Bernoulli **24**(3), 2018, 1860–1909
https://doi.org/10.3150/16-BEJ914

Strong consistency of multivariate spectral variance estimators in Markov chain Monte Carlo

DOOTIKA VATS[1,*] JAMES M. FLEGAL[2,**] and GALIN L. JONES[1,†]

[1]*School of Statistics, University of Minnesota, Minneapolis, MN 55455, USA.*
E-mail: [*] *vatsx007@umn.edu;* [†] *galin@umn.edu*
[2]*Department of Statistics, University of California, Riverside, Riverside, CA 92521, USA.*
E-mail: [**] *jflegal@ucr.edu*

Markov chain Monte Carlo (MCMC) algorithms are used to estimate features of interest of a distribution. The Monte Carlo error in estimation has an asymptotic normal distribution whose multivariate nature has so far been ignored in the MCMC community. We present a class of multivariate spectral variance estimators for the asymptotic covariance matrix in the Markov chain central limit theorem and provide conditions for strong consistency. We examine the finite sample properties of the multivariate spectral variance estimators and its eigenvalues in the context of a vector autoregressive process of order 1.

Keywords: Markov chain; Monte Carlo; spectral methods; standard errors

1. Introduction

Markov chain Monte Carlo (MCMC) methods are often required for parameter estimation in the statistical models encountered in modern applications. The typical MCMC experiment consists of simulating a Markov chain in order to estimate a vector of quantities, such as moments or quantiles, associated with the target distribution. However, the multivariate nature of the estimation has only rarely been acknowledged in the MCMC literature. We consider the situation where estimation of a vector of means is of interest. Given a multivariate Markov chain central limit theorem (CLT) for the sample mean vector, we show that a class of multivariate spectral variance estimators (MSVEs) are strongly consistent estimators of the covariance matrix in the asymptotic normal distribution. We also establish strong consistency of the eigenvalues of any strongly consistent estimator of the asymptotic covariance matrix.

We know of no other comparable work in the context of MCMC. Kosorok [41] did propose estimators of the asymptotic covariance matrix which generalized work in the univariate case by Geyer [23]. However, these estimators are asymptotically conservative and are based on the properties of reversible Markov chains, an assumption we do not make. There has been a substantial amount of work in the univariate setting. In particular, Atchadé [4] and Flegal and Jones [19] established strong consistency of certain univariate spectral variance estimators, but the multivariate problem is more complicated and requires much new work. Moreover, our work represents a substantial generalization of the univariate results and requires much weaker conditions on the Markov chain. Thus, we also improve the current results in the univariate setting.

1350-7265

We will give a more formal description of the problem studied here. Let F be a probability distribution with support X, equipped with a countably generated σ-field $\mathcal{B}(\mathsf{X})$ and let $g : \mathsf{X} \to \mathbb{R}^p$ be an F-integrable function such that

$$\theta := \mathrm{E}_F g = \int_{\mathsf{X}} g(x)\, dF$$

is the p-dimensional vector of interest. Note that X and θ often have different dimensions. It is common to resort to MCMC methods to estimate θ when it is difficult to obtain θ analytically or to produce independent samples from F. MCMC is popular because it is straightforward to simulate a Harris ergodic (i.e., aperiodic, F-irreducible, and positive Harris recurrent) Markov chain having invariant distribution F (Geyer [24], Liu [43], Robert and Casella [50]). Letting $X = \{X_1, X_2, X_3, \ldots\}$ denote such a Markov chain, estimation is easy since, for any initial distribution, with probability 1,

$$\theta_n := \frac{1}{n} \sum_{t=1}^{n} g(X_t) \to \theta \qquad \text{as } n \to \infty. \tag{1.1}$$

Of course, for any n there will be an unknown *Monte Carlo error* in estimation, $\theta_n - \theta$, and assessment of this Monte Carlo error is critical to the reliability of the simulation results (Flegal *et al.* [18], Flegal and Jones [20], Geyer [23], Jones and Hobert [37]). However, the multivariate nature of the Monte Carlo error has been largely ignored in the MCMC literature (but see Gong and Flegal [26]).

Instead, the primary focus has been on assessing the univariate Monte Carlo error. Let $g^{(i)}$, $\theta_n^{(i)}$, and $\theta^{(i)}$, denote the ith components of g, θ_n, and θ, respectively. Then $\theta_n^{(i)} - \theta^{(i)}$ is the unknown Monte Carlo error of the ith component. The approximate sampling distribution of this error is available via a Markov chain CLT if there exists $0 < \sigma_i^2 < \infty$ such that, as $n \to \infty$,

$$\sqrt{n}\big(\theta_n^{(i)} - \theta^{(i)}\big) \xrightarrow{d} \mathrm{N}\big(0, \sigma_i^2\big). \tag{1.2}$$

(See Jones [35] and Roberts and Rosenthal [52] for a discussion of the conditions for (1.2).) Due to serial correlation in X, $\mathrm{Var}_F g^{(i)} \neq \sigma_i^2$, except in trivial cases. Nevertheless, consistent estimation of σ_i^2 is key to constructing asymptotically valid confidence intervals for $\theta^{(i)}$ and hence in assessing the reliability of the simulation results (Flegal and Gong [17], Flegal *et al.* [18], Glynn and Whitt [25], Jones *et al.* [36], Jones and Hobert [37]). Thus consistent estimation of σ_i^2 has received significant attention; Atchadé [4], Damerdji [10], and Flegal and Jones [19] studied spectral variance estimators, Hobert *et al.* [28] and Mykland *et al.* [46] investigated estimators based on regenerative simulation, and Jones *et al.* [36] studied nonoverlapping batch means. Geyer [23] introduced asymptotically conservative estimators based on the spectral properties of reversible Markov chains. Doss *et al.* [13] considered univariate estimators in the context of estimating quantiles.

In the multivariate setting, the approximate sampling distribution of the Monte Carlo error is available via a Markov chain CLT if there exists a positive definite $p \times p$ matrix Σ such that

$$\sqrt{n}(\theta_n - \theta) \xrightarrow{d} \mathrm{N}_p(0, \Sigma) \qquad \text{as } n \to \infty. \tag{1.3}$$

We consider a class of MSVEs of Σ and provide conditions for strong consistency. Our main assumption on the process is the existence of a multivariate *strong invariance principle* (SIP); that is, we assume that the centered and appropriately scaled partial sum process is similar to a Brownian motion. Specifically, an SIP holds for $\{g(X_t)\}_{t\geq 1}$ if there exists a $p \times p$ lower triangular matrix L and an increasing function ψ on the integers such that, with probability 1,

$$n(\theta_n - \theta) = LB(n) + O\big(\psi(n)\big) \qquad \text{as } n \to \infty,$$

where $B(n)$ denotes a p-dimensional standard Brownian motion and $LL^T = \Sigma$. If ψ is such that $\psi(n)/\sqrt{n} \to 0$ as $n \to \infty$, the SIP implies a strong law, a CLT, and a functional CLT for θ_n. Under moment conditions on g, an SIP with $\psi(n) = n^{1/2-\lambda}$ for some $\lambda > 0$ holds for polynomially ergodic Markov chains.

