larger class of mSV estimators. We also provide guidance on optimal bandwidth selection in multivariate settings. Simulation studies show the optimal bandwidth improves performance over using $n^{1/2}$ and $n^{1/3}$.

Bandwidth estimation for BM and SV estimators is related to the bandwidth selection in nonparametric kernel density estimation where many data based methods have been proposed. An interested reader is directed to Bowman (1984), Jones et al. (1996), Silverman (1986), Woodroofe (1970), Hall (1980), Sheather (1983), Sheather and Jones (1991). In the context of kernel density estimation, the bandwidth is often chosen so that the asymptotic mean integrated squared error of the kernel density estimator is minimized. The proportionality constant in bandwidth expression can be approached by plugging in a pilot estimate, which again requires a proper bandwidth. Our work is motivated by flat top pilot estimates see Politis (2003, 2009) and iterative plug-in method see Brockmann et al. (1993); Bühlmann (1996). These methods were addressed in the context of spectral estimation that is closely related to the estimation of Σ. Hence we consider similar approaches when estimating the proportionality constant of $n^{1/3}$ for mBM and mSV estimator in MCMC. Flat top pilot estimates in quantile estimation have been considered by Liu et al. (2016) to determine the number of draws needed for Bayesian credible interval.

The rest of this chapter is organized as follows. Section 2.1 derives asymptotic variance and bias results for a class of mSV estimators. Then an optimal bandwidth selection procedure is given based on asymptotic mean squared error. Section 2.2 introduces two pilot estimates for the proportional constant of the optimal bandwidth. Section 2.3 considers three examples to compare performances of suggested bandwidth with $n^{1/2}$ and $n^{1/3}$.

2.1 Mean squared consistency

Recall $\theta = E_F g$. Let $Y_t = g(X_t),\ t = 1,2,3,...$, and denote $\gamma(s) = \mathrm{E}_F[(Y_t - \theta)(Y_{t+s} - \theta)^T]$. We are interested in estimating

$$\Sigma = \mathrm{Var}_F(Y_1) + \sum_{s=1}^{\infty}[\gamma(s) + \gamma(s)^T].$$

Let $\bar{Y} = n^{-1}\sum_{t=1}^{n} Y_t$, mSV estimator Vats et al. (2015b) is defined as

$$\hat{\Sigma}_s = \hat{\gamma}(0) + \sum_{s=1}^{b-1} w_n(s)[\hat{\gamma}(s) + \hat{\gamma}(s)^T],$$

where

$$\hat{\gamma}(s) = \frac{1}{n}\sum_{t=1}^{n-s}(Y_t - \bar{Y})(Y_{t+s} - \bar{Y})^T,$$

$w_n(s)$ is the *lag window* and b is the *truncation point*. We are interested in the asymptotic mean squared error of $\hat{\Sigma}_s$, which are based on the leading terms in asymptotic bias and variance of $\hat{\Sigma}_s$. Denote Euclidean norm by $\|\cdot\|$. A fundamental condition required in this thesis is the *strong invariant principle* established by Vats et al. (2015b).

Condition 1.(*strong invariant principle*) There exists a $p-$dimensional vector θ, a $p \times p$ lower triangular matrix L, an increasing function ψ on integers, a finite random variable D and a sufficiently rich probability space Ω such that for almost all $\omega \in \Omega$ and for all $n > n_0$,

$$\left\|\sum_{t=1}^{n} Y_t - n\theta - LB(n)\right\| < D(\omega)\psi(n) \quad \text{w.p. 1.} \tag{2.1.1}$$

For polynomial ergodic Markov chains, under certain moment conditions on g, *Condition 1* holds with $\psi(n) = n^{1/2-\lambda}$ for some $\lambda > 0$ see Kuelbs and Philipp (1980); Vats et al. (2015a,b). If $\psi(n)$ satisfies that $\psi(n)/\sqrt{n} \to 0$ as $n \to \infty$, *Condition 1* also implies a strong law, a Markov chain CLT and a functional CLT.

Condition 2. The lag window $w_n(\cdot)$ is an even function defined on $\mathbb{Z}$ such that

$$|w_n(s)| \leq 1 \quad \text{for all } n \text{ and } s,$$

$$w_n(0) = 1 \quad \text{for all } n,$$

$$w_n(s) = 0 \quad \text{for all } |s| \geq b.$$

Condition 3. b is an integer sequence such that $b \to \infty$ and $n/b \to \infty$ as $n \to \infty$, where b and n/b are monotonically non-decreasing.

To approach means squared error of $\hat{\Sigma}_s$, we define an asymptotically equivalent expression. Let $\bar{Y}_l(k) = k^{-1}\sum_{t=1}^{k} Y_{l+t}$ for $l = 0, ..., (n-k)$, define

$$\hat{\Sigma}_w = \frac{1}{n}\sum_{k=1}^{b}\sum_{l=0}^{n-k} k^2 \Delta_2 w_n(k)[\bar{Y}_l(k) - \bar{Y}][\bar{Y}_l(k) - \bar{Y}]^T.$$

Suppose $\hat{\Sigma}_w = \hat{\Sigma}_s - d$, the following proposition shows that the two expression are asymptotically equivalent.

Proposition 1. Assume condition 2 and 3 hold and condition 1 holds for both g and h. If $n > 2b$, $b_n^{-1} logn = O(1)$ and $b^{-1}\psi(n) \to 0$, $b^{-1}\psi_h(n) \to 0$ as $n \to \infty$, further

$$bn^{-1}\sum_{k=1}^{b} k|\Delta_1 w_n(k)| \to 0,$$

then $d \to 0$ $w.p.1$ as $n \to \infty$.

Proof. See Appendix B.

Given *Proposition 1*, we will focus on mean squared error of $\hat{\Sigma}_w$ to approach mean squared error of $\hat{\Sigma}_s$, which is challenging to obtain directly. Let $\hat{\Sigma}_{w,ij}$ be the ijth entry of $\hat{\Sigma}_w$, then the following theorem provides the asymptotic variance of each entry for $\hat{\Sigma}_w$.

Theorem 1. Define $h(X_t) = (g(X_t) - E_F g)^2$, $t = 1, 2, 3, ...$ and suppose $\|E_F h\| < \infty$. Suppose *condition 1* holds for g with L, D and ψ and for h with L_h, D_h, ψ_h. Let *condition 2* and *3* hold. If

1. $\sum_{k=1}^{b}(\Delta_2 w_k)^2 \leq O\left(\frac{1}{b^2}\right)$,

2. $b\psi^2(n) logn(\sum_{k=1}^{b}|\Delta_2 w_n(k)|)^2 \to 0$.

3. $\psi^2(n)\sum_{k=1}^{b}|\Delta_2 w_n(k)| \to 0$.

Then

$$Var[\hat{\Sigma}_{w,ij}] = [\Sigma_{ii}\Sigma_{jj} + \Sigma_{ij}^2] \cdot \left[\frac{2}{3}\sum_{k=1}^{b}(\Delta_2 w_k)^2 k^3 \cdot \frac{1}{n} + 2\sum_{t=1}^{b-1}\sum_{u=1}^{b-t}\Delta_2 w_u \cdot \Delta_2 w_{t+u} \cdot \left(\frac{2}{3}u^3 + u^2 t\right) \cdot \frac{1}{n}\right] + o\left(\frac{b}{n}\right)$$
$$:= (S + o(1)) \cdot \frac{b}{n}, \tag{2.1.2}$$

Proof. See Appendix C.

Remark. It can be shown that condition 1 in theorem 1 is satisfied for *Bartlett* window

$$w_{bt}(s) = (1 - s/b)I(|s| < b),$$

where $\Delta_2 w_{bt}(b) = b^{-1}$ and $\Delta_2 w_{bt}(k) = 0$ for $k = 1, 2, ..., b-1$. Hence $\sum_{k=1}^{b}(\Delta_2 w_k) = 1/b^2$. Consider *flat top window* discussed in section 2.3, $\Delta_2 w_n(b/2) = -2/b$, $\Delta_2 w_{ft}(b) = 2/b$ and $\Delta_2 w_{ft}(k) = 0$ for $k \neq b/2,\ b$. Then $\sum_{k=1}^{b}(\Delta_2 w_k)^2 = 8/b^2$, condition 1 is also satisfied. It also follows the fact that w_{ft} is equivalent to the difference of two w_{bt}.

Bias of the diagonal entries $\hat{\Sigma}_{w,ii}$, $i = 1, 2, ..., p$ has previously been discussed by Chien et al. (1997); Goldsman and Meketon (1986); Song and Schmeiser (1995). Bias of off-diagonal entries follows a similar expression.

Theorem 2. Let X be a stationary uniformly ergodic Markov chain, if $E_F g^{12} < \infty$ and *condition 3* holds, and further

$$\sum_{k=1}^{b} k\Delta_2 w_n(k) = 1,$$

then

$$Bias[\hat{\Sigma}_{w,ij}] = \sum_{k=1}^{b} \Delta_2 w_n(k) \cdot \Gamma_{ij} + o\left(\frac{b}{n}\right) + o\left(\frac{1}{b}\right),$$

where

$$\Gamma = -\sum_{s=1}^{\infty} s[\gamma(s) + \gamma(s)^T]$$

and

$$\Gamma_{ij} = -\sum_{s=1}^{\infty} s[\gamma_{ij}(s) + \gamma_{ji}(s)].$$

Proof. See Appendix D.

Corollary 1. Under conditions of theorem 1 and theorem 2,

$$\text{MSE}[\hat{\Sigma}_{w,ij}] \to 0 \quad \text{as } n \to \infty.$$

Proof. Follows directly from theorem 1 and theorem2.

Besides *corollary 1*, we are also interested in the expression of $\text{MSE}[\hat{\Sigma}_{w,ij}]$, which is a function of bandwidth b. By minimizing the mean squared error of $\hat{\Sigma}_{w,ij}$, an optimal bandwidth that minimizes the asymptotic mean squared error of $\hat{\Sigma}_{w,ij}$ can be achieved.

Corollary 2. Suppose conditions in theorem 1 and theorem 2 hold. if $S \neq 0$ in theorem 1 and

$$\sum_{k=1}^{b} \Delta_2 w_n(k) = \frac{C}{b} \tag{2.1.3}$$

for a constant $C \neq 0$, then

$$\text{MSE}[\hat{\Sigma}_{w,ij}] = \left[\sum_{k=1}^{b} \Delta_2 w_n(k)\right]^2 \Gamma_{ij}^2 + \frac{Sb}{n} + o\left(\frac{b}{n}\right) + o\left(\frac{1}{b}\right)$$

$$= \frac{C^2\Gamma_{ij}^2}{b^2} + \frac{Sb}{n} + o\left(\frac{b}{n}\right) + o\left(\frac{1}{b}\right),$$

and

$$\hat{b}_{opt,ij} = \left(\frac{2C^2\Gamma_{ij}^2 n}{S}\right)^{1/3} \tag{2.1.4}$$

is an optimal bandwidth $\hat{b}_{opt,ij}$ that minimizes mean squared error.

Remark. General expression of S and C are not given but can be easily derived for a specific window. *Corollary 2* requires non-zero S and C. In other words, variance and bias in theorem 1 and theorem 2 are restricted to estimators with a corresponding die down rate no faster than the dominant terms in those theorems. Nevertheless, the results apply to an important family of estimators including *Bartlett* estimator.

Remark. Corollary 2 shows an optimal bandwidth for each entry of $\hat{\Sigma}_w$. We would like to choose a bandwidth that benefits all elements of $\hat{\Sigma}_w$. Based on simulation result,

$$\hat{b}_{opt} = \frac{1}{p^2}\sum_i\sum_j \hat{b}_{opt,ij}$$

is used as the optimal bandwidth for $\hat{\Sigma}_w$ that minimizes mean squared error of all entries on average.

As an example, consider mSV with *Bartlett* window $\hat{\Sigma}_{bt}$. From theorem 2 and the proof of theorem 1, it is easy to obtain bias and variance

$$\text{Bias}[\hat{\Sigma}_{bt,ij}] = \frac{\Gamma_{ij}}{b} + o\left(\frac{b}{n}\right) + o\left(\frac{1}{b}\right),$$

$$\text{Var}[\hat{\Sigma}_{bt,ij}] = \frac{2}{3}(\Sigma_{ii}\Sigma_{jj} + \Sigma_{ij}^2)\frac{b}{n} + o\left(\frac{b}{n}\right).$$

Mean-squared error of the (i, j)th element is

$$\text{MSE}(\hat{\Sigma}_{bt,ij}^2) = \frac{\Gamma_{ij}^2}{b^2} + \frac{2}{3}(\Sigma_{ii}\Sigma_{jj} + \Sigma_{ij}^2)\frac{b}{n} + o\left(\frac{b}{n}\right) + o\left(\frac{1}{b^2}\right),$$

hence

$$\hat{b}_{bt,ij} = \left(\frac{3\Gamma_{ij}^2 n}{\Sigma_{ii}\Sigma_{jj} + \Sigma_{ij}^2} \right)^{1/3}.$$

The average of $\hat{b}_{bt,ij}$ is then used as bandwidth of $\hat{\Sigma}_{bt}$.

A potential issue with suggested bandwidth is that it involves unknown quantities determined by Markov chain, such as Γ and Σ. Flegal and Jones (2010) suggested using $n^{1/2}$ in univariate setting, in order to adjust for the unknown constant of $n^{1/3}$ due to the poor performance of $n^{1/3}$. We would like to improve this convention of bandwidth selection by considering the coefficient of $n^{1/3}$.

2.2 Estimation of coefficient

This section considers pilot estimation of Σ and Γ in order to estimate the coefficient of $n^{1/3}$. The idea of plugging in pilot estimates has been applied in density estimation see e.g. Jones et al. (1996); Loader (1999); Politis (2003); Woodroofe (1970). Since estimation of Σ is related to spectral density estimation at its origin and Γ is related to derivative of the spectral density, it is natural to consider pilot estimates of the form

$$\hat{\Sigma}^{(0)} = \hat{\gamma}(0) + \sum_{s=1}^{b-1} w_n(s)[\hat{\gamma}(s) + \hat{\gamma}(s)^T],$$

and

$$\hat{\Gamma}^{(0)} = -\sum_{s=1}^{b-1} w_n(s) \cdot s[\hat{\gamma}(s) + \hat{\gamma}(s)^T],$$

where $w_n(\cdot)$ is a window function. More specifically, *Tukey-Hanning* window and flat top window Politis and Romano (1995) are used in the above pilot estimates. These window functions again require optimal bandwidths to achieve an accurate coefficient estimate. In this section, we consider coefficient estimation and bandwidth selection in pilot estimates.

2.2.1 Flat top pilot estimate

Politis and Romano (1995) introduced a family of spectral variance estimators using flat top window functions. These window functions are constructed from windows such as *Bartlett* and *Tukey-Hanning* window by letting the function equal to 1 near zero. Bias reduction of flat top estimator was illustrated by Politis and Romano (1995). It was suggested by Politis (2003) to use flat top windows as pilot estimates in nonparametric density estimation. We apply a similar idea under MCMC context when estimating Σ and Γ.

Consider the following mSV with w_{ft} window for pilot estimate of Σ and Γ

$$\hat{\Sigma}_{ft}^{(0)} = \hat{\gamma}(0) + \sum_{s=1}^{b-1} w_{ft}(s)[\hat{\gamma}(s) + \hat{\gamma}(s)^T],$$

and

$$\hat{\Gamma}_{ft}^{(0)} = -\sum_{s=1}^{b-1} w_{ft}(s) \cdot s[\hat{\gamma}(s) + \hat{\gamma}(s)^T].$$

Asymptotic variance and bias of $\hat{\Sigma}_{ft}^{(0)}$ and $\hat{\Gamma}_{ft}^{(0)}$ are discussed by Politis (2003). Bandwidth of pilot estimate $\hat{\Sigma}_{ft}^{(0)}$ and $\hat{\Gamma}_{ft}^{(0)}$ follows a simplified version of *Empirical Rule* suggested byPolitis (2009).

Bandwidth of flat top pilot. Denote $\hat{\rho}_{ij}(s) = \hat{\gamma}_{ij}(s)/\sqrt{\hat{\gamma}_{ii}(0)\hat{\gamma}_{jj}(0)}$. Let $\rho(s) = \max_{1\leq i,j\leq p} |\hat{\rho}_{ij}(s)|$ and b_0 be the smallest positive integer such that $|\hat{\rho}(b_0+k)| < c\sqrt{\log n/n}$, for $k = 1,2,...K_n$, where $c > 0$ is a fixed constant and K_n is a positive, non-decreasing integer-valued function of n such that $K_n = o(\log n)$. Then $b = 2b_0$ is the bandwidth of $w_{ft}(\cdot)$.

Remark. Any $c > 0$ and $1 \leq K_n \leq n$ would work for the asymptotic result but for finite samples, it is suggested to chose $c = 2$ and $K_n = 5$ Politis (2003).

Remark. The rule aims to find the point b_0 after which sample autocorrelation are negligible. The original *Empirical rule* Politis (2009) suggests an optimal bandwidth for each entry of the matrix $\hat{\Sigma}_{ft}$ which should yield better performance for density estimation than using the same bandwidth for all entries. Due to the heavy computation encountered in MCMC context, this simplified version is applied when choosing b_0, which is more conservative.

2.2.2 Iterative plug-in pilot estimate

Iterative plug-in method was introduced by Brockmann et al. (1993) in nonparametric regression for independent observations. Bühlmann (1996) also applied the idea in nonparametric spectral density estimation on frequency domain of stationary time series. The idea is based on the circular relationship of $\hat{\Sigma}$ and its bandwidth b. In other words, b is required to estimate $\hat{\Sigma}$, while $\hat{\Sigma}$ is required in order to obtain an optimal b. If we starts from a b to estimate $\hat{\Sigma}$, this $\hat{\Sigma}$ can be used to obtain a better b, which again results in a better $\hat{\Sigma}$. Therefore an optimal bandwidth can be achieved by iteratively repeating the above procedure. A univariate iterative plug-in procedure was applied by Bühlmann (1996) to select bandwidth for spectral density at its origin, which is equivalent to the diagonal entries of Σ adjusted by a constant. Therefore we discuss this method when estimating bandwidth for the diagonal entries of $\hat{\Sigma}$ for illustration purpose. A multivariate iterative plug-in procedure should work in a similar way.

Suppose the goal is to estimate coefficient of the optimal bandwidth for $\hat{\Sigma}_{bt,ii}$. Then an optimal bandwidth that minimizes mean squared error of $\hat{\Sigma}_{bt,ii}$ is

$$\hat{b}_{bt,ii} = \left(\frac{3\Gamma_{ii}^2 n}{2\Sigma_{ii}^2} \right)^{1/3}.$$

Bühlmann (1996) discussed estimating Γ_{ii} and Σ_{ii} by

$$\hat{\Sigma}_{ii,\tilde{w}}^{(0)} = \hat{\gamma}_{ii}(0) + \sum_{s=1}^{b-1} \tilde{w}(s)[\hat{\gamma}_{ii}(s) + \hat{\gamma}_{ii}(s)],$$

and

$$\hat{\Gamma}_{ii,\bar{w}}^{(0)} = -\sum_{s=1}^{b-1} \bar{w}(s) \cdot s[\hat{\gamma}_{ii}(s) + \hat{\gamma}_{ii}(s)],$$

where

$$\tilde{w}(s) = \begin{cases} (1+\cos(\pi|s|/b))/2 & \text{for } \leq b \\ 0 & \text{for } |s| > b \end{cases}$$

and

$$\bar{w}(s) = \begin{cases} 1 & \text{for } |s| \leq 0.8b \\ (1+\cos\{5(|s|/b-0.8)\pi\})/2 & \text{for } 0.8b < |s| \leq b \\ 0 & \text{for } |s| > b \end{cases}$$

are *Tukey-Haning* window and flat top *Tukey-Hannig* window (split rectangular-cosine window in Bühlmann (1996)). Notice $\hat{b}_{bt,ii}$ is the same as optimal bandwidth for univariate *Bartlett* estimator derived by Flegal and Jones (2010).

Although *Empirical Rule* can be applied again to select bandwidth of $\bar{w}(\cdot)$, it is still challenging to obtain bandwidth for *Tukey-Hanning* window. Therefore pilot bandwidth of $\hat{\Sigma}^{(0)}_{ii,\tilde{w}}$ and $\hat{\Gamma}^{(0)}_{ii,\bar{w}}$ are both chosen according to the following iterative plug-in procedure Bühlmann (1996).

Bandwidth of iterative plug-in pilot. Let $b_0 = n$ be the starting bandwidth. For $t = 1,\ldots,4$, iterate

$$b_t = \Big[\frac{6\sum_{s=-b_{t-1}n^{-4/21}}^{b_{t-1}n^{-4/21}} \bar{w}^2(s)s^2\hat{\gamma}^2_{ii}(s)}{\sum_{s=-b_{t-1}n^{-4/21}}^{b_{t-1}n^{-4/21}} \hat{\gamma}^2_{ii}(s)}\Big]^{1/3} \cdot n^{1/3}. \tag{2.2.1}$$

The bandwidth for pilot estimate is $b = b_4 n^{-4/21}$.

Remark. (2.2.1) is a global step iteratively searching for an global optimal bandwidth

$$b_{global} = \Big(\frac{3\cdot\int_{-\pi}^{\pi}\{f^{(1)}(\lambda)\}^2 d\lambda}{\int_{-\pi}^{\pi}\{f(\lambda)\}^2 d\lambda}\Big)\cdot n^{1/3}$$

that minimizes mean integrated squared error $\text{MISE} = \text{E}[\int_{-\pi}^{\pi}(\hat{f}(\lambda)-f(\lambda))^2 d\lambda]$ for *Bartlett* window function at λ, where

$$f(\lambda) = \frac{1}{2\pi}\sum_{s=-\infty}^{\infty}\gamma_{ii}(s)e^{-is\lambda},$$

and

$$f^{(1)}(\lambda) = \frac{1}{2\pi} \sum_{s=-\infty}^{\infty} |s| \gamma_{ii}(s) e^{-is\lambda}.$$

If we estimate $\int_{-\pi}^{\pi} \{f(\lambda)\}^2$ and $\int_{-\pi}^{\pi} \{f^{(1)}(\lambda)\}^2$ in b_{global} by

$$\frac{1}{2} \int_{-\pi}^{\pi} \left[\frac{1}{2\pi} \sum_{s=-b}^{b} \hat{\gamma}_{ii}(s) e^{-is\lambda} \right]^2 d\lambda$$

and

$$\int_{-\pi}^{\pi} \left[\frac{1}{2\pi} \sum_{s=-b}^{b} \bar{w}(s) |s| \hat{\gamma}_{ii}(s) e^{-is\lambda} \right]^2 d\lambda$$

for a certain bandwidth b, and apply Parseval's identity

$$\int_{-\pi}^{\pi} \left[\sum_{s=-c}^{c} h(s) e^{-i\lambda s} \right]^2 d\lambda = \sum_{s=-c}^{c} h^2(s),$$

bandwidth expression of (2.2.1) can be obtained. Although Σ_{ii} and Γ_{ii} are related to local results of $f(0)$ and $f^{(1)}(0)$, the global step stabilizes the procedure and achieves correct asymptotic order, with $n^{-4/21}$ being an inflation factor that adjusts for an optimal rate. An interested reader is directed to Bühlmann (1996) for more details.

Remark. We considered iterative plug-in method for *Bartlett* estimator as a motivative example. Bühlmann (1996) also discussed iterative plug-in method for *Tukey-Hanning* SV estimator. Same idea should as well work for SV estimator with other window functions.

2.3 Examples

Three examples are considered to evaluate performance of the optimal bandwidth as a comparison to conventional bandwidth of $n^{1/3}$ and $n^{1/2}$. In the first univariate auto-regressive example, both iterative plug-in and flat top pilots are evaluated when estimating coefficient of bandwidth.

Multivariate examples based on real data and vectorized auto-regressive process are then considered to show the benefit of proposed optimal bandwidth.

2.3.1 Univariate auto-regressive example

Recall ε_i are i.i.d N(0,1). Consider the following autoregressive process of order 1 (AR(1)):

$$X_i = \phi X_{i-1} + \varepsilon_i \quad \text{for } i = 1, 2, ...$$

For $|\phi| < 1$, the Markov chain is geometrically ergodic with invariant distribution $\mathrm{N}(0, 1/(1-\phi)^2)$. Consider approximating $\theta = E[X_i]$ by $\theta_n = \bar{X}_n$. We would like to estimate

$$\sigma^2 = \mathrm{Var}[X_1] + 2\sum_{s=1}^{\infty} \mathrm{Cov}(X_t, X_{t+s}) = 1/(1-\phi)^2.$$

Since $\mathrm{cov}(X_1, X_i) = \phi^{i-1}/(1-\phi^2)$, large ϕ results in a Markov chain with high auto-correlation therefore we consider a range of ϕ. To calculate true coefficient of the optimal bandwidth, we need the following result

$$2\sum_{s=1}^{\infty} |s| \mathrm{Cov}(X_t, X_{t+s}) = \frac{2\phi}{(1-\phi^2)(1-\phi)^2}.$$

SV estimator with *Bartlett* window $\hat{\sigma}^2_{bt}$ and BM estimator $\hat{\sigma}^2_{BM}$ are applied to estimate σ^2 since optimal batch size of BM estimator also involves unknown quantities of the Markov chain Flegal and Jones (2010). We consider flat top and iterative plug-in pilot estimate of the optimal bandwidth for $\hat{\sigma}^2_{bt}$ and $\hat{\sigma}^2_{BM}$.

Consider 500 replications for each ϕ from 0.4 to 0.9. In each replication, generate 1e4 of AR(1) sample and apply flat top and iterative plug-in pilot to estimate coefficient of optimal bandwidth for $\hat{\sigma}^2_{bt}$ and $\hat{\sigma}^2_{BM}$. Figure 2.1 and 2.2 shows the coefficient results for $\hat{\sigma}_{BM}$ and $\hat{\sigma}_{bt}$. If $n^{1/2}$ is used to adjust for the unknown constant of $n^{1/3}$, it is equivalent to choose a constant of $n^{1/6}$ that is determined by sample size n, regardless of ϕ. For n={1e4, 1e5, 1e6}, the coefficient of $n^{1/6}$ is denoted by the dotted lines.

To see how bandwidth affects the estimation of σ^2, $\hat{\sigma}^2_{bt}$ and $\hat{\sigma}^2_{BM}$ are applied to AR(1) sample of lenght 1e4. Four different bandwidths are used, $n^{1/3}$, $n^{1/2}$, optimal bandwidth with flat top pilot and optimal bandwidth with iterative plug-in pilot. Results of 500 replications are also shown in figure 2.1 and 2.2 for ϕ from 0.7 to 0.96.

Coefficient plots in figure 2.1 and 2.2 indicate that both pilots estimates are close to the true optimal coefficient. Iterative plug-in method has a smaller variance. As a compensation, it requires more computation and has slower convergence rate compared with flat top pilot Politis (2003), hence we will use flat top pilot in the following examples. Nevertheless, the idea of optimal bandwidth is open to all legitimate pilot estimation.

With optimal bandwidth, performances of $\hat{\sigma}^2_{bt}$ and $\hat{\sigma}^2_{BM}$ are improved compared with a bandwidth of $n^{1/3}$ or $n^{1/2}$. Although $n^{1/3}$ is the correct asymptotic rate, when the coefficient is ignored, it results in poor performance especially for highly correlated Markov chain. In fact, figure 2.1 and 2.2 show how the true coefficient inflates as correlation increases, which explains why $n^{1/3}$ is poorly behaved for higher ϕ.

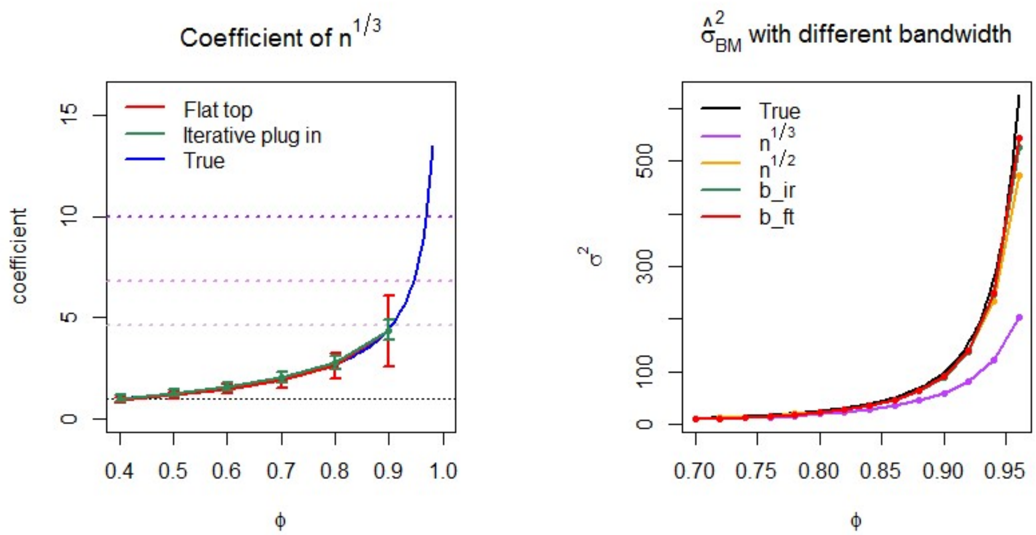


Figure 2.1 Left plot is average of estimated coefficient for $\hat{\sigma}^2_{BM}$ over 500 iterations with 95% CI. True value of coefficients are calculated as a comparison. From high to low, top three dotted lines denote $n^{1/6}$ when $n =$ 1e6, 1e5, 1e4. The bottom dotted line is reference when coefficient equals to 1. Right plot is average of 500 $\hat{\sigma}^2_{BM}$ with different bandwidth.

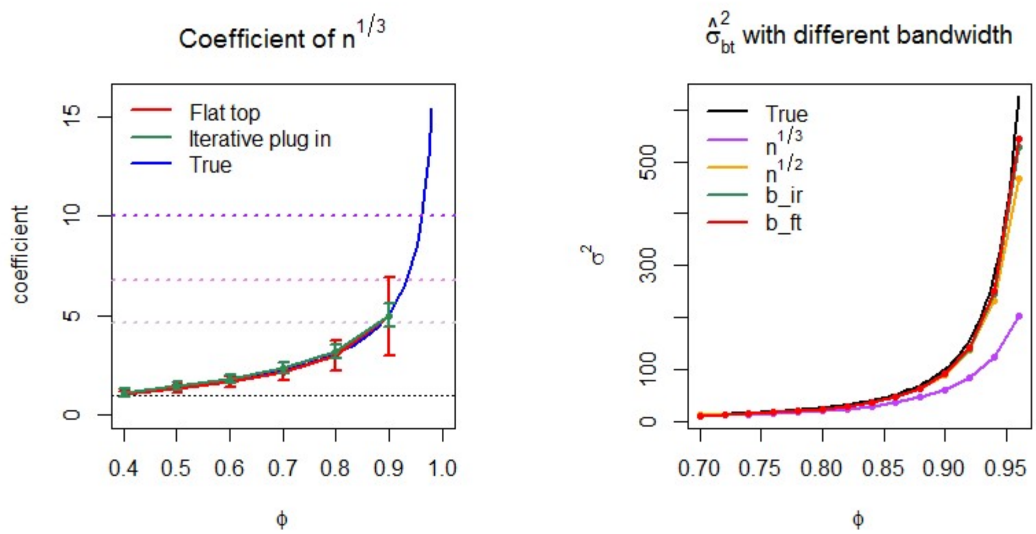


Figure 2.2 Left plot is average of estimated coefficient for $\hat{\sigma}^2_{bt}$ over 500 iterations with 95% CI. True value of coefficients are calculated as a comparison. From high to low, top three dotted lines denote $n^{1/6}$ when $n = $ 1e6, 1e5, 1e4. The bottom dotted line is reference when coefficient equals to 1. Right plot is average of 500 $\hat{\sigma}^2_{bt}$ with different bandwidth.

2.3.2 Vector auto-regressive example

For $i = 1, 2, ...$, recall p-dimensional vector autoregressive process of order 1 (VAR(1))

$$X_i = \Phi X_{i-1} + \varepsilon_i,$$

where $X_i \in \mathbb{R}^p$, ε_i are i.i.d $N_p(0, I_p)$ and Φ is a $p \times p$ matrix. Let $\otimes$ be the Kronecker product. When the largest eigenvalue of Φ in absolute value is less than 1, the Markov chain is geometrically ergodic Tjøstheim (1990) with invariant distribution $N_p(0,V)$, where $vec(V) = (I_{p^2} - \Phi \otimes \Phi)^{-1} vec(I_p)$. Consider approximating $\theta = EX_i$ by $\theta_n = \bar{X}_n$, we would like to estimate

$$\Sigma = \text{Var}[X_1] + 2\sum_{s=1}^{\infty} \text{Cov}(X_1, X_1 + s)$$

$$= (I_p - \Phi)^{-1} V + V (I_p - \Phi)^{-1} - V.$$

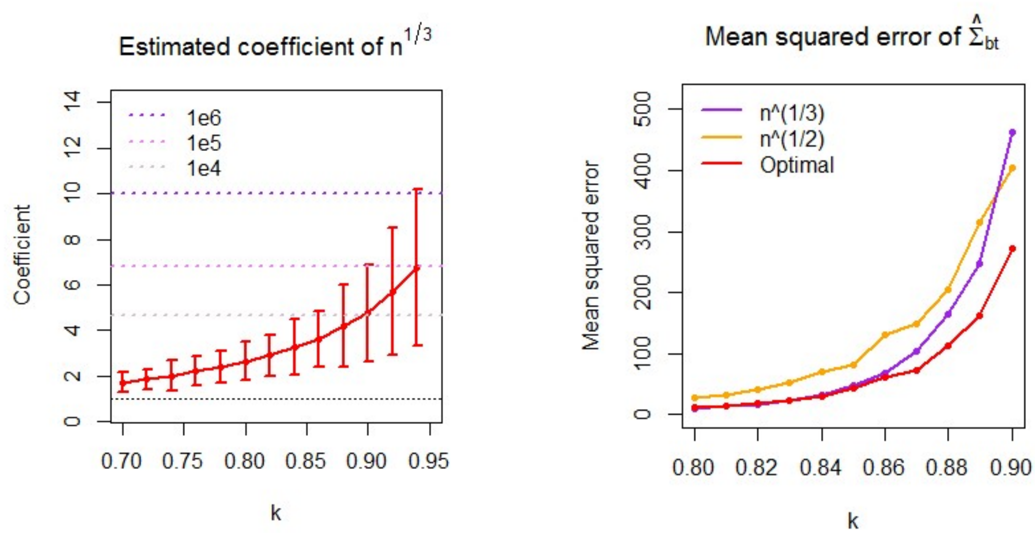


Figure 2.3 Left plot is average of estimated coefficient for $\hat{\Sigma}^2_{bt}$ over 500 iterations with 95% CI. From high to low, top three dotted lines denote $n^{1/6}$ when $n =$ 1e6, 1e5, 1e4. The bottom dotted line is reference when coefficient equals to 1. Right plot is average of 500 *mse* with different bandwidth.