There has been a substantial amount of work in the context of MCMC on establishing that Markov chains are at least polynomially ergodic. An incomplete list is given by Acosta *et al.* [1], Doss and Hobert [14], Fort and Moulines [21], Hobert and Geyer [27], Jarner and Hansen [30], Jarner and Roberts [31,32], Johnson and Geyer [34], Johnson and Jones [33], Jones *et al.* [38], Marchev and Hobert [44], Petrone *et al.* [47], Roberts and Rosenthal [51], Roberts and Tweedie [53], Rosenthal [54], Roy and Hobert [55], Tan and Hobert [57], Tan *et al.* [58], and Tierney [59]. While establishing that a Markov chain is at least polynomially ergodic can be challenging, it is not the obstacle that it once was.

1.1. Motivating example

As motivation for the use of multivariate methods, we present a simple Bayesian logistic regression model. For $i = 1, \ldots, K$, let Y_i be a binary response variable. For the ith observation, let $X_i = (x_{i1}, x_{i2}, \ldots, x_{i5})$ be the observed vector of predictors, then

$$Y_i | X_i, \beta \overset{\text{ind}}{\sim} \text{Bernoulli}\left(\frac{1}{1 + e^{-X_i\beta}}\right) \quad \text{and} \quad \beta \sim N_5(0, I_5). \tag{1.4}$$

The resulting posterior F is intractable and hence MCMC is used to obtain estimates of the regression coefficient, β. We use the `logit` dataset in the `mcmc` R package which contains four predictors and 100 observations. The goal is to estimate the posterior mean of $\beta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)^T$. Thus, g here is the identity function mapping to $\mathbb{R}^5$.

To sample from the posterior, we use the Polya-Gamma Gibbs sampler of Polson *et al.* [49] (see the `R` package `BayesLogit`) which was shown to be uniformly ergodic by Choi and Hobert [7]. Although the chain mixes fairly quickly as seen in the autocorrelation plot for β_0 in Figure 1, the cross-correlation plot between β_0 and β_2 indicates correlation across these components that is ignored by univariate methods. As a result in Figure 2, the multivariate confidence ellipse is oriented along non-standard axes (see Vats *et al.* [61] for details on how to construct such confidence regions). The ellipse is compared to two univariate confidence boxes; the smaller

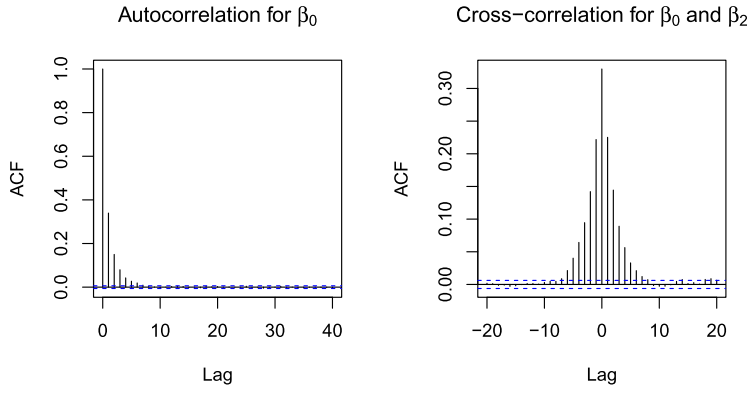


Figure 1. Autocorrelation plot for β_0 and the cross-correlation plot between β_0 and β_2 for a Monte Carlo sample size of $n = 10^5$.

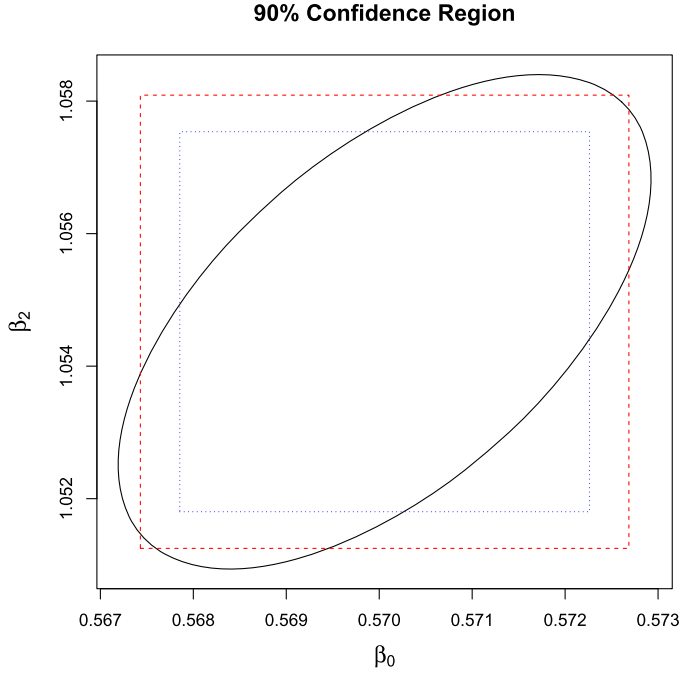


Figure 2. 90% confidence regions constructed using univariate and multivariate methods. The solid ellipse is constructed using an MSVE, the dotted smaller box is constructed using an uncorrected univariate spectral variance estimator and the dashed larger box is constructed using a univariate spectral variance estimator corrected by Bonferroni.

Table 1. Volume to the pth ($p = 5$) root and coverage probabilities for 90% confidence regions constructed using MSVE, uncorrected univariate spectral estimators and Bonferroni corrected univariate spectral estimators. Replications = 1000 and standard errors are indicated in parenthesis

n	MSVE	Bonferroni corrected	Uncorrected
		Volume to the 5th root	
1e3	0.0574 (4.93e–05)	0.0687 (7.02e–05)	0.0483 (4.93e–05)
1e4	0.0189 (7.50e–06)	0.0226 (1.12e–05)	0.0160 (7.90e–06)
1e5	0.0061 (1.10e–06)	0.0073 (1.50e–06)	0.0051 (1.10e–06)
		Coverage probabilities	
1e3	0.853 (0.0112)	0.871 (0.0106)	0.549 (0.0157)
1e4	0.882 (0.0102)	0.904 (0.0093)	0.612 (0.0154)
1e5	0.895 (0.0097)	0.910 (0.0090)	0.602 (0.0155)

uncorrected for multiple testing and the larger corrected for two tests using a Bonferroni correction.