We consider optimal bandwidth with flat top pilot when estimating Σ by $\hat{\Sigma}_{bt}$. Φ_0 is chosen as follows to guarantee geometric ergodicity of VAR(1) samples. Consider a $p \times p$ matrix A with each entry generated from standard normal distribution, let $B = AA^T$ be a symmetric matrix with the largest eigenvalue m, then $\Phi_0 = B/(m+0.001)$ leads to geometric ergodic chains. Then we evaluate a series of $\Phi = k \cdot \Phi_0$, where $k = \{0.8,\ 0.81, ...,\ 0.90\}$. Larger k implies stronger auto-covariance and cross auto-covariance of the chain.

For each Φ, optimal coefficient with flat top pilot is computed for VAR(1) of length 1e4. The averages over 500 replications are plotted in figure 2.3 with 95% confidence interval. $n^{1/6}$ for $n =$ 1e4, 1e5 and 1e6 are provided as references for bandwidth of $n^{1/2}$. Further more, $\hat{\Sigma}_{bt}$ is computed with $n^{1/3}$, $n^{1/2}$ and optimal bandwidth. Let $E = \hat{\Sigma}_{bt} - \Sigma$, mean squared error across entries of E can be reflected by

$$mse = \frac{1}{p^2} \sum_i \sum_j e_{ij}^2.$$

To compare performance of different bandwidth, mse of corresponding $\hat{\Sigma}_{bt}$ is calculated. Average over 500 replications in 2.3 shows that by using suggested optimal bandwidth, mse is significantly for various strength of autocorrelations.

2.3.3 Bayesian dynamic space-time example

This example is applied to monthly temperature data collected at 10 nearby station in northeastern United States in 2000, which is a subset of `NETemp` data described in `R` package `spBayes` Finley et al. (2013). A Bayesian dynamic model proposed by Gelfand et al. (2005) is fitted to the data and the model treats time as discrete and space as continuous variable.

Suppose y_t denote the temperature observed at location s and time t for $s = 1, 2, ..., N_s$ and $t = 1, 2, ..., N_t$. Let $x_t(s)$ be a $k \times 1$ vector of predictors and β_t be a $k \times 1$ coefficient vector, which is a purely time component. $u_t(s)$ denotes a space-time component. The model is

$$y_t(s) = \boldsymbol{x}_t(s)^T \boldsymbol{\beta}_t + u_t(s) + \varepsilon_t(s), \;\; \varepsilon_t \sim N(0, \tau_t^2),$$

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \boldsymbol{\eta}_t; \;\; \boldsymbol{\eta}_t \sim N_p(0, \Sigma_\eta),$$

$$u_t(s) = u_{t-1}(s) + w_t(s); \;\; w_t(s) \sim GP(0, C_t(\cdot, \sigma_t^2, \phi_t)).$$

$GP(0, C_t(\cdot, \sigma_t^2, \phi_t))$ is a spatial Gaussian process where $C_t(s_1, s_2; \sigma_t^2, \phi_t) = \sigma_t^2 \rho(s_1, s_2; \phi_t)$. $\rho(\cdot; \phi)$ is an exponential correlation function with ϕ controlling the correlation decay, and σ_t^2 represents the spatial variance components. The Gaussian spacial process allows closer location to have higher correlation. Time effect for both $\boldsymbol{\beta}_t$ and $u_t(s)$ are characterized by transition equations to achieve reasonable dependence structure. We are interested in estimating posterior expectation of 185 parameters $\theta = (\boldsymbol{\beta}_t,\, u_t(s),\, \sigma_t^2,\, \Sigma_\eta,\, \tau_t^2,\, \phi_t)$, their prior follows `spDynlM` function in `spBayes` package.

Elevation is the only predictor in this example, hence $\boldsymbol{\beta}_t = (\beta_t^{(0)},\, \beta_t^{(1)})^T$ for $t = 1, 2, ..., 12$, where $\beta_t^{(0)}$ is intercept and $\beta_t^{(1)}$ is coefficient of covariate elevation. Consider the intercept parameter for two consecutive month $\beta_1^{(0)}$ and $\beta_2^{(0)}$. Let $\{X_i = (\beta_{1,i}^{(0)}, \beta_{2,i}^{(0)})^T\}_{i=1}^{\infty}$ be posterior sample of $(\beta_1^{(0)},\, \beta_2^{(0)})$ generated from `spDynlM` function. We are interested in estimating

$$\Sigma = \mathrm{Var}(X_i) + \sum_{s=1}^{\infty} [\gamma(s) + \gamma(s)^T]$$

by $\hat{\Sigma}_{bt}$ with different bandwidth.

An average over 20 chains of length 1e7 is used to approximate true value of $\beta_1^{(0)}$ and $\beta_2^{(0)}$. Then consider 1000 replications of the following. For each replication, $\hat{\Sigma}_{bt}$ with bandwidth $n^{1/3}$, $n^{1/2}$ and $\hat{b}_{opt}$ are calculated for chain length n={1e4, 5e4, 1e5, 2e5}, then 90% confidence intervals based on $\hat{\Sigma}_{bt}$ is used to check whether the true $(\beta_1^{(0)},\, \beta_2^{(0)})$ is captured. Capture rates from 1000 replications in table 2.1 show that $\hat{\Sigma}_{bt}$ with bandwidth $\hat{b}_{opt}$ leads to captures rates closer to 90%

compared with $n^{1/3}$ and $n^{1/2}$, suggesting a better estimate of Σ. Figure 2.4 plots confidence regions for $(\beta_1^{(0)},\ \beta_2^{(0)})$ and $(\beta_1^{(0)},\ \sigma_1^2)$ constructed from $\hat{\Sigma}_{bt}$. Bandwidth of $n^{1/3}$ ignores the proportionality constants and leads to a much smaller ellipse.

Table 2.1 Capture rate of 90% confidence region for $(\beta_1^{(0)},\ \beta_2^{(0)})$ based on $\hat{\Sigma}_{bt}$ over 1000 replications.

Bandwidth	n=1e4	=5e4	n=1e5	n=2e5
$n^{1/3}$	0.371	0.479	0.496	0.561
$n^{1/2}$	0.625	0.747	0.778	0.843
$\hat{b}_{opt}$	0.763	0.845	0.865	0.902

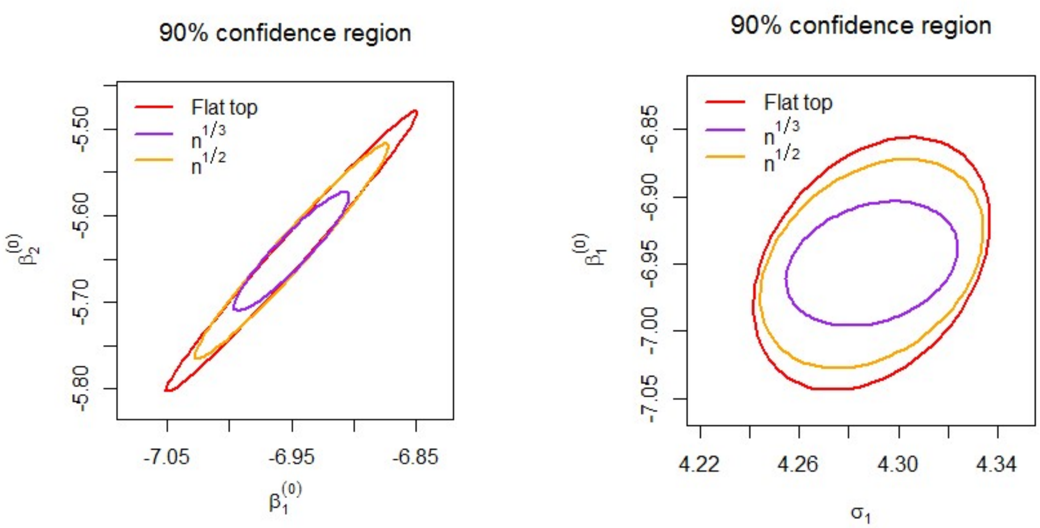


Figure 2.4 Confidence regions for $(\beta_1^{(0)},\ \beta_2^{(0)})$ and $(\beta_1^{(0)},\ \sigma_1)$ based on $\hat{\Sigma}_{bt}$ and a chain length of 1e5.

2.4 Discussion

In this chapter, we consider asymptotic mean squared consistency and optimal bandwidth selection for a class of multivariate spectral variance estimator, under conditions that are easy to check under MCMC setting. Although mainly focused on mSV estimator, the suggested optimal bandwidth can also be applied to other estimate of Σ, such as batch means and overlapping batch means method. Diagonal element of Σ is equivalent to univariate variance of CLT, hence univariate optimal bandwidth selection of SV, BM and OBM follows the same procedure.

To reduce computational effort, we used the first 1e4 MCMC samples for flat top pilot estimats regardless of the chain length used to estimate θ and Σ. As a result, performance of the estimator is significantly improved without adding much computation, which is crucial in practice especially for multivariate problems. One can also use the last 1e4 samples for better performances. We choose 1e4 as a compromise of computation and accuracy based on simulation. In fact, there is a major improvement even if only 1e3 pilot samples are used. Hence the length of pilot samples can be determined according to computation resources available.

Asymptotic variance of mSV requires the chain to be polynomial ergodic, while the bias results are based on uniformly ergodicity. Bias results under polynomial ergodicity is of interest and requires future work.

2.A Appendix for Chapter 2

A Brownian motion and propositions

Denote

$$\hat{\gamma}(s) = \frac{1}{n}\sum_t (Y_t - \bar{Y})(Y_{t+s} - \bar{Y})^T = \frac{1}{n}\sum_t V_t V_{t+s}^T,$$

later we will show that the difference between $\hat{\Sigma}_s$ and $\hat{\Sigma}_w$ is

$$d = \frac{1}{n}\left[\sum_{h=1}^{b}\Delta_1 w_n(h)\sum_{r=1}^{h-1}V_rV_r^T + \sum_{s=1}^{b-1}\sum_{h=1}^{b-s}\Delta_1 w_n(s+h)\left(\sum_{r=1}^{h-1}(V_rV_{r+s}^T + V_{r+s}V_r^T)\right)\right]$$
$$+\frac{1}{n}\left[\sum_{h=1}^{b}\sum_{r=n-b+h+1}^{n}\Delta_1 w_n(n-r+h+1)V_rV_r^T + \sum_{s=1}^{b-1}\sum_{h=1}^{b-s}\sum_{r=n-b+h+1}^{n-s}\Delta_1 w_n(n-r+h+1)(V_rV_{r+s}^T + V_{r+s}V_r^T)\right].$$

Let $B = \{B(t), t \geq 0\}$ be a p-dimensional standard Brownian motion. Denote $\tilde{\gamma}(s)$, $\tilde{\Sigma}_s$, $\tilde{\Sigma}_w$ and $\tilde{d}$ the Brownian motion analog of $\hat{\gamma}(s)$, $\hat{\Sigma}_s$, $\hat{\Sigma}_w$ and $\tilde{d}$. Specifically, define Brownian motion increments $U_t = B(t) - B(t-1)$ for $t = 1, ..., n$, then $U_1, ...U_n \sim N_p(0,\ I_p)$. Let $\bar{B} = n^{-1}B(n)$ and $T_t = U_t - \bar{B}$. then

$$\tilde{\gamma}(s) = \frac{1}{n}\sum_t (U_t - \bar{B})(U_{t+s} - \bar{B})^T = \frac{1}{n}\sum_t T_t T_{t+s}^T,$$

$$\tilde{\Sigma}_s = \tilde{\gamma}(0) + \sum_{s=1}^{b-1} w_n(s)[\tilde{\gamma}(s) + \tilde{\gamma}(s)^T],$$

$$\tilde{\Sigma}_w = \frac{1}{n}\sum_{k=1}^{b}\sum_{l=0}^{n-k} k^2\Delta_2 w_n(k)[\bar{B}_l(k) - \bar{B}][\bar{B}_l(k) - \bar{B}]^T,$$

where $\bar{B}_l(k) = k^{-1}(B(l+k) - B(l))$ for $l = 0, ..., (n-k)$, and

$$\tilde{d} = \frac{1}{n}\left[\sum_{h=1}^{b}\Delta_1 w_n(h)\sum_{r=1}^{h-1}T_rT_r^T + \sum_{s=1}^{b-1}\sum_{h=1}^{b-s}\Delta_1 w_n(s+h)\left(\sum_{r=1}^{h-1}(T_rT_{r+s}^T + T_{r+s}T_r^T)\right)\right]$$
$$+\frac{1}{n}\left[\sum_{h=1}^{b}\sum_{r=n-b+h+1}^{n}\Delta_1 w_n(n-r+h+1)T_rT_r^T + \sum_{s=1}^{b-1}\sum_{h=1}^{b-s}\sum_{r=n-b+h+1}^{n-s}\Delta_1 w_n(n-r+h+1)(T_rT_{r+s}^T + T_{r+s}T_r^T)\right].$$

Also define the following matrix:

$$\tilde{\Sigma}_{w,L} = \frac{1}{n}\sum_{k=1}^{b}\sum_{l=0}^{n-k} k^2\Delta_2 w_n(k)L[\bar{B}_l(k) - \bar{B}][\bar{B}_l(k) - \bar{B}]^T L^T,$$

where L is the lower triangular matrix satisfying $\Sigma = LL^T$. Let $B^{(i)}(t)$ be the ith component of vector $B(t)$, we need some properties of Brownian Motion.

Proposition 2. Csörgo and Révész (2014) Suppose *Condition 3* holds, then for all $\varepsilon > 0$ and for almost all sample paths, there exists $n_0(\varepsilon)$ such that for all $n \geq n_0$ and all $i = 1, ..., p$

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

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

$$|B^{(i)}(n)| < (1+\varepsilon)\sqrt{2nloglogn}.$$

Recall $\Sigma = LL^T$ where L is a lower triangular matrix. Define $C(t) := LB(t)$, let $C^{(i)}(t)$ be the ith component of $C(t)$ and define $\bar{C}_l^{(i)}(k) = k^{-1}(C^{(i)}(l+k) - C^{(i)}(l))$, $\bar{C}^{(i)} = n^{-1}C^{(i)}(n)$. We have the following propositions.

Proposition 3. Vats et al. (2015b) For all $\varepsilon > 0$ and for almost all sample paths, there exists $n_0(\varepsilon)$ such that for all $n \geq n_0$ and all $i = 1, ..., p$

$$|C^{(i)}(n)| < (1+\varepsilon)(2n\Sigma_{ii}loglogn)^{1/2},$$

where Σ_{ii} is the ith diagonal entry of Σ.

Proposition 4. Vats et al. (2015b) If *condition 3* holds, then for all $\varepsilon > 0$ and for almost all sample paths, there exists $n_0(\varepsilon)$ such that for all $n \geq n_0$ and all $i = 1, ..., p$

$$|\bar{C}_l^{(i)}(k)| \leq \frac{1}{k} \sup_{0\leq l\leq n-b} \sup_{0\leq s\leq b} |C^{(i)}(l+s) - C^{(i)}(l)| < \frac{1}{k} 2(1+\varepsilon)(b\Sigma_{ii}logn)^{1/2},$$

where Σ_{ii} is the ith diagonal entry of Σ.

Proposition 5. If variable X and Y are jointly normally distributed with

$$\begin{bmatrix} X \\ Y \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} l_{11} & l_{12} \\ l_{12} & l_{22} \end{bmatrix}\right),$$

then $E[X^2Y^2] = 2l_{12}^2 + l_{11}l_{22}$.

Proof. Let

$$\begin{bmatrix} l_{11} & l_{12} \\ l_{12} & l_{22} \end{bmatrix} = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}.$$

Suppose Z_1 and Z_2 are two independent standard normal variables, rewrite X and Y by $X = \sigma_x Z_1$ and $Y = \sigma_Y[\rho Z_1 + \sqrt{1-\rho^2}Z_2]$, then

$$\begin{aligned} E[X^2Y^2] &= E[\sigma_x^2 Z_1^2[\sigma_y(\rho Z_1 + \sqrt{1-\rho^2}Z_2)]^2] \\ &= E[\rho^2\sigma_x^2\sigma_y^2 Z_1^4] + (1-\rho^2)\sigma_x^2\sigma_y^2 E[Z_1^2Z_2^2] + 2\rho\sqrt{1-\rho^2}\sigma_x^2\sigma_y^2 E[Z_1^3]E[Z_2] \\ &= l_{12}^2 E[Z_1^4] + (l_{11}l_{22} - l_{12}^2)E[Z_1^2]E[Z_2^2] \\ &= 2l_{12}^2 + l_{11}l_{22}. \end{aligned}$$

Proposition 6. Janssen and Stoica (1987) If X_1, X_2, X_3 and X_4 are jointly normally distributed with mean 0, then

$$E[X_1X_2X_3X_4] = E[X_1X_2]E[X_3X_4] + E[X_1X_3]E[X_2X_4] + E[X_1X_4]E[X_2X_3].$$

B Proof of proposition 1

We will first show that $\tilde{\Sigma}_w = \tilde{\Sigma}_s - \tilde{d}$ and $\hat{\Sigma}_w = \hat{\Sigma}_s - d$. *Proposition* 1 can be proved by showing that $\tilde{d} \to 0$ and $d \to 0$.

Lemma 1(**). Under *condition 2*, $\tilde{\Sigma}_w = \tilde{\Sigma}_s - \tilde{d}$.

Proof. The proof is similar as the proof for theorem 3.1 in Damerdji (1991). We need the following straightforward results.

$$\Delta_1 w_n(l) = \sum_{k=l}^{b} \Delta_2 w_n(k). \tag{2.A.1}$$

$$\sum_{l=s+1}^{b} \Delta_1 w_n(l) = w_n(s), \ \sum_{l=1}^{b} \Delta_1 w_n(l) = 1. \tag{2.A.2}$$

We will prove that for $i, j = 1, ..., p$, $\tilde{\Sigma}_{w,ij} = \tilde{\Sigma}_{s,ij} - \tilde{d}_{ij}$. Notice that

$$\begin{aligned}
&(\bar{B}_l^{(i)}(k) - \bar{B}^{(i)})(\bar{B}_l^{(j)}(k) - \bar{B}^{(j)}) \\
&= \left(\frac{1}{k}[(U_{l+1}^{(i)} + U_{l+2}^{(i)} + \cdots + U_{l+k}^{(i)}) - k\bar{B}^{(i)}]\right)\left(\frac{1}{k}[(U_{l+1}^{(j)} + U_{l+2}^{(j)} + \cdots + U_{l+k}^{(j)}) - k\bar{B}^{(j)}]\right) \\
&= \frac{1}{k^2}\left(\sum_{h=1}^{k} T_{l+h}^{(i)}\right)\left(\sum_{h=1}^{k} T_{l+h}^{(j)}\right) \\
&= \frac{1}{k^2}\left(\sum_{h=1}^{k} T_{l+h}^{(i)} T_{l+h}^{(j)} + \sum_{s=1}^{k-1}\sum_{h=1}^{k-s} T_{l+h}^{(i)} T_{l+h+s}^{(j)} + \sum_{s=1}^{k-1}\sum_{h=1}^{k-s} T_{l+h}^{(j)} T_{l+h+s}^{(i)}\right).
\end{aligned}$$

For simplicity, denote $w_n(k)$ by w_k. Plug in the above expression,

$$\begin{aligned}
\tilde{\Sigma}_{w,ij} &= \frac{1}{n}\sum_{k=1}^{b}\sum_{l=0}^{n-k} k^2 \Delta_2 w_k (\bar{B}_l^{(i)}(k) - \bar{B}^{(i)})(\bar{B}_l^{(j)}(k) - \bar{B}^{(j)}) \\
&= \frac{1}{n}\sum_{k=1}^{b}\sum_{l=0}^{n-k} \Delta_2 w_k \sum_{h=1}^{k} T_{l+h}^{(i)} T_{l+h}^{(j)} + \frac{1}{n}\sum_{k=1}^{b}\sum_{l=0}^{n-k} \Delta_2 w_k \sum_{s=1}^{k-1}\sum_{h=1}^{k-s} (T_{l+h}^{(i)} T_{l+h+s}^{(j)} + T_{l+h}^{(j)} T_{l+h+s}^{(i)}) \\
&:= \mathrm{I} + \mathrm{II}_{ij} + \mathrm{II}_{ji}.
\end{aligned}$$

Recall that

$$\tilde{\gamma}_{ij}(s) = \frac{1}{n}\sum_t (U_t^{(i)} - \bar{B}^{(i)})(U_t^{(j)} - \bar{B}^{(j)}) = \frac{1}{n}\sum_t T_t^{(i)} T_t^{(j)}.$$

Change the order of sums in I and apply (2.A.1)

$$\begin{aligned}
\mathrm{I} &= \frac{1}{n}\sum_{k=1}^{b}\sum_{l=0}^{n-k}\sum_{h=1}^{k} \Delta_2 w_k T_{l+h}^{(i)} T_{l+h}^{(j)} = \frac{1}{n}\sum_{k=1}^{b}\sum_{h=1}^{k}\sum_{l=0}^{n-k} \Delta_2 w_k T_{l+h}^{(i)} T_{l+h}^{(j)} = \frac{1}{n}\sum_{h=1}^{b}\sum_{k=h}^{b}\sum_{l=0}^{n-k} \Delta_2 w_k T_{l+h}^{(i)} T_{l+h}^{(j)} \\
&= \frac{1}{n}\sum_{h=1}^{b}\sum_{l=0}^{n-h}\sum_{k=h}^{b} \Delta_2 w_k T_{l+h}^{(i)} T_{l+h}^{(j)} - \frac{1}{n}\sum_{h=1}^{b}\sum_{l=n-b+1}^{n-h}\sum_{k=n-l+1}^{b} \Delta_2 w_k T_{l+h}^{(i)} T_{l+h}^{(j)} \\
&= \frac{1}{n}\sum_{h=1}^{b}\sum_{l=0}^{n-h} T_{l+h}^{(i)} T_{l+h}^{(j)} \sum_{k=h}^{b} \Delta_2 w_k - \frac{1}{n}\sum_{h=1}^{b}\sum_{l=n-b+1}^{n-h} T_{l+h}^{(i)} T_{l+h}^{(j)} \sum_{k=n-l+1}^{b} \Delta_2 w_k \\
&= \frac{1}{n}\sum_{h=1}^{b}\sum_{l=0}^{n-h} T_{l+h}^{(i)} T_{l+h}^{(j)} \Delta_1 w_h - \frac{1}{n}\sum_{h=1}^{b}\sum_{l=n-b+1}^{n-h} T_{l+h}^{(i)} T_{l+h}^{(j)} \Delta_1 w_{n-l+1} \\
&= \frac{1}{n}\sum_{h=1}^{b} \Delta_1 w_h \sum_{l=0}^{n-h} T_{l+h}^{(i)} T_{l+h}^{(j)} - \frac{1}{n}\sum_{h=1}^{b}\sum_{l=n-b+1}^{n-h} T_{l+h}^{(i)} T_{l+h}^{(j)} \Delta_1 w_{n-l+1}
\end{aligned}$$

$$= \sum_{h=1}^{b} \Delta_1 w_h \left[\tilde{\gamma}_{ij}(0) - n^{-1}(T_1^{(i)}T_1^{(j)} + \cdots + T_{h-1}^{(i)}T_{h-1}^{(j)})\right] - \frac{1}{n}\sum_{h=1}^{b}\sum_{l=n-b+1}^{n-h} T_{l+h}^{(i)}T_{l+h}^{(j)}\Delta_1 w_{n-l+1}$$

$$= \tilde{\gamma}_{ij}(0)\sum_{h=1}^{b}\Delta_1 w_h - \frac{1}{n}\sum_{h=1}^{b}\Big[\Delta_1 w_h(T_1^{(i)}T_1^{(j)} + T_2^{(i)}T_2^{(j)} + \ldots T_{h-1}^{(i)}T_{h-1}^{(j)})$$

$$+\Delta_1 w_b T_{n-b+h+1}^{(i)}T_{n-b+h+1}^{(j)} + \Delta_1 w_{b-1}T_{n-b+h+2}^{(i)}T_{n-b+h+2}^{(j)} + \ldots + \Delta_1 w_{h+1}T_n^{(i)}T_n^{(j)}\Big]$$

$$= \tilde{\gamma}_{ij}(0) - \frac{1}{n}\sum_{h=1}^{b}\left[\Delta_1 w_h \sum_{r=1}^{h-1} T_r^{(i)}T_r^{(j)} + \sum_{r=n-b+h+1}^{n}\Delta_1 w_{n-r+h+1}T_r^{(i)}T_r^{(j)}\right].$$

For II_{ij}, apply (2.A.1) and ,

$$\mathrm{II}_{ij} = \frac{1}{n}\sum_{k=1}^{b}\sum_{l=0}^{n-k}\sum_{s=1}^{k-1}\sum_{h=1}^{k-s}\Delta_2 w_k T_{l+h}^{(i)}T_{l+h+s}^{(j)} = \frac{1}{n}\sum_{k=1}^{b}\sum_{s=1}^{k-1}\sum_{l=0}^{n-k}\sum_{h=1}^{k-s}\Delta_2 w_k T_{l+h}^{(i)}T_{l+h+s}^{(j)}$$

$$= \frac{1}{n}\sum_{k=1}^{b}\sum_{s=1}^{k-1}\sum_{h=1}^{k-s}\sum_{l=0}^{n-k}\Delta_2 w_k T_{l+h}^{(i)}T_{l+h+s}^{(j)} = \frac{1}{n}\sum_{s=1}^{b-1}\sum_{k=s+1}^{b}\sum_{h=1}^{k-s}\sum_{l=0}^{n-k}\Delta_2 w_k T_{l+h}^{(i)}T_{l+h+s}^{(j)}$$

$$= \frac{1}{n}\sum_{s=1}^{b-1}\sum_{h=1}^{b-s}\sum_{k=h+s}^{b}\sum_{l=0}^{n-k}\Delta_2 w_k T_{l+h}^{(i)}T_{l+h+s}^{(j)}$$

$$= \frac{1}{n}\sum_{s=1}^{b-1}\sum_{h=1}^{b-s}\sum_{l=0}^{n-h-s}\sum_{k=h+s}^{b}\Delta_2 w_k T_{l+h}^{(i)}T_{l+h+s}^{(j)} - \frac{1}{n}\sum_{s=1}^{b-1}\sum_{h=1}^{b-s}\sum_{l=n-b+1}^{n-h-s}\sum_{k=n-l+1}^{b}\Delta_2 w_k T_{l+h}^{(i)}T_{l+h+s}^{(j)}$$

$$= \frac{1}{n}\sum_{s=1}^{b-1}\sum_{h=1}^{b-s}\sum_{l=0}^{n-h-s}T_{l+h}^{(i)}T_{l+h+s}^{(j)}\sum_{k=h+s}^{b}\Delta_2 w_k - \frac{1}{n}\sum_{s=1}^{b-1}\sum_{h=1}^{b-s}\sum_{l=n-b+1}^{n-h-s}T_{l+h}^{(i)}T_{l+h+s}^{(j)}\sum_{k=n-l+1}^{b}\Delta_2 w_k$$

$$= \frac{1}{n}\sum_{s=1}^{b-1}\sum_{h=1}^{b-s}\sum_{l=0}^{n-h-s}T_{l+h}^{(i)}T_{l+h+s}^{(j)}\Delta_1 w_{h+s} - \frac{1}{n}\sum_{s=1}^{b-1}\sum_{h=1}^{b-s}\sum_{l=n-b+1}^{n-h-s}T_{l+h}^{(i)}T_{l+h+s}^{(j)}\Delta_1 w_{n-l+1}$$

$$= \frac{1}{n}\sum_{s=1}^{b-1}\sum_{h=1}^{b-s}\Delta_1 w_{h+s}\sum_{l=0}^{n-h-s}T_{l+h}^{(i)}T_{l+h+s}^{(j)} - \frac{1}{n}\sum_{s=1}^{b-1}\sum_{h=1}^{b-s}\sum_{l=n-b+1}^{n-h-s}\Delta_1 w_{n-l+1}T_{h-1}^{(i)}T_{h-1+s}^{(j)}$$

$$= \sum_{s=1}^{b-1}\sum_{h=1}^{b-s}\Delta_1 w_{h+s}\left[\tilde{\gamma}_{ij}(s) - n^{-1}(T_1^{(i)}T_{1+s}^{(j)} + \ldots + T_{h+1}^{(i)}T_{h+1+s}^{(j)})\right]$$

$$- \frac{1}{n}\sum_{s=1}^{b-1}\sum_{h=1}^{b-s}\sum_{l=n-b+1}^{n-h-s}\Delta_1 w_{n-l+1}T_{l+h}^{(i)}T_{l+h+s}^{(j)}$$

$$= \sum_{s=1}^{b-1}\sum_{h=1}^{b-s}\Delta_1 w_{h+s}\tilde{\gamma}_{ij}(s) - \frac{1}{n}\sum_{s=1}^{b-1}\sum_{h=1}^{b-s}\Big[\Delta_1 w_{h+s}(T_1^{(i)}T_{1+s}^{(j)} + \ldots + T_{h-1}^{(i)}T_{h-1+s}^{(j)})$$

$$+\Delta_1 w_b T_{n-b+h+1}^{(i)}T_{n-b+h+1+s}^{(j)} + \Delta_1 w_{b-1}T_{n-b+h+2}^{(i)}T_{n-b+h+2+s}^{(j)} + \ldots + \Delta_1 w_{h+s+1}T_{n-s}^{(i)}T_n^{(j)}\Big]$$

$$= \sum_{s=1}^{b-1} w_s\tilde{\gamma}_{ij}(s) - \frac{1}{n}\sum_{s=1}^{b-1}\sum_{h=1}^{b-s}\left[\Delta_1 w_{h+s}\sum_{r=1}^{h-1}T_r^{(i)}T_{r+s}^{(j)} + \sum_{r=n-b+h+1}^{n-s} w_{n-r+h+1}T_r^{(i)}T_{r+s}^{(j)}\right].$$

Notice $\tilde{\gamma}_{ij}(s) = \tilde{\gamma}_{ji}(-s)$,

$$\begin{aligned}
\mathrm{II}_{ji} &= \sum_{s=1}^{b-1}\sum_{h=1}^{b-s}\Delta_1 w_{h+s}\tilde{\gamma}_{ji}(s) - \frac{1}{n}\sum_{s=1}^{b-1}\sum_{h=1}^{b-s}\Big[\Delta_1 w_{h+s}(T_1^{(j)}T_{1+s}^{(i)} + \ldots + T_{h-1}^{(j)}T_{h-1+s}^{(i)}) \\
&\quad + \Delta_1 w_b T_{n-b+h+1}^{(j)}T_{n-b+h+1+s}^{(i)} + \Delta_1 w_{b-1}T_{n-b+h+2}^{(j)}T_{n-b+h+2+s}^{(i)} + \ldots + \Delta_1 w_{h+s+1}T_{n-s}^{(j)}T_n^{(i)}\Big] \\
&= \sum_{s=1}^{b-1} w_s\tilde{\gamma}_{ji}(s) - \frac{1}{n}\sum_{s=1}^{b-1}\sum_{h=1}^{b-s}\left[\Delta_1 w_{h+s}\sum_{r=1}^{h-1}T_r^{(j)}T_{r+s}^{(i)} + \sum_{r=n-b+h+1}^{n-s} w_{n-r+h+1}T_r^{(j)}T_{r+s}^{(i)}\right].
\end{aligned}$$

Combine I, II_{ij} and II_{ji},

$$\begin{aligned}
\tilde{\Sigma}_{w,ij} &= \mathrm{I} + \mathrm{II}_{ij} + \mathrm{II}_{ji} = \left[\tilde{\gamma}_{ij}(0) + \sum_{s=1}^{b-1} w_s\tilde{\gamma}_{ij}(s) + \sum_{s=1}^{b-1} w_s\tilde{\gamma}_{ji}(s)\right] \\
&\quad - \frac{1}{n}\sum_{h=1}^{b}\left[\Delta_1 w_h\sum_{r=1}^{h-1}T_r^{(i)}T_r^{(j)} + \sum_{r=n-b+h+1}^{n}\Delta_1 w_{n-r+h+1}T_r^{(i)}T_r^{(j)}\right] \\
&\quad - \frac{1}{n}\sum_{s=1}^{b-1}\sum_{h=1}^{b-s}\left[\Delta_1 w_{h+s}\sum_{r=1}^{h-1}(T_r^{(i)}T_{r+s}^{(j)} + T_r^{(j)}T_{r+s}^{(i)}) + \sum_{r=n-b+h+1}^{n-s} w_{n-r+h+1}(T_r^{(i)}T_{r+s}^{(j)} + T_r^{(j)}T_{r+s}^{(i)})\right] \\
&= \tilde{\Sigma}_{s,ij} - \tilde{d}_{ij}.
\end{aligned}$$

The next three lemmas show that $\tilde{d}_{ij} \to 0$ as $n \to \infty$.

Lemma 2. (Vats et al. (2015b) lemma 6) Under condition 2, 3, if as $n \to \infty$,

$$bn^{-1}\sum_{k=1}^{b} k|\Delta_1 w_n(k)| \to 0,$$

then

$$\frac{b}{n}\left(\sum_{t=1}^{b}|\Delta_1 w_n(t)| + 2\sum_{s=1}^{b-1}\sum_{t=1}^{b-s}|\Delta_1 w_n(s+t)|\right) \to 0.$$

*Lemma 3**. Let condition 2, 3 hold and $n > 2b$. If as $n \to \infty$,

$$bn^{-1}\sum_{k=1}^{b} k|\Delta_1 w_n(k)| \to 0,$$

then $\tilde{d} \to 0$ *w.p.* 1.

Proof. For $i, j = 1, ..., p$, we prove that $\tilde{d}_{ij} \to 0$ $w.p.1$. Using the inequality $|ab| \leq (a^2 + b^2)/2$, for $h = 1, ..., b$,

$$\sum_{r=1}^{h-1} |T_r^{(i)} T_r^{(j)}| \leq \frac{1}{2} \sum_{r=1}^{h-1} [(T_r^{(i)})^2 + (T_r^{(j)})^2] \leq \frac{1}{2} \sum_{r=1}^{2b} (T_r^{(i)})^2 + \frac{1}{2} \sum_{r=1}^{2b} (T_r^{(j)})^2.$$

Similarly,

$$\begin{aligned}
&\sum_{r=1}^{h-1} (|T_r^{(i)} T_{r+s}^{(j)}| + |T_r^{(j)} T_{r+s}^{(i)}|) \\
&\leq \frac{1}{2} \sum_{r=1}^{2b} (T_r^{(i)})^2 + \frac{1}{2} \sum_{r=1}^{b} (T_{r+s}^{(j)})^2 + \frac{1}{2} \sum_{r=1}^{2b} (T_r^{(j)})^2 + \frac{1}{2} \sum_{r=1}^{b} (T_{r+s}^{(i)})^2 \\
&\leq \frac{1}{2} \sum_{r=1}^{2b} (T_r^{(i)})^2 + \frac{1}{2} \sum_{r=1}^{2b} (T_r^{(j)})^2 + \frac{1}{2} \sum_{r=1}^{2b} (T_r^{(j)})^2 + \frac{1}{2} \sum_{r=1}^{2b} (T_r^{(i)})^2 \\
&= \sum_{r=1}^{2b} (T_r^{(i)})^2 + \sum_{r=1}^{2b} (T_r^{(j)})^2,
\end{aligned}$$

therefore

$$\begin{aligned}
|\tilde{d}_{ij}| &= \left| \frac{1}{n} \left(\sum_{h=1}^{b} \Delta_1 w_n(h) \sum_{r=1}^{h-1} T_r^{(i)} T_r^{(j)} + \sum_{s=1}^{b-1} \sum_{h=1}^{b-s} \Delta_1 w_n(s+h) \sum_{r=1}^{h-1} [T_r^{(i)} T_{r+s}^{(j)} + T_r^{(j)} T_{r+s}^{(i)}] \right) \right| \\
&\leq \frac{1}{n} \left(\sum_{h=1}^{b} |\Delta_1 w_n(h)| \cdot \sum_{r=1}^{h-1} |T_r^{(i)} T_r^{(j)}| + \sum_{s=1}^{b-1} \sum_{h=1}^{b-s} |\Delta_1 w_n(s+h)| \cdot \sum_{r=1}^{h-1} (|T_r^{(i)} T_{r+s}^{(j)}| + |T_r^{(j)} T_{r+s}^{(i)}|) \right) \\
&\leq \frac{1}{n} \sum_{h=1}^{b} |\Delta_1 w_n(h)| \cdot \left(\frac{1}{2} \sum_{r=1}^{2b} (T_r^{(i)})^2 + \frac{1}{2} \sum_{r=1}^{2b} (T_r^{(j)})^2 \right) \\
&\quad + \frac{1}{n} \sum_{s=1}^{b-1} \sum_{h=1}^{b-s} |\Delta_1 w_n(s+h)| \cdot \left(\sum_{r=1}^{2b} (T_r^{(i)})^2 + \sum_{r=1}^{2b} (T_r^{(j)})^2 \right) \\
&= \frac{1}{b} \left(\frac{1}{2} \sum_{r=1}^{2b} (T_r^{(i)})^2 + \frac{1}{2} \sum_{r=1}^{2b} (T_r^{(j)})^2 \right) \times \frac{b}{n} \left(\sum_{h=1}^{b} |\Delta_1 w_n(h)| + 2 \sum_{s=1}^{b-1} \sum_{h=1}^{b-s} |\Delta_1 w_n(s+h)| \right) \\
&:= H \times I.
\end{aligned}$$

By *lemma* 2, $I \to 0$ as $n \to \infty$. We will show that H stays bounded $w.p.1$ which guarantees $|\tilde{d}_{ij}| \to 0$ as $n \to \infty$. Recall that $T_r = U_r - \bar{B}_n$, apply *proposition* 2,

$$\frac{1}{2b} \sum_{r=1}^{2b} (T_r^{(i)})^2 = \frac{1}{2b} \sum_{r=1}^{2b} (U_r^{(i)} - \bar{B}_n^{(i)})^2$$

$$
\begin{aligned}
&= \frac{1}{2b}\sum_{r=1}^{2b}(U_r^{(i)})^2 - \frac{1}{2b}2\bar{B}_n^{(i)} \cdot \sum_{r=1}^{2b} U_r^{(i)} + (\bar{B}_n^{(i)})^2 \\
&\leq \left| \frac{1}{2b}\sum_{r=1}^{2b}(U_r^{(i)})^2 \right| + |2\bar{B}_n^{(i)}| \cdot \left| \frac{1}{2b}\sum_{r=1}^{2b} U_r^{(i)} \right| + |(\bar{B}_n^{(i)})^2| \\
&< \left| \frac{1}{2b}\sum_{r=1}^{2b}(U_r^{(i)})^2 \right| + (\frac{2}{n}(1+\varepsilon)(2nloglogn)^{1/2}) \cdot \left| \frac{1}{2b}\sum_{r=1}^{2b} U_r^{(i)} \right| \\
&\quad + \left(\frac{1}{n}(1+\varepsilon)(2nloglogn)^{1/2} \right)^2 \\
&= \left| \frac{1}{2b}\sum_{r=1}^{2b}(U_r^{(i)})^2 \right| + \left| \frac{1}{2b}\sum_{r=1}^{2b}(U_r^{(i)}) \right| \cdot O((n^{-1}logn)^{1/2}) + O(n^{-1}logn).
\end{aligned}
$$

For $r = 1,...,2b$, $U_r^{(i)} \sim N(0,\ 1)$ and $(U_r^{(i)})^2 \sim \chi^2(1)$, by classical strong law of large numbers, both

$$
\frac{1}{2b}\sum_{r=1}^{2b} U_r^{(i)} \quad \text{and} \quad \frac{1}{2b}\sum_{r=1}^{2b}(U_r^{(i)})^{1/2}
$$

stay bounded $w.p.1$, hence H stays bounded and we proved $\tilde{d_{ij}} \to 0$ as $n \to \infty$.

Lemma 4. (Vats et al. (2015b) Lemma 8) Set $h(X_t) = [g(X_t) - E_F g]^2$ for $t = 1,2,3,...$ and assume $\|E_F h\| < \infty$. Let condition 1 hold for h such that there exists a nonnegative increasing function ψ_h on the positive integers, a lower triangular matrix L_h, a finite random variable D_h and an $n_0 \in N$ such that $w.p.1$, for $n \geq n_0$,

$$
\left\| \sum_{k=1}^{n} h(X_k) - nE_F h - L_h B(n) \right\| < D_h \psi_h(n).
$$

Also assume condition 3 hold and as $n \to \infty$,

$$
b^{-1}\psi_h(n) \to 0 \quad \text{and} \quad b^{-1}logn = O(1),
$$

then

$$
\frac{1}{b}\sum_{k=1}^{b} h(X_k)
$$

stays bounded $w.p.1$ as $n \to \infty$. By *lemma 3* and *lemma 4*, it can be proved that $d \to 0$ using a similar proof as (Vats et al. (2015b) lemma 9), hence *proposition 1* is proved.

C Proof of theorem 1

We first define Brownian motion analog of the multivariate flat-top spectral variance estimator:

$$\tilde{\Sigma}_{ft} = \frac{1}{1-c}\frac{b}{n}\sum_{l=0}^{n-b}[\bar{B}_l(b)-\bar{B}][\bar{B}_l(b)-\bar{B}]^T - \frac{c}{1-c}\frac{cb}{n}\sum_{l=0}^{n-b}[\bar{B}_l(cb)-\bar{B}][\bar{B}_l(cb)-\bar{B}]^T.$$

Let $\tilde{\Sigma}_{ft,L} = L\tilde{\Sigma}_{ft}L^T$ where L is the lower triangular matrix such that $\Sigma = LL^T$. Then

$$\begin{aligned}
\tilde{\Sigma}_{ft,L} = & \frac{1}{1-c}\frac{b}{n}\sum_{l=0}^{n-b} L[\bar{B}_l(b)-\bar{B}][\bar{B}_l(b)-\bar{B}]^T L^T \\
& - \frac{c}{1-c}\frac{b}{n}\sum_{l=0}^{n-cb} L[\bar{B}_l(cb)-\bar{B}][\bar{B}_l(cb)-\bar{B}]^T L^T \\
= & \frac{1}{1-c}\frac{b}{n}\sum_{l=0}^{n-b}[L\bar{B}_l(b)-L\bar{B}][L\bar{B}_l(b)-L\bar{B}]^T \\
& - \frac{c}{1-c}\frac{b}{n}\sum_{l=0}^{n-cb}[L\bar{B}_l(cb)-L\bar{B}][L\bar{B}_l(cb)-L\bar{B}]^T \\
= & \frac{1}{1-c}\frac{b}{n}\sum_{l=0}^{n-b}[\bar{C}_l(b)-\bar{C}][\bar{C}_l(b)-\bar{C}]^T \\
& - \frac{c}{1-c}\frac{b}{n}\sum_{l=0}^{n-cb}[\bar{C}_l(cb)-\bar{C}][\bar{C}_l(cb)-\bar{C}]^T.
\end{aligned}$$

Let $\tilde{\Sigma}_{ft,L,ij}$ be the (i,j)th entry of $\tilde{\Sigma}_{ft,L}$. Then

$$\begin{aligned}
\tilde{\Sigma}_{ft,L,ij} = & \frac{1}{1-c}\frac{b}{n}\sum_{l=0}^{n-b}[\bar{C}_l^{(i)}(b)-\bar{C}^{(i)}][\bar{C}_l^{(j)}(b)-\bar{C}^{(j)}] \\
& - \frac{c}{1-c}\frac{b}{n}\sum_{l=0}^{n-cb}[\bar{C}_l^{(i)}(cb)-\bar{C}^{(i)}][\bar{C}_l^{(j)}(cb)-\bar{C}^{(j)}].
\end{aligned}$$

Lemma 5. $(**)$ The variance of $\tilde{\Sigma}_{ft,L,ij}$ satisfies

$$\frac{n}{b}\text{Var}[\tilde{\Sigma}_{ft,L,ij}] = \left(\frac{4c}{3}+\frac{2}{3}\right)[\Sigma_{ii}\Sigma_{jj}+\Sigma_{ij}^2]+o(1).$$

Proof. $\text{Var}[\tilde{\Sigma}_{ft,L,ij}] = \text{E}[\tilde{\Sigma}^2_{ft,L,ij}] - (\text{E}[\tilde{\Sigma}_{ft,L,ij}])^2$. First, consider $\text{E}[\tilde{\Sigma}^2_{ft,L,ij}]$.

$$E[\tilde{\Sigma}^2_{ft,L,ij}] = E\Bigg[\Bigg(\frac{1}{1-c}\frac{b}{n}\sum_{l=0}^{n-b}(\bar{C}_l^{(i)}(b)-\bar{C}^{(i)})(\bar{C}_l^{(j)}(b)-\bar{C}^{(j)})$$

$$-\frac{c}{1-c}\frac{cb}{n}\sum_{l=0}^{n-cb}(\bar{C}_l^{(i)}(cb)-\bar{C}^{(i)})(\bar{C}_l^{(j)}(cb)-\bar{C}^{(j)})\Bigg)^2\Bigg]$$

$$=E\left[\left(\frac{1}{1-c}\right)^2\frac{b^2}{n^2}\left[\sum_{l=0}^{n-b}(\bar{C}_l^{(i)}(b)-\bar{C}^{(i)})(\bar{C}_l^{(j)}(b)-\bar{C}^{(j)})\right]^2\right.$$

$$+\left(\frac{c}{1-c}\right)^2\frac{(cb)^2}{n^2}\left[\sum_{l=0}^{n-cb}(\bar{C}_l^{(i)}(cb)-\bar{C}^{(i)})(\bar{C}_l^{(j)}(cb)-\bar{C}^{(j)})\right]^2$$

$$\left.-\frac{2c^2}{(1-c)^2}\frac{b^2}{n^2}\left[\sum_{l=0}^{n-b}(\bar{C}_l^{(i)}(b)-\bar{C}^{(i)})(\bar{C}_l^{(j)}(b)-\bar{C}^{(j)})\right]\left[\sum_{l=0}^{n-cb}(\bar{C}_l^{(i)}(cb)-\bar{C}^{(i)})(\bar{C}_l^{(j)}(cb)-\bar{C}^{(j)})\right]\right]$$

$$:=A_1+A_2+A_3$$

where

$$A_1=E\left[\left(\frac{1}{1-c}\right)^2\frac{b^2}{n^2}\left[\sum_{l=0}^{n-b}(\bar{C}_l^{(i)}(b)-\bar{C}^{(i)})(\bar{C}_l^{(j)}(b)-\bar{C}^{(j)})\right]^2\right],$$

$$A_2=E\left[\left(\frac{c}{1-c}\right)^2\frac{(cb)^2}{n^2}\left[\sum_{l=0}^{n-cb}(\bar{C}_l^{(i)}(cb)-\bar{C}^{(i)})(\bar{C}_l^{(j)}(cb)-\bar{C}^{(j)})\right]^2\right],$$

$$A_3=E\left[-\frac{2c^2}{(1-c)^2}\frac{b^2}{n^2}\left[\sum_{l=0}^{n-b}(\bar{C}_l^{(i)}(b)-\bar{C}^{(i)})(\bar{C}_l^{(j)}(b)-\bar{C}^{(j)})\right]\left[\sum_{l=0}^{n-cb}(\bar{C}_l^{(i)}(cb)-\bar{C}^{(i)})(\bar{C}_l^{(j)}(cb)-\bar{C}^{(j)})\right]\right].$$

Consider the calculation of A_1:

$$A_1=\frac{1}{(1-c)^2}\frac{b^2}{n^2}E\left[\sum_{l=0}^{n-b}(\bar{C}_l^{(i)}(b)-\bar{C}^{(i)})^2(\bar{C}_l^{(j)}(b)-\bar{C}^{(j)})^2\right.$$

$$+2\sum_{s=1}^{b-1}\sum_{l=0}^{n-b-s}(\bar{C}_l^{(i)}(b)-\bar{C}^{(i)})(\bar{C}_l^{(j)}(b)-\bar{C}^{(j)})(\bar{C}_{l+s}^{(i)}(b)-\bar{C}^{(i)})(\bar{C}_{l+s}^{(j)}(b)-\bar{C}^{(j)})$$

$$\left.+2\sum_{s=b}^{n-b}\sum_{l=0}^{n-b-s}(\bar{C}_l^{(i)}(b)-\bar{C}^{(i)})(\bar{C}_l^{(j)}(b)-\bar{C}^{(j)})(\bar{C}_{l+s}^{(i)}(b)-\bar{C}^{(i)})(\bar{C}_{l+s}^{(j)}(b)-\bar{C}^{(j)})\right].$$

Denote A, B, C the first sum and the two double sums in the above equation, respectively. First calculate the expectation of A. Define $U_t^{(i)}=B^{(i)}(t)-B^{(i)}(t-1)$, then $U_t^{(i)}\sim N(0,1)$ for $t=1,2,...,n$ and

$$\bar{B}_l^{(i)}(b)-\bar{B}^{(i)}=\frac{(n-b)}{nb}\sum_{t=l+1}^{l+b}U_t^{(i)}-\frac{1}{n}\sum_{t=1}^{l}U_t^{(i)}-\frac{1}{n}\sum_{t=l+b+1}^{n}U_t^{(i)}.$$

Notice that $E[\bar{B}_l^{(i)}(b) - \bar{B}^{(i)}] = 0$ for $l = 0, ..., (n-b)$ and

$$Var[\bar{B}_l^{(i)}(b) - \bar{B}^{(i)}] = \left(\frac{n-b}{nb}\right)^2 b + \frac{n-b}{n^2} = \frac{n-b}{bn},$$

therefore

$$\bar{B}_l^{(i)} - \bar{B}^{(i)} \sim N\left(0,\ \frac{n-b}{bn}\right)$$

and

$$\bar{B}_l(b) - \bar{B}_n \sim N\left(0,\ \frac{n-b}{bn} I_p\right),$$

resulting

$$\bar{C}_l(b) - \bar{C}_n = L(\bar{B}_l - \bar{B}_n) \sim N\left(0,\ \frac{n-b}{bn} LL^T\right). \tag{2.A.3}$$

Now consider $E[(\bar{C}_l^{(i)}(b) - \bar{C}^{(i)})^2 (\bar{C}_l^{(j)}(b) - \bar{C}^{(j)})^2] := E[Z_i^2 Z_j^2]$ where $Z_i = \bar{C}_l^{(i)}(b) - \bar{C}^{(i)}$ and $Z_j = \bar{C}_l^{(j)}(b) - \bar{C}^{(j)}$. Recall $\Sigma = LL^T$, then

$$\begin{bmatrix} Z_i \\ Z_j \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix},\ \frac{n-b}{bn} \begin{bmatrix} \Sigma_{ii} & \Sigma_{ij} \\ \Sigma_{ij} & \Sigma_{jj} \end{bmatrix} \right).$$

Apply *proposition 5*,

$$\begin{aligned} & E[(\bar{C}_l^{(i)}(b) - \bar{C}^{(i)})^2 (\bar{C}_l^{(j)}(b) - \bar{C}^{(j)})^2] = E[Z_i^2 Z_j^2] \\ &= 2\left(\frac{n-b}{bn}\Sigma_{ij}\right)^2 + \left(\frac{n-b}{bn}\Sigma_{ii}\right)\left(\frac{n-b}{bn}\Sigma_{jj}\right) \\ &= \left(\frac{n-b}{bn}\right)^2 [\Sigma_{ij}^2 + \Sigma_{ii}\Sigma_{jj}] + \left(\frac{n-b}{bn}\right)^2 \Sigma_{ij}^2. \end{aligned}$$

Therefore

$$\begin{aligned} EA &= \sum_{l=0}^{n-b} E[(\bar{C}_l^{(i)}(b) - \bar{C}^{(i)})(\bar{C}_l^{(j)}(b) - \bar{C}^{(j)})] \\ &= (n-b+1)\left(\frac{n-b}{bn}\right)^2 (\Sigma_{ij}^2 + \Sigma_{ii}\Sigma_{jj}) + \sum_{l=0}^{n-b}\left(\frac{n-b}{bn}\right)^2 \Sigma_{ij}^2 \end{aligned}$$

$$= o\left(\frac{n}{b}\right) + \sum_{l=0}^{n-b}\left(\frac{n-b}{bn}\right)^2 \Sigma_{ij}^2.$$

To calculate EB, for $s = 1, 2, ..., (b-1)$, consider

$$E[(\bar{C}_l^{(i)}(b) - \bar{C}^{(i)})(\bar{C}_l^{(j)}(b) - \bar{C}^{(j)})(\bar{C}_{l+s}^{(i)}(b) - \bar{C}^{(i)})(\bar{C}_{l+s}^{(j)}(b) - \bar{C}^{(j)})].$$

We first derive the joint distribution of the following vector of length $2p$

$$\begin{bmatrix} \bar{C}_l(b) - \bar{C} \\ \bar{C}_{l+s}(b) - \bar{C} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \left(\frac{n-b}{bn}\right)\Sigma & \left(\frac{n-b}{bn} - \frac{s}{b^2}\right)\Sigma \\ \left(\frac{n-b}{bn} - \frac{s}{b^2}\right)\Sigma & \left(\frac{n-b}{bn}\right)\Sigma \end{bmatrix}\right).$$

Recall (2.3.1),

$$\bar{C}_l(b) - \bar{C} \to N\left(0,\ \frac{n-b}{bn}\Sigma\right)$$

$$\bar{C}_{l+s}(b) - \bar{C} \to N\left(0,\ \frac{n-b}{bn}\Sigma\right).$$

We will show that

$$\text{Cov}(\bar{C}_l(b) - \bar{C},\ \bar{C}_{l+s}(b) - \bar{C}) = \left(\frac{n-b}{bn} - \frac{s}{b^2}\right)\Sigma.$$

Notice that

$$\begin{aligned} &\text{Cov}(\bar{C}_l(b) - \bar{C},\ \bar{C}_{l+s}(b) - \bar{C}) \\ &= E[(\bar{C}_l(b) - \bar{C})(\bar{C}_{l+s}(b) - \bar{C})^T] \\ &= E[L \cdot (\bar{B}_l(b) - \bar{B})(\bar{B}_{l+s}(b) - \bar{B})^T \cdot L^T] \\ &= L \cdot E[(\bar{B}_l(b) - \bar{B})(\bar{B}_{l+s}(b) - \bar{B})^T] \cdot L^T, \end{aligned}$$

we first consider $E[(\bar{B}_l(b) - \bar{B})(\bar{B}_{l+s}(b) - \bar{B})^T]$ by considering each entry of the matrix.

For $i \neq j$,

$$E[\bar{B}_l^{(i)}(b) - \bar{B}^{(i)}][\bar{B}_{l+s}^{(j)}(b) - \bar{B}^{(j)}] = E[\bar{B}_l^{(i)}(b) - \bar{B}^{(i)}] \cdot E[\bar{B}_{l+s}^{(j)}(b) - \bar{B}^{(j)}] = 0.$$

For $i = j$, we need to calculate $E[\bar{B}_l^{(i)}(b) - \bar{B}^{(i)}][\bar{B}_{l+s}^{(i)}(b) - \bar{B}^{(i)}]$. Let $Z_1^{(i)} = \bar{B}_l^{(i)}(b) - \bar{B}^{(i)}$ and $Z_2^{(i)} = \bar{B}_{l+s}^{(i)}(b) - \bar{B}^{(i)}$, then the joint distribution of Z is

$$\begin{bmatrix} Z_1^{(i)} \\ Z_2^{(i)} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{n-b}{bn} & \frac{nb-ns-b^2}{nb^2} \\ \frac{nb-ns-b^2}{nb^2} & \frac{n-b}{bn} \end{bmatrix} \right),$$

and

$$Z_1^{(i)} | Z_2^{(i)} \sim N\left(\frac{b(n-b)-ns}{b(n-b)} Z_2^{(i)},\ \frac{2bs(n-b)-ns^2}{b^3(n-b)} \right),$$

$$Z_2^{(i)} \sim N\left(0,\ \frac{n-b}{bn}\right).$$

Hence

$$\begin{aligned}
E[Z_i^{(i)} Z_2^{(i)}] &= E_{Z_2^{(i)}}[E_{Z_1^{(i)}|Z_2^{(i)}}[Z_1^{(i)} Z_2^{(i)} | Z_2^{(i)}]] \\
&= E_{Z_2^{(i)}}\left[Z_2^{(i)} \left(\frac{b(n-s)-ns}{b(n-b)} Z_2^{(i)} \right) \right] \\
&= \frac{b(n-s)-ns}{b(n-b)} E[(Z_2^{(i)})^2] \\
&= \frac{b(n-s)-ns}{b(n-b)} \cdot \frac{n-b}{bn} \\
&= \frac{n-b}{bn} - \frac{s}{b^2}.
\end{aligned}$$

Combine the above results, we have

$$\mathrm{Cov}(\bar{C}_l(b) - \bar{C},\ \bar{C}_{l+s}(b) - \bar{C}) = L \cdot \left(\frac{n-b}{bn} - \frac{s}{b^2} \right) I_p \cdot L^T = \left(\frac{n-b}{bn} - \frac{s}{b^2} \right) \cdot \Sigma.$$

Let $Z_1 = \bar{C}_l(b)^{(i)} - \bar{C}^{(i)}$, $Z_2 = \bar{C}_l(b)^{(j)} - \bar{C}^{(j)}$, $Z_3 = \bar{C}_{l+s}(b)^{(i)} - \bar{C}^{(i)}$, $Z_4 = \bar{C}_{l+s}(b)^{(j)} - \bar{C}^{(j)}$. From the joint distribution of the $2p$ vector, we have

$$\begin{bmatrix} Z_1 \\ Z_2 \\ Z_3 \\ Z_4 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \left(\frac{n-b}{bn}\right)\Sigma_{ii} & \left(\frac{n-b}{bn}\right)\Sigma_{ij} & \left(\frac{n-b}{bn} - \frac{s}{b^2}\right)\Sigma_{ii} & \left(\frac{n-b}{bn} - \frac{s}{b^2}\right)\Sigma_{ij} \\ & \left(\frac{n-b}{bn}\right)\Sigma_{jj} & \left(\frac{n-b}{bn} - \frac{s}{b^2}\right)\Sigma_{ij} & \left(\frac{n-b}{bn} - \frac{s}{b^2}\right)\Sigma_{jj} \\ & & \left(\frac{n-b}{bn}\right)\Sigma_{ii} & \left(\frac{n-b}{bn}\right)\Sigma_{ij} \\ & & & \left(\frac{n-b}{bn}\right)\Sigma_{jj} \end{bmatrix} \right),$$

we only show the upper triangle entries of the covariance matrix and the lower triangle can be obtained by symmetry. Apply *proposition 6*,

$$\begin{aligned} E[Z_1Z_2Z_3Z_4] &= E[Z_1Z_2]\cdot E[Z_3Z_4] + E[Z_1Z_3]\cdot E[Z_2Z_4] + E[Z_1Z_4]\cdot E[Z_2Z_3] \\ &= \left(\frac{n-b}{bn}\right)^2 \Sigma_{ij}^2 + \left(\frac{n-b}{bn} - \frac{s}{b^2}\right)^2 \Sigma_{ii}\Sigma_{jj} + \left(\frac{n-b}{bn} - \frac{s}{b^2}\right)^2 \Sigma_{ij}^2, \end{aligned}$$

resulting

$$\begin{aligned} EB &= \sum_{s=1}^{b-1}\sum_{l=0}^{n-b-s} E[(\bar{C}_l^{(i)}(b) - \bar{C}^{(i)})(\bar{C}_l^{(j)}(b) - \bar{C}^{(j)})(\bar{C}_{l+s}^{(i)}(b) - \bar{C}^{(i)})(\bar{C}_{l+s}^{(j)}(b) - \bar{C}^{(j)})] \\ &= \sum_{s=1}^{b-1}\sum_{l=0}^{n-b-s} E[Z_1Z_2Z_3Z_4] \\ &= \sum_{s=1}^{b-1}\sum_{l=0}^{n-b-s} \left[\left(\frac{n-b}{bn}\right)^2 \Sigma_{ij}^2 + \left(\frac{n-b}{bn} - \frac{s}{b^2}\right)^2 \Sigma_{ii}\Sigma_{jj} + \left(\frac{n-b}{bn} - \frac{s}{b^2}\right)^2 \Sigma_{ij}^2 \right]. \end{aligned}$$

Consider the sum of the second term in the above expression.

$$\begin{aligned} &\sum_{s=1}^{b-1}\sum_{l=0}^{n-b-s} \left(\frac{n-b}{bn} - \frac{s}{b^2}\right)^2 \\ &= \sum_{s=1}^{b-1}\sum_{l=0}^{n-b-s} \left[\frac{s^2}{b^4} + \left(\frac{2}{b^2 n} - \frac{2}{b^3}\right) s + \left(\frac{1}{b^2} + \frac{1}{n^2} - \frac{2}{bn}\right) \right] \\ &= \sum_{s=1}^{b-1} \left[-\frac{s^3}{b^4} + \left(\frac{n}{b^4} + \frac{1}{b^3} + \frac{1}{b^4} - \frac{2}{b^2 n}\right) s^2 \right. \\ &\qquad + \left(\frac{3}{b^2} - \frac{2n}{b^3} + \frac{2}{b^2 n} - \frac{2}{b^3} - \frac{1}{n^2}\right) s \end{aligned}$$

$$
\begin{aligned}
&\quad + \left(\frac{n}{b^2} + \frac{3}{n} - \frac{3}{b} + \frac{1}{b^2} + \frac{1}{n^2} - \frac{2}{bn} - \frac{b}{n^2}\right)\Bigg] \\
&= -\frac{1}{b^4}\left(\frac{b^4}{4} - \frac{b^3}{2} + \frac{b^2}{4}\right) + \left(\frac{n}{b^4} + \frac{1}{b^3} + \frac{1}{b^4} - \frac{2}{b^2 n}\right)\left(\frac{b^3}{3} - \frac{b^2}{2} + \frac{b}{6}\right) \\
&\quad + \left(\frac{3}{b^2} - \frac{2n}{b^3} + \frac{2}{b^2 n} - \frac{2}{b^3} - \frac{1}{n^2}\right)\left(\frac{b^2}{2} - \frac{b}{2}\right) \\
&\quad + \left(\frac{n}{b^2} + \frac{3}{n} - \frac{3}{b} + \frac{1}{b^2} + \frac{1}{n^2} - \frac{2}{bn} - \frac{b}{n^2}\right)(b-1) \\
&= \frac{n}{b^4} \cdot \frac{b^3}{3} - \frac{2n}{b^3} \cdot \frac{b^2}{2} + \frac{n}{b^2} \cdot b \\
&= \frac{1}{3}\frac{n}{b} + o\left(\frac{n}{b}\right).
\end{aligned}
$$

Plug in the above results, we have

$$
EB = \Sigma_{ij}^2 \sum_{s=1}^{b-1} \sum_{l=0}^{n-b-s} \left(\frac{n-b}{bn}\right)^2 + (\Sigma_{ii}\Sigma_{jj} + \Sigma_{ij}^2)\left[\frac{1}{3}\frac{n}{b} + o\left(\frac{n}{b}\right)\right].
$$

Similarly as EB, we calculate EC by first consider

$$
E[(\bar{C}_l^{(i)}(b) - \bar{C}^{(i)})(\bar{C}_l^{(j)}(b) - \bar{C}^{(j)})(\bar{C}_{l+s}^{(i)}(b) - \bar{C}^{(i)})(\bar{C}_{l+s}^{(j)}(b) - \bar{C}^{(j)})]
$$

for $s = b, ..., (n-b)$. We show that the joint distribution of the $2p$ vector is

$$
\begin{bmatrix} \bar{C}_l(b) - \bar{C} \\ \bar{C}_{l+s}(b) - \bar{C} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \left(\frac{n-b}{bn}\right)\Sigma & -\frac{1}{n}\Sigma \\ -\frac{1}{n}\Sigma & \left(\frac{n-b}{bn}\right)\Sigma \end{bmatrix}\right).
$$

For $i \neq j$,

$$
E[\bar{B}_l^{(i)}(b) - \bar{B}^{(i)}][\bar{B}_{l+s}^{(j)}(b) - \bar{B}^{(j)}] = E[\bar{B}_l^{(i)}(b) - \bar{B}^{(i)}] \cdot E[\bar{B}_{l+s}^{(j)}(b) - \bar{B}^{(j)}] = 0.
$$

For $i = j$, we need to calculate $E[\bar{B}_l^{(i)}(b) - \bar{B}^{(i)}][\bar{B}_{l+s}^{(i)}(b) - \bar{B}^{(i)}]$ for $s = b, ..., (n-b)$.. Let $Z_1^{(i)} = \bar{B}_l^{(i)}(b) - \bar{B}^{(i)}$ and $Z_2^{(i)} = \bar{B}_{l+s}^{(i)}(b) - \bar{B}^{(i)}$. Since

$$
\begin{bmatrix} Z_1^{(i)} \\ Z_2^{(i)} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{n-b}{bn} & -\frac{1}{n} \\ -\frac{1}{n} & \frac{n-b}{bn} \end{bmatrix}\right),
$$

then

$$Z_1^{(i)}|Z_2^{(i)} \sim N\left(\frac{-b}{n-b}Z_2^{(i)},\ \left[\frac{n-b}{bn} - \frac{bn}{n^2(n-b)}\right]\right),$$
$$Z_2^{(i)} \sim N\left(0,\ \frac{n-b}{bn}\right).$$

Hence

$$\begin{aligned}
E[Z_i^{(i)}Z_2^{(i)}] &= E_{Z_2^{(i)}}[E_{Z_1^{(i)}|Z_2^{(i)}}[Z_1^{(i)}Z_2^{(i)}|Z_2^{(i)}]] \\
&= E_{Z_2^{(i)}}\left((Z_2^{(i)})^2 \cdot \frac{-b}{n-b}\right) \\
&= \frac{-b}{n-b} \cdot \frac{n-b}{bn} \\
&= -\frac{1}{n}.
\end{aligned}$$