We assess the performance of these confidence regions by comparing their coverage probabilities and volumes over 1000 independent replications for varying Monte Carlo sample sizes. In particular we look at the volume to the pth root ($p = 5$ in this example). The "true" posterior mean is determined by obtaining a Monte Carlo estimate from a sample of length 10^9. Results are presented in Table 1. Note that as the Monte Carlo sample size increases, the multivariate methods produce confidence regions with the nominal coverage probability of 90% with significantly lower volume compared to the Bonferroni corrected regions. The uncorrected regions have far from desirable coverage probabilities.

One reason for the reduction in volume of the ellipsoid is that multivariate methods capture information ignored by univariate analysis. This also leads to a better understanding of the effective samples obtained in an MCMC sample. Vats *et al.* [61] provide the following estimator of effective sample size

$$n\left(\frac{|\widehat{\Lambda}|}{|\widehat{\Sigma}|}\right)^{1/p},$$

where $\widehat{\Lambda}$ is the sample covariance matrix for $g(X_t)$, $\widehat{\Sigma}$ is a strongly consistent estimator of Σ, and $|\cdot|$ denotes determinant. They demonstrate the superiority of this estimator to the univariate estimator of effective sample size of Kass *et al.* [39] and Gong and Flegal [26].

The rest of the paper is organized as follows. In Section 2, we formally define the MSVE and present conditions for strong consistency. We also establish strong consistency of the eigenvalues. Section 3 contains a simulation study where we investigate the finite sample properties of the MSVE in the context of a vector autoregressive process. Finally, we present a discussion in Section 4. Many technical details of the proofs from Section 2 are deferred to the Appendices.

2. Spectral estimators and results

2.1. Definition of MSVE

Let $Y_t = g(X_t) - \theta$, $t = 1, 2, 3, \ldots$ and define the lag s, $s \geq 0$, autocovariance matrix as

$$\gamma(s) = \gamma(-s)^T = \mathrm{E}_F\big[Y_t Y_{t+s}^T\big].$$

Define I_s as $I_s = \{1, \ldots, (n-s)\}$ for $s \geq 0$ and as $I_s = \{(1-s), \ldots, n\}$ for $s < 0$. Let $\bar{Y}_n = n^{-1} \sum_{t=1}^n Y_t$ and define the lag s sample autocovariance as

$$\gamma_n(s) = \frac{1}{n} \sum_{t \in I_s} (Y_t - \bar{Y}_n)(Y_{t+s} - \bar{Y}_n)^T. \tag{2.1}$$

The MSVE is a weighted and truncated sum of the lag s sample autocovariances,

$$\widehat{\Sigma}_S = \sum_{s=-(b_n-1)}^{b_n-1} w_n(s)\gamma_n(s), \tag{2.2}$$

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

2.2. Strong consistency of MSVE

2.2.1. *Strong invariance principle*

While Markov chains are our primary interest, we only require $\{X_t\}_{t \geq 1}$ to be a stochastic process which satisfies a strong invariance principle or SIP. In the interest of clarity, the SIP was stated somewhat loosely in Section 1. What follows is a formal statement of our assumption.

Recall that F is a distribution having support X, $g : \mathsf{X} \to \mathbb{R}^p$, and we are interested in estimating $\theta = \mathrm{E}_F g$. We assume g^2 (where the square is element-wise) is an F-integrable function. Set $h(X_t) = [g(X_t) - \theta]^2$, let $\|\cdot\|$ denote the Euclidean norm, and let $B(t)$ denote a p-dimensional standard Brownian motion.

We will require an SIP for the partial sums of both g and h. We assume there exists a $p \times p$ lower triangular matrix L, an increasing function ψ on the integers, a finite random variable D, and a sufficiently rich probability space such that, with probability 1,

$$\left\| \sum_{t=1}^n g(X_t) - n\theta - LB(n) \right\| < D\psi(n). \tag{2.3}$$

We also assume there exists a finite p-vector θ_h, a $p \times p$ lower triangular matrix L_h, an increasing function ψ_h on the integers, a finite random variable D_h, and a sufficiently rich probability space such that, with probability 1,

$$\left\| \sum_{t=1}^n h(X_t) - n\theta_h - L_h B(n) \right\| < D_h \psi_h(n). \tag{2.4}$$

Remark 1. Strong invariance principles have attracted much research interest and have been shown to hold for a wide variety of processes; see Section 4 for some discussion on this point. Results from Kuelbs and Philipp [42] show that for the Markov chains commonly encountered in MCMC settings, (2.3) and (2.4) hold with $\psi(n) = \psi_h(n) = n^{1/2-\lambda}$ for some $\lambda > 0$. The correlation of the process is measured indirectly by ψ (Philipp and Stout [48]); a large serial correlation implies λ is closer to 0 while for less correlated processes λ is closer to 1/2.

2.2.2. *Strong consistency*

In (2.2), we define the MSVE as the weighted and truncated sum of the lag s sample autocovariances. We make the following assumptions on the lag window $w_n(\cdot)$ and the truncation point b_n.

Condition 1. The lag window $w_n(\cdot)$ is an even function defined on $\mathbb{Z}$ such that

(a) $|w_n(s)| \leq 1$ for all n and s,
(b) $w_n(0) = 1$ for all n, and
(c) $w_n(s) = 0$ for all $|s| \geq b_n$.

Anderson [2] gives a list of lag windows that satisfy Condition 1. We will consider some of these further in Section 2.2.4.

The following Conditions 2 and 3 are technical conditions ensuring that b_n grows at the right rate compared to n.

Condition 2. Let b_n be an integer sequence such that $b_n \to \infty$ and $n/b_n \to \infty$ as $n \to \infty$ where b_n and n/b_n are non-decreasing.

Condition 3. Let b_n be an integer sequence such that

(a) there exists a constant $c \geq 1$ such that $\sum_n (b_n/n)^c < \infty$,
(b) $b_n n^{-1} \log n \to 0$ as $n \to \infty$,
(c) $b_n^{-1} \log n = O(1)$, and
(d) $n > 2b_n$.

If $b_n = \lfloor n^\nu \rfloor$, where $0 < \nu < 1$, then Condition 3 is satisfied if $n > 2^{1/(1-\nu)}$.

Define

$$\Delta_1 w_n(k) = w_n(k-1) - w_n(k)$$

and

$$\Delta_2 w_n(k) = w_n(k-1) - 2w_n(k) + w_n(k+1).$$

Condition 4. Let b_n be an integer sequence, w_n be the lag window, and $\psi(n)$ and $\psi_h(n)$ be positive functions on the integers such that,

(a) $b_n n^{-1} \sum_{k=1}^{b_n} k|\Delta_1 w_n(k)| \to 0$ as $n \to \infty$,
(b) $b_n \psi(n)^2 \log n (\sum_{k=1}^{b_n} |\Delta_2 w_n(k)|)^2 \to 0$ as $n \to \infty$,

(c) $\psi(n)^2 \sum_{k=1}^{b_n} |\Delta_2 w_n(k)| \to 0$ as $n \to \infty$,
(d) $b_n^{-1}\psi_h(n) \to 0$ as $n \to \infty$, and
(e) $b_n^{-1}\psi(n) \to 0$ as $n \to \infty$.

Condition 4(a) connects the truncation point b_n to the lag window w_n. In Section 2.2.4, we will present examples of lag windows that satisfy this condition. The functions $\psi(n)$ and $\psi_h(n)$ in Conditions 4(b), (c), (d), and (e) correspond to the functions described in (2.3) and (2.4) and thus these four conditions connect the truncation point b_n, the lag window w_n, and the correlation of the process, measured indirectly by $\psi(n)$ and $\psi_h(n)$. In Lemma 1, we present sufficient conditions for Conditions 4(a), (b), and (c).

Theorem 1. *Suppose the strong invariance principles* (2.3) *and* (2.4) *hold. If Conditions* 1, 2, 3, *and* 4 *hold, then* $\widehat{\Sigma}_S \to \Sigma$, *with probability* 1, *as* $n \to \infty$.

Outline of proof. The proof is split into several lemmas; see the Appendix for details. Define for $l = 0, \ldots, (n - b_n)$, $\bar{Y}_l(k) = k^{-1} \sum_{t=1}^{k} Y_{l+t}$ and

$$\widehat{\Sigma}_{w,n} = \frac{1}{n} \sum_{l=0}^{n-b_n} \sum_{k=1}^{b_n} k^2 \Delta_2 w_n(k) \big[\bar{Y}_l(k) - \bar{Y}_n\big]\big[\bar{Y}_l(k) - \bar{Y}_n\big]^T.$$

For $t = 1, \ldots, n$, define $Z_t = Y_t - \bar{Y}_n$. Then, in Lemma 4 we show that $\widehat{\Sigma}_{w,n} = \widehat{\Sigma}_S - d_n$, where

$$\begin{aligned} d_n = \frac{1}{n} \Bigg\{ & \sum_{t=1}^{b_n} \Delta_1 w_n(t) \Bigg(\sum_{l=1}^{t-1} Z_l Z_l^T + \sum_{l=n-b_n+t+1}^{n} Z_l Z_l^T \Bigg) \\ & + \sum_{s=1}^{b_n-1} \Bigg[\sum_{t=1}^{b_n-s} \Delta_1 w_n(s+t) \\ & \times \Bigg(\sum_{l=1}^{t-1} \big(Z_l Z_{l+s}^T + Z_{l+s} Z_l^T \big) + \sum_{l=n-b_n+t+1}^{n-s} \big(Z_l Z_{l+s}^T + Z_{l+s} Z_l^T \big) \Bigg) \Bigg] \Bigg\}. \end{aligned} \tag{2.5}$$

Notice that in (2.5) we use the convention that empty sums are zero. In Lemma 9, we show that $d_n \to 0$ as $n \to \infty$ with probability 1. Thus, $\widehat{\Sigma}_{w,n} - \widehat{\Sigma}_S \to 0$, with probability 1, as $n \to \infty$. In Lemma 14, we show that $\widehat{\Sigma}_{w,n} \to \Sigma$, with probability 1, as $n \to \infty$, and the result follows. □

We use Theorem 1 to give conditions for the strong consistency of $\widehat{\Sigma}_S$ when the underlying stochastic process is a Harris ergodic Markov chain having invariant distribution F, but first we need a couple of definitions. Recall that F has support X and $\mathcal{B}(\mathsf{X})$ is a countably generated σ-field. For $n \in \mathbb{N} = \{1, 2, 3, \ldots\}$, let the n-step Markov kernel associated with X starting at $x \in \mathsf{X}$ be $P^n(x, dy)$. Then if $A \in \mathcal{B}(\mathsf{X})$ and $r \in \{1, 2, 3, \ldots\}$, $P^n(x, A) = \Pr(X_{r+n} \in A | X_r = x)$. Let $\|\cdot\|_{\mathrm{TV}}$ denote the total variation norm. The Markov chain is *polynomially ergodic of order* ξ

where $\xi > 0$ if there exists $M : \mathsf{X} \to \mathbb{R}^+$ with $E_F M < \infty$ such that

$$\left\| P^n(x, \cdot) - F(\cdot) \right\|_{\mathrm{TV}} \leq M(x) n^{-\xi}. \tag{2.6}$$

Notice that polynomial ergodicity is weaker than geometric or uniform ergodicity; see Meyn and Tweedie [45].

Remark 2. Polynomial ergodicity is often proved by establishing the following drift condition. For a function $V : \mathsf{X} \to [1, \infty)$ there exists $d > 0$, $b < \infty$, and $0 \leq \tau < 1$ such that for $x \in \mathsf{X}$

$$E\big[V(X_{n+1}) | X_n = x\big] - V(x) \leq -d\big[V(x)\big]^{\tau} + bI(x \in C),$$

where C is a small set. In order to verify that $E_F M < \infty$, it is sufficient to show that $E_F V < \infty$ by Theorem 14.3.7 in Meyn and Tweedie [45].

Theorem 2. *Suppose $E_F \|g\|^{4+\delta} < \infty$ for some $\delta > 0$. Let X be a polynomially ergodic Markov chain of order $\xi \geq (1+\varepsilon)(1+2/\delta)$ for some $\varepsilon > 0$. Then (2.3) and (2.4) hold with*

$$\psi(n) = \psi_h(n) = n^{1/2-\lambda},$$

for some $\lambda > 0$ that depends on p, ε, and δ. If Conditions 1, 2, 3, and 4 hold, then $\widehat{\Sigma}_S \to \Sigma$, with probability 1, as $n \to \infty$.

Proof. See Appendix A.4. □

Remark 3. We rely on results provided by Kuelbs and Philipp [42] to establish the existence of (2.3) and (2.4) in Theorem 2. The precise relationship of λ with p, ε, and δ is not investigated in Kuelbs and Philipp [42] and remains an open problem.