Then we have

$$\mathrm{Cov}(\bar{C}_l(b) - \bar{C},\ \bar{C}_{l+s}(b) - \bar{C}) = L \cdot \left(-\frac{1}{n}\right) I_p \cdot L^T = -\frac{1}{n} \cdot \Sigma,$$

which yields the joint distribution of the $2p$ vector. Again denote $Z_1 = \bar{C}_l(b)^{(i)} - \bar{C}^{(i)}$, $Z_2 = \bar{C}_l(b)^{(j)} - \bar{C}^{(j)}$, $Z_3 = \bar{C}_{l+s}(b)^{(i)} - \bar{C}^{(i)}$, $Z_4 = \bar{C}_{l+s}(b)^{(j)} - \bar{C}^{(j)}$. The expectation of C is

$$\begin{aligned}
EC &= \sum_{s=b}^{n-b}\sum_{l=0}^{n-b-s} E[(\bar{C}_l^{(i)}(b) - \bar{C}^{(i)})(\bar{C}_l^{(j)}(b) - \bar{C}^{(j)})(\bar{C}_{l+s}^{(i)}(b) - \bar{C}^{(i)})(\bar{C}_{l+s}^{(j)}(b) - \bar{C}^{(j)})] \\
&= \sum_{s=b}^{n-b}\sum_{l=0}^{n-b-s} E[Z_1Z_2Z_3Z_4] \\
&= \sum_{s=b}^{n-b}\sum_{l=0}^{n-b-s} \left[\left(\frac{n-b}{bn}\right)^2 \Sigma_{ij}^2 + \frac{1}{n^2}\Sigma_{ii}\Sigma_{jj} + \frac{1}{n^2}\Sigma_{ij}^2\right].
\end{aligned}$$

Consider the second summation in the above expression first.

$$\begin{aligned}
&\sum_{s=b}^{n-b}\sum_{l=1}^{n-b+1-s} \frac{1}{n^2} \\
&= \frac{1}{n^2} \cdot \sum_{s=b}^{n-b}(n-b-s+1) \\
&= -\frac{1}{n^2}\left(\frac{n^2}{2} - bn + \frac{n}{2}\right) + \left(\frac{1}{n} - \frac{b}{n^2} + \frac{1}{n^2}\right)(n-2b+1)
\end{aligned}$$

$$= o\left(\frac{n}{b}\right).$$

Plug in EC, we have

$$EC = \Sigma_{ij}^2 \sum_{s=b}^{n-b} \sum_{l=0}^{n-b-s} \left(\frac{n-b}{bn}\right)^2 + o\left(\frac{n}{b}\right).$$

Combine EA, EB and EC, we can calculate A_1 as follows.

$$\begin{aligned}
A_1 &= E\left[\left(\left(\frac{1}{1-c}\right)^2 \frac{b^2}{n^2}\left[\sum_{l=0}^{n-b}(\bar{C}_l^{(i)}(b)-\bar{C}^{(i)})(\bar{C}_l^{(j)}(b)-\bar{C}^{(j)})\right]^2\right]\right. \\
&= \frac{1}{(1-c)^2}\cdot\frac{b^2}{n^2}\cdot[EA+2EB+2EC] \\
&= \frac{1}{(1-c)^2}\cdot\frac{b^2}{n^2}\cdot\left[\frac{2}{3}(\Sigma_{ii}\Sigma_{jj}+\Sigma_{ij}^2)\cdot\frac{n}{b}+o\left(\frac{n}{b}\right)\right. \\
&\left.+\Sigma_{ij}^2\left(\sum_{l=0}^{n-b}\left(\frac{n-b}{bn}\right)^2+2\sum_{s=1}^{b-1}\sum_{l=0}^{n-b-s}\left(\frac{n-b}{bn}\right)^2+2\sum_{s=b}^{n-b}\sum_{l=0}^{n-b-s}\left(\frac{n-b}{bn}\right)^2\right)\right] \\
&= \frac{1}{(1-c)^2}\cdot\frac{b^2}{n^2}\cdot\left[\frac{2}{3}(\Sigma_{ii}\Sigma_{jj}+\Sigma_{ij}^2)\cdot\frac{n}{b}+o\left(\frac{n}{b}\right)+\Sigma_{ij}^2\left(\frac{n-b}{bn}\right)^2(n-b+1)^2\right] \\
&= \frac{1}{(1-c)^2}\cdot\frac{b^2}{n^2}\cdot\left[\frac{2}{3}(\Sigma_{ii}\Sigma_{jj}+\Sigma_{ij}^2)\cdot\frac{n}{b}+o\left(\frac{n}{b}\right)+\Sigma_{ij}^2\left(\frac{n^2}{b^2}-\frac{4n}{b}\right)\right] \\
&= \frac{1}{(1-c)^2}\cdot\left[\frac{2}{3}(\Sigma_{ii}\Sigma_{jj}+\Sigma_{ij}^2)\cdot\frac{b}{n}+\Sigma_{ij}^2-4\Sigma_{ij}^2\cdot\frac{b}{n}\right]+o\left(\frac{b}{n}\right). \qquad (2.A.4)
\end{aligned}$$

Notice the similar structure of A_1 and A_2, we have

$$A_2 = \frac{c^2}{(1-c)^2}\cdot\left[\frac{2}{3}(\Sigma_{ii}\Sigma_{jj}+\Sigma_{ij}^2)\cdot\frac{cb}{n}+\Sigma_{ij}^2-4\Sigma_{ij}^2\cdot\frac{cb}{n}\right]+o\left(\frac{b}{n}\right).$$

Now we calculate A_3.

$$\begin{aligned}
A_3 &= -\frac{2c^2}{(1-c)^2}\cdot\frac{b^2}{n^2}\left[\sum_{l=0}^{n-b}(\bar{C}_l^{(i)}(b)-\bar{C}_n^{(i)})(\bar{C}_l^{(j)}(b)-\bar{C}_n^{(j)})\right]\left[\sum_{l=0}^{n-cb}(\bar{C}_l^{(i)}(cb)-\bar{C}_n^{(i)})(\bar{C}_l^{(j)}(cb)-\bar{C}_n^{(j)})\right] \\
&= -\frac{2c^2}{(1-c)^2}\cdot\frac{b^2}{n^2}\cdot E[((1-c)b+1)(n-b+1)\cdot OL^{(i)}OL^{(j)} \\
&+2\sum_{s=1}^{cb-1}\sum_{l=0}^{n-b-s}(\bar{C}_l(b)^{(i)}-\bar{C}^{(i)})(\bar{C}_l(b)^{(j)}-\bar{C}^{(j)})(\bar{C}_{l+(1-c)b+s}^{(i)}(cb)-\bar{C}^{(i)})(\bar{C}_{l+(1-c)b+s}^{(j)}(cb)-\bar{C}^{(j)})
\end{aligned}$$

$$+2\sum_{s=cb}^{n-b}\sum_{l=0}^{n-b-s}(\bar{C}_l(b)^{(i)}-\bar{C}^{(i)})(\bar{C}_l(b)^{(j)}-\bar{C}^{(j)})(\bar{C}^{(i)}_{l+(1-c)b+s}(cb)-\bar{C}^{(i)})(\bar{C}^{(j)}_{l+(1-c)b+s}(cb)-\bar{C}^{(j)})],$$

where

$$OL^{(i)}=(\bar{C}^{(i)}_p(b)-\bar{C}^{(i)})(\bar{C}^{(i)}_q(bc)-\bar{C}^{(i)}),$$

$$OL^{(j)}=(\bar{C}^{(j)}_p(b)-\bar{C}^{(j)})(\bar{C}^{(j)}_q(bc)-\bar{C}^{(j)}),$$

for p, q satisfying $q\geq p$ and $q+cb\leq p+b$. In the above equation, denote the two double sums by D and G:

$$D=\sum_{s=1}^{cb-1}\sum_{l=0}^{n-b-s}(\bar{C}_l(b)^{(i)}-\bar{C}^{(i)})(\bar{C}_l(b)^{(j)}-\bar{C}^{(j)})(\bar{C}^{(i)}_{l+(1-c)b+s}(cb)-\bar{C}^{(i)})(\bar{C}^{(j)}_{l+(1-c)b+s}(cb)-\bar{C}^{(j)}),$$

$$G=\sum_{s=cb}^{n-b}\sum_{l=0}^{n-b-s}(\bar{C}_l(b)^{(i)}-\bar{C}^{(i)})(\bar{C}_l(b)^{(j)}-\bar{C}^{(j)})(\bar{C}^{(i)}_{l+(1-c)b+s}(cb)-\bar{C}^{(i)})(\bar{C}^{(j)}_{l+(1-c)b+s}(cb)-\bar{C}^{(j)}).$$

First, consider $E[OL^{(i)}OL^{(j)}]$. We first derive the joint distribution of the following $2p$ vector.

$$\begin{bmatrix}\bar{C}_p(b)-\bar{C}_n\\ \bar{C}_q(cb)-\bar{C}_n\end{bmatrix}\sim N\left(\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}\left(\frac{n-b}{bn}\right)\Sigma & \left(\frac{n-b}{bn}\right)\Sigma\\ \left(\frac{n-b}{bn}\right)\Sigma & \left(\frac{n-cb}{cbn}\right)\Sigma\end{bmatrix}\right).$$

By (2.A.3),

$$\bar{C}_p(b)-\bar{C}_n\sim N\left(0,\frac{n-b}{bn}\Sigma\right),$$
$$\bar{C}_q(cb)-\bar{C}_n\sim N\left(0,\frac{n-cb}{cbn}\Sigma\right).$$

For $i\neq j$,

$$E[\bar{B}^{(i)}_p(b)-\bar{B}^{(i)}][\bar{B}^{(j)}_q(cb)-\bar{B}^{(j)}]=E[\bar{B}^{(i)}_p(b)-\bar{B}^{(i)}]\cdot E[\bar{B}^{(j)}_q(cb)-\bar{B}^{(j)}]=0.$$

For $i=j$, we need to calculate $E[\bar{B}^{(i)}_p(b)-\bar{B}^{(i)}_n][\bar{B}^{(i)}_q(cb)-\bar{B}^{(i)}_n]$ for p, q satisfying $q\geq p$ and $q+cb\leq p+b$. Let $Z^{(i)}_1=\bar{B}^{(i)}_p(b)-\bar{B}^{(i)}_n$ and $Z^{(i)}_2=\bar{B}^{(i)}_q(cb)-\bar{B}^{(i)}_n$, then

$$\begin{bmatrix}Z^{(i)}_1\\ Z^{(i)}_2\end{bmatrix}\sim N\left(\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}\frac{n-b}{bn} & \frac{n-b}{bn}\\ \frac{n-b}{bn} & \frac{n-cb}{cbn}\end{bmatrix}\right),$$

thus

$$Z_1^{(i)}|Z_2^{(i)} \sim N\left(\frac{c(n-b)}{n-cb}Z_2^{(i)},\ \frac{(1-c)n+(c-1)b}{b(n-cb)}\right),$$
$$Z_2^{(i)} \sim N\left(0,\ \frac{n-cb}{cbn}\right).$$

Hence

$$\begin{aligned}
E[Z_i^{(i)}Z_2^{(i)}] &= E_{Z_2^{(i)}}[E_{Z_1^{(i)}|Z_2^{(i)}}[Z_1^{(i)}Z_2^{(i)}|Z_2^{(i)}]] \\
&= E_{Z_2^{(i)}}\left[(Z_2^{(i)})^2 \cdot \frac{c(n-b)}{n-cb}\right] \\
&= \frac{c(n-b)}{n-cb} \cdot \frac{n-cb}{cbn} \\
&= \frac{n-b}{bn}.
\end{aligned}$$

Then we have

$$\text{Cov}(\bar{C}_p(b)-\bar{C}_n,\ \bar{C}_q(cb)-\bar{C}_n) = L \cdot \left(\frac{n-b}{bn}\right) I_p \cdot L^T = \frac{n-b}{bn} \cdot \Sigma,$$

which yields the joint distribution of the $2p$ vector. Denote $Z_1 = \bar{C}_p(b)^{(i)} - \bar{C}^{(i)}$, $Z_2 = \bar{C}_p(b)^{(j)} - \bar{C}^{(j)}$, $Z_3 = \bar{C}_q(cb)^{(i)} - \bar{C}^{(i)}$, $Z_4 = \bar{C}_q(cb)^{(j)} - \bar{C}^{(j)}$. Then

$$\begin{aligned}
&E[((1-c)b+1)(n-b+1)OL^{(i)}OL^{(j)}] \\
&= ((1-c)b+1)(n-b+1) \cdot E[(\bar{C}_p^{(i)}(b)-\bar{C}^{(i)})(\bar{C}_p^{(j)}(b)-\bar{C}^{(j)})(\bar{C}_q^{(i)}(cb)-\bar{C}^{(i)})(\bar{C}_q^{(j)}(cb)-\bar{C}^{(j)})] \\
&= ((1-c)b+1)(n-b+1) \cdot E[Z_1Z_2Z_3Z_4] \\
&= ((1-c)b+1)(n-b+1) \cdot \left(\frac{n-b}{bn}\right)^2 (\Sigma_{ij}^2 + \Sigma_{ii}\Sigma_{jj}) \\
&\quad + ((1-c)b+1)(n-b+1) \cdot \left(\frac{n-b}{bn}\right)\left(\frac{n-cb}{cbn}\right)\Sigma_{ij}^2.
\end{aligned}$$

Notice that

$$((1-c)b+1)(n-b+1) \cdot \left(\frac{n-b}{bn}\right)^2$$

$$= ((1-c)b+1)(n-b+1) \cdot \left(\frac{1}{b^2} + \frac{1}{n^2} - \frac{2}{bn}\right)$$
$$= (1-c)bn\frac{1}{b^2} + o\left(\frac{n}{b}\right)$$
$$= (1-c)\frac{n}{b} + o\left(\frac{n}{b}\right),$$

and

$$((1-c)b+1)(n-b+1) \cdot \left(\frac{n-b}{bn}\right)\left(\frac{n-cb}{cbn}\right)$$
$$= ((1-c)b+1)(n-b+1) \cdot \left(\frac{1}{cb^2} + \frac{1}{n^2} - \frac{(c+1)}{cbn}\right)$$
$$= (1-c)bn\frac{1}{cb^2} + o\left(\frac{n}{b}\right)$$
$$= \frac{1-c}{c}\frac{n}{b} + o\left(\frac{n}{b}\right).$$

We have

$$E[((1-c)b+1)(n-b+1)OL^{(i)}OL^{(j)}]$$
$$= \left[(1-c)(\Sigma_{ii}\Sigma_{jj} + \Sigma_{ij}^2) + \frac{1-c}{c}\Sigma_{ij}^2\right] \cdot \frac{n}{b} + o\left(\frac{n}{b}\right).$$

We calculate ED by first derive the following joint distribution.

$$\begin{bmatrix} \bar{C}_l(b) - \bar{C} \\ \bar{C}_{l+(1-c)b+s}(cb) - \bar{C} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \left(\frac{n-b}{bn}\right)\Sigma & \left(\frac{1}{b} - \frac{1}{n} - \frac{s}{cb^2}\right)\Sigma \\ \left(\frac{1}{b} - \frac{1}{n} - \frac{s}{cb^2}\right)\Sigma & \left(\frac{n-cb}{cbn}\right)\Sigma \end{bmatrix}\right).$$

For $i \neq j$,

$$E[\bar{B}_l^{(i)}(b) - \bar{B}^{(i)}][\bar{B}_{l+(1-c)b+s}^{(j)}(cb) - \bar{B}^{(j)}] = E[\bar{B}_l^{(i)}(b) - \bar{B}^{(i)}] \cdot E[\bar{B}_{l+(1-c)b+s}^{(j)}(cb) - \bar{B}^{(j)}] = 0.$$

For $i = j$, we need to calculate $E[\bar{B}_l^{(i)}(b) - \bar{B}^{(i)}][\bar{B}_{l+(1-c)b+s}^{(i)}(cb) - \bar{B}^{(i)}]$ for $s = 1, ..., (cb-1)$. Let $Z_1^{(i)} = \bar{B}_l^{(i)}(b) - \bar{B}^{(i)}$, $Z_2 = \bar{B}_{l+(1-c)b+s}^{(i)}(cb) - \bar{B}^{(i)}$, then

$$\begin{bmatrix} Z_1^{(i)} \\ Z_2^{(i)} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{n-b}{bn} & \frac{cbn - cb^2 - sn}{cb^2n} \\ \frac{cbn - cb^2 - sn}{cb^2n} & \frac{n-cb}{cbn} \end{bmatrix}\right),$$

hence

$$Z_1^{(i)}|Z_2^{(i)} \sim N\left(\frac{cbn-cb^2-sn}{b(n-cb)}Z_2^{(i)},\ \frac{(c-c^2)b^2(n-b)-s^2n+2cb(n-b)s}{cb^3(n-cb)}\right),$$

$$Z_2^{(i)} \sim N\left(0,\ \frac{n-cb}{cbn}\right).$$

Hence

$$\begin{aligned}
E[Z_i^{(i)}Z_2^{(i)}] &= E_{Z_2^{(i)}}[E_{Z_1^{(i)}|Z_2^{(i)}}[Z_1^{(i)}Z_2^{(i)}|Z_2^{(i)}]] \\
&= E_{Z_2^{(i)}}\left[(Z_2^{(i)})^2 \cdot \frac{cbn-cb^2-sn}{b(n-cb)}\right] \\
&= \frac{cbn-cb^2-sn}{b(n-cb)} \cdot \frac{n-cb}{cbn} \\
&= \frac{cbn-sb^2-sn}{cb^2n} \\
&= \frac{1}{b}-\frac{1}{n}-\frac{s}{cb^2}.
\end{aligned}$$

Then we have

$$\mathrm{Cov}(\bar{C}_l(b)-\bar{C},\ \bar{C}_{l+s}(b)-\bar{C}) = L\cdot\left(\frac{1}{b}-\frac{1}{n}-\frac{s}{cb^2}\right)I_p\cdot L^T = \left(\frac{1}{b}-\frac{1}{n}-\frac{s}{cb^2}\right)\cdot\Sigma,$$

which yields the joint distribution of the $2p$ vector. Again denote $Z_1 = \bar{C}_l^{(i)}(b)-\bar{C}^{(i)}$, $Z_2 = \bar{C}_l^{(j)}(b)-\bar{C}^{(j)}$, $Z_3 = \bar{C}^{(i)}_{l+(1-c)b+s}(cb)-\bar{C}^{(i)}$, $Z_4 = \bar{C}^{(j)}_{l+(1-c)b+s}(cb)-\bar{C}^{(j)}$. The expectation of D is

$$\begin{aligned}
ED &= \sum_{s=1}^{cb-1}\sum_{l=0}^{n-b-s} E[(\bar{C}_l^{(i)}(b)-\bar{C}^{(i)})(\bar{C}_l^{(j)}(b)-\bar{C}^{(j)})(\bar{C}^{(i)}_{l+(1-c)b+s}(cb)-\bar{C}^{(i)})(\bar{C}^{(j)}_{l+(1-c)b+s}(cb)-\bar{C}^{(j)})] \\
&= \sum_{s=1}^{cb-1}\sum_{l=0}^{n-b-s} E[Z_1Z_2Z_3Z_4] \\
&= \sum_{s=1}^{cb-1}\sum_{l=0}^{n-b-s}\left[\left(\frac{n-b}{bn}\right)\left(\frac{n-cb}{cbn}\right)\Sigma_{ij}^2 + \left(\frac{1}{b}-\frac{1}{n}-\frac{s}{cb^2}\right)^2\Sigma_{ii}\Sigma_{jj} + \left(\frac{1}{b}-\frac{1}{n}-\frac{s}{cb^2}\right)^2\Sigma_{ij}^2\right].
\end{aligned}$$

Consider the first summation in the above expression.

$$\sum_{s=1}^{cb-1}\sum_{l=0}^{n-b-s}\left(\frac{n-b}{bn}\right)\left(\frac{n-cb}{cbn}\right)$$

$$
\begin{aligned}
&= \sum_{s=1}^{cb-1} \left[\left(\frac{1}{cb^2} - \frac{c+1}{c} \frac{1}{bn} + \frac{1}{n^2} \right)(n-b+1) - \left(\frac{1}{cb^2} - \frac{c+1}{c} \frac{1}{bn} + \frac{1}{n^2} \right) s \right] \\
&= \left(\frac{1}{cb^2} - \frac{c+1}{c} \frac{1}{bn} + \frac{1}{n^2} \right)(n-b+1)(cb-1) - \left(\frac{1}{cb^2} - \frac{c+1}{c} \frac{1}{bn} + \frac{1}{n^2} \right) \left(\frac{c^2b^2}{2} - \frac{cb}{2} \right) \\
&= \frac{1}{cb^2} \cdot n \cdot cb + o\left(\frac{n}{b}\right) \\
&= \frac{n}{b} + o\left(\frac{n}{b}\right).
\end{aligned}
$$

Consider the second summation in ED.

$$
\begin{aligned}
&\sum_{s=1}^{cb-1} \sum_{l=0}^{n-b-s} \left(\frac{1}{b} - \frac{1}{n} - \frac{s}{cb^2} \right)^2 \\
&= \sum_{s=1}^{cb-1} \sum_{l=0}^{n-b-s} \left[\frac{s^2}{c^2b^4} + \left(\frac{2}{cb^2n} - \frac{2}{cb^3} \right) s + \left(\frac{1}{b^2} + \frac{1}{n^2} - \frac{2}{bn} \right) \right] \\
&= \sum_{s=1}^{cb-1} \left[-\frac{s^3}{c^2b^4} + \left(\frac{n}{c^2b^4} + \left(\frac{2}{c} - \frac{1}{c^2} \right) \frac{1}{b^3} + \frac{1}{c^2} \frac{1}{b^4} - \frac{2}{c} \frac{1}{b^2n} \right) s^2 \right. \\
&\qquad + \left[\left(\frac{4}{c} - 1 \right) \frac{1}{b^2} - \frac{2}{c} \frac{n}{b^3} + \left(2 - \frac{2}{c} \right) \frac{1}{bn} + \frac{2}{c} \frac{1}{b^2n} - \frac{2}{c} \frac{1}{b^3} - \frac{1}{n^2} \right] s \\
&\qquad \left. + \left(\frac{n}{b^2} + \frac{3}{n} - \frac{3}{b} - \frac{b}{n^2} + \frac{1}{b^2} + \frac{1}{n^2} - \frac{2}{bn} \right) \right] \\
&= -\frac{1}{c^2b^4} \left(\frac{c^2b^4}{4} - \frac{c^3b^3}{2} + \frac{c^2b^2}{4} \right) \\
&\quad + \left(\frac{n}{c^2b^4} + \left(\frac{2}{c} - \frac{1}{c^2} \right) \frac{1}{b^3} + \frac{1}{c^2} \frac{1}{b^4} - \frac{2}{c} \frac{1}{b^2n} \right) \cdot \left(\frac{c^3b^3}{3} - \frac{c^2b^2}{2} \frac{cb}{6} \right) \\
&\quad + \left(\left(\frac{4}{c} - 1 \right) \frac{1}{b^2} - \frac{2}{c} \frac{n}{b^3} + \left(2 - \frac{2}{c} \right) \frac{1}{bn} + \frac{2}{c} \frac{1}{b^2n} - \frac{2}{c} \frac{1}{b^3} - \frac{1}{n^2} \right) \cdot \left(\frac{c^2b^2}{2} - \frac{cb}{2} \right) \\
&\quad + \left(\frac{n}{b^2} + \frac{3}{n} - \frac{3}{b} - \frac{b}{n^2} + \frac{1}{b^2} + \frac{1}{n^2} - \frac{2}{bn} \right) \cdot (cb-1) \\
&= \frac{1}{c^2} \cdot \frac{n}{b^4} \cdot \frac{c^3b^3}{3} - \frac{2}{c} \cdot \frac{n}{b^3} \cdot \frac{c^2b^2}{2} + \frac{n}{b^2} \cdot cb + o\left(\frac{n}{b}\right) \\
&= \frac{c}{3} \cdot \frac{n}{b} + o\left(\frac{n}{b}\right).
\end{aligned}
$$

Plug in the above result, we have

$$
ED = \frac{c}{3}(\Sigma_{ii}\Sigma_{jj} + \Sigma_{ij}^2)\frac{n}{b} + \Sigma_{ij}^2 \frac{n}{b} + o\left(\frac{n}{b}\right).
$$

Finally, we calculate EG. For $s = cb, ..., (n-b)$. We show that the joint distribution of the $2p$ vector is

$$\begin{bmatrix} \bar{C}_l(b) - \bar{C} \\ \bar{C}_{l+(1-c)b+s}(cb) - \bar{C} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \left(\frac{n-b}{bn}\right)\Sigma & -\frac{1}{n}\Sigma \\ -\frac{1}{n}\Sigma & \left(\frac{n-cb}{cbn}\right)\Sigma \end{bmatrix} \right).$$

For $i \neq j$,

$$E[\bar{B}_l^{(i)}(b) - \bar{B}^{(i)}][\bar{B}_{l+(1-c)b+s}^{(j)}(cb) - \bar{B}^{(j)}] = E[\bar{B}_l^{(i)}(b) - \bar{B}^{(i)}] \cdot E[\bar{B}_{l+(1-c)b+s}^{(j)}(cb) - \bar{B}^{(j)}] = 0.$$

For $i = j$, we need to calculate $E[\bar{B}_l^{(i)}(b) - \bar{B}^{(i)}][\bar{B}_{l+(1-c)b+s}^{(i)}(cb) - \bar{B}^{(i)}]$ for $s = cb, ..., (n-b)$.. Let $Z_1^{(i)} = \bar{B}_l^{(i)}(b) - \bar{B}^{(i)}$, $Z_2 = \bar{B}_{l+(1-c)b+s}^{(i)}(cb) - \bar{B}^{(i)}$, then

$$\begin{bmatrix} Z_1^{(i)} \\ Z_2^{(i)} \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{n-b}{bn} & -\frac{1}{n} \\ -\frac{1}{n} & \frac{n-cb}{cbn} \end{bmatrix} \right),$$

hence

$$Z_1^{(i)} | Z_2^{(i)} \sim N\left(\frac{cb}{cb-n} Z_2^{(i)},\ \left(\frac{n-b}{bn} - \frac{cb}{n(n-cb)} \right) \right),$$
$$Z_2^{(i)} \sim N\left(0,\ \frac{n-cb}{cbn}\right).$$

Hence

$$\begin{aligned} E[Z_i^{(i)} Z_2^{(i)}] &= E_{Z_2^{(i)}}[E_{Z_1^{(i)}|Z_2^{(i)}}[Z_1^{(i)} Z_2^{(i)} | Z_2^{(i)}]] \\ &= E_{Z_2^{(i)}}\left[(Z_2^{(i)})^2 \cdot \frac{cb}{cb-n} \right] \\ &= \frac{cb}{cb-n} \cdot \frac{n-cb}{cbn} \\ &= -\frac{1}{n}. \end{aligned}$$

Then we have

$$\mathrm{Cov}(\bar{C}_l(b) - \bar{C},\ \bar{C}_{l+s}(b) - \bar{C}) = L \cdot \left(-\frac{1}{n}\right) I_p \cdot L^T = -\frac{1}{n} \cdot \Sigma,$$

which yields the joint distribution of the $2p$ vector. Again denote $Z_1 = \bar{C}_l^{(i)}(b) - \bar{C}^{(i)}$, $Z_2 = \bar{C}_l^{(j)}(b) - \bar{C}^{(j)}$, $Z_3 = \bar{C}_{l+(1-c)b+s}^{(i)}(cb) - \bar{C}^{(i)}$, $Z_4 = \bar{C}_{l+(1-c)b+s}^{(j)}(cb) - \bar{C}^{(j)}$. The expectation of G is

$$
\begin{aligned}
EG &= \sum_{s=cb}^{n-b} \sum_{l=0}^{n-b-s} E[(\bar{C}_l^{(i)}(b) - \bar{C}^{(i)})(\bar{C}_l^{(j)}(b) - \bar{C}^{(j)})(\bar{C}_{l+(1-c)b+s}^{(i)}(cb) - \bar{C}^{(i)})(\bar{C}_{l+(1-c)b+s}^{(j)}(cb) - \bar{C}^{(j)})] \\
&= \sum_{s=cb}^{n-b} \sum_{l=0}^{n-b-s} E[Z_1 Z_2 Z_3 Z_4] \\
&= \sum_{s=cb}^{n-b} \sum_{l=0}^{n-b-s} \left[\left(\frac{n-b}{bn}\right)\left(\frac{n-cb}{cbn}\right)\Sigma_{ij}^2 + \frac{1}{n^2}\Sigma_{ii}\Sigma_{jj} + \frac{1}{n^2}\Sigma_{ij}^2 \right].
\end{aligned}
$$

Consider the first summation in the above expression.

$$
\begin{aligned}
&\sum_{s=cb}^{n-b} \sum_{l=0}^{n-b-s} \left(\frac{n-b}{bn}\right)\left(\frac{n-cb}{cbn}\right) \\
&= \sum_{s=cn}^{n-cb} \left(\frac{1}{cb^2} - \frac{c+1}{c}\frac{1}{bn} + \frac{1}{n^2}\right)(n-b+1) - \left(\frac{1}{cb^2} - \frac{c+1}{c}\frac{1}{bn} + \frac{1}{n^2}\right)s \\
&= \left(\frac{1}{cb^2} - \frac{c+1}{c}\frac{1}{bn} + \frac{1}{n^2}\right)(n-b+1)[n-(1+c)b+1] \\
&\quad - \left(\frac{1}{cb^2} - \frac{c+1}{c}\frac{1}{bn} + \frac{1}{n^2}\right)\frac{[(c-1)b+n][n-(1+c)b+1]}{2} \\
&= \left(\frac{1}{cb^2} - \frac{c+1}{c}\frac{1}{bn} + \frac{1}{n^2}\right)[n^2-(2+c)bn+2n+(1+c)b^2-(2+c)b+1] \\
&\quad \left(\frac{1}{cb^2} - \frac{c+1}{c}\frac{1}{bn} + \frac{1}{n^2}\right)\cdot\frac{n^2-(1+c)bn+n+(c-1)bn-(1+c)(c-1)b^2+(c-1)b}{2} \\
&= \left(\frac{1}{cb^2}\cdot n^2 - \frac{1}{cb^2}(2+c)\cdot bn - \frac{1+c}{c}\frac{1}{bn}\cdot n^2\right) \\
&\quad - \left(\frac{1}{cb^2}\cdot\frac{n^2}{2} - \frac{1}{cb^2}\cdot bn - \frac{1+c}{c}\cdot\frac{1}{bn}\cdot\frac{n^2}{2}\right) + o\left(\frac{n}{b}\right) \\
&= \frac{1}{2c}\frac{n^2}{b^2} - \left(\frac{3}{2}+\frac{3}{2c}\right)\frac{n}{b} + o\left(\frac{n}{b}\right).
\end{aligned}
$$

Consider the second summation in EG.

$$
\begin{aligned}
&\sum_{s=cb}^{n-b} \sum_{l=0}^{n-b-s} \frac{1}{n^2} \\
&= \sum_{s=cb}^{n-b} -\frac{s}{n^2} + \left(\frac{1}{n} - \frac{b}{n^2} + \frac{1}{n^2}\right)
\end{aligned}
$$

$$= -\frac{1}{n^2}\left(\frac{n^2}{2} - bn\right) + \left(\frac{1}{n} - \frac{b}{n^2} + \frac{1}{n^2}\right)\cdot[n-(1+c)b+1]$$

$$= o\left(\frac{n}{b}\right).$$

Plug in the above result, we have

$$EG = \Sigma_{ij}^2 \frac{1}{2c}\cdot\frac{n^2}{b^2} - \Sigma_{ij}^2\left(\frac{3}{2}+\frac{3}{2c}\right)\cdot\frac{n}{b} + o\left(\frac{n}{b}\right).$$

Combine $E[((1-c)b+1)(n-b+1)OL^{(i)}OL^{(j)}]$, ED and EG, we can calculate A_3.

$$\begin{aligned}
A_3 &= E\left[-\frac{2c^2}{(1-c)^2}\frac{b^2}{n^2}\left[\sum_{l=0}^{n-b}(\bar{C}_l^{(i)}(b)-\bar{C}^{(i)})(\bar{C}_l^{(j)}(b)-\bar{C}^{(j)})\right]\left[\sum_{l=0}^{n-cb}(\bar{C}_l^{(i)}(cb)-\bar{C}^{(i)})(\bar{C}_l^{(j)}(cb)-\bar{C}^{(j)})\right]\right] \\
&= -\frac{2c^2}{(1-c)^2}\cdot\frac{n^2}{b^2}\left[E[((1-c)b+1)(n-b+1)OL^{(i)}OL^{(j)}]+2ED+2EG\right] \\
&= -\frac{2c^2}{(1-c)^2}\cdot\left[(1-c)(\Sigma_{ii}\Sigma_{jj}+\Sigma_{ij}^2)\cdot\frac{b}{n}+\frac{1-c}{c}\Sigma_{ij}^2\cdot\frac{b}{n}\right. \\
&\qquad +\frac{2c}{3}(\Sigma_{ii}\Sigma_{jj}+\Sigma_{ij}^2)\cdot\frac{b}{n}+2\Sigma_{ij}^2\cdot\frac{b}{n} \\
&\qquad \left.+\frac{1}{c}\Sigma_{ij}^2-\frac{3c+3}{c}\Sigma_{ij}^2\cdot\frac{b}{n}\right]+o\left(\frac{b}{n}\right) \\
&= \frac{2c^2(c-3)}{3(1-c)^2}(\Sigma_{ii}\Sigma_{jj}+\Sigma_{ij}^2)\cdot\frac{b}{n}-\frac{2c^2}{(1-c)^2}\left(\frac{1-c+2c-3c-3}{c}\right)\cdot\Sigma_{ij}^2\cdot\frac{b}{n} \\
&\quad -\frac{2c}{(1-c)^2}\cdot\Sigma_{ij}^2+o\left(\frac{b}{n}\right) \\
&= \frac{2c^2(c-3)}{3(1-c)^2}(\Sigma_{ii}\Sigma_{jj}+\Sigma_{ij}^2)\cdot\frac{b}{n}-\frac{2c^2}{(1-c)^2}\left(\frac{-2c-2}{c}\right)\cdot\Sigma_{ij}^2\cdot\frac{b}{n} \\
&\quad -\frac{2c}{(1-c)^2}\cdot\Sigma_{ij}^2+o\left(\frac{b}{n}\right).
\end{aligned} \tag{2.A.5}$$

Now we calculate $E[\tilde{\Sigma}_{ft,L,ij}^2]$ by combining A_1, A_2 and A_3.

$$\begin{aligned}
E[\tilde{\Sigma}_{ft,L,ij}^2] &= A_1+A_2+A_3 \\
&= \frac{4c^3-6c^2+2}{3(1-c)^2}(\Sigma_{ii}\Sigma_{jj}+\Sigma_{ij}^2)\cdot\frac{b}{n}+\frac{1+c^2-2c}{(1-c)^2}\Sigma_{ij}^2 \\
&\quad +\left(-\frac{4}{(1-c)^2}-\frac{4c^3}{(1-c)^2}+\frac{2c(2c+2)}{(1-c)^2}\right)\Sigma_{ij}^2\cdot\frac{b}{n}+o\left(\frac{b}{n}\right)
\end{aligned}$$

$$= \frac{4c^3-6c^2+2}{3(1-c)^2}(\Sigma_{ii}\Sigma_{jj}+\Sigma_{ij}^2)\cdot\frac{b}{n}+\Sigma_{ij}^2+\frac{-4c^3+4c^2+4c-4}{(1-c)^2}\Sigma_{ij}^2\cdot\frac{b}{n}+o\left(\frac{b}{n}\right).$$

We also need $E[\tilde{\Sigma}_{ft,L,ij}]^2$ in order to calculate $Var[\tilde{\Sigma}_{ft,L,ij}]$. Consider $E[(\bar{C}_l^{(i)}(b)-\bar{C}^{(i)})(\bar{C}_l^{(j)}(b)-\bar{C}^{(j)})]$ and $E[(\bar{C}_l^{(i)}(cb)-\bar{C}^{(i)})(\bar{C}_l^{(j)}(cb)-\bar{C}^{(j)})]$. By (2.3.1),

$$\bar{C}_l(b)-\bar{C}\sim N\left(0,\frac{n-b}{bn}\Sigma\right),$$

$$\bar{C}_l(cb)-\bar{C}\sim N\left(0,\frac{n-cb}{cbn}\Sigma\right).$$

Therefore

$$E[(\bar{C}_l^{(i)}(b)-\bar{C}^{(i)})(\bar{C}_l^{(j)}(b)-\bar{C}^{(j)})]=\frac{n-b}{bn}\Sigma_{ij},$$

$$E[(\bar{C}_l^{(i)}(cb)-\bar{C}^{(i)})(\bar{C}_l^{(j)}(cb)-\bar{C}^{(j)})]=\frac{n-cb}{cbn}\Sigma_{ij},$$

resulting

$$\begin{aligned}(E[\tilde{\Sigma}_{ft,L,ij}])^2 &= \Bigg(\frac{1}{1-c}\cdot\frac{b}{n}\sum_{l=0}^{n-b}E[(\bar{C}_l^{(i)}(b)-\bar{C}_n^{(i)})(\bar{C}_l^{(j)}(b)-\bar{C}_n^{(j)})]\\ &\quad -\frac{c}{1-c}\cdot\frac{cb}{n}\sum_{l=0}^{n-cb}E[(\bar{C}_l^{(i)}(cb)-\bar{C}_n^{(i)})(\bar{C}_l^{(j)}(cb)-\bar{C}_n^{(j)})]\Bigg)^2\\ &=\left(\frac{1}{1-c}\cdot\frac{b}{n}\cdot(n-b+1)\cdot\frac{n-b}{bn}\Sigma_{ij}-\frac{c}{1-c}\cdot\frac{cb}{n}(n-cb+1)\cdot\frac{n-cb}{cbn}\Sigma_{ij}\right)^2\\ &=\Sigma_{ij}^2\left(\frac{1}{1-c}\cdot\frac{b}{n}\cdot(n-b+1)\cdot\frac{n-b}{bn}-\frac{c}{1-c}\cdot\frac{cb}{n}(n-cb+1)\cdot\frac{n-cb}{cbn}\right)^2\\ &=\Sigma_{ij}^2\left(1+\frac{4c^2+4c-4c^3-4}{(1-c)^2}\cdot\frac{b}{n}+o\left(\frac{b}{n}\right)\right)\\ &=\Sigma_{ij}^2-\frac{4c^3-4c^2-4c+4}{(1-c)^2}\Sigma_{ij}^2\cdot\frac{b}{n}+o\left(\frac{b}{n}\right).\end{aligned}$$

Plug in the result from univariate case. The variance can be calculated as

$$\begin{aligned}Var[\tilde{\Sigma}_{ft,L,ij}] &= E[\tilde{\Sigma}_{ft,L,ij}^2]-E[\tilde{\Sigma}_{ft,L,ij}]^2\\ &=\left(\frac{4c^3-6c^2+2}{3(1-c)^2}(\Sigma_{ii}\Sigma_{jj}+\Sigma_{ij}^2)\cdot\frac{b}{n}+\Sigma_{ij}^2+\frac{-4c^3+4c^2+4c-4}{(1-c)^2}\Sigma_{ij}^2\cdot\frac{b}{n}\right)\end{aligned}$$

$$
\begin{aligned}
&-\left(\Sigma_{ij}^2 - \frac{4c^3-4c^2-4c+4}{(1-c)^2}\Sigma_{ij}^2 \cdot \frac{b}{n}\right) + o\left(\frac{b}{n}\right) \\
&= \frac{4c^3-6c^2+2}{3(1-c)^2}(\Sigma_{ii}\Sigma_{jj}+\Sigma_{ij}^2)\cdot\frac{b}{n} + o\left(\frac{b}{n}\right) \\
&= \frac{4c[(1-c)^2+\frac{1}{2c}(c-1)^2]}{3(1-c)^2}(\Sigma_{ii}\Sigma_{jj}+\Sigma_{ij}^2)\cdot\frac{b}{n} + o\left(\frac{b}{n}\right) \\
&= \left(\frac{4c}{3}+\frac{2}{3}\right)(\Sigma_{ii}\Sigma_{jj}+\Sigma_{ij}^2)\cdot\frac{b}{n} + o\left(\frac{b}{n}\right).
\end{aligned}
$$

Lemma 6($**$). Under *condition 2* and *conditions 3*, if

$$\sum_{k=1}^{b}(\Delta_2 w_k)^2 \leq O\left(\frac{1}{b^2}\right),$$

then

$$Var[\tilde{\Sigma}_{w,L,ij}] = (\Sigma_{ii}\Sigma_{jj}+\Sigma_{ij}^2)\left[\frac{2}{3}\sum_{k=1}^{b}(\Delta_2 w_k)^2 k^3\cdot\frac{1}{n} + 2\sum_{t=1}^{b-1}\sum_{u=1}^{b-t}\Delta_2 w_u \Delta_2 w_{t+u}\left(\frac{2}{3}u^3+u^2 t\right)\cdot\frac{1}{n}\right] + o\left(\frac{b}{n}\right).$$

Proof. Note

$$(\Delta_2 w_k)^2 \leq \sum_{k=1}^{b}(\Delta_2 w_k)^2 \leq O\left(\frac{1}{b^2}\right),$$

hence

$$a_k = b\cdot\Delta_2 w_k \leq O(1).$$

Consider

$$\tilde{\Sigma}_{w,L,ij} = \frac{1}{n}\sum_{k=1}^{b}\sum_{l=0}^{n-k} k^2 \Delta_2 w_n(k)[\bar{C}_l^{(1)}(k)-\bar{C}^{(i)}][\bar{C}_l^{(j)}(k)-\bar{C}^{(j)}].$$

Let $c_k = k/b$ for $k = 1,...,b$, also denote $a_k = b\cdot\Delta_2 w_k$ for simplicity. Consider:

$$
\begin{aligned}
\tilde{\Sigma}_{w,L,ij} &= \frac{1}{n}\sum_{k=1}^{b}\sum_{l=0}^{n-k} k^2 \Delta_2 w_n(k)[\bar{C}_l^{(i)}(k)-\bar{C}^{(i)}][\bar{C}_l^{(j)}(k)-\bar{C}^{(j)}] \\
&= \frac{1}{n}\sum_{k=1}^{b}\sum_{l=0}^{n-k} c_k^2 b^2 \Delta_2 w_n(k)[\bar{C}_l^{(i)}(k)-\bar{C}^{(i)}][\bar{C}_l^{(j)}(k)-\bar{C}^{(j)}] \\
&= \sum_{k=1}^{b} c_k a_k \left(\frac{c_k b}{n}\sum_{l=0}^{n-c_k b}[\bar{C}_l^{(i)}(k)-\bar{C}^{(i)}][\bar{C}_l^{(j)}(k)-\bar{C}^{(j)}]\right).
\end{aligned}
$$

Recall (2.A.4) and (2.A.5) from flat-top estimate.

$$E\left[\frac{b^2}{n^2}\cdot\left(\sum_{l=0}^{n-b}(\bar{C}_l^{(i)}(b)-\bar{C}^{(i)})(\bar{C}_l^{(j)}(b)-\bar{C}^{(j)})\right)^2\right]=\Sigma_{ij}^2+\frac{2}{3}(\Sigma_{ii}\Sigma_{jj}+\Sigma_{ij}^2)\cdot\frac{b}{n}-4\Sigma_{ij}^2\cdot\frac{b}{n}+o\left(\frac{b}{n}\right),$$

$$E\left[\frac{b^2}{n^2}\cdot\left(\sum_{p=0}^{n-c_tb}(\bar{C}_p^{(i)}(c_tb)-\bar{C}^{(i)})(\bar{C}_p^{(j)}(c_tb)-\bar{C}^{(j)})\right)\cdot\left(\sum_{q=0}^{n-c_{u+t}b}(\bar{C}_q^{(i)}(c_{t+u}b)-\bar{C}^{(i)})(\bar{C}_q^{(j)}(c_{t+u}b)-\bar{C}^{(j)})\right)\right]$$
$$=\frac{1}{c}\Sigma_{ij}^2+\frac{3-c}{3}(\Sigma_{ii}\Sigma_{jj}+\Sigma_{ij}^2)\frac{b}{n}-\frac{2c+2}{c}\Sigma_{ij}^2\frac{b}{n}+o\left(\frac{b}{n}\right).$$

Define $A_{1,ij}^{(k)}$ and $A_{2,ij}^{(ut)}$ as below then we have

$$\begin{aligned}A_{1,ij}^{(k)}&=E\left[\frac{(c_kb)^2}{n^2}\cdot\left(\sum_{k=0}^{n-c_kb}(\bar{C}_l^{(i)}(c_kb)-\bar{C}^{(i)})(\bar{C}_l^{(j)}(c_kb)-\bar{C}^{(j)})\right)^2\right]\\&=\left(\frac{2}{3}(\Sigma_{ii}\Sigma_{jj}+\Sigma_{ij}^2)-4\Sigma_{ij}^2\right)\cdot\frac{c_kb}{n}+\Sigma_{ij}^2+o\left(\frac{b}{n}\right).\end{aligned}$$

and

$$\begin{aligned}A_{2,ij}^{(ut)}&=E\left[\frac{(c_{u+t}b)^2}{n^2}\cdot\left(\sum_{p=0}^{n-c_tb}(\bar{C}_p^{(i)}(c_tb)-\bar{C}^{(i)})(\bar{C}_p^{(j)}(c_tb)-\bar{C}^{(j)})\right)\right.\\&\qquad\left.\cdot\left(\sum_{q=0}^{n-c_{u+t}b}(\bar{C}_q^{(i)}(c_{t+u}b)-\bar{C}^{(i)})(\bar{C}_q^{(j)}(c_{t+u}b)-\bar{C}^{(j)})\right)\right]\\&=\left[\left(1-\frac{c_u}{3c_{u+t}}\right)(\Sigma_{ii}\Sigma_{jj}+\Sigma_{ij}^2)-\left(2+\frac{2c_{u+t}}{c_u}\right)\Sigma_{ij}^2\right]\frac{c_{u+t}b}{n}+\frac{c_{u+t}}{c_u}\Sigma_{ij}^2+o\left(\frac{b}{n}\right)\\&=\left[\left(c_{u+t}-\frac{c_u}{3}\right)(\Sigma_{ii}\Sigma_{jj}+\Sigma_{ij}^2)-\left(2c_{u+t}+\frac{2c_{u+t}^2}{c_u}\right)\Sigma_{ij}^2\right]\frac{b}{n}+\frac{c_{u+t}}{c_u}\Sigma_{ij}^2+o\left(\frac{b}{n}\right).\end{aligned}$$

In order to calculate $Var[\tilde{\Sigma}_{w,L,ij}]$, we will calculate $E[\tilde{\Sigma}_{w,L,ij}^2]$ and $(E[\tilde{\Sigma}_{w,L,ij}])^2$.

$$\begin{aligned}E[\tilde{\Sigma}_{w,L,ij}^2]&=E\left[\left(\sum_{k=1}^{b}c_ka_k\cdot\left[\frac{c_kb}{n}\sum_{l=0}^{n-c_kb}(\bar{C}_l^{(i)}(k)-\bar{C}^{(i)})(\bar{C}_l^{(j)}(k)-\bar{C}^{(j)})\right]\right)^2\right]\\&=E\left[\sum_{k=1}^{b}\left(c_ka_k\cdot\left[\frac{c_kb}{n}\sum_{l=0}^{n-c_kb}(\bar{C}_l^{(i)}(k)-\bar{C}^{(i)})(\bar{C}_l^{(j)}(k)-\bar{C}^{(j)})\right]\right)^2\right.\end{aligned}$$

$$
\begin{aligned}
&+2\sum_{t=1}^{b-1}\sum_{u=1}^{b-t} c_u^2 a_u c_{t+u}^2 a_{t+u}\cdot\frac{b^2}{n^2}\left(\sum_{p=0}^{n-c_t b}(\bar{C}_p^{(i)}(c_t b)-\bar{C}^{(i)})(\bar{C}_p^{(j)}(c_t b)-\bar{C}^{(j)})\right)\\
&\qquad\qquad\cdot\left(\sum_{q=0}^{n-c_{t+u}b}(\bar{C}_q^{(i)}(c_{t+u}b)-\bar{C}^{(i)})(\bar{C}_q^{(j)}(c_{t+u}b)-\bar{C}^{(j)})\right)\Bigg]\\
&=\sum_{k=1}^{b} c_k^2 a_k^2\cdot E\left[\frac{(c_k b)^2}{n^2}\cdot\left(\sum_{l=0}^{n-c_k b}(\bar{C}_l^{(i)}(k)-\bar{C}^{(i)})(\bar{C}_l^{(j)}(k)-\bar{C}^{(j)})\right)^2\right]\\
&\quad+2\sum_{t=1}^{b-1}\sum_{u=1}^{b-t} c_u^2 a_u a_{t+u}\cdot E\left[\frac{(c_{u+t}b)^2}{n^2}\cdot\left(\sum_{p=0}^{n-c_t b}(\bar{C}_p^{(i)}(c_t b)-\bar{C}^{(i)})(\bar{C}_p^{(j)}(c_t b)-\bar{C}^{(j)})\right)\right.\\
&\qquad\qquad\left.\cdot\left(\sum_{q=0}^{n-c_{t+u}b}(\bar{C}_q^{(i)}(c_{t+u}b)-\bar{C}^{(i)})(\bar{C}_q^{(j)}(c_{t+u}b)-\bar{C}^{(j)})\right)\right]\\
&=\sum_{k=1}^{b} c_k^2 a_k^2 A_{1,ij}^{(k)}+2\sum_{t=1}^{b-1}\sum_{u=1}^{b-t} c_u^2 a_u a_{u+t} A_{2,ij}^{(ut)}\\
&=o\left(\frac{b}{n}\right)+\sum_{k=1}^{b} c_k^2 a_k^2\left[\left(\frac{2}{3}(\Sigma_{ii}\Sigma_{jj}+\Sigma_{ij}^2)-4\Sigma_{ij}^2\right)\cdot\frac{c_k b}{n}+\Sigma_{ij}^2\right]\\
&+2\sum_{t=1}^{b-1}\sum_{u=1}^{b-t} c_u^2 a_u a_{u+t}\left[\left[\left(c_{u+t}-\frac{c_u}{3}\right)(\Sigma_{ii}\Sigma_{jj}+\Sigma_{ij}^2)-\left(2c_{u+t}+\frac{2c_{u+t}^2}{c_u}\right)\Sigma_{ij}^2\right]\cdot\frac{b}{n}+\frac{c_{u+t}}{c_u}\Sigma_{ij}^2\right]\\
&=\left[\sum_{k=1}^{b} c_k^2 a_k^2\cdot\Sigma_{ij}^2+2\sum_{t=1}^{b-1}\sum_{u=1}^{b-t} c_u^2 a_u a_{u+t}\frac{c_{u+t}}{c_u}\cdot\Sigma_{ij}^2\right]+\left[\sum_{k=1}^{b} c_k^3 a_k^2\left(\frac{2}{3}[\Sigma_{ii}\Sigma_{jj}+\Sigma_{ij}^2]-4\Sigma_{ij}^2\right)\cdot\frac{b}{n}\right.\\
&\quad\left.+2\sum_{t=1}^{b-1}\sum_{u=1}^{b-t} c_u^2 a_u a_{u+t}\left[\left(c_{u+t}-\frac{c_u}{3}\right)(\Sigma_{ii}\Sigma_{jj}+\Sigma_{ij}^2)-\left(2c_{u+t}+\frac{2c_{u+t}^2}{c_u}\right)\Sigma_{ij}^2\right]\frac{b}{n}\right]+o\left(\frac{b}{n}\right)\\
&=(\sum_{k=1}^{b} a_k c_k)^2\Sigma_{ij}^2+\sum_{k=1}^{b} c_k^3 a_k^2\left(\frac{2}{3}(\Sigma_{ii}\Sigma_{jj}+\Sigma_{ij}^2)-4\Sigma_{ij}^2\right)\cdot\frac{b}{n}\\
&\quad+2\sum_{t=1}^{b-1}\sum_{u=1}^{b-t} c_u^2 a_u a_{u+t}\left[\left(c_{u+t}-\frac{c_u}{3}\right)(\Sigma_{ii}\Sigma_{jj}+\Sigma_{ij}^2)-\left(2c_{u+t}+\frac{2c_{u+t}^2}{c_u}\right)\Sigma_{ij}^2\right]\cdot\frac{b}{n}+o\left(\frac{b}{n}\right).
\end{aligned}
$$

Now we calculate $(E[\tilde{\Sigma}_{w,L,ij}])^2$. Since

$$
\bar{C}_l(c_k b)-\bar{C}\sim N\left(0,\ \frac{n-c_k b}{c_k bn}\Sigma\right),
$$

$$
E[(C_l^{(i)}(c_k b)-\bar{C}^{(i)})(C_l^{(j)}(c_k b)-\bar{C}^{(j)})]=\frac{n-c_k b}{c_k bn}\Sigma_{ij},
$$

then

$$
(E[\tilde{\Sigma}_{w,L,ij}])^2)=\left(\frac{1}{n}\sum_{k=1}^{b}\sum_{l=0}^{n-k}k^2\Delta_2 w_k E[(C_l^{(i)}(c_k b)-\bar{C}^{(i)})(C_l^{(j)}(c_k b)-\bar{C}^{(j)})]\right)^2
$$

$$= \left(\sum_{k=1}^{b} c_k a_k \left[\frac{c_k b}{n} \sum_{l=0}^{n-c_k b} E[(C_l^{(i)}(c_k b) - \bar{C}^{(i)})(C_l^{(j)}(c_k b) - \bar{C}^{(j)})] \right] \right)^2$$

$$= \left(\sum_{k=1}^{b} c_k a_k \left[\frac{c_k b}{n} \cdot (n - c_k b + 1) \cdot \frac{n - c_k b}{c_k b n} \cdot \Sigma_{ij} \right] \right)^2 \text{ apply (1.4.3)}$$

$$= \Sigma_{ij}^2 \left[(\sum_{k=1}^{b} a_k c_k)^2 - \sum_{k=1}^{b} 4 a_k^2 c_k^3 \cdot \frac{b}{n} - 2 \sum_{t=1}^{b-1} \sum_{u=1}^{b-t} a_u a_{u+t} (2c_u^2 c_{u+t} + 2 c_u c_{u+t}^2) \cdot \frac{b}{n} \right] + o\left(\frac{b}{n}\right),$$

resulting

$$\begin{aligned}
& Var[\tilde{\Sigma}_{w,L,ij}] = E[\tilde{\Sigma}_{w,L,ij}^2] - (E[\tilde{\Sigma}_{w,L,ij}])^2 \\
= & \sum_{k=1}^{b} c_k^3 a_k^2 \left(\left[\frac{2}{3} (\Sigma_{ii}\Sigma_{jj} + \Sigma_{ij}^2) - 4\Sigma_{ij}^2 \right] + 4\Sigma_{ij}^2 \right) \cdot \frac{b}{n} \\
& + 2 \sum_{t=1}^{b-1} \sum_{u=1}^{b-t} \left(c_u^2 a_u a_{u+t} \left[\left(c_{u+t} - \frac{c_u}{3} \right) (\Sigma_{ii}\Sigma_{jj} + \Sigma_{ij}^2) - \left(2c_{u+t} + \frac{2c_{u+t}^2}{c_u} \right) \Sigma_{ij}^2 \right] \right. \\
& \left. + a_u a_{u+t} (2c_u^2 c_{u+t} + 2c_u c_{u+t}^2) \Sigma_{ij}^2 \right) \cdot \frac{b}{n} + o\left(\frac{b}{n}\right) \\
= & \sum_{k=1}^{b} \frac{2}{3} c_k^3 a_k^2 (\Sigma_{ii}\Sigma_{jj} + \Sigma_{ij}^2) \cdot \frac{b}{n} + 2 \sum_{t=1}^{b} \sum_{u=1}^{b-t} \left(c_u^2 c_{u+t} - \frac{1}{3} c_u^3 \right) a_u a_{u+t} (\Sigma_{ii}\Sigma_{jj} + \Sigma_{ij}^2) \cdot \frac{b}{n} + o\left(\frac{b}{n}\right) \\
= & \sum_{k=1}^{b} \frac{2}{3} \left(\frac{k}{b}\right)^3 (b\Delta_2 w_k)^2 (\Sigma_{ii}\Sigma_{jj} + \Sigma_{ij}^2) \cdot \frac{b}{n} \\
& + 2 \sum_{t=1}^{b-1} \sum_{u=1}^{b-t} \left[\left(\left(\frac{u}{b}\right)^2 \frac{u+t}{b} - \frac{1}{3} \left(\frac{u}{b}\right)^3 \right) b\Delta_2 w_u \cdot b\Delta_2 w_{u+t} \right] (\Sigma_{ii}\Sigma_{jj} + \Sigma_{ij}^2) \cdot \frac{b}{n} + o\left(\frac{b}{n}\right) \\
= & (\Sigma_{ii}\Sigma_{jj} + \Sigma_{ij}^2) \cdot \left[\frac{2}{3} \sum_{k=1}^{b} (\Delta_2 w_k)^2 k^3 \cdot \frac{1}{n} + 2 \sum_{t=1}^{b-1} \sum_{u=1}^{b-t} \Delta_2 w_u \cdot \Delta_2 w_{t+u} \cdot \left(\frac{2}{3} u^3 + u^2 t \right) \cdot \frac{1}{n} \right] + o\left(\frac{b}{n}\right).
\end{aligned}$$

Lemma 7(***). Let *condition 1* hold for g and *condition 2, 3* hold. If as $n \to \infty$,

$$b\psi(n)^2 logn \left(\sum_{k=1}^{b} |\Delta_2 w_n(k)| \right)^2 \to 0, \tag{2.A.6}$$

and

$$\psi(n)^2 \sum_{k=1}^{b} |\Delta_2 w_n(k)| \to 0, \tag{2.A.7}$$

then $\hat{\Sigma}_w \to L\tilde{\Sigma}_w L^T = \tilde{\Sigma}_{w,L}$ $w.p.$ 1.

Proof. We will prove for $i, j = 1, ..., p$, $\hat{\Sigma}_{w,ij} \to \tilde{\Sigma}_{w,L,ij}$. In this proof, let $Y_i = g(x_i) - E_F g$.

Recall

$$\hat{\Sigma}_{w,ij} - \tilde{\Sigma}_{w,L,ij} = \frac{1}{n}\sum_{k=1}^{b}\sum_{l-0}^{n-k} k^2 \Delta_2 w_k [(\bar{Y}_l^{(i)}(k) - \bar{Y}^{(i)})(\bar{Y}_l^{(j)}(k) - \bar{Y}^{(j)}) - (\bar{C}_l^{(i)}(k) - \bar{C}^{(i)})(\bar{C}_l^{(j)}(k) - \bar{C}^{(j)})].$$

Let

$$A_k = k(\bar{Y}_l(k) - \bar{C}_l(k)),$$

$$D_k = B(l+k) - B(l),$$

$$E_{n,k} = k\bar{B},$$

$$F_{n,k} = k(\bar{Y} - \bar{C}),$$

since

$$\begin{aligned} k(\bar{Y}_l^{(i)}(k) - \bar{Y}^{(i)}) &= k(\bar{Y}_l^{(i)}(k) - \bar{Y}^{(i)} + \bar{C}_l^{(i)}(k) - \bar{C}_l^{(i)}(k) + \bar{C}^{(i)} - \bar{C}^{(i)}) \\ &= k(\bar{Y}_l^{(i)}(k) - \bar{C}_l^{(i)}(k)) + (k\bar{C}_l^{(i)}(k) - k\bar{C}^{(i)}) - k(\bar{Y}^{(i)} - \bar{C}^{(i)}) \\ &= A_k^{(i)} + (LD_k)^{(i)} - (LE_{n,k})^{(i)} - F_{n,k}^{(i)}, \end{aligned}$$

$$\begin{aligned} |\hat{\Sigma}_{w,ij} - \tilde{\Sigma}_{w,L,ij}| \le \frac{1}{n}\sum_{k=1}^{b}\sum_{l=0}^{n-k} |\Delta_2 w_k| \cdot |[A_k^{(i)} + (LD_k)^{(i)} - (LE_{n,k})^{(i)} - F_{n,k}^{(i)}] \cdot [A_k^{(j)} + (LD_k)^{(j)} - (LE_{n,k})^{(j)} - F_{n,k}^{(j)}] \\ - [(LD_k)^{(i)} - (LE_{n,k})^{(i)}] \cdot [(LD_k)^{(j)} - (LE_{n,k})^{(j)}]| \\ = \frac{1}{n}\sum_{k=1}^{b}\sum_{l=0}^{n-k} |\Delta_2 w_k| \cdot |A_k^{(i)}A_k^{(j)} + A_k^{(i)}(LD_k)^{(j)} - A_k^{(i)}(LE_{n,k})^{(j)} - A_k^{(i)}F_{n,k}^{(j)} \\ + (LD_k)^{(i)}A_k^{(j)} - (LD_k)^{(i)}F_{n,k}^{(j)} - (LE_{n,k})^{(i)}A_k^{(j)} + (LE_{n,k})^{(i)}F_{n,k}^{(j)} \\ - F_{n,k}^{(i)}A_k^{(j)} - F_{n,k}^{(i)}(LD_k)^{(j)} + F_{n,k}^{(i)}(LE_{n,k})^{(j)} + F_{n,k}^{(i)}F_{n,k}^{(j)}| \\ = \frac{1}{n}\sum_{k=1}^{b}\sum_{l=0}^{n-k} |\Delta_2 w_k| \cdot |A_k^{(i)}A_k^{(j)} + F_{n,k}^{(i)}F_{n,k}^{(j)} + [A_k^{(i)}(LD_k)^{(j)} + (LD_k)^{(i)}A_k^{(j)}] \\ - [A_k^{(i)}(LE_{n,k})^{(j)} + (LE_{n,k})^{(i)}A_k^{(j)}] - [A_k^{(i)}F_{n,k}^{(j)} + F_{n,k}^{(i)}A_k^{(j)}] \end{aligned}$$

$$- [(LD_k)^{(i)} F_{n,k}^{(j)} + F_{n,k}^{(i)} (LD_k)^{(j)}] + [(LE_{n,k})^{(i)} F_{n,k}^{(j)} + F_{n,k}^{(i)} (LE_{n,k})^{(j)}]|.$$

Now we prove each term in the above expression goes to 0 as $n \to \infty$.

1. $\frac{1}{n} \sum_{k=1}^{b} \sum_{l=0}^{n-k} |\Delta_2 w_k| \cdot |A_k^{(i)} A_k^{(j)}|$. By *condition 1*, and note $l + k \leq n$,

$$\begin{aligned}
|A_k^{(i)}| &= |k(\bar{Y}_l(k) - \bar{C}_l(k))| \\
&= |[\sum_{t=1}^{k+l} Y_t^{(i)} - \sum_{t=1}^{l} Y_t^{(i)}] - [C^{(i)}(k+l) - C^{(i)}(l)]| \\
&\leq |\sum_{t=1}^{k+l} Y_t^{(i)} - C^{(i)}(k+l)| + |\sum_{t=1}^{l} Y_t^{(i)} - C^{(i)}(l)| \\
&\leq D \cdot \psi(l+k) + D \cdot \psi(l) \\
&\leq 2D\psi(n),
\end{aligned}$$

therefore by the assumption of the lemma, as $n \to \infty$,

$$\frac{1}{n} \sum_{k=1}^{b} \sum_{l=1}^{n-k} |\Delta_2 w_k| \cdot |A_k^{(i)} A_k^{(j)}| \leq 4D^2 \cdot \psi^2(n) \sum_{k=1}^{b} |\Delta_2 w_k| \left(\frac{n-k+1}{n} \right) \to 0.$$

2. $\frac{1}{n} \sum_{k=1}^{b} \sum_{l=0}^{n-k} |\Delta_2 w_k| \cdot |F_{n,k}^{(i)} F_{n,k}^{(j)}|$. By *condition 1*,

$$\begin{aligned}
|F_{n,k}^{(i)}| &= |k(\bar{Y}^{(i)} - \bar{C}^{(i)})| \\
&= |\frac{k}{n} (\sum_{t=1}^{n} Y_t^{(i)} - C^{(i)}(n))| \\
&\leq \frac{k}{n} D\psi(n),
\end{aligned}$$

hence by *condition 3* and (2.A.7), as $n \to \infty$,

$$\frac{1}{n} \sum_{k=1}^{b} \sum_{l=0}^{n-k} |\Delta_2 w_k| \cdot |F_{n,k}^{(i)} F_{n,k}^{(j)}| \leq \frac{b^2}{n^2} \cdot D^2 \cdot \psi^2(n) \sum_{k=1}^{b} |\Delta_2 w_k| \left(\frac{n-k+1}{n} \right) \to 0.$$

3. $\frac{1}{n} \sum_{k=1}^{b} \sum_{l=0}^{n-k} |\Delta_2 w_k| \cdot |A_k^{(i)} (LD_k)^{(j)} + (LD_k)^{(i)} A_k^{(j)}|$. First consider $LD_k^{(i)}$, by *proposition 4*,

$$|(LD_k)^{(i)}| = |[LB(l+k)]^{(i)} - [LB(l)]^{(i)}| = |C^{(i)}(l+k) - C^{(i)}(l)|$$

$$\leq \sup_{0\leq l\leq n-b}\sup_{0\leq s\leq b}|C^{(i)}(l+s)-C^{(i)}(l)|$$

$$\leq 2(1+\varepsilon)(b\Sigma_{ii}logn)^{1/2},$$

by (2.A.6), as $n\to\infty$,

$$\frac{1}{n}\sum_{k=1}^{b}\sum_{l=0}^{n-k}|\Delta_2 w_k|\cdot|A_k^{(i)}(LD_k)^{(j)}+(LD_k)^{(i)}A_k^{(j)}|$$

$$\leq 2\left[[2(1+\varepsilon)(b\Sigma_{ii}logn)^{1/2}]\cdot[2D\psi(n)]\cdot\sum_{k=1}^{b}|\Delta_2 w_k|\left(\frac{n-k+1}{n}\right)\right]$$

$$=8D(1+\varepsilon)\Sigma_{ii}^{1/2}\cdot[b\psi^2(n)logn]^{1/2}\sum_{k=1}^{b}|\Delta_2 w_k|\left(\frac{n-k+1}{n}\right)\to 0.$$

4. $\frac{1}{n}\sum_{k=1}^{b}\sum_{l=0}^{n-k}|\Delta_2 w_k|\cdot|A_k^{(i)}(LE_{n,k})^{(j)}+(LE_{n,k})^{(i)}A_k^{(j)}|$. First consider $LE_{n,k}^{(i)}$, by *proposition 3*,

$$|(LE_{n,k})^{(i)}|\leq\frac{k}{n}(1+\varepsilon)(2n\Sigma_{ii}loglogn)^{1/2},$$

therefore by *condition 3* and (2.A.6), as $n\to\infty$,

$$\frac{1}{n}\sum_{k=1}^{b}\sum_{l=0}^{n-k}|\Delta_2 w_k|\cdot|A_k^{(i)}(LE_{n,k})^{(j)}+(LE_{n,k})^{(i)}A_k^{(j)}|$$

$$\leq 2\left[\sum_{k=1}^{b}\left[|\Delta_2 w_k|\left(\frac{n-k+1}{n}\right)\right]\cdot[2D\psi(n)]\cdot\left[\frac{b}{n}(1+\varepsilon)(2n\Sigma_{ii}loglogn)^{1/2}\right]\right]$$

$$=4\sqrt{2}\Sigma_{ii}^{1/2}\sum_{k=1}^{b}\left[|\Delta_2 w_k|\left(\frac{n-k+1}{n}\right)\right]\cdot[D\psi(n)](1+\varepsilon)\sqrt{\frac{b}{n}}(bloglogn)^{1/2}$$

$$\leq 4\sqrt{2}D\Sigma_{ii}^{1/2}(1+\varepsilon)\cdot\sqrt{\frac{b}{n}}\cdot(blogn)^{1/2}\psi(n)\sum_{k=1}^{b}|\Delta_2 w_k|\left(\frac{n-k+1}{n}\right)\to 0.$$

5. $\frac{1}{n}\sum_{k=1}^{b}\sum_{l=0}^{n-k}|\Delta_2 w_k|\cdot|A_k^{(i)}(LF_{n,k})^{(j)}+(LF_{n,k})^{(i)}A_k^{(j)}|$. By *condition 3* and (2.A.7),

$$\frac{1}{n}\sum_{k=1}^{b}\sum_{l=0}^{n-k}|\Delta_2 w_k|\cdot|A_k^{(i)}(LF_{n,k})^{(j)}+(LF_{n,k})^{(i)}A_k^{(j)}|$$

$$\leq 2\left[\sum_{k=1}^{b}\left[|\Delta_2 w_k|\left(\frac{n-k+1}{n}\right)\right]\cdot\left(\frac{b}{n}D\psi(n)\right)\cdot[2D\psi(n)]\right]$$

$$=4D^2\frac{b}{n}\cdot\psi^2(n)\sum_{k=1}^{b}|\Delta_2 w_k|\left(\frac{n-k+1}{n}\right)\to 0.$$

6. $\frac{1}{n}\sum_{k=1}^{b}\sum_{l=0}^{n-k}|\Delta_2 w_k|\cdot|(LD_k)^{(i)}(F_{n,k})^{(j)}+(F_{n,k})^{(i)}(LD_k)^{(j)}|$. By *condition 3* and (2.A.6),

$$\begin{aligned}
&\frac{1}{n}\sum_{k=1}^{b}\sum_{l=0}^{n-k}|\Delta_2 w_k|\cdot|(LD_k)^{(i)}(F_{n,k})^{(j)}+(F_{n,k})^{(i)}(LD_k)^{(j)}| \\
&\leq 2\left[\sum_{k=1}^{b}\left[|\Delta_2 w_k|\left(\frac{n-k+1}{n}\right)\right]\cdot[2(1+\varepsilon)(b\Sigma_{ii}logn)^{1/2}]\cdot\left(\frac{b}{n}D\psi(n)\right)\right] \\
&= 4D(1+\varepsilon)\Sigma_{ii}^{1/2}\cdot\frac{b}{n}\cdot(blogn)^{1/2}\psi(n)\sum_{k=1}^{b}|\Delta_2 w_k|\left(\frac{n-k+1}{n}\right)\to 0.
\end{aligned}$$

7. $\frac{1}{n}\sum_{k=1}^{b}\sum_{l=0}^{n-k}|\Delta_2 w_k|\cdot|(LE_{n,k})^{(i)}(F_{n,k})^{(j)}+(F_{n,k})^{(i)}(LE_{n,k})^{(j)}|$. By *condition 3* and (2.A.6),

$$\begin{aligned}
&\frac{1}{n}\sum_{k=1}^{b}\sum_{l=0}^{n-k}|\Delta_2 w_k|\cdot|(LE_{n,k})^{(i)}(F_{n,k})^{(j)}+(F_{n,k})^{(i)}(LE_{n,k})^{(j)}| \\
&\leq 2\left[\sum_{k=1}^{b}\left[|\Delta_2 w_k|\left(\frac{n-k+1}{n}\right)\right]\cdot\left(\frac{b}{n}(1+\varepsilon)(2n\Sigma_{ii}loglogn)^{1/2}\right)\cdot\left(\frac{b}{n}D\psi(n)\right)\right] \\
&\leq 2\sqrt{2}D(1+\varepsilon)\Sigma_{ii}^{1/2}\cdot\frac{b}{n}\cdot\sqrt{\frac{b}{n}}\cdot(blogn)^{1/2}\psi(n)\sum_{k=1}^{b}|\Delta_2 w_k|\left(\frac{n-k+1}{n}\right)\to 0.
\end{aligned}$$

Combine 1 to 7, we have $|\hat{\Sigma}_{w,ij}-\tilde{\Sigma}_{w,L,ij}|\to 0$.