Remark 4. When $p = 1$, the MSV estimator reduces to the spectral variance estimator (SVE) considered by Atchadé [4], Damerdji [10], and Flegal and Jones [19]. However, our result requires weaker conditions. First notice that Flegal and Jones [19] required weaker conditions than Damerdji [10]. Thus we only need to compare Theorem 2 to the results in Atchadé [4] and Flegal and Jones [19], both of whom required the Markov chains to be geometrically ergodic and to satisfy a one-step minorization condition. Thus Theorem 2 substantially weakens the conditions on the underlying Markov chain, while extending the results to the $p \geq 1$ setting.

2.2.3. *Strong consistency of eigenvalues*

Having obtained a strongly consistent estimator of Σ, it is natural to consider the eigenvalues of the estimator.

Theorem 3. *Let $\widehat{\Sigma}$ be any strongly consistent estimator of Σ and let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p > 0$ be the eigenvalues of Σ. Let $\hat{\lambda}_1, \ldots, \hat{\lambda}_p$ be the p eigenvalues of $\widehat{\Sigma}$ such that $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_p$, then $\hat{\lambda}_k \to \lambda_k$, with probability 1, as $n \to \infty$ for all $1 \leq k \leq p$.*

Proof. Let $\|\cdot\|_F$ denote the Frobenius norm. By Weyl's inequality (Franklin [22]), for $\varepsilon > 0$, if $\|\widehat{\Sigma} - \Sigma\|_F \le \varepsilon$, then for all $1 \le k \le p$, $|\hat{\lambda}_k - \lambda_k| \le \varepsilon$, which gives the desired result. □

Remark 5. Theorem 3 immediately implies that under the conditions of either Theorem 1 or Theorem 2 the sample eigenvalues of the MSVE are consistent for the population eigenvalues. That is, $\hat{\lambda}_k \to \lambda_k$, with probability 1, as $n \to \infty$ for all $1 \le k \le p$.

Sample eigenvalues can play an important role in multivariate analyses. For example, the length of any axis of the confidence region constructed from $\widehat{\Sigma}_S$ is determined by the magnitude of the relevant estimated sample eigenvalue. Thus the largest eigenvalue is associated with the axis having the largest estimated Monte Carlo error. This also suggests that dimension reduction methods could be useful in assessing the reliability of the simulation effort. Although this is a potentially interesting research direction it is beyond the scope of this paper.

2.2.4. *Lag window conditions*

The following generalization of Lemma 7 in Flegal and Jones [19] is useful for checking that a lag window satisfies the conditions of Theorem 1.

Lemma 1. *Reparameterize w_n such that w_n is defined on* $[0, 1]$ *and* $w_n(0) = 1$ *and* $w_n(1) = 0$. *Further assume that w_n is twice continuously differentiable and that there exists finite constants D_1 and D_2 such that $|w_n'(x)| \le D_1$ and $|w_n''(x)| < D_2$. Then as $n \to \infty$,*

1. *Condition* 4(a) *holds if* $b_n^2 n^{-1} \to 0$,
2. *Conditions* 4(b) *and* (c) *holds if* $b_n^{-1}\psi(n)^2 \log n \to 0$.

Proof. The argument is the same as that of Lemma 7 in Flegal and Jones [19] and hence is omitted. □

Remark 6. It is common to use $b_n = \lfloor n^\nu \rfloor$ in which case Conditions 4(a), (b), and (c) hold, if we choose $0 < \nu < 1/2$ such that $n^{-\nu}\psi(n)^2 \log n \to 0$ as $n \to \infty$.

Remark 7. We now consider some examples of lag windows which satisfy Condition 1 and consider whether Conditions 4(a), (b), and (c) hold.

1. *Simple truncation*: $w_n(k) = I(|k| < b_n)$. Using this window the estimator obtained is truncated at b_n but weighted identically. In this case, $\Delta_2 w_n(k) = 0$ for $k = 1, \ldots, b_n - 2$, $\Delta_2 w_n(b_n - 1) = -1$ and $\Delta_2 w_n(b_n) = 1$. It is easy to see that Condition 4(c) is not satisfied.
2. *Blackman–Tukey*: $w_n(k) = [1 - 2a + 2a\cos(\pi|k|/b_n)]I(|k| < b_n)$ where $a > 0$. This is a generalization for the *Tukey–Hanning* window where $a = 1/4$. For fixed a, the Blackman–Tukey window satisfies the conditions of Lemma 1, thus Conditions 4(a), (b), and (c) hold if $b_n^2 n^{-1} \to 0$ and $b_n^{-1}\psi(n)^2 \log n \to 0$ as $n \to \infty$.
3. *Parzen*: $w_n(k) = [1 - |k|^q / b_n^q]I(|k| < b_n)$ for $q \in \mathbb{Z}^+$. When $q = 1$ this is the *modified Bartlett* window. It is easy to show that the Parzen window satisfies the conditions for

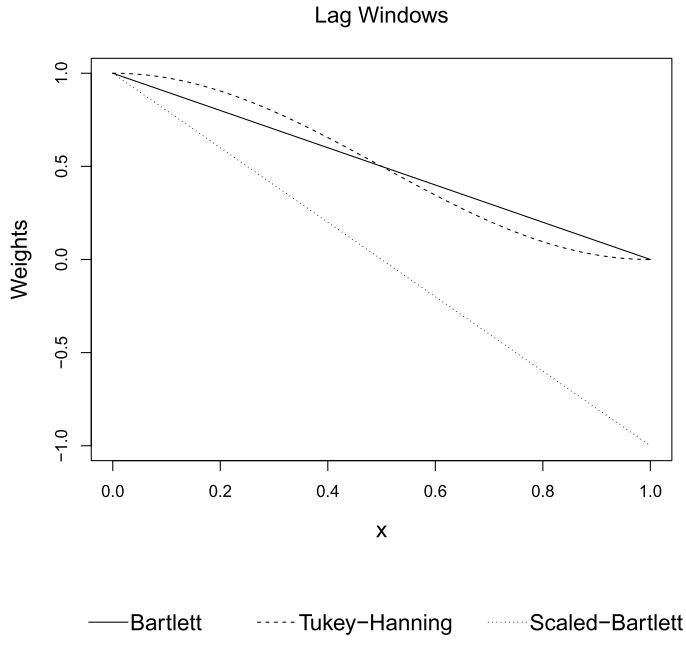


Figure 3. Plot of three lag windows, modified Bartlett (Bartlett), Tukey–Hanning and the scale-parameter Bartlett with scale parameter 2 (Scaled-Bartlett).