Lemma 8.

$$E[\hat{\Sigma}_{w,ij}-\tilde{\Sigma}_{w,L,ij}]\to 0 \quad \text{and} \quad E[\hat{\Sigma}_{w,ij}-\tilde{\Sigma}_{w,L,ij}]^2\to 0.$$

Proof of Theorem 1. Recall *lemma 6*,

$$\begin{aligned}
Var[\tilde{\Sigma}_{w,L,ij}] &= [\Sigma_{ii}\Sigma_{jj}+\Sigma_{ij}^2]\cdot\left[\frac{2}{3}\sum_{k=1}^{b}(\Delta_2 w_k)^2k^3\cdot\frac{1}{n}+2\sum_{t=1}^{b-1}\sum_{u=1}^{b-t}\Delta_2 w_u\cdot\Delta_2 w_{t+u}\cdot\left(\frac{2}{3}u^3+u^2t\right)\cdot\frac{1}{n}\right]+o\left(\frac{b}{n}\right) \\
&:= (S+o(1))\cdot\frac{b}{n},
\end{aligned}$$

where S is a constant. Define

$$\eta = Var[\hat{\Sigma}_{w,ij}-\tilde{\Sigma}_{w,L,ij}]+2E[(\hat{\Sigma}_{w,ij}-\tilde{\Sigma}_{w,L,ij})(\tilde{\Sigma}_{w,L,ij}-E\tilde{\Sigma}_{w,L,ij})],$$

we first show that $\eta\to 0$ as $n\to\infty$. Under *condition 3*, use Cauchy-Schwarz inequality and $Var[X]\leq EX^2$.

$$
\begin{aligned}
|\eta| &= |Var[\hat{\Sigma}_{w,ij} - \tilde{\Sigma}_{w,L,ij}] + 2E[(\hat{\Sigma}_{w,ij} - \tilde{\Sigma}_{w,L,ij})(\tilde{\Sigma}_{w,L,ij} - E\tilde{\Sigma}_{w,L,ij})]| \\
&\leq E[\hat{\Sigma}_{w,ij} - \tilde{\Sigma}_{w,L,ij}]^2 + 2\sqrt{E[\hat{\Sigma}_{w,ij} - \tilde{\Sigma}_{w,L,ij}]^2 \cdot E[\tilde{\Sigma}_{w,L,ij} - E\tilde{\Sigma}_{w,L,ij}]^2} \\
&= o(1) + 2\sqrt{o(1) \cdot [(S + o(1)) \cdot \frac{b}{n}]} \\
&= o(1) + 2\left(\frac{b}{n}\right)^{1/2} [o(1) \cdot (S + o(1))]^{1/2} \\
&= o(1).
\end{aligned}
$$

Now consider $Var[\hat{\Sigma}_{w,ij}]$, by lemma 8

$$
\begin{aligned}
Var[\hat{\Sigma}_{w,ij}] &= E[\hat{\Sigma}_{w,ij} - E\hat{\Sigma}_{w,ij}]^2 \\
&= E[\hat{\Sigma}_{w,ij} - \tilde{\Sigma}_{w,L,ij} + \tilde{\Sigma}_{w,L,ij} - E\tilde{\Sigma}_{w,L,ij} + E\tilde{\Sigma}_{w,L,ij} - E\hat{\Sigma}_{w,ij}]^2 \\
&= E[(\hat{\Sigma}_{w,ij} - \tilde{\Sigma}_{w,L,ij}) + (\tilde{\Sigma}_{w,L,ij} - E\tilde{\Sigma}_{w,L,ij}) - (E\hat{\Sigma}_{w,ij} - E\tilde{\Sigma}_{w,L,ij})]^2 \\
&= E[(\hat{\Sigma}_{w,ij} - \tilde{\Sigma}_{w,L,ij}) - E(\hat{\Sigma}_{w,ij} - \tilde{\Sigma}_{w,L,ij})]^2 + E[\tilde{\Sigma}_{w,L,ij} - E\tilde{\Sigma}_{w,L,ij}]^2 \\
&\quad + 2E[[(\hat{\Sigma}_{w,ij} - \tilde{\Sigma}_{w,L,ij}) - E(\hat{\Sigma}_{w,ij} - \tilde{\Sigma}_{w,L,ij})] \cdot [\tilde{\Sigma}_{w,ij} - E\tilde{\Sigma}_{w,ij}]] \\
&= E[(\hat{\Sigma}_{w,ij} - \tilde{\Sigma}_{w,L,ij}) - E(\hat{\Sigma}_{w,ij} - \tilde{\Sigma}_{w,L,ij})]^2 + E[\tilde{\Sigma}_{w,L,ij} - E\tilde{\Sigma}_{w,L,ij}]^2 \\
&\quad + 2E[(\hat{\Sigma}_{w,ij} - \tilde{\Sigma}_{w,L,ij}) \cdot (\tilde{\Sigma}_{w,ij} - E\tilde{\Sigma}_{w,ij})] \\
&= E[\tilde{\Sigma}_{w,L,ij} - E\tilde{\Sigma}_{w,L,ij}]^2 + \eta \\
&= Var[\tilde{\Sigma}_{w,L,ij}] + \eta \\
&= S \cdot \frac{b}{n} + o\left(\frac{b}{n}\right) + o(1).
\end{aligned}
$$

Similar as the proof of theorem 4 in Flegal and Jones (2010), *lemma 8* results in

$$
\frac{n}{b}\mathrm{Var}[\hat{\Sigma}_{w,ij}] = S + o(1),
$$

that is

$$Var[\tilde{\Sigma}_{w,ij}] = [\Sigma_{ii}\Sigma_{jj} + \Sigma_{ij}^2] \cdot \left[\frac{2}{3}\sum_{k=1}^{b}(\Delta_2 w_k)^2 k^3 \cdot \frac{1}{n} + 2\sum_{t=1}^{b-1}\sum_{u=1}^{b-t}\Delta_2 w_u \cdot \Delta_2 w_{t+u} \cdot \left(\frac{2}{3}u^3 + u^2 t\right) \cdot \frac{1}{n}\right] + o\left(\frac{b}{n}\right).$$

D Proof of theorem 2

When $i = j$, for a stationary and uniformly ergodic, if $E_F g^{12} < \infty$ and *Condition 3* holds, then

$$\mathrm{Var}[\bar{Y}_l^{(i)}(k)] - \mathrm{Var}[\bar{Y}^{(i)}(k)] = \frac{n-k}{kn}\left(\Sigma_{ii} + \frac{n+k}{kn}\Gamma_{ii} + o\left(\frac{1}{k^2}\right)\right),$$

followed by univariate results discussed by (Chien et al. (1997), Song and Schmeiser (1995) proposition 2, Goldsman and Meketon (1986)). Similarly,

$$Cov[\bar{Y}_l^{(i)}(k), \bar{Y}_l^{(j)}(k)] - Cov[\bar{Y}_n^{(i)}, \bar{Y}_n^{(j)}] = \frac{n-k}{kn}\left(\Sigma_{ij} + \frac{n+k}{kn}\Gamma_{ij} + o\left(\frac{1}{k^2}\right)\right).$$

To obtain the expression of $Bias(\hat{\Sigma}_{w,ij})$, we consider $E[\hat{\Sigma}_{w,ij}]$.

$$\begin{aligned}
E[\hat{\Sigma}_{w,ij}] &= \frac{1}{n}\sum_{k=1}^{b} k^2 \Delta_2 w_k \cdot E\left(\sum_{l=0}^{n-k}(\bar{Y}_l^{(i)}(k) - \bar{Y}^{(i)})(\bar{Y}_l^{(j)}(k) - \bar{Y}^{(j)})\right) \\
&= \frac{1}{n}\sum_{k=1}^{b} k^2 \Delta_2 w_k \cdot \left(\sum_{l=0}^{n-k} E[\bar{Y}_l^{(i)}(k) \cdot \bar{Y}_l^{(i)}(k)] - (n-k+1)E[\bar{Y}^{(i)} \cdot \bar{Y}^{(j)}]\right) + o\left(\frac{b}{n}\right) + o\left(\frac{1}{b}\right) \\
&= \frac{1}{n}\sum_{k=1}^{b} k^2 \Delta_2 w_k (n-k+1) \cdot \frac{n-k}{kn}\left(\Sigma_{ij} + \frac{n+k}{kn}\Gamma_{ij} + o\left(\frac{1}{k^2}\right)\right) + o\left(\frac{b}{n}\right) + o\left(\frac{1}{b}\right) \\
&= \sum_{k=1}^{b}\frac{(n-k+1)(n-k)k\Delta_2 w_n(k)}{n^2} \cdot \Sigma_{ij} + \sum_{k=1}^{b}\frac{(n-k+1)(n^2-k^2)\Delta_2 w_n(k)}{n^3} \cdot \Gamma_{ij} \\
&\quad + o\left(\sum_{k=1}^{b}\frac{(n-k+1)(n-k)\Delta_2 w_n(k)}{n^2 k}\right) + o\left(\frac{b}{n}\right) + o\left(\frac{1}{b}\right) \\
&= \sum_{k=1}^{b}\frac{(n-k+1)(n-k)k\Delta_2 w_n(k)}{n^2} \cdot \Sigma_{ij} + \sum_{k=1}^{b}\frac{(n-k+1)(n^2-k^2)\Delta_2 w_n(k)}{n^3} \cdot \Gamma_{ij} + o\left(\frac{b}{n}\right) + o\left(\frac{1}{b}\right).
\end{aligned}$$

If

$$\sum_{k=1}^{b} k\Delta_2 w_n(k) = 1,$$

then

$$E[\hat{\Sigma}_{w,ij}] = \Sigma_{ij} + \sum_{k=1}^{b} \frac{(n-k+1)(n^2-k^2)\Delta_2 w_n(k)}{n^3} \cdot \Gamma_{ij} + o\left(\frac{b}{n}\right) + o\left(\frac{1}{b}\right)$$
$$= \Sigma_{ij} + \sum_{k=1}^{b} \Delta_2 w_n(k) \cdot \Gamma_{ij} + o\left(\frac{b}{n}\right) + o\left(\frac{1}{b}\right),$$

and

$$Bias[\hat{\Sigma}_{w,ij}] = \sum_{k=1}^{b} \Delta_2 w_n(k) \cdot \Gamma_{ij} + o\left(\frac{b}{n}\right) + o\left(\frac{1}{b}\right).$$

E Equivalence of $\hat{\sigma}_s^2$ and $\hat{\sigma}_w^2$

Next consider the variance of $\hat{\Sigma}_w$ and $\hat{\Sigma}_s$. Let's consider univariate case for illustration purpose. Multivariate case can be shown in a similar manner. Recall $V_t = Y_t - \bar{Y}$. Then

$$(\bar{Y}_l(k) - \bar{Y})^2 = \frac{1}{k^2}[(Y_{l+1} + Y_{l+2} + \cdots + Y_{l+k}) - k\bar{Y}]^2$$
$$= \frac{1}{k^2}[\sum_{l=1}^{k} V_{l+k}]^2 = \frac{1}{k^2}[\sum_{h=1}^{k} V_{l+h}^2 + 2\sum_{s=1}^{k-1}\sum_{h=1}^{k-s} V_{l+h}V_{l+h+s}].$$

Plugging in the above expression,

$$\hat{\sigma}_w = \frac{1}{n}\sum_{k=1}^{b_n}\sum_{l=0}^{n-k} k^2 \Delta_2 w_k (\bar{Y}_l(k) - \bar{Y})^2$$
$$= \frac{1}{n}\sum_{k=1}^{b_n}\sum_{l=0}^{n-k} \Delta_2 w_k \cdot \sum_{h=1}^{k} V_{l+h}^2 + \frac{2}{n}\sum_{k=1}^{b_n}\sum_{l=0}^{n-k} \Delta_2 w_k \cdot \sum_{s=1}^{k-1}\sum_{h=1}^{k-s} V_{l+h}V_{l+h+s}$$
$$:= \mathrm{I} + \mathrm{II}$$

Changing the order of sums in I and apply result 1

$$\mathrm{I} = \frac{1}{n}\sum_{k=1}^{b}\sum_{l=0}^{n-k}\sum_{h=1}^{k} \Delta_2 w_k V_{l+h}^2 = \frac{1}{n}\sum_{k=1}^{b}\sum_{h=1}^{k}\sum_{l=0}^{n-k} \Delta_2 w_k V_{l+h}^2 = \frac{1}{n}\sum_{h=1}^{b}\sum_{k=h}^{b_n}\sum_{l=0}^{n-k} \Delta_2 w_k V_{l+h}^2$$
$$= \frac{1}{n}\sum_{h=1}^{b}\sum_{l=0}^{n-h}\sum_{k=h}^{b} \Delta_2 w_k V_{l+h}^2 - \frac{1}{n}\sum_{h=1}^{b}\sum_{l=n-b+1}^{n-h}\sum_{k=n-l+1}^{b} \Delta_2 w_k V_{l+h}^2$$

$$
\begin{aligned}
&= \frac{1}{n}\sum_{h=1}^{b}\sum_{l=0}^{n-h} V_{l+h}^2 \sum_{k=h}^{b}\Delta_2 w_k - \frac{1}{n}\sum_{h=1}^{b}\sum_{l=n-b+1}^{n-h} V_{l+h}^2 \sum_{k=n-l+1}^{b}\Delta_2 w_k \\
&= \frac{1}{n}\sum_{h=1}^{b}\sum_{l=0}^{n-h} V_{l+h}^2 \Delta_1 w_h - \frac{1}{n}\sum_{h=1}^{b}\sum_{l=n-b+1}^{n-h} V_{l+h}^2 \Delta_1 w_{n-l+1} \\
&= \frac{1}{n}\sum_{h=1}^{b}\Delta_1 w_h \sum_{l=0}^{n-h} V_{l+h}^2 - \frac{1}{n}\sum_{h=1}^{b}\sum_{l=n-b+1}^{n-h} V_{l+h}^2 \Delta_1 w_{n-l+1} \\
&= \frac{1}{n}\sum_{h=1}^{b}\Delta_1 w_h \sum_{t=h}^{n} V_t^2 - \frac{1}{n}\sum_{h=1}^{b}\sum_{t=n-b+h+1}^{n} V_t^2 \Delta_1 w_{n-t+h+1}
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
\mathrm{II} &= \frac{2}{n}\sum_{s=1}^{b-1}\sum_{h=1}^{b-s}\Delta_1 w_{h+s}\sum_{l=0}^{n-h-s} V_{l+h}V_{l+h+s} - \frac{2}{n}\sum_{s=1}^{b-1}\sum_{h=1}^{b-s}\sum_{l=n-b+1}^{n-h-s}\Delta_1 w_{n-l+1}V_{l+h}V_{l+h+s} \\
&= \frac{2}{n}\sum_{s=1}^{b-1}\sum_{h=1}^{b-s}\Delta_1 w_{h+s}\sum_{t=h}^{n-s} V_t V_{t+s} - \frac{2}{n}\sum_{s=1}^{b-1}\sum_{h=1}^{b-s}\sum_{t=n-b+h+1}^{n-s}\Delta_1 w_{n-t+h+1}V_t V_{t+s}
\end{aligned}
$$

therefore

$$
\begin{aligned}
\hat{\sigma}_w &= \left(\frac{1}{n}\sum_{h=1}^{b}\Delta_1 w_h \sum_{t=h}^{n} V_t^2 + \frac{2}{n}\sum_{s=1}^{b-1}\sum_{h=1}^{b-s}\Delta_1 w_{h+s}\sum_{t=h}^{n-s} V_t V_{t+s}\right) \\
&\quad - \left(\frac{1}{n}\sum_{h=1}^{b}\sum_{t=n-b+h+1}^{n} V_t^2 \Delta_1 w_{n-t+h+1} + \frac{2}{n}\sum_{s=1}^{b-1}\sum_{h=1}^{b-s}\sum_{t=n-b+h+1}^{n-s}\Delta_1 w_{n-t+h+1}V_t V_{t+s}\right) \\
&= \left(\frac{1}{n}\sum_{t=1}^{n}\sum_{h=1}^{t}\Delta_1 w_h V_t^2 - \frac{1}{n}\sum_{t=n-b+2}^{n}\sum_{h=1}^{t-(n-b+1)}\Delta_1 w_{n-t+h+1}V_t^2\right) \\
&\quad + \left(\frac{2}{n}\sum_{s=1}^{b-1}\sum_{t=1}^{n-s}\sum_{h=1}^{t}\Delta_1 w_{h+s}V_t V_{t+s} - \frac{2}{n}\sum_{s=1}^{b-1}\sum_{t=n-b+2}^{n-s}\sum_{h=1}^{t-(n-b+1)}\Delta_1 w_{n-t+h+1}V_t V_{t+s}\right).
\end{aligned}
$$

Recall that

$$
\begin{aligned}
\hat{\sigma}_s &= \hat{\gamma}(0) + 2\sum_{s=1}^{b_n-1} w_n(s)\hat{\gamma}(s) \\
&= \frac{1}{n}\sum_{t=1}^{n}(Y_t - \bar{Y}_n)^2 + 2\sum_{s=1}^{b-1} w_n(s)\frac{1}{n}\sum_{t=1}^{n-s}(Y_t - \bar{Y}_n)(Y_{t+s} - \bar{Y}_n) \\
&= \frac{1}{n}\sum_{t=1}^{n} V_t^2 + \frac{2}{n}\sum_{s=1}^{b-1} w_n(s)\sum_{t=1}^{n-s} V_t V_{t+s}.
\end{aligned}
$$

Notice that

$$
\sum_{h=1}^{b}\Delta_1 w_h = 1 \quad \text{and} \quad \sum_{h=s+1}^{b}\Delta_1 w_n(h) = w_n(s),
$$

in the expression of $\hat{\sigma}_w$, t goes form 1 to n, when $t \geq b, \sum_{h=1}^{t} w_n(s) = 1$

we can rewrite $\hat{\sigma}_s$ and $\hat{\sigma}_w$ as

$$\hat{\sigma}_s = \left(\frac{1}{n} \sum_{t=1}^{b} a_0 V_t^2 + \frac{1}{n} \sum_{t=b+1}^{n-b} a_0 V_t^2 + \frac{1}{n} \sum_{t=n-b+1}^{n} a_0 V_t^2 \right) + \left(\frac{2}{n} \sum_{s=1}^{b-1} \sum_{t=1}^{b-s} a_s V_t V_{t+s} + \frac{2}{n} \sum_{s=1}^{b-1} \sum_{t=b-s+1}^{n-b} a_s V_t V_{t+s} + \frac{2}{n} \sum_{s=1}^{b-1} \sum_{t=n-b+1}^{n-s} a_s V_t V_{t+s} \right),$$

and

$$\hat{\sigma}_w = \left(\frac{1}{n} \sum_{t=1}^{b} c_{0,t} V_t^2 + \frac{1}{n} \sum_{t=b+1}^{n-b} a_0 V_t^2 + \frac{1}{n} \sum_{t=n-b+1}^{n} d_{0,t} V_t^2 \right) + \left(\frac{2}{n} \sum_{s=1}^{b-1} \sum_{t=1}^{b-s} c_{s,t} V_t V_{t+s} + \frac{2}{n} \sum_{s=1}^{b-1} \sum_{t=b-s+1}^{n-b} a_s V_t V_{t+s} + \frac{2}{n} \sum_{s=1}^{b-1} \sum_{t=n-b+1}^{n-s} d_{s,t} V_t V_{t+s} \right),$$

where a_0, a_s, $c_{0,t}$, $c_{s,t}$, $d_{0,t}$ and $d_{s,t}$ are constants. Denote

$$C_0 = \begin{bmatrix} c_{0,1} \\ c_{0,2} \\ \vdots \\ c_{0,b} \end{bmatrix} \qquad A_0 = \begin{bmatrix} a_0 \\ a_0 \\ \vdots \\ a_0 \end{bmatrix} \qquad D_0 = \begin{bmatrix} d_{0,n-b+1} \\ d_{0,n-b+2} \\ \vdots \\ d_{0,n} \end{bmatrix}$$

$$C_s = \begin{bmatrix} c_{s,1} \\ c_{s,2} \\ \vdots \\ c_{s,b} \end{bmatrix} \qquad A_s = \begin{bmatrix} a_s \\ a_s \\ \vdots \\ a_s \end{bmatrix} \qquad D_s = \begin{bmatrix} d_{s,n-b+1} \\ d_{s,n-b+2} \\ \vdots \\ d_{s,n} \end{bmatrix},$$

and for $k = 1, 2, ...(n-b)/b$,

$$V_0^{(0)} = \begin{bmatrix} V_1^2 \\ V_2^2 \\ \vdots \\ V_b^2 \end{bmatrix} \qquad V_0^{(k)} = \begin{bmatrix} V_{kb+1} \\ V_{kb+2} \\ \vdots \\ V_{kb+b} \end{bmatrix} \qquad V_s^{(0)} = \begin{bmatrix} V_1 V_{1+s} \\ V_2 V_{2+s} \\ \vdots \\ V_{b-s} V_b \end{bmatrix} \qquad V_s^{(k)} = \begin{bmatrix} V_{k(b-s)+1} V_{k(b-s)+1+s} \\ V_{k(b-s)+2} V_{k(b-s)+2+s} \\ \vdots \\ V_{(k+1)(b-s)} V_{(k+1)(b-s)+s} \end{bmatrix},$$

then we have

$$\hat{\sigma}_s = \frac{1}{n}\sum_{k=0}^{\frac{n}{b}-1} A_0^T V_0^{(k)} + \frac{2}{n}\sum_{s=1}^{b-1}\sum_{k=0}^{\frac{n-s}{b-s}-1} A_s^T V_s^{(k)},$$

and

$$\hat{\sigma}_w^2 = \frac{1}{n}\left(C_0^T V_0^{(0)} + \sum_{k=1}^{\frac{n}{b}-2} A_0^T V_0^{(k)} + D_0^T V_0^{(\frac{n}{b}-1)}\right) + \frac{2}{n}\sum_{s=1}^{b-1}\left(C_s^T V_s^{(0)} + \sum_{k=1}^{\frac{n-b}{b-s}-2} A_s^T V_s^{(k)} + D_s^T V_s^{(\frac{n-b}{b-s}-1)}\right).$$

For a big enough b, if

$$Var(V_{b+1}+V_{b+2}+...,V_{2b}) \approx Var(V_{2b+1}+V_{2b+2}+...V_{3b}) \approx Var(V_{3b+1}+V_{3b+2}+...,+V_{4b}) \approx ...,$$

and $Var(g(X_t)) < \infty$, then

$$\begin{aligned}
Var(\hat{\sigma}_w) &\approx \frac{1}{n^2}\Bigg[\left(\frac{n}{b}-2\right)^2 A_0^T Var(V_0^{(1)})A_0 + 4\sum_{s=1}^{b-1}\left(\frac{n-b}{b-s}-2\right)^2 A_s^T Var(V_s^{(1)})A_s \\
&\quad + 2\sum_{s=1}^{b-1} Cov(\left(\frac{n}{b}-2\right)A_0^T V_0^{(1)}, 2\left(\frac{n-b}{b-s}-2\right)A_s^T V_s^{(1)})\Bigg] + o\left(\frac{1}{b^2}\right) \\
&= \frac{1}{n^2}\Bigg[\left(\frac{n}{b}\right)^2 A_0^T Var(V_0^{(1)})A_0 + 4\sum_{s=1}^{b-1}\left(\frac{n-b}{b-s}\right)^2 A_s^T Var(V_s^{(1)})A_s \\
&\quad + 4\sum_{s=1}^{b-1}\left(\frac{n}{b}\right)\left(\frac{n-b}{b-s}\right) A_0^T Cov(V_0^{(1)}, V_s^{(1)})A_s\Bigg] + o\left(\frac{1}{b^2}\right),
\end{aligned}$$

similarly,

$$\begin{aligned}
Var(\hat{\sigma}_S) &\approx \frac{1}{n^2}\Bigg[\left(\frac{n}{b}\right)^2 A_0^T Var(V_0^{(1)})A_0 + 4\sum_{s=1}^{b-1}\left(\frac{n-b}{b-s}\right)^2 A_s^T Var(V_s^{(1)})A_s \\
&\quad + 4\sum_{s=1}^{b-1}\left(\frac{n}{b}\right)\left(\frac{n-b}{b-s}\right) A_0^T Cov(V_0^{(1)}, V_s^{(1)})A_s\Bigg] + o\left(\frac{1}{b^2}\right),
\end{aligned}$$

then as $n \to \infty$, $Var(\hat{\sigma}_s^2) = Var(\hat{\sigma}_w^2)$. A similar argument applies for biases of the two estimators, which shows that the two estimators are asymptotically equivalent in the sense of mean squared error. A similar argument applies for the multivariate case.

Chapter 3

Efficient estimator

As mentioned in chapter one, Bayesian analysis tends to involves multiple parameters hence we usually face high-dimensional problems. Although mSV estimators can be used to obtain reliable estimates of Σ especially when using the optimal bandwidth suggested in chapter 2, they are computational expensive. The computational burden may discourage practitioners to keep track of $\hat{\Sigma}$ and further terminate the chain in a sensible way. Estimators that are accurate and fast to compute are highly desired from a practical point of view.

In this chapter, we propose a family of variation estimator of multivariate sample mean. These estimators are especially convenient under MCMC context but they can also be applied in other fields such as time series and nonparametric analysis. The new estimators enjoy the same flexibility of choosing various window functions as mSV. Simulation shows that by choosing flat top window function, the resulting estimator are superior to mBM, which is related to *Bartlett* window. In the meantime, the new estimators have significant reduction of computation time compared with mSV, hence provides an applicable solution to the problem faced by multivariate MCMC methods. We prove the strong consistency of the new estimators, followed with the discussion of their minor sacrifice on the convergence rate compared with mSV. The performance of the new estimators are illustrated by univariate and multivariate auto-regressive models. These simulations coincide with the theoretical results, showing that the new estimators converge to the correct value, and as dimension or chain length increases, the new estimators save significant amount of time compared with mSV. The variance of the new estimators are slightly larger than mSV, but the ratio between the variance of a new estimator and the corresponding mSV estimator with the same window function are usually less than two, which seems to be negligible since the actual variance of these estimators are already small given a chain with a reasonable length. We also consider a Bayesian spatial-temporal model applied to temperature data collected from ten nearby weather station in the year 2010. It takes a reasonable amount of time to compute the proposed estimator for the 185 parameters in the example while mSV requires much longer time for such high dimensional problem .

The rest of the chapter is organized as follows. The new estimator is defined in Section 3.1, together with its strong consistency and some discussion regarding variance. Section 3.2 contains three examples, including a Bayesian model applied to real data.

3.1 Efficient Spectral Variance Estimator (EFSV)

We first establish notations needed for the main results. Let F be a probability distribution with support $\mathsf{X} \in \mathbb{R}^d$ and $g : \mathsf{X} \to \mathbb{R}^p$ be a F-integrable function. We are interested in estimating the p-dimensional vector

$$\theta := E_F g = \int_{\mathsf{X}} g(x) dF.$$

by MCMC methods. Let $X = \{X_t, t \geq 1\}$ be a Harris ergodic Markov chain with invariant distribution F, then with probability 1,

$$\theta_n := \frac{1}{n}\sum_{t=1}^{n} g(X_t) \to \theta \qquad \text{as} \qquad n \to \infty.$$

The sampling distribution for $\theta_n - \theta$ is available via a Markov chain Central Limit Theorem if there exists a positive definite symmetric matrix Σ such that

$$\sqrt{n}(\theta_n - \theta) \xrightarrow{d} N_p(0, \Sigma) \qquad \text{as} \qquad n \to \infty,$$

where

$$\Sigma = \text{Var}_F(g(X_1)) + \sum_{s=1}^{\infty}[\text{Cov}_F(g(X_1), g(X_{1+s})) + \text{Cov}_F(g(X_1), g(X_{1+s}))^T],$$

and it is usually unknown. If an estimation $\hat{\Sigma}$ is available, it can be used to access the variation of the estimator θ_n. Denote $Y_t = g(X_t)$ and $\bar{Y} = n^{-1}\sum_{t=1}^{n}$, Vats et al. (2015b) introduced the mSV estimator of Σ with the expression

$$\hat{\Sigma}_{sv} = \hat{\Gamma}(0) + \sum_{s=1}^{b} w_n(s)[\hat{\Gamma}(s) + \hat{\Gamma}(s)^T],$$

where

$$\hat{\Gamma}(s) = \frac{1}{n}\sum_{t=1}^{n-s}(Y_t - \bar{Y})(Y_{t+s} - \bar{Y})^T,$$

and *truncation point b* is a finite number that increases as n and $w_n(\cdot)$ is the *lag window*. The strong consistency of the estimator was also provided by Vats et al. (2015b). $\hat{\Sigma}_{sv}$ with a flat top window $w_n(\cdot)$ was shown to have superior performance in nonparametric estimation of multivariate density function. Politis and Romano (1999). However, computational of $\hat{\Sigma}_{sv}$ is too expensive.

The mBM estimator introduced by Vats et al. (2015a)) is cheaper to calculate compared with mSV. Recall that for $n = ab$ and $l = 0, 1, ..., a-1$, the mean vector for batch l is denoted by $\bar{Y}_l(b) = b^{-1}\sum_{t=1}^{b} Y_{lb+t}$, where b is the batch size. mBM is based on totally a non-overlapping batches, then sample variance of batch means is used to estimate Σ

$$\hat{\Sigma}_{bm} = \frac{b}{a-1}\sum_{l=0}^{a-1}(\bar{Y}_l(b) - \bar{Y})(\bar{Y}_l(b) - \bar{Y})^T.$$

We consider OBM in multivariate setting. Recall $\bar{Y}_l(b) = b^{-1}\sum_{t=1}^{b} Y_{l+t}$, there are $n-b$ batches of length b in multivariate OBM estimator with the following expression

$$\hat{\Sigma}_{obm} = \frac{nb}{(n-b)(n-b+1)}\sum_{l=0}^{n-b+1}(\bar{Y}_l(b) - \bar{Y})(\bar{Y}_l(b) - \bar{Y})^T.$$

$\hat{\Sigma}_{obm}$ has a similar structure as $\hat{\Sigma}_{bm}$ by considering the sample variance of multiple batch means. Since the bathes in $\hat{\Sigma}_{bm}$ do not overlap, its sparseness in bathes yields a faster computation while remaining a consistent estimator of Σ. $\hat{\Sigma}_{bm}$ and $\hat{\Sigma}_{obm}$ are related to mSV with *Bartlett* window $\hat{\Sigma}_{bt}$. $\hat{\Sigma}_{obm}$ is equivalent to mSV estimator with *Bartlett* window apart from some end effects. Interested readers are directed to (Welch (1987), **?**, Meketon and Schmeiser (1984), Song and Schmeiser (1993)).

Our goal is to construct a mBM version of mSV for a given window function, so that the nonoverlapping structure of the new estimator allows for cheaper calculation while inheriting the desired property from a certain window function. If we are able to rewrite any mSV in terms of overlapping batches as Σ_{obm}, then it is possible to reduce number of batches by keeping those non-overlapping batches, thus reducing computing time. $\hat{\Sigma}_{OBM}$ version of mSV has previously been addressed by various authors including (**?**, Flegal and Jones (2010), Damerdji (1991)). Chapter 2 suggested an OBM version of mSV that is closer to mSV compared with the expression used in previous literature. Let $\bar{Y}_l(k) = k^{-1}\sum_{t=1}^{k} Y_{l+t}$ for $l = 0, ..., (n-k)$, denote $\Delta_1 w_n(k) = w_n(k-1) -$

$w_n(k)$ and $\Delta_2 w_n(k) = w_n(k-1) - 2w_n(k) + w_n(k+1)$. The expression is as follows

$$\hat{\Sigma}_w = \frac{1}{n}\sum_{k=1}^{b}\sum_{l=0}^{n-k} k^2 \Delta_2 w_n(k)[\bar{Y}_l(k) - \bar{Y}][\bar{Y}_l(k) - \bar{Y}]^T.$$

Suppose $d = \hat{\Sigma}_{sv}$-$\hat{\Sigma}_w$, it was shown that $d \to 0$ as $n \to \infty$ (see chapter 2), hence $\hat{\Sigma}_{sv}$ has an asymptotically equivalent expression in terms of overlapping batches. From $\hat{\Sigma_w}$, we suggest an efficient spectral variance (EFSV) estimator $\hat{\Sigma}_{ef}$. Denote $\bar{Y}_l(k) = k^{-1}\sum_{t=1}^{k} Y_{lk+t}$ for $l = 0, 1, ..., a_k - 1$, and $k = 1, 2, ..., b$ where $a_k = \lfloor (n/k) \rfloor$,

$$\hat{\Sigma}_{ef} = \sum_{k=1}^{b} \frac{1}{a_k - 1} \sum_{l=0}^{a_k - 1} k^2 \Delta_2 w_n(k)(\bar{Y}_l(k) - \bar{Y})(\bar{Y}_l(k) - \bar{Y})^T.$$

mBM is included in the family of $\hat{\Sigma}_{ef}$, with a $w_{bt}(\cdot)$ window function. For *Bartlett* window, $\Delta_2 w_{bt}(s) = 0$ for $s = 1, 2, ..., (b-1)$, $\Delta_2 w_{bt}(b) = 1/b$, hence $\hat{\Sigma}_{bt}$ is equivalent to mBM.

Notice that $\Delta_2 w_n(\cdot)$ has a similar expression as $w_n''(\cdot)$ in a sense. For instance, if a window function is twice continuously differentiable on $(0,\ b)$, it can be seen from the proof of (Flegal and Jones (2010) *lemma* 7) that the two quantities are closely related. If the window function has a simple structure so that $\Delta_2 w_n(s) = 0$ for certain s, then the first summation in the expression of $\hat{\Sigma}_{ef}$ can be further simplified, resulting a shorter computation time. One example is *Bartlett* window where $\Delta_2 w_n(s)$ and $w_n''(s)$ are 0 for $s = 1, 2, ..., (b-1)$. Form this point view, the new estimator tends to be more beneficial for window function with $\Delta_2 w_n(s) = 0$ for certain s.

Politis and Romano (1995) introduced a class of flat top window functions that modify existing window by letting $w_n(s) = 1$ for s near 0. Here we consider the flat top window function constructed from *Bartlett* window $w_{ft}(\cdot)$, which is equivalent to the the difference of two *Bartlett* spectral variance estimators in the following way

$$\hat{\Sigma}_{ft} = \hat{\Gamma}(0) + 2\sum_{s=1}^{b} w_{ft}(s)\hat{\Gamma}(s) = 2\hat{\Sigma}_{bt}^{(1)} - \hat{\Sigma}_{bt}^{(2)},$$

where bandwidth of $\hat{\Sigma}_{ft}$ and $\hat{\Sigma}_{bt}^{(1)}$ equals to b and $\hat{\Sigma}_{bt}^{(1)}$ has a bandwidth of $b/2$. (Politis and Romano (1995)) demonstrated that flat top structure near 0 contributes to bias reduction of the SV estimator. For $\hat{\Sigma}_{ef}$, we suggest using $w_{ft}(\cdot)$ constructed from *Bartlett* window for bias reduction purpose. Further more, $\Delta_2 w_{ft}(b/2) = -2/b$, $\Delta_2 w_{ft}(b) = 2/b$, for other s, $\Delta_2 w_{ft}(s) = 0$. Using this window, the first summation in $\hat{\Sigma}_{ef}$ becomes a summation of two terms, which is ideal from a time saving aspect. We will see the bias and computational advantages of $\hat{\Sigma}_{ef}$ with flat top window.

One disadvantage of the flat top window, as pointed out by **?**, is that $\hat{\Sigma}_{ft}$ is not guaranteed to be positive semi-definite, which is required in order to estimate Σ. Modification based on eigenvalues of $\hat{\Sigma}_{ft}$ is discussed by **?** to correct the estimation when positive semi-definite is an issue.

3.1.1 Theoretical results

In this section, we establish conditions under which strong consistency of $\hat{\Sigma}_{ef}$ holds. Denote Euclidean norm by $\|\cdot\|$, the following conditions are needed.

Condition 1.(*strong invariant principle*) There exists a $p-$dimensional vector θ, a $p \times p$ lower triangular matrix L, an increasing function ψ on integers, a finite random variable D and a sufficiently rich probability space Ω such that for almost all $\omega \in \Omega$ and for all $n > n_0$,

$$\left\| \sum_{t=1}^{n} g(X_t) - n\theta - LB(n) \right\| < D(\omega)\psi(n) \quad \text{w.p. 1.} \tag{3.1.1}$$

Condition 2. The lag window $w_n(\cdot)$ is an even function defined on $\mathbb{Z}$ such that

$$|w_n(s)| \leq 1 \quad \text{for all } n \text{ and } s,$$

$$w_n(0) = 1 \quad \text{for all } n,$$

$$w_n(s) = 0 \quad \text{for all } |s| \geq b.$$

Condition 3. b is an integer sequence such that $b \to \infty$ and $n/b \to \infty$ as $n \to \infty$, where b and n/b are monotonically nondecreasing.

Theorem 1. Let *condition 1* hold for g and *condition 2, 3* hold. Suppose there exists a constant $c \geq 1$ such that $\sum_n (b/n)^c < \infty$. If

$$\sum_{k=1}^{b} k\Delta_2 w_n(k) = 1, \tag{3.1.2}$$

$$b\psi(n)^2 logn \left(\sum_{k=1}^{b} |\Delta_2 w_n(k)| \right)^2 \to 0, \tag{3.1.3}$$

and

$$\psi(n)^2 \sum_{k=1}^{b} |\Delta_2 w_n(k)| \to 0, \tag{3.1.4}$$

then *with probability 1*, $\hat{\Sigma}_{ef} \to \Sigma$ as $n \to \infty$.

Proof. See appendix A.

Remark. Condition (3.1.1) is also necessary to guarantee that the estimator is asymptotically unbiased, see chapter 2. It can be shown that *Bartlett* window satisfies since

$$\sum_{k=1}^{b} \Delta_2 w_{bt}(k) = b \cdot \frac{1}{b} = 1.$$

Condition is also satisfied by flat top window $w_{ft}(\cdot)$ where

$$\sum_{k=1}^{b} \Delta_2 w_{ft}(k) = -\frac{2}{b} \cdot \frac{b}{2} + \frac{2}{b} \cdot b = 1.$$

Remark. The following result provides conditions under which *condition 1* holds.

(Vats et al. (2015b) Corollary 1) Let X be a polynomial ergodic Markov chain of order k. In addition, let $\|E_F g\|^{2+\delta} < \infty$ for some $0 < \delta \leq 1$, then for $k > (1+\varepsilon)(1+2/\delta),\ \varepsilon > 0$, (3.1.1) holds with $\psi(n) = n^{1/2-\lambda}$ for some $\lambda > 0$ that depends on $p,\ \varepsilon$ and δ.

Since $\hat{\Sigma}_{ef}$ are based on less batches compared with $\hat{\Sigma}_{sv}$, a variance inflation is expected. We are interested in how the sparseness in batches may influence variance of $\hat{\Sigma}^2_{ef}$. Univariate estimators are equivalent to diagonal entries of multivariate estimators. The off-diagonal entries are expected to behave in the same manner by observing the proof in chapter 2. Therefore we consider variance of univariate estimators for illustration purpose. This should also shed lights on multivariate estimators due to their intrinsic similarities.

Let $\hat{\sigma}^2_{bm}$ and $\hat{\sigma}^2_{bt}$ be BM and SV estimator with *Bartlett* window. Flegal and Jones (2010) showed that the ratio of asymptotic variances between OBM and BM is 1.5. Variance of $\hat{\sigma}^2_{bt}$ were disccused by Politis and White (2004), Lahiri (1999). Under certain moment and mixing conditions, it is equivalent to the variance of $\hat{\sigma}^2_{obm}$, which implies that as $n \to \infty$,

$$\text{Var}[\hat{\sigma}^2_{bm}]/\text{Var}[\hat{\sigma}^2_{bt}] = 1.5.$$

We would like to explore the variance ratio between SV and EFSV with flat top window $w_{ft}(\cdot)$. Denote $\hat{\sigma}^2_{ft}$ and $\hat{\sigma}^2_{efft}$ the corresponding univariate SV and EFSV estimator, the following theorem holds.

Theorem 2. If *condition 1* and *condition 3* hold, then

$$\frac{n}{b}\text{Var}[\hat{\sigma}^2_{ft}] = \frac{8}{3}\sigma^4_g + o(1),$$

and

$$\frac{n}{b}\text{Var}[\hat{\sigma}^2_{efft}] = 5\sigma^4_g + o(1).$$

Proof. See appendix B.

The little o notation $f(n) = o(g(n))$ means $\lim_{n\to\infty} f(n)/g(n) = 0$. It follows from theorem 2 that the ratio of the asymptotic variances is

$$\text{Var}[\hat{\sigma}^2_{efft}]/\text{Var}[\hat{\sigma}^2_{ft}] = 1.875.$$

Remark. A general expression of ratio between SV and EFSV is challenging to obtain but it can be derived given a specific lag window or approximated by simulation. In section 3, we will approximate by simulation the variances ratio for *Tukey-Hanning* window

$$w_{th}(s) = \begin{cases} (1+\cos(\pi|s|/b))/2 & \text{for } \leq b \\ 0 & \text{for } |s| > b. \end{cases}$$

Remark. Bias of $\hat{\Sigma}_{ef}$ is equivalent to bias of $\hat{\Sigma}_{sv}$. The result can be obtain by (Meketon and Schmeiser (1984) proposition 2).

3.2 Examples

In this section, three examples are considered to estimate θ by θ_n, then SV and EFSV estimators are applied to estimate variance of θ_n. In the first example, geometrically ergodic Markov chains are generated from univariate auto-regressive models. The example aims to evaluate performances of EFSV and compare variances between EFSV and SV estimators for three window functions $w_{bt}(\cdot)$, $w_{ft}(\cdot)$ and $w_{th}(\cdot)$. Then a multivariate vector auto-regressive model is considered to show computational gains of EFSV over SV for a range of dimensions and chain lengths. The last example compares performances of the two estimators on high dimensional real dataset.

3.2.1 Univariate auto-regressive example

Suppose ε_i are i.i.d N(0,1). Consider the following autoregressive process of order 1 (AR(1)):

$$X_i = \phi X_{i-1} + \varepsilon_i \quad \text{for } i = 1, 2, \ldots$$

For $|\phi| < 1$, the Markov chain is geometrically ergodic with invariant distribution $\mathrm{N}(0, 1/(1-\phi)^2)$. Consider approximating $\theta = E[X_i]$ by $\theta_n = \bar{X}_n$. We would like to estimate

$$\sigma^2 = \mathrm{Var}[X_1] + 2\sum_{s=1}^{\infty} \mathrm{Cov}(X_1, X_{1+s}) = 1/(1-\phi)^2.$$

Since $\mathrm{cov}(X_1, X_i) = \phi^{i-1}/(1-\phi^2)$, large ϕ results in a Markov chain with high auto-correlation therefore we consider a range of ϕ. The true value of σ^2 are used to evaluate performances of estimators.

More specifically, consider ϕ from 0.2 to 0.9. For each ϕ, generate AR(1) sample of length 1e5 and calculate the sample mean, then compute EFSV and SV using three window functions. Repeat the procedure 500 times to obtain sample variance of each estimator. $n^{1/3}$ is chosen to be the batch size or truncation point for all six estimators. The average of each estimator over

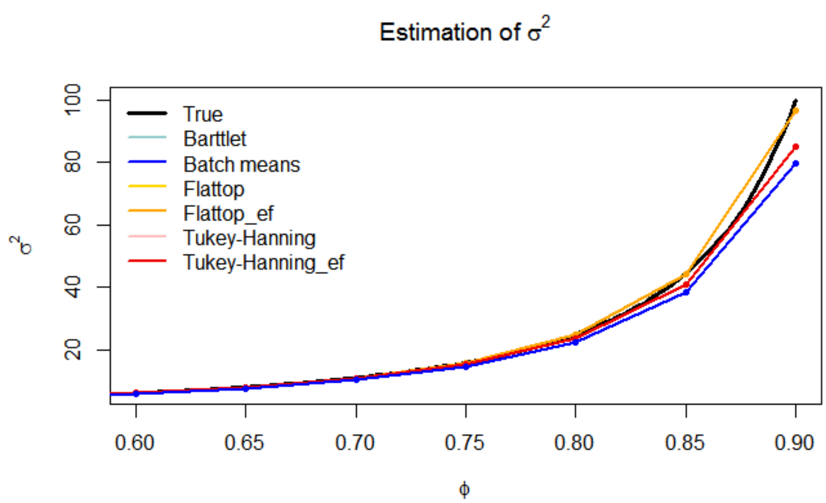


Figure 3.1 Estimation of σ^2 for AR(1) model with ϕ between 0.6 and 0.9. *Bartlett*, Flattop and *Tukey-Hanning* window are used for both SV and EFSV method. Results are based on the average over 500 replications, with a chain length of 1e5 for each replication.

500 replications is plotted against the true value of σ^2 as shown in figure 3.1. All six estimators

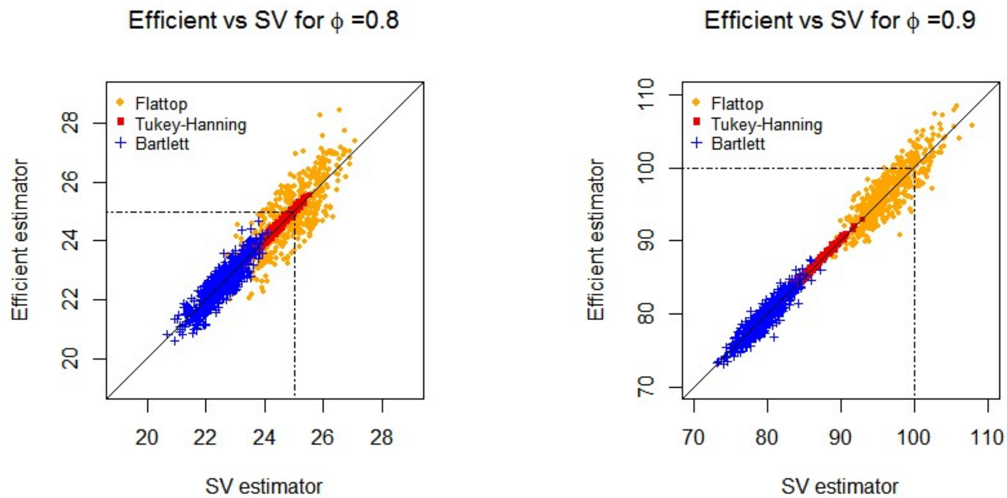


Figure 3.2 EFSV and SV estimators for $\phi = 0.8,\ 0.9$, true values of ϕ are denoted by dashed lines.

perform well when auto-covariance is low. As ϕ increases, SV and EFSV with flat top window seem to have better estimate compared with the other two window functions. Estimates between SV and EFSV with the same window are very close and they mostly overlap in figure 3.1. To have a closer observation, figure 3.2 plots 500 SV estimates against EFSV estimates for three windows when $\rho = 0.8,\ 0.9$. Flat top and *Bartlett* window have slightly larger discrepancies between SV and EFSV than *Tukey-Hanning* window but still are close to the line, showing that EFSV estimates are close to SV. When $\phi = 0.9$, flat top window has the smallest bias. Variance ratios between EFSV and SV are plotted in figure 3.3. Ratios for *Bartlett* and flat top window are close to the theoretical values of 1.5 and 1.875. *Tukey-Hanning* window has a ratio that is very close to 1, showing little inflation of variance.

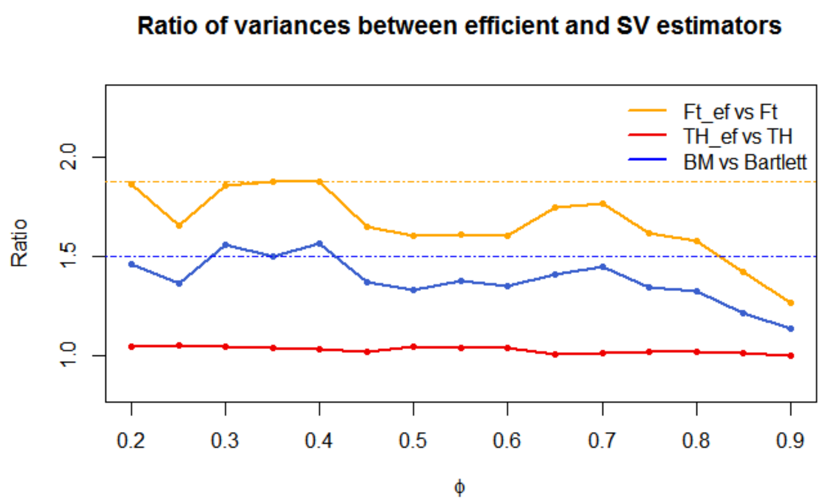


Figure 3.3 Variance ratio between EFSV and SV for AR(1) model using $w_{bt}(\cdot)$, $w_{ft}(\cdot)$ and $w_{th}(\cdot)$. Theoretical ratios of *Bartlett* and flat top window are 1.5 and 1.875, as shown by the dashed lines. Variances are calculated based over 500 replications, with a chain length of 1e5 for each replication.

3.2.2 Vector auto-regressive example

For $i = 1, 2, ...,$ consider p-dimensional vector autoregressive process of order 1 (VAR(1))

$$X_i = \Phi X_{i-1} + \varepsilon_i,$$

where $X_i \in \mathbb{R}^p$, ε_i are i.i.d $N_p(0, I_p)$ and Φ is a $p \times p$ matrix. Let $\otimes$ be the Kronecker product. When the largest eigenvalue of Φ in absolute value is less than 1, the Markov chain is geometrically ergodic Tjøstheim (1990) with invariant distribution $N_p(0, V)$, where $vec(V) = (I_{p^2} - \Phi \otimes \Phi)^{-1} vec(I_p)$. Consider approximating $\theta = EX_i$ by $\theta_n = \bar{X}_n$, we would like to estimate

$$\begin{aligned}\Sigma &= \text{Var}[X_1] + 2\sum_{s=1}^{\infty} \text{Cov}(X_1, X_1 + s) \\ &= (I_p - \Phi)^{-1} V + V(I_p - \Phi)^{-1} - V.\end{aligned}$$

We are interested in how chain length n and dimension p affect the computation time of EVSV and SV. For each combination of $p = 10,\ 20,\ 30$ and $n =$1e5, 1e6, 5e6, a geometrically Markov chain from VAR(1) is generated. Φ is chosen as follows to guarantee geometrically ergodicity. Consider a $p \times p$ matrix A with each entry generated from standard normal distribution, let $B = AA^T$ be a symmetric matrix with the largest eigenvalue m, then $\Phi = B/(m+1)$ is used in VAR(1). EFSV and SV estimators with $w_{bt}(\cdot)$, $w_{ft}(\cdot)$ and $w_{th}(\cdot)$ are applied to estimate Σ. We consider a total of 50 replications.

Table **??** shows ratios of average computational time between SV and EFSV for three window functions. There is significant computation gain for EFSV using flat top and *Bartlett* window. There is not much computation gain for *Tukey-Hanning* window due to the first summation in the expression of $\hat{\Sigma}_{th}$.

Besides computing time, one may be interested in the accuracy of an estimator. For an estimator $\hat{\Sigma}$, consider $E = \hat{\Sigma} - \Sigma$, and mean squared error across entries of E

$$mse = \frac{1}{p^2} \sum_i \sum_j e_{ij}^2$$

is used as a measurement of accuracy. Consider $\Phi_0 = B/(m+0.1)$ where B and m are constructed in the same way as above. Then we evaluate a series of $\Phi = k \cdot \Phi_0$, where $k = \{0.001, 0.01, 0.1, 0.5, 0.8\}$. Larger k implies stronger auto-covariance and cross auto-covariance of the chain. *mse* of EFSV and SV with $w_{bt}(\cdot)$, $w_{ft}(\cdot)$ and $w_{th}(\cdot)$ are calculated for each of the 500 replications.

Figure 3.4 shows the ratio of average *mse* between EFSV and SV for three windows. All ratios are below 2 for a range of Φ, indicating that EFSV has an less than 2 times inflated *mse* across various auto-covariances. Ratios of *Bartlett* and *Tukey-Hanning* window have a significant drop for $k = 0.8$ while flat top estimators have less of the trend. It is because both EFSV and SV using *Bartlett* and *Tukey-Hanning* window have poor performances when $k = 0.8$, resulting big *mse* for

n	Flat top 5e4	1e5	5e5	*Bartlett* 5e4	1e5	5e5	*Tukey-Hanning* 5e4	1e5	5e5
p=10	33.3	22.9	32.8	27.4	48.9	62.8	0.24	0.57	0.88
p=20	33.3	53.8	70.4	66.5	92.0	110.0	0.42	0.87	1.30
p=30	46.1	72.2	94.1	88.7	118.8	144.0	0.47	0.96	1.47

Table 3.1 Time ratio of SV and efficient estimator

both estimators a and lower ratio. Flat top estimators maintain better estimates, therefore the ratio is relatively stable when $k = 0.8$, which is in favor of flat top window estimators.

Truncation point for all estimators is $n^{1/3}$. Combine the computation and accuracy information, EFSV with flat top estimator exhibits superior performance compared with existing SV estimators.

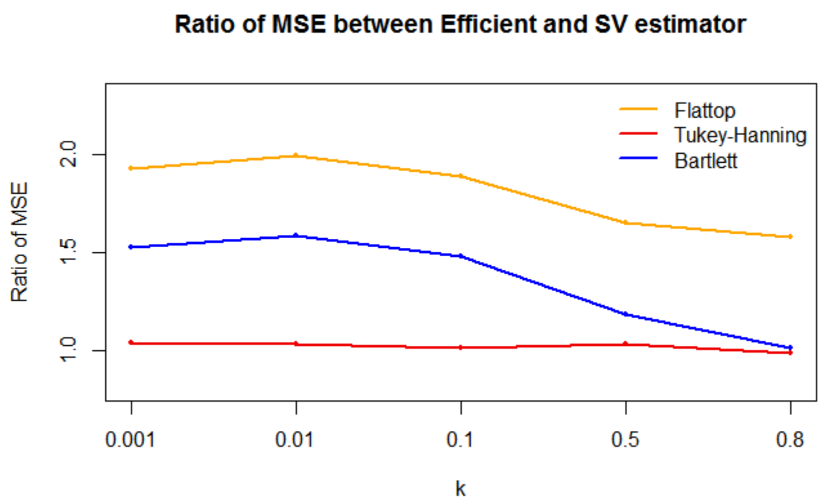


Figure 3.4 Ratio of MSE

3.2.3 Bayesian dynamic space-time example

This example is applied to monthly temperature data collected at 10 nearby station in northeastern United States in 2000, which is a subset of `NETemp` data described in `R` package spBayes Finley et al. (2013). A Bayesian dynamic model proposed by Gelfand et al. (2005) is fitted to the data and the model treats time as discrete and space as continuous variable.

Suppose y_t denote the temperature observed at location s and time t for $s = 1,2,...,N_s$ and $t = 1,2,...,N_t$. Let $x_t(s)$ be a $k \times 1$ vector of predictors and β_t be a $k \times 1$ coefficient vector, which is a purely time component. $u_t(s)$ denotes a space-time component. The model is

$$y_t(s) = \boldsymbol{x}_t(s)^T \boldsymbol{\beta}_t + u_t(s) + \varepsilon_t(s), \;\; \varepsilon_t \sim N(0, \tau_t^2),$$

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \boldsymbol{\eta}_t; \;\; \boldsymbol{\eta}_t \sim N_p(0, \Sigma_\eta),$$

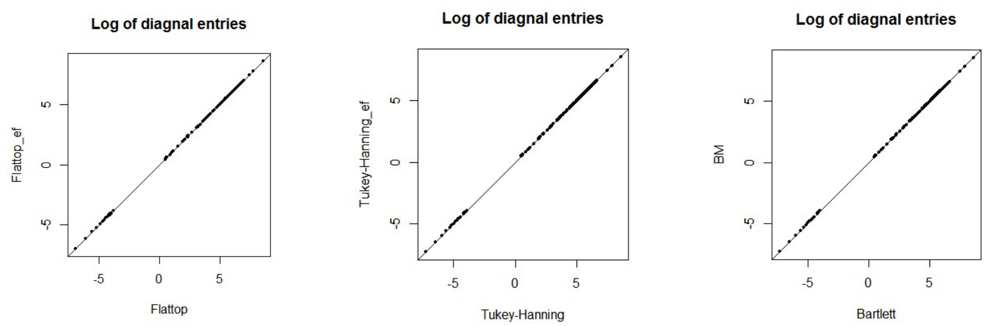
$$u_t(s) = u_{t-1}(s) + w_t(s); \;\; w_t(s) \sim GP(0, C_t(\cdot, \sigma_t^2, \phi_t)).$$

$GP(0, C_t(\cdot, \sigma_t^2, \phi_t))$ is a spatial Gaussian process where $C_t(s_1, s_2; \sigma_t^2, \phi_t) = \sigma_t^2 \rho(s_1, s_2; \phi_t)$. $\rho(\cdot; \phi)$ is an exponential correlation function with ϕ controlling the correlation decay, and σ_t^2 represents the spatial variance components. The Gaussian spacial process allows closer location to have higher correlation. Time effect for both $\boldsymbol{\beta}_t$ and $u_t(s)$ are characterized by transition equations, delivering a reasonable dependence structure. We are interested in estimating posterior expectation of 185 parameters $\theta = (\boldsymbol{\beta}_t,\; u_t(s),\; \sigma_t^2,\; \Sigma_\eta,\; \tau_t^2,\; \phi_t)$, their prior follows `spDynlM` function in `spBayes` package.

Consider Markov chains of length 5e4, 1e5 and 2e5. $\hat{\Sigma}_{efft}$ and $\hat{\Sigma}_{ft}$ are computed and ratios computation time between SV and EFSV are in table 3.2. For a high dimensional Bayesian analysis, the suggested EFSV estimator are much cheaper to compute using *Bartlett* and flat top window. With the advantage of bias reduction, EFSV with flat top window produces reliable results within a reasonable amount of time. Figure 3.5 are the log of diagonal elements form EFSV and SV estimators. The two methods have similar estimates.

Table 3.2 Time ratio of SV and EFSV for three windows

	N=5e4	N=1e5	N=2e5
Flattop	111.6	140.8	172.6
Tukey-Hanning	4.7	5.2	6.0
Bartlett	183.3	231.4	265.1



(a) Flattop (b) Tukey-Hanning (c) Bartlett

Figure 3.5 Log of diagonal entries of SV and EFSV estimators for three window functions

This chapter considers estimating asymptotic variance in Markov chain CLT by EFSV estimator. The method is especially useful in high dimensional problem where computation time is a major concern. Under various scenarios, EFSV estimates are comparable to SV with a slightly larger variance, but the ratio of variances are no more than two for windows considered in this chapter.

3.A Appendix of Chapter 3

We first introduce some notations and propositions. Let $B = \{B(t), t \geq 0\}$ be a p-dimensional standard Brownian motion. Denote $\bar{B} = n^{-1}B(n)$ and $\bar{B}_l(k) = k^{-1}[B(lk+k) - B(lk)]$. The Brownian motion counterpart of $\hat{\Sigma}_{ef}$ is

$$\tilde{\Sigma}_{ef} = \sum_{k=1}^{b} \frac{1}{a_k - 1} \sum_{l=0}^{a_k-1} k^2 \Delta_2 w_n(k)(\bar{B}_l(k) - \bar{B})(\bar{B}_l(k) - \bar{B})^T.$$

Define the following matrix:

$$\tilde{\Sigma}_{w,L} = \frac{1}{n} \sum_{k=1}^{b} \sum_{l=0}^{n-k} k^2 \Delta_2 w_n(k) L[\bar{B}_l(k) - \bar{B}][\bar{B}_l(k) - \bar{B}]^T L^T,$$

where L is the lower triangular matrix satisfying $\Sigma = LL^T$. Define $C(t) := LB(t)$, let $C^{(i)}(t)$ be the ith component of $C(t)$ and define $\bar{C}_l^{(i)}(k) = k^{-1}(C^{(i)}(l+k) - C^{(i)}(l))$, $\bar{C}^{(i)} = n^{-1}C^{(i)}(n)$.

Proposition 1. Vats et al. (2015b) For all $\varepsilon > 0$ and for almost all sample paths, there exists $n_0(\varepsilon)$ such that for all $n \geq n_0$ and all $i = 1, ..., p$

$$|C^{(i)}(n)| < (1+\varepsilon)(2n\Sigma_{ii} log log n)^{1/2},$$

where Σ_{ii} is the ith diagonal entry of Σ.

Proposition 2. Vats et al. (2015b) If *condition 3* holds, then for all $\varepsilon > 0$ and for almost all sample paths, there exists $n_0(\varepsilon)$ such that for all $n \geq n_0$ and all $i = 1, ..., p$

$$|\bar{C}_l^{(i)}(k)| \leq \frac{1}{k} \sup_{0 \leq l \leq n-b} \sup_{0 \leq s \leq b} |C^{(i)}(l+s) - C^{(i)}(l)| < \frac{1}{k} 2(1+\varepsilon)(b\Sigma_{ii} log n)^{1/2},$$

where Σ_{ii} is the ith diagonal entry of Σ.

A Proof of Theorem 1

Lemma 1. Suppose *condition 3* holds. If there exists a constant $c \geq 1$ such that $\sum_n (b/n)^c < \infty$ and

$$\sum_{k=1}^{b} k \Delta_2 w_n(k) = 1,$$

then $\tilde{\Sigma}_{ef} \to I_p$, *with probability 1*, where I_p is the $p \times p$ identity matrix.