Lemma 1, and thus Conditions 4(a), (b), and (c) hold if $b_n^2 n^{-1} \to 0$ and $b_n^{-1}\psi(n)^2 \log n \to 0$ as $n \to \infty$.

4. *Scale-parameter modified Bartlett*: $w_n(k) = [1 - \eta|k|/b_n]I(|k| < b_n)$ where η is a positive constant not equal to 1. Then $\Delta_1 w_n(k) = \eta b_n^{-1}$ for $k = 1, 2, \ldots, b_n - 1$ and $\Delta_1 w_n(b_n) = 1 - \eta + \eta b_n^{-1}$ so that Condition 4(a) is satisfied when $b_n^2 n^{-1} \to 0$ as $n \to \infty$. Also, $\Delta_2 w_n(k) = 0$ for $k = 1, 2, \ldots, b_n - 2$, $\Delta_2 w_n(b_n - 1) = \eta - 1$ and $\Delta_2 w_n(b_n) = 1 - \eta + \eta b_n^{-1}$. We conclude that $\sum_{k=1}^{b_n} |\Delta_2 w_n(k)|$ does not converge to 0 and hence Condition 4(c) is not satisfied.

Figure 3 provides a graph of the three lag windows we consider in the next section, specifically, the modified Bartlett, Tukey–Hanning, and scale-parameter modified Bartlett windows. It is evident that the modified Bartlett and Tukey–Hanning windows are similar and the scale-parameter modified Bartlett window weighs the lags more severely.

3. Simulation

We consider some finite sample properties of the MSVE in the context of a vector autoregressive process of order 1 or VAR(1). Let

$$y_t = \Phi y_{t-1} + \varepsilon_t, \tag{3.1}$$

Table 2. Simulation settings 1 through 6

Setting	p	Eigenvalues of Φ for $i = 0, \ldots, p-1$
1	10	$\lambda_i = 0.01 + i(0.20 - 0.01)/(p-1)$
2	10	$\lambda_i = 0.40 + i(0.60 - 0.40)/(p-1)$
3	10	$\lambda_i = 0.70 + i(0.90 - 0.70)/(p-1)$
4	50	$\lambda_i = 0.01 + i(0.20 - 0.01)/(p-1)$
5	50	$\lambda_i = 0.40 + i(0.60 - 0.40)/(p-1)$
6	50	$\lambda_i = 0.70 + i(0.90 - 0.70)/(p-1)$

where $y_t \in \mathbb{R}^p$ for all t, Φ is a $p \times p$ matrix, $\varepsilon_t \overset{\text{i.i.d.}}{\sim} N_p(0, W)$, and y_0 is the zero vector. While this is a simple model, it is useful to study since we can control the correlation of the process.

We assume that the largest eigenvalue of Φ, $\phi_{\max}$, is less than 1 in absolute value, in which case the stationary distribution for the process is $F = N_p(0, V)$ where $\operatorname{vec}(V) = (I_{p^2} - \Phi \otimes \Phi)^{-1} \operatorname{vec}(W)$. Here $\otimes$ denotes Kronecker product and I_{p^2} is the $p^2 \times p^2$ identity matrix. With some algebra it can be shown that the lag s autocovariance matrix for $s > 0$ is

$$\gamma(s) = \Phi^s V \quad \text{and} \quad \gamma(-s) = V\bigl(\Phi^T\bigr)^s.$$

Consider estimating $\mathrm{E}_F y$ with $\bar{y}_n$, the Monte Carlo estimate. Tjøstheim [60] showed that the process is geometrically ergodic as long as $|\phi_{\max}| < 1$. In fact, the smaller the largest eigenvalue, the faster the process mixes. Since F has a moment generating function, a CLT holds with

$$\begin{aligned}
\Sigma &= \sum_{s=-\infty}^{\infty} \gamma(s) \\
&= \sum_{s=0}^{\infty} \gamma(s) + \sum_{s=-\infty}^{0} \gamma(s) - V \\
&= \sum_{s=0}^{\infty} \Phi^s V + \sum_{s=-\infty}^{0} V\bigl(\Phi^T\bigr)^s - V \\
&= (1 - \Phi)^{-1} V + V\bigl(1 - \Phi^T\bigr)^{-1} - V.
\end{aligned} \tag{3.2}$$

For this process, we investigate the performance of the class of MSVE in estimating Σ. We set W to be the first order autoregressive covariance matrix with correlation $\rho = 0.5$ and present simulation results for different settings of Φ and p. These settings are presented in Table 2. For Settings 1 and 4, $\phi_{\max} = 0.2$, Settings 2 and 5, $\phi_{\max} = 0.6$ and Settings 3 and 6, $\phi_{\max} = 0.9$. Thus, these three pairs of settings yield processes with different mixing rates.

We compare the performance of three lag windows: modified Bartlett, Tukey–Hanning, and scale-parameter modified Bartlett with scale $= 2$. In Section 2, we showed that the modified Bartlett and the Tukey–Hanning windows satisfy the conditions of Theorem 1 while the scale-parameter modified Bartlett does not.

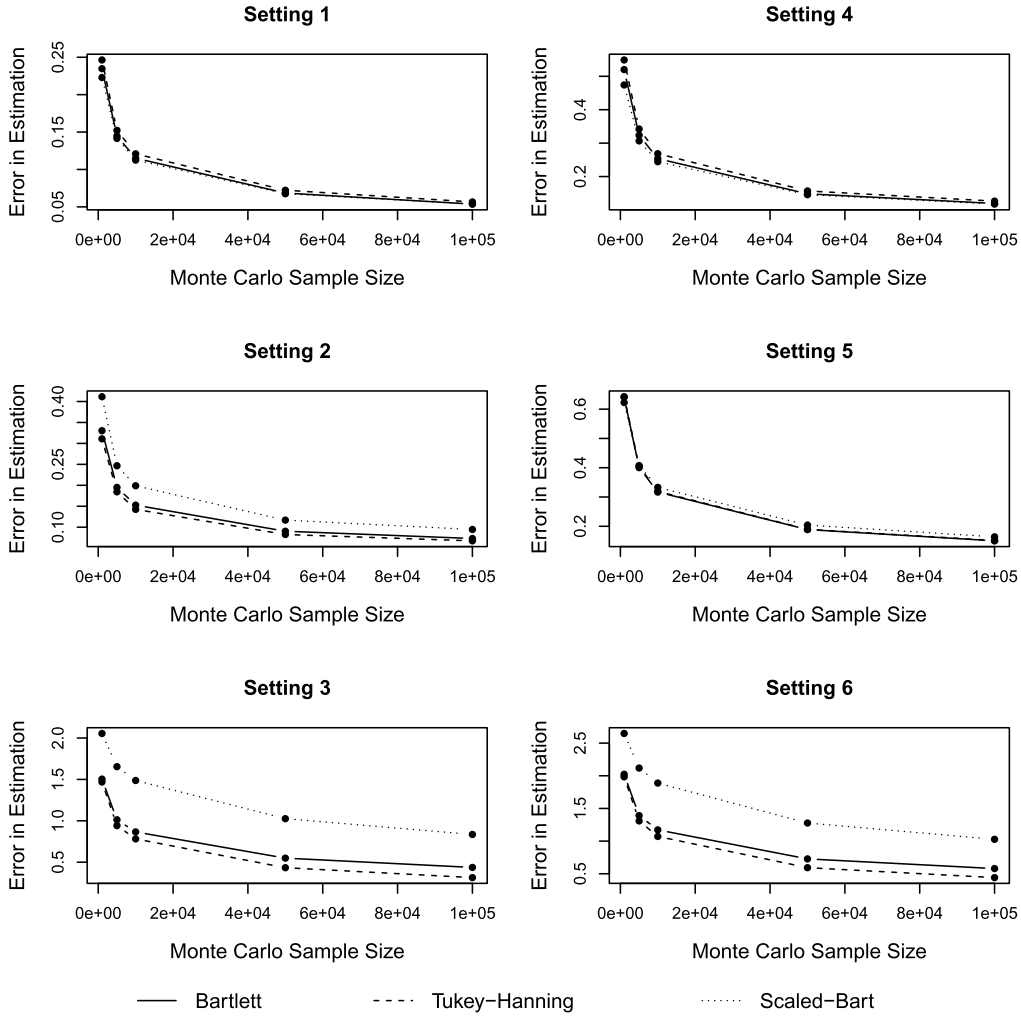


Figure 4. $\|\widehat{\Sigma}_S - \Sigma\|_F / \|\Sigma\|_F$ for the three lag windows at different Monte Carlo sample sizes for all six settings averaged over 100 iterations.

For each setting, we do the following in each of 100 independent replications. We observe the process for a Monte Carlo sample size of $1e5$, and calculate the three MSVEs at samples {1e3, 5e3, 1e4, 5e4, 1e5} with $b_n = \lfloor n^{1/3} \rfloor$. The error in estimation is determined by calculating the average relative difference in Frobenius norm, i.e. $\|\widehat{\Sigma} - \Sigma\|_F / \|\Sigma\|_F$ for each of the three windows at all five Monte Carlo sample sizes.

In Figure 4, we plot the results for all settings for all three lag windows. For Settings 1 and 4, all three lag windows perform equally well while for Settings 3 and 6, the scale parameter modified Bartlett window performs poorly. In all settings, the modified Bartlett and the Tukey–Hanning

windows perform similarly, but the Tukey–Hanning window is slightly better when the chain mixes more slowly. The plots also indicate that as $\phi_{\max}$ increases, a larger Monte Carlo sample size is required for a desired error in estimation threshold. This is as expected since we know for higher values of $\phi_{\max}$, the process mixes more slowly.

In Section 2, we presented the proof for the convergence of the eigenvalues of the MSVE in Remark 5. To study the finite sample properties of the maximum eigenvalue, we observe its behavior for the three different lag windows at different Monte Carlo sample sizes over each of 100 independent replications. At each replication, we observe the relative error in estimation, $|\hat{\lambda}_1 - \lambda_1|/\lambda_1$. The results are presented in Figure 5 and are similar to what was observed for the convergence of the MSVEs. For Settings 2, 3, 5 and 6, the scale-parameter modified Bartlett window performs significantly worse than the Tukey–Hanning and the modified Bartlett windows. When the chain mixes more slowly, the Tukey–Hanning window appears to give slightly better results.

It is natural to investigate the stability of estimation of the largest eigenvalue. We study this empirically for Setting 1 by observing the shape of the distribution of the maximum eigenvalue for the estimates of Σ obtained through the three lag windows at varying Monte Carlo sample sizes over the 100 independent replications. Using (3.2), the true maximum eigenvalue for this setting is 2.683. In Figure 6, we notice that as the Monte Carlo sample size increases, the shape of the density of the largest eigenvalue is increasingly symmetric and centered at this true value. In addition, as the Monte Carlo sample size increases, the variance of the largest estimated eigenvalue decreases. This is observed for all three lag windows.

4. Discussion

Estimation of the asymptotic covariance matrix in the CLT as in (1.3) has received little attention in the MCMC literature thus far. Due to the results of this paper, practitioners are now equipped with a class of strongly consistent multivariate spectral variance estimators of Σ.

However, multivariate spectral variance estimators are also encountered outside of the MCMC context. For example, they are often used for heteroscedastic and autocorrelation consistent (HAC) estimation of covariance matrices which, for example, arise in the study of generalized method of moments and autoregressive processes with heteroscedastic errors. See Andrews [3] for motivating examples. In the context of HAC estimation, DeJong [11] obtained conditions under which the class of MSVEs are strongly consistent. However, these conditions are restrictive in the context of MCMC. In particular, his Assumption 2 (DeJong [11], page 264) will not be satisfied in many typical MCMC applications. Additionally, we require weaker mixing conditions on the underlying stochastic process. That is, although Markov chains are the primary focus for us, our results hold for much more general stochastic processes as we explain below.

Our main assumption on the underlying stochastic process are the SIPs as stated in (2.3) and (2.4). The existence of an SIP has attracted much research interest. Consider the univariate case. For independent and identically distributed (i.i.d.) processes, the first result of this kind is due to Strassen [56] who showed $\psi(n) = \sqrt{n \log\log n}$. Komlós *et al.* [40] found that