Proof. We show that the diagonal elements of $\tilde{\Sigma}_{ef}$ goes to 1 and the off-diagonal elements goes to 0. For $i = j$,

$$\begin{aligned}
\tilde{\Sigma}_{ef,ii} &= \sum_{k=1}^{b} \frac{a_k}{a_k - 1} \left(\frac{1}{a_k} \sum_{l=0}^{a_k-1} k^2 \Delta_2 w_n(k) (\bar{B}_l^{(i)}(k) - \bar{B}^{(i)})^2 \right) \\
&= \sum_{k=1}^{b} \frac{a_k}{a_k - 1} \left(\frac{1}{a_k} \sum_{l=0}^{a_k-1} k^2 \Delta_2 w_n(k) (\bar{B}_l^{(i)}(k)^2 + (\bar{B}^{(i)})^2 - 2\bar{B}_l^{(i)}(k)\bar{B}^{(i)}) \right) \\
&= \sum_{k=1}^{b} \frac{a_k}{a_k - 1} \left(\frac{1}{a_k} \sum_{l=0}^{a_k-1} k^2 \Delta_2 w_n(k) \bar{B}_l^{(i)}(k)^2 + \frac{1}{a_k} a_k (\bar{B}^{(i)})^2 k^2 \Delta_2 w_n(k) - \frac{2}{a_k} \bar{B}^{(i)} k^2 \Delta_2 w_n(k) \sum_{l=0}^{a_k-1} \bar{B}_l^{(i)}(k) \right) \\
&= \sum_{k=1}^{b} \frac{a_k}{a_k - 1} \left(\frac{1}{a_k} \sum_{l=0}^{a_k-1} k^2 \Delta_2 w_n(k) \bar{B}_l^{(i)}(k)^2 + (\bar{B}^{(i)})^2 k^2 \Delta_2 w_n(k) - \frac{2}{a_k} \bar{B}^{(i)} k^2 \Delta_2 w_n(k) \frac{n}{k} \bar{B}^{(i)} \right) \\
&= \sum_{k=1}^{b} \frac{a_k}{a_k - 1} \left(\frac{1}{a_k} \sum_{l=0}^{a_k-1} k^2 \Delta_2 w_n(k) \bar{B}_l^{(i)}(k)^2 - (\bar{B}^{(i)})^2 k^2 \Delta_2 w_n(k) \right) \\
&= \sum_{k=1}^{b} \frac{a_k}{a_k - 1} \left[k \Delta_2 w_n(k) \left(\frac{1}{a_k} \sum_{l=0}^{a_k-1} k \bar{B}_l^{(i)}(k)^2 - k (\bar{B}^{(i)})^2 \right) \right].
\end{aligned}$$

By (Damerdji (1994) proof of proposition 3.1),

$$\frac{1}{a_k} \sum_{l=0}^{a_k-1} k \bar{B}_l^{(i)}(k)^2 \to 1 \quad \text{and} \quad k(\bar{B}^{(i)})^2 \to 0 \quad \text{as} \quad n \to \infty.$$

Therefore

$$\tilde{\Sigma}_{ef,ii} \to \sum_{k=1}^{b} k \Delta_2 w_n(k) = 1 \quad \text{as} \quad n \to \infty.$$

When $i \neq j$,

$$\begin{aligned}
\tilde{\Sigma}_{ef,ij} &= \sum_{k=1}^{b} \frac{1}{a_k - 1} \sum_{l=0}^{a_k-1} k^2 \Delta_2 w_n(k) (\bar{B}_l^{(i)}(k) - \bar{B}^{(i)})(\bar{B}_l^{(j)}(k) - \bar{B}^{(j)}) \\
&= \sum_{k=1}^{b} k \Delta_2 w_n(k) \frac{k}{a_k - 1} \sum_{l=0}^{a_k-1} [\bar{B}_l^{(i)}(k) \bar{B}_l^{(j)}(k) - \bar{B}_l^{(i)}(k) \bar{B}^{(j)} - \bar{B}^{(i)} \bar{B}_l^{(j)}(k) + \bar{B}^{(i)} \bar{B}^{(j)}].
\end{aligned}$$

By proof of (*lemma 3* Vats et al. (2015a)),

$$\frac{k}{a_k - 1} \sum_{l=0}^{a_k-1} [\bar{B}_l^{(i)}(k) \bar{B}_l^{(j)}(k) - \bar{B}_l^{(i)}(k) \bar{B}^{(j)} - \bar{B}^{(i)} \bar{B}_l^{(j)}(k) + \bar{B}^{(i)} \bar{B}^{(j)}] \to 0.$$

Together with (3.1.1), $\tilde{\Sigma}_{ef,ij} \to 0$, resulting $\tilde{\Sigma}_{ef} \to I_p$ *with probability 1.*

Lemma 2. Let *condition 1* hold for g and *condition 2, 3* hold. If as $n \to \infty$,

$$b\psi(n)^2 logn \left(\sum_{k=1}^{b} |\Delta_2 w_n(k)| \right)^2 \to 0,$$

and

$$\psi(n)^2 \sum_{k=1}^{b} |\Delta_2 w_n(k)| \to 0,$$

then $\hat{\Sigma}_{ef} \to L\tilde{\Sigma}_{ef}L^T = \tilde{\Sigma}_{ef,L}$ *with probability 1.*

Proof. we will show that for $i, j = 1, 2, ...p$, $\hat{\Sigma}_{ef,ij} \to \tilde{\Sigma}_{ef,L,ij}$. Let $Y_i = g(X_i) - E_F g$, $\bar{C}_l(k) = L\bar{B}_l(k)$ and $\bar{C} = L\bar{B}$. Then

$$\begin{aligned}
\tilde{\Sigma}_{ef,L} &= \sum_{k=1}^{b} \frac{1}{a_k - 1} \sum_{l=0}^{a_k-1} k^2 \Delta_2 w_n(k) L(\bar{B}_l(k) - \bar{B})(\bar{B}_l(k) - \bar{B})^T L^T \\
&= \sum_{k=1}^{b} \frac{1}{a_k - 1} \sum_{l=0}^{a_k-1} k^2 \Delta_2 w_n(k) (\bar{C}_l(k) - \bar{C})(\bar{C}_l(k) - \bar{C})^T,
\end{aligned}$$

yielding

$$\hat{\Sigma}_{ef,ij} - \tilde{\Sigma}_{ef,L,ij} = \sum_{k=1}^{b} \frac{1}{a_k - 1} \sum_{l-0}^{n-k} k^2 \Delta_2 w_k [(\bar{Y}_l^{(i)}(k) - \bar{Y}^{(i)})(\bar{Y}_l^{(j)}(k) - \bar{Y}^{(j)}) - (\bar{C}_l^{(i)}(k) - \bar{C}^{(i)})(\bar{C}_l^{(j)}(k) - \bar{C}^{(j)})].$$

Let

$$A_k = k(\bar{Y}_l(k) - \bar{C}_l(k)),$$

$$D_k = B(l+k) - B(l),$$

$$E_{n,k} = k\bar{B},$$

$$F_{n,k} = k(\bar{Y} - \bar{C}),$$

since

$$k(\bar{Y}_l^{(i)}(k) - \bar{Y}^{(i)}) = k(\bar{Y}_l^{(i)}(k) - \bar{Y}^{(i)} + \bar{C}_l^{(i)}(k) - \bar{C}_l^{(i)}(k) + \bar{C}^{(i)} - \bar{C}^{(i)})$$

$$= k(\bar{Y}_l^{(i)}(k) - \bar{C}_l^{(i)}(k)) + (k\bar{C}_l^{(i)}(k) - k\bar{C}^{(i)}) - k(\bar{Y}^{(i)} - \bar{C}^{(i)})$$

$$= A_k^{(i)} + (LD_k)^{(i)} - (LE_{n,k})^{(i)} - F_{n,k}^{(i)},$$

$$\begin{aligned}
|\hat{\Sigma}_{ef,ij} - \tilde{\Sigma}_{ef,L,ij}| \leq & \sum_{k=1}^{b} \frac{1}{a_k - 1} \sum_{l=0}^{n-k} |\Delta_2 w_k| \cdot |[A_k^{(i)} + (LD_k)^{(i)} - (LE_{n,k})^{(i)} - F_{n,k}^{(i)}] \cdot [A_k^{(j)} + (LD_k)^{(j)} - (LE_{n,k})^{(j)} - F_{n,k}^{(j)}] \\
& - [(LD_k)^{(i)} - (LE_{n,k})^{(i)}] \cdot [(LD_k)^{(j)} - (LE_{n,k})^{(j)}]| \\
= & \sum_{k=1}^{b} \frac{1}{a_k - 1} \sum_{l=0}^{n-k} |\Delta_2 w_k| \cdot |A_k^{(i)} A_k^{(j)} + A_k^{(i)} (LD_k)^{(j)} - A_k^{(i)} (LE_{n,k})^{(j)} - A_k^{(i)} F_{n,k}^{(j)} \\
& + (LD_k)^{(i)} A_k^{(j)} - (LD_k)^{(i)} F_{n,k}^{(j)} - (LE_{n,k})^{(i)} A_k^{(j)} + (LE_{n,k})^{(i)} F_{n,k}^{(j)} \\
& - F_{n,k}^{(i)} A_k^{(j)} - F_{n,k}^{(i)} (LD_k)^{(j)} + F_{n,k}^{(i)} (LE_{n,k})^{(j)} + F_{n,k}^{(i)} F_{n,k}^{(j)}| \\
= & \sum_{k=1}^{b} \frac{1}{a_k - 1} \sum_{l=0}^{n-k} |\Delta_2 w_k| \cdot |A_k^{(i)} A_k^{(j)} + F_{n,k}^{(i)} F_{n,k}^{(j)} + [A_k^{(i)} (LD_k)^{(j)} + (LD_k)^{(i)} A_k^{(j)}] \\
& - [A_k^{(i)} (LE_{n,k})^{(j)} + (LE_{n,k})^{(i)} A_k^{(j)}] - [A_k^{(i)} F_{n,k}^{(j)} + F_{n,k}^{(i)} A_k^{(j)}] \\
& - [(LD_k)^{(i)} F_{n,k}^{(j)} + F_{n,k}^{(i)} (LD_k)^{(j)}] + [(LE_{n,k})^{(i)} F_{n,k}^{(j)} + F_{n,k}^{(i)} (LE_{n,k})^{(j)}]|.
\end{aligned} \tag{3.A.1}$$

We can show each of the twelve sums in the above expression goes to 0. Apply *condition 1*,

$$\begin{aligned}
|A_k^{(i)}| &= k[\bar{Y}_l^{(i)}(k) - \bar{C}_l^{(i)}(k)] \\
&= k\left[k^{-1} \sum_{t=1}^{k} Y_{lk+t}^{(i)} - k^{-1}(C^{(i)}(lk+k) - C^{(i)}(lk))\right] \\
&= \left[\sum_{t=1}^{(lk+k)} Y_t^{(i)} - \sum_{t=1}^{lk} Y_t^{(i)}\right] - [C^{(i)}(lk+k) - C^{(i)}(lk)] \\
&= \left[\sum_{t=1}^{lk+k} Y_t^{(i)} - C^{(i)}(lk+k)\right] - \left[\sum_{t=1}^{lk} Y_t^{(i)} - C^{(i)}(lk)\right] \\
&\leq D \cdot \psi(l+k) + D \cdot \psi(l) \\
&\leq 2D\psi(n).
\end{aligned}$$

Hence by the assumption of the lemma, as $n \to \infty$,

$$\sum_{k=1}^{b} \frac{1}{a_k - 1} \sum_{l=0}^{a_k-1} |\Delta_2 w_n(k)||A_k^{(i)} A_k^{(j)}| \leq 4D^2 \psi^2(n) \sum_{k=1}^{b} \frac{a_k}{a_k - 1} |\Delta_2 w_n(k)| \to 0.$$

Similarly, by results from chapter 2 and *proposition 1, 2*, the other terms in (3.A.1) goes to 0, and $\hat{\Sigma}_{ef} \to L\tilde{\Sigma}_{ef}L^T$ *with probability 1*.

Proof of theorem 1 follows from *lemma 1* and *lemma 2*.

B Proof of Theorem 2

Variance of $\hat{\sigma}^2_{ft}$ follows from the proof of theorem in chapter 2 and taking $c = 1/2$ in *lemma 1* in chapter 2. We will derive the variance of $\hat{\sigma}^2_{ef-ft}$.

Lemma 3. Under *condition 3*,

$$\frac{n}{b}\text{Var}[\tilde{\sigma}^2_{ef-ft}] = 5 + o(1).$$

Proof.

$$\tilde{\sigma}^2_{ef-ft} = \frac{2b}{a-1} \sum_{l=0}^{a-1} (\bar{B}_l(b) - \bar{B})^2 - \frac{b/2}{2a-1} \sum_{l=0}^{2a-1} (\bar{B}_l(b/2) - \bar{B})^2.$$

It can be shown that

$$E[\tilde{\sigma}^4_{ef-ft}] = \frac{4a^4 + 8a^3 - 12a^2 + 4a}{(a-1)^2(2a-1)^2} + o(\frac{1}{a}),$$

and

$$\left(E[\tilde{\sigma}^2_{ef-ft}]\right)^2 = 1,$$

resulting

$$\begin{aligned}\text{Var}[\tilde{\sigma}^2_{ef-ft}] &= E[\tilde{\sigma}^4_{ef-ft}] - \left(E[\tilde{\sigma}^2_{ef-ft}]\right)^2 \\ &= \frac{5}{a} + o(\frac{1}{a}).\end{aligned}$$

Note $n = ab$, *lemma 3* is proved.

Similar as *lemma 5* and theorem 1 in chapter 2, we can show that as $n \rightarrow$,

$$E[\hat{\sigma}^2_{ef-ft} - \sigma^2_g \tilde{\sigma}^2_{ef-ft}] \rightarrow 0,$$

and

$$\frac{n}{b}\mathrm{Var}[\hat{\sigma}^2_{ef-ft}] = 5\sigma^4_g + o(1).$$

Theorem 2 is proved.

Chapter 4

Conclusions

My thesis focuses on estimating Σ in Markov chain central limit theorem for high-dimensional MCMC method. I have mainly two contributions. First, I proposed a procedure to select optimal bandwidth for a family of mSV estimator, which received no previous attention under MCMC context. The proposed bandwidth significantly improved the commonly used bandwidth of $n^{1/3}$. Other related results include asymptotic variance and bias, as well as mean squared consistency of the estimators considered. Second, I try to ease the computational burden for high-dimensional problem by considering a new family of estimators that are fast to compute yet delivers comparative results as existing mSV methods. The establishment of this work answers a crucial practical question, that is, can we monitor Σ in an accurate yet computationally affordable way so that one can terminate Markov chain in a sensible way for high-dimensional problem.

It needs to be pointed out that bias and variance conclusions in chapter 2 are established under different ergodicity conditions. Bias results requires uniformly ergodicity, which is stronger than polynomial ergodicity required by variance results. It is of interest to explore asymptotic bias under weaker ergodicity conditions for a more uniformed results.

Recall (2.2.1) in chapter 2, bandwidth in iterative plug-in pilot estimates was obtained by

$$b_t = \Big[\frac{6\sum_{s=-b_{t-1}n^{-4/21}}^{b_{t-1}n^{-4/21}} \bar{w}^2(s)s^2\hat{\gamma}_{ii}^2(s)}{\sum_{s=-b_{t-1}n^{-4/21}}^{b_{t-1}n^{-4/21}} \hat{\gamma}_{ii}^2(s)}\Big]^{1/3} \cdot n^{1/3}$$

for $t = 1, ..., 4$, and bandwidth for pilot estimate is $b = b_4 n^{-4/21}$. The pilot estimates are then used in local step where $\lambda = 0$ to obtain the optimal bandwidth for $\hat{\sigma}_{bt}^2$ estimator. (2.2.1) is based on the global optimal bandwidth

$$b_{global} = \Big(\frac{3 \cdot \int_{-\pi}^{\pi}\{f^{(1)}(\lambda)\}^2 d\lambda}{\int_{-\pi}^{\pi}\{f(\lambda)\}^2 d\lambda}\Big) \cdot n^{1/3}$$

that minimizes mean integrated squared error $\text{MISE} = \text{E}[\int_{-\pi}^{\pi}(\hat{f}(\lambda) - f(\lambda))^2 d\lambda]$ for *Bartlett* window. Although the goal is to obtain an optimal bandwidth for $\hat{\sigma}_{bt}^2$, the iterative updating procedure in global step estimates $3 \cdot \int_{-\pi}^{\pi}\{f^{(1)}(\lambda)\}^2 d\lambda$ by a flat top *Tukey-Hanning* window $\bar{w}(s)$. It is appealing to consider a b_{global} that incorporates the optimal rate of flat top *Tukey-Hanning* window since in each of the four iteration, the updated b_t is used as the bandwidth for the flat top *Tukey-Hanning* window in the next iteration, and b is the bandwidth for pilot estimates based on *Tukey-Hanning* and flat top *Tukey-Hanning* window. However, b_t is updated according to the optimal bandwidth for *Bartlett* window. A starting point is to consider how the bandwidth for $\hat{\sigma}_{bt}^2$ in local step will be affected when b_t is updated according to the optimal bandwidth of flat top *Tukey-Hanning* window in global step.

Simulation results in 2.3.1 shows that iterative plug-in pilot estimates have smaller variance compared with flat top pilot estimates. When dimension is low and computation resource is available, one may prefer iterative plug-in pilot estimate to achieve estimates with less variance. However, there are two restrictions on the iterative plug-in pilot estimates discussed in chapter 2.2. Firstly, only the iterative plug-in method for $\hat{\sigma}_{bt}^2$ estimator is introduced in this dissertation. For more general results, see Bühlmann (1996). Secondly, the method is established under univariate settings. It is attempting to explore the iterative plug-in method under multivariate settings, and the idea of iterative plug-in should work regardless of dimensions.

The mBM estimator Vats et al. (2015a) belongs to the family of EFSV estimator. It is EFSV estimator with *Bartlett* window function. To our best knowledge, only mBM is addressed as an applicable estimator of Σ in multivariate MCMC context. But the performance of mBM estimator is tightly attached to the performance of *Bartlett* window, which can be sub-optimal compared with flat top window as shown in chapter 3. The EFSV estimators are extensions of mBM from this point of view. By choosing window function such as flat top window, the estimates are more accurate compared with mBM. Meanwhile they have tremendous computational gains compared with mSV estimators with flat top window.

As shown in chapter 3, EFSV with *Tukey-Hanning* does not reduce computation time as dramatically as the other two windows. By rearranging the order of terms in the expression of EFSV estimators, they can be viewed as a linear combination of mBM estimators with various bandwidths. Hence the computing time is shorter for those EFSV estimators with less components in the linear combination. In other words, EFSV estimator is more beneficial to window functions with $\Delta_2 w_n(s) = 0$ for certain s. Nevertheless, as dimension and chain length increases, EFSV could still save computing time as shown in the 3.2.2. More research can be done regarding the efficient coding of EFSV with *Tukey-Hanning* window if one prefers to use this window.

The proposed EFSV estimators can be more influential than illustrated in this dissertation mainly because when used in a stopping rule, multiple checking of $\text{Var}[\bar{X}_n]$ is usually required as n increases. The computational gains of EFSV then becomes tremendous especially for high-dimensional problem. More importantly, the resulting convenience and efficiency can determine whether or not a practitioner keeps track of Var$[\bar{X}_n]$. Hence fast computing or not directly affects whether a chain is stopped in a sensible way or not.

All the estimators in chapter 3 use a bandwidth of $n^{1/3}$ for fair comparisons, but $n1/3$ is not necessarily the best choice. In fact, there is room to improve their performances by using a better bandwidth. It was shown in chapter 2 that estimating the coefficient of $n^{1/3}$ results in a significant improvement for the estimators. Similar technique can be applied to EFSV estimators. If we want to omit further exploration, the same bandwidth for SV estimator can be used for the corresponding EFSV estimator, since optimal bandwidth expression of SV and EFSV usually have similar components. (see Flegal and Jones (2010) for an example regarding optimal bandwidth of BM and OBM). As for flat top window, it is suggested by Politis (2003) to use an *empirical rule* for the optimal batch size of flattop SV estimator. We suggest using the same *empirical rule* for EFSV with flat top window, but more work can be done regarding the bandwidth selection for EFSV estimators.

Results in chapter 3 can also be applied to the optimal truncation point methods disccused in chapter 2. The flat top pilot estimates could be replaced by EFSV estimators using flat top window function, and the resulting $\hat{\Sigma}^{(0)}_{efft}$ is a consistent estimate of Σ according to theorem 1 in chapter 3. The consistency of $\hat{\Gamma}^{(0)}_{efft}$ needs to be further established to complete the EFSV pilot estimates but should be obtainable as the strong consistency of $\hat{\Gamma}^{(0)}_{ft}$ was given by Politis (2003). In fact, using EFSV as pilot estimates is an attractive idea since obtaining pilot estimates is a preliminary step to achieve a good $\hat{\Sigma}_{sv}$, and a fast estimate will be appreciated. Computational concern is also the reason why flat top pilot estimates are computed based on a portion of the sample.

Lastly, methods established in this dissertation are under multivariate settings which is also determinant to a reasonable termination of the chain, given that Bayesian analysis almost always involves multiple parameters. Results restricted to univariate settings are theoretically important but can hardly be applied in practice.

Bibliography

Anderson, T. W. (2011). *The statistical analysis of time series*, volume 19. John Wiley & Sons.

Atchadé, Y. F. (2016). Markov chain Monte Carlo confidence intervals. *Bernoulli*, 22(3):1808–1838.

Bednorz, W. and LatuszyŃski, K. (2012). A few remarks on fixed-width output analysis for Markov chain Monte Carlo by jones et al. *Journal of the American Statistical Association*.

Bowman, A. W. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, 71(2):353–360.

Bradley, R. C. (1985). Basic properties of strong mixing conditions. Technical report, DTIC Document.

Bratley, P., Fox, B. L., and Schrage, L. E. (2011). *A guide to simulation*. Springer Science & Business Media.

Brockmann, M., Gasser, T., and Herrmann, E. (1993). Locally adaptive bandwidth choice for kernel regression estimators. *Journal of the American Statistical Association*, 88(424):1302–1309.

Brooks, S. (1998a). Markov chain Monte Carlo method and its application. *Journal of the royal statistical society: series D (the Statistician)*, 47(1):69–100.

Brooks, S. (1998b). Markov chain Monte Carlo method and its application. *Journal of the royal statistical society: series D (the Statistician)*, 47(1):69–100.

Brooks, S., Gelman, A., Jones, G., and Meng, X.-L. (2011). *Handbook of Markov Chain Monte Carlo*. CRC Press.

Bühlmann, P. (1996). Locally adaptive lag-window spectral estimation. *Journal of Time Series Analysis*, 17(3):247–270.

Bühlmann, P. and Künsch, H. R. (1999). Block length selection in the bootstrap for time series. *Computational Statistics & Data Analysis*, 31(3):295–310.

Chan, K. S. and Geyer, C. J. (1994). Discussion: Markov chains for exploring posterior distributions. *The Annals of Statistics*, 22(4):1747–1758.

Charnes, J. M. (1995). Analyzing multivariate output. In *Proceedings of the 27th conference on Winter simulation*, pages 201–208. IEEE Computer Society.

Chen, D.-F. R. and Seila, A. F. (1987). Multivariate inference in stationary simulation using batch means. In *Proceedings of the 19th conference on Winter simulation*, pages 302–304. ACM.

Chien, C., Goldsman, D., and Melamed, B. (1997). Large-sample results for batch means. *Management Science*, 43(9):1288–1295.

Csáki, E. and Csörgő, M. (1995). On additive functionals of Markov chains. *Journal of Theoretical Probability*, 8(4):905–919.

Csörgo, M. and Révész, P. (2014). *Strong approximations in probability and statistics*. Academic Press.

Damerdji, H. (1991). Strong consistency and other properties of the spectral variance estimator. *Management Science*, 37(11):1424–1440.

Damerdji, H. (1994). Strong consistency of the variance estimator in steady-state simulation output analysis. *Mathematics of Operations Research*, 19(2):494–512.

Damerdji, H. (1995). Mean-square consistency of the variance estimator in steady-state simulation output analysis. *Operations Research*, 43(2):282–291.

Efron, B. and Morris, C. (1975). Data analysis using stein's estimator and its generalizations. *Journal of the American Statistical Association*, 70(350):311–319.

Finley, A. O., Banerjee, S., and Gelfand, A. E. (2013). spbayes for large univariate and multivariate point-referenced spatio-temporal data models. *arXiv preprint arXiv:1310.8192*.

Fishman, G. S. (1978). Principles of discrete event simulation.[book review].

Flegal, J. M. and Gong, L. (2013). Relative fixed-width stopping rules for Markov chain Monte Carlo simulations. *arXiv preprint arXiv:1303.0238*.

Flegal, J. M., Haran, M., and Jones, G. L. (2008). Markov chain Monte Carlo: Can we trust the third significant figure? *Statistical Science*, pages 250–260.

Flegal, J. M. and Jones, G. L. (2010). Batch means and spectral variance estimators in Markov chain Monte Carlo. *The Annals of Statistics*, 38(2):1034–1070.

Gelfand, A. E., Banerjee, S., and Gamerman, D. (2005). Spatial process modelling for univariate and multivariate dynamic spatial data. *Environmetrics*, 16(5):465–479.

Gelfand, A. E., Kim, H.-J., Sirmans, C., and Banerjee, S. (2003). Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association*, 98(462):387–396.

Geyer, C. J. (1999). Likelihood inference for spatial point processes. *Stochastic geometry: likelihood and computation*, 80:79–140.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (1996). Introducing Markov chain Monte Carlo. In *Markov chain Monte Carlo in practice*, pages 1–19. Springer.

Givens, G. H. and Hoeting, J. A. (2012). *Computational statistics*, volume 708. John Wiley & Sons.

Glynn, P. W. and Iglehart, D. L. (1990). Simulation output analysis using standardized time series. *Mathematics of Operations Research*, 15(1):1–16.

Glynn, P. W. and Whitt, W. (1990). Estimating the asymptotic variance with batch means. Technical report, DTIC Document.

Glynn, P. W. and Whitt, W. (1992). The asymptotic validity of sequential stopping rules for stochastic simulations. *The Annals of Applied Probability*, pages 180–198.

Goldsman, D. and Meketon, M. (1986). *A comparison of several variance estimators*. Georgia Institute of Technology.

Gong, L. and Flegal, J. M. (2015). A practical sequential stopping rule for high-dimensional Markov chain Monte Carlo. *Journal of Computational and Graphical Statistics*, (just-accepted):00–00.

Gong, L., Flegal, J. M., Spindler, S. R., and Mote, P. L. (2015). Bayesian model selection on linear mixed-effects models for comparisons between multiple treatments and a control. *arXiv preprint arXiv:1509.07510*.

Häggström, O. (2005). On the central limit theorem for geometrically ergodic Markov chains. *Probability theory and related fields*, 132(1):74–82.

Hall, P. (1980). Objective methods for the estimation of window size in the nonparametric estimation of a density. *Unpublished manuscript*.

Harris, B. (1967). *Spectral analysis of time series*. Wiley.

Hobert, J. P. and Geyer, C. J. (1998). Geometric ergodicity of gibbs and block gibbs samplers for a hierarchical random effects model. *Journal of Multivariate Analysis*, 67(2):414–430.

Hobert, J. P., Jones, G. L., Presnell, B., and Rosenthal, J. S. (2002). On the applicability of regenerative simulation in Markov chain Monte Carlo. *Biometrika*, pages 731–743.

Janssen, P. H. and Stoica, P. (1987). *On the expectation of the product of four matrix-valued Gaussian random variables*. Eindhoven University of Technology.

Jarner, S. F. and Hansen, E. (2000). Geometric ergodicity of Metropolis algorithms. *Stochastic processes and their applications*, 85(2):341–361.

Jones, G. L. (2004). On the Markov chain central limit theorem. *Probability surveys*, 1(299-320):5–1.

Jones, G. L., Haran, M., Caffo, B. S., and Neath, R. (2006). Fixed-width output analysis for Markov chain Monte Carlo. *Journal of the American Statistical Association*, 101(476):1537–1547.

Jones, G. L. and Hobert, J. P. (2001). Honest exploration of intractable probability distributions via Markov chain Monte Carlo. *Statistical Science*, pages 312–334.

Jones, G. L. and Hobert, J. P. (2004). Sufficient burn-in for gibbs samplers for a hierarchical random effects model. *Annals of statistics*, pages 784–817.

Jones, M. C., Marron, J. S., and Sheather, S. J. (1996). A brief survey of bandwidth selection for density estimation. *Journal of the American Statistical Association*, 91(433):401–407.

Kosorok, M. R. (2000). Monte Carlo error estimation for multivariate Markov chains. *Statistics & probability letters*, 46(1):85–93.

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.

Lahiri, S. N. (1999). Theoretical comparisons of block bootstrap methods. *Annals of Statistics*, pages 386–404.

Lindvall, T. (2002). *Lectures on the coupling method*. Courier Corporation.

Liu, J., Nordman, D. J., and Meeker, W. Q. (2016). The number of MCMC draws needed to compute bayesian credible bounds. *The American Statistician*, 70(3):275–284.

Liu, J. S. (2008). *Monte Carlo strategies in scientific computing*. Springer Science & Business Media.

Loader, C. R. (1999). Bandwidth selection: classical or plug-in? *Annals of Statistics*, pages 415–438.

Marchev, D. and Hobert, J. P. (2004). Geometric ergodicity of van dyk and meng's algorithm for the multivariate student's t model. *Journal of the American Statistical Association*, 99(465):228–238.

Meketon, M. S. and Schmeiser, B. (1984). Overlapping batch means: Something for nothing? In *Proceedings of the 16th conference on Winter simulation*, pages 226–230. IEEE Press.

Mengersen, K. L. and Tweedie, R. L. (1996). Rates of convergence of the hastings and Metropolis algorithms. *The Annals of Statistics*, 24(1):101–121.

Meyn, S. P. and Tweedie, R. L. (1994). Computable bounds for geometric convergence rates of Markov chains. *The Annals of Applied Probability*, pages 981–1011.

Meyn, S. P. and Tweedie, R. L. (2009). *Markov chains and stochastic stability*. Cambridge university press.

Meyn, S. P. and Tweedie, R. L. (2012). *Markov chains and stochastic stability*. Springer Science & Business Media.

Muñoz, D. F. and Glynn, P. W. (2001). Multivariate standardized time series for steady-state simulation output analysis. *Operations Research*, 49(3):413–422.

Mykland, P., Tierney, L., and Yu, B. (1995). Regeneration in Markov chain samplers. *Journal of the American Statistical Association*, 90(429):233–241.

Neath, R. C. and Jones, G. L. (2009). Variable-at-a-time implementations of Metropolis-Hastings. *arXiv preprint arXiv:0903.0664*.

Newey, W. K. and West, K. D. (1986). A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix.

Parzen, E. (1957a). On choosing an estimate of the spectral density function of a stationary time series. *The Annals of Mathematical Statistics*, pages 921–932.

Parzen, E. (1957b). On consistent estimates of the spectrum of a stationary time series. *The Annals of Mathematical Statistics*, pages 329–348.

Philipp, W. and Stout, W. F. (1975). *Almost sure invariance principles for partial sums of weakly dependent random variables*, volume 161. American Mathematical Soc.

Politis, D. N. (2003). Adaptive bandwidth choice. *Journal of Nonparametric Statistics*, 15(4-5):517–533.

Politis, D. N. (2009). Higher-order accurate, positive semi-definite estimation of large-sample covariance and spectral density matrices. *Department of Economics, UCSD*.

Politis, D. N. and Romano, J. P. (1995). Bias-corrected nonparametric spectral estimation. *Journal of time series analysis*, 16(1):67–103.

Politis, D. N. and Romano, J. P. (1996). On flat-top kernel spectral density estimators for homogeneous random fields. *Journal of Statistical Planning and Inference*, 51(1):41–53.

Politis, D. N. and Romano, J. P. (1999). Multivariate density estimation with general flat-top kernels of infinite order. *Journal of Multivariate Analysis*, 68(1):1–25.

Politis, D. N. and White, H. (2004). Automatic block-length selection for the dependent bootstrap. *Econometric Reviews*, 23(1):53–70.

Robert, C. and Casella, G. (2013). *Monte Carlo statistical methods*. Springer Science & Business Media.

Roberts, G. O., Gelman, A., and Gilks, W. R. (1997). Weak convergence and optimal scaling of random walk Metropolis algorithms. *The annals of applied probability*, 7(1):110–120.

Roberts, G. O. and Polson, N. G. (1994). On the geometric convergence of the gibbs sampler. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 377–384.

Roberts, G. O. and Rosenthal, J. S. (1997). Geometric ergodicity and hybrid Markov chains. *Electron. Comm. Probab*, 2(2):13–25.

Roberts, G. O. and Rosenthal, J. S. (1999). Convergence of slice sampler Markov chains. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(3):643–660.

Roberts, G. O. and Rosenthal, J. S. (2004). General state space Markov chains and MCMC algorithms. *Probability Surveys*, 1:20–71.

Roberts, G. O. and Rosenthal, J. S. (2009). Examples of adaptive MCMC. *Journal of Computational and Graphical Statistics*, 18(2):349–367.

Romano, D. N. P. J. P. (1994). Bias-corrected nonparametric spectral estimation ii.

Rosenblatt, M. (1971). *Markov processes: structure and asymptotic behavior*. Springer Berlin-Heidelberg-New York.

Rosenthal, J. S. (1995). Minorization conditions and convergence rates for Markov chain Monte Carlo. *Journal of the American Statistical Association*, 90(430):558–566.

Rosenthal, J. S. (1996). Analysis of the gibbs sampler for a model related to james-stein estimators. *Statistics and Computing*, 6(3):269–275.

Roy, V. and Hobert, J. P. (2007). Convergence rates and asymptotic standard errors for Markov chain Monte Carlo algorithms for bayesian probit regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(4):607–623.

Rudemo, M. (1982). Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics*, pages 65–78.

Sheather, S. (1983). A data-based algorithm for choosing the window width when estimating the density at a point. *Computational Statistics & Data Analysis*, 1:229–238.

Sheather, S. J. and Jones, M. C. (1991). A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 683–690.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*, volume 26. CRC press.

Song, W. T. (1996). On the estimation of optimal batch sizes in the analysis of simulation output. *European Journal of Operational Research*, 88(2):304–319.

Song, W. T. and Schmeiser, B. (1988). Minimal-mse linear combinations of variance estimators of the sample mean. In *Proceedings of the 20th conference on Winter simulation*, pages 414–421. ACM.

Song, W. T. and Schmeiser, B. W. (1993). Variance of the sample mean: Properties and graphs of quadratic-form estimators. *Operations Research*, 41(3):501–517.

Song, W. T. and Schmeiser, B. W. (1995). Optimal mean-squared-error batch sizes. *Management Science*, 41(1):110–123.

Tan, A. and Hobert, J. P. (2012). Block gibbs sampling for bayesian random effects models with improper priors: Convergence and regeneration. *Journal of Computational and Graphical Statistics*.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *the Annals of Statistics*, pages 1701–1728.

Tjøstheim, D. (1990). Non-linear time series and Markov chains. *Advances in Applied Probability*, pages 587–611.

Vats, D., Flegal, J. M., and Jones, G. L. (2015a). Multivariate output analysis for Markov chain Monte Carlo. *arXiv preprint arXiv:1512.07713*.

Vats, D., Flegal, J. M., and Jones, G. L. (2015b). Strong consistency of multivariate spectral variance estimators in Markov chain Monte Carlo. *arXiv preprint arXiv:1507.08266*.

Welch, P. D. (1987). On the relationship between batch means, overlapping means and spectral estimation. In *Proceedings of the 19th conference on Winter simulation*, pages 320–323. ACM.

Woodroofe, M. (1970). On choosing a delta-sequence. *The Annals of Mathematical Statistics*, 41(5):1665–1671.

Zeidler, E. (2013). *Nonlinear Functional Analysis and Its Applications: III: Variational Methods and Optimization*. Springer Science & Business Media.