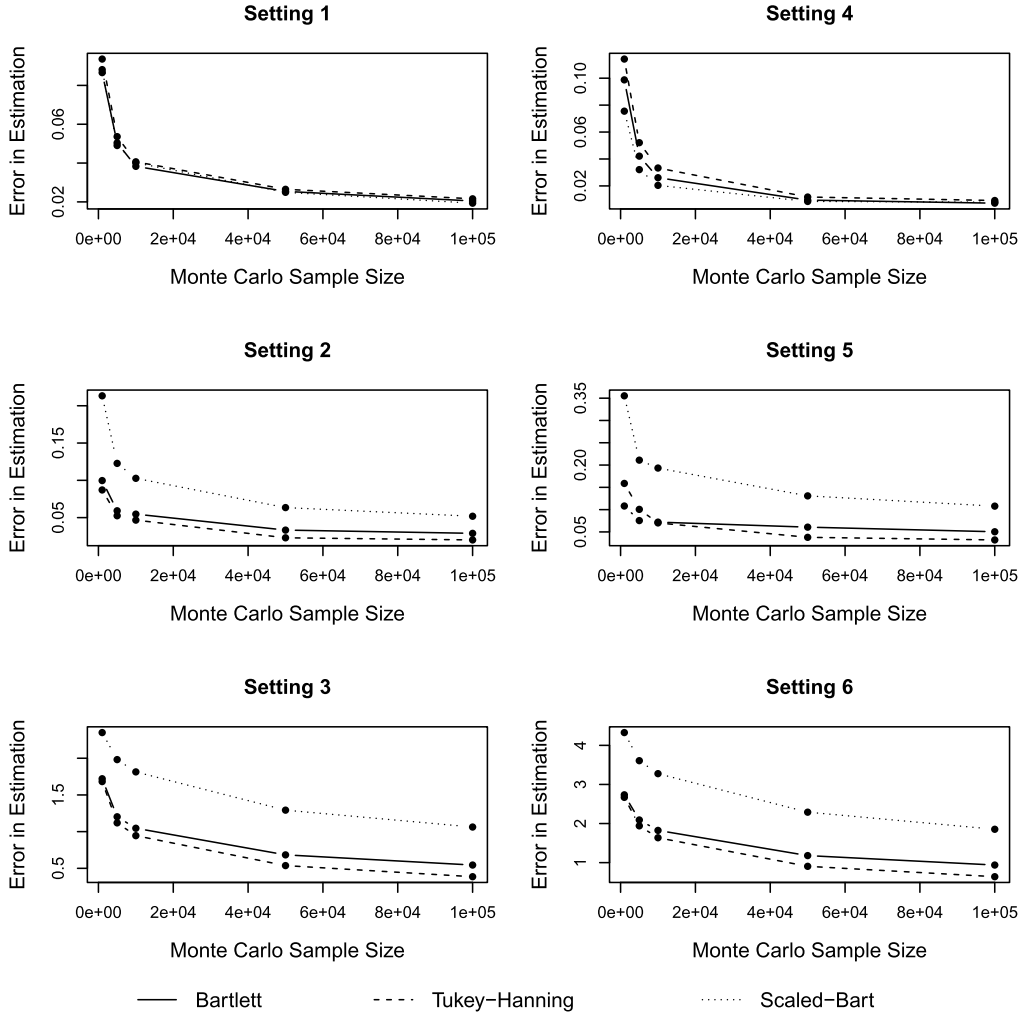
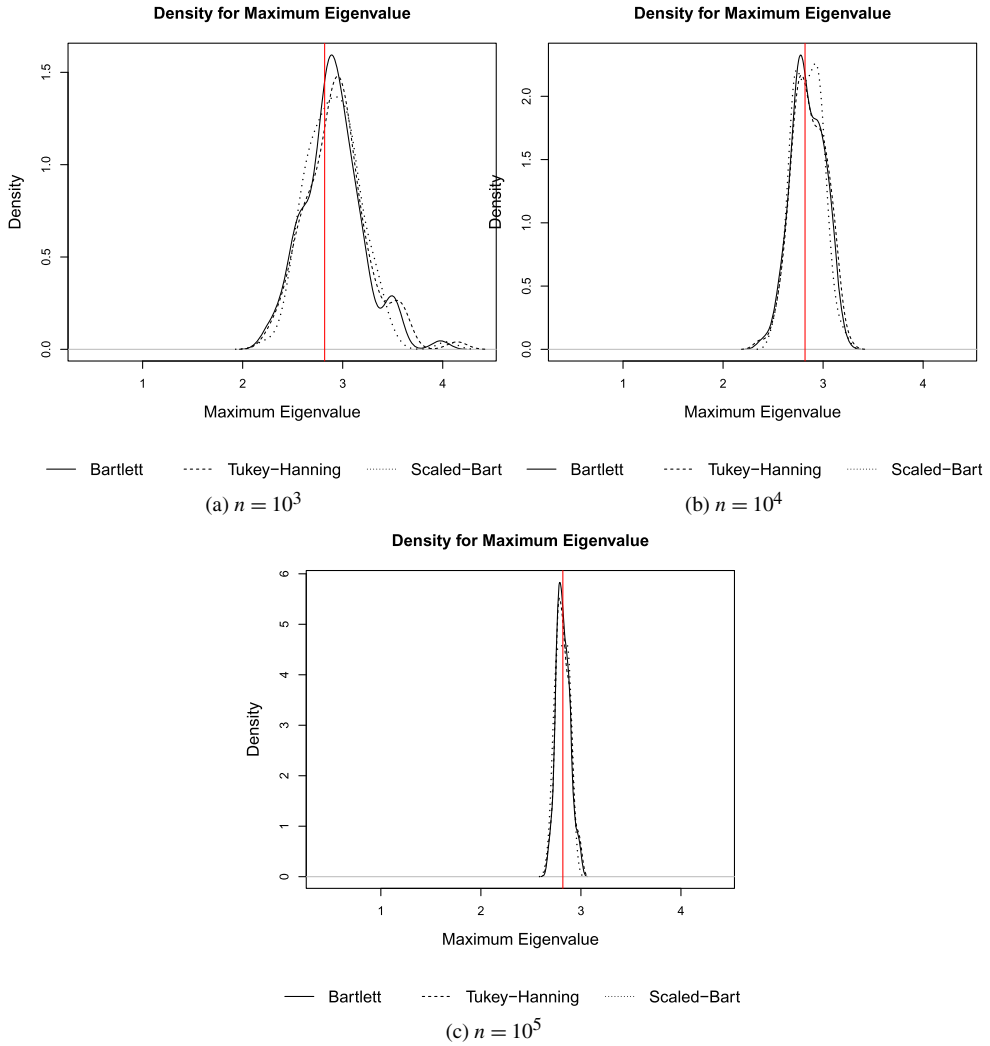


Figure 5. $|\hat{\lambda}_1 - \lambda_1|/\lambda_1$ for the three lag windows at different Monte Carlo sample sizes for all six settings averaged over 100 iterations.

if $\mathrm{E}_F|g|^{2+\delta} < \infty$, then $\psi(n) = n^{1/2-\lambda}$ for $\lambda > 0$ (often called the KMT bound). Komlós *et al.* [40] also showed that if g has all moments in a neighborhood of 0, then $\psi(n) = \log n$. The results of Komlós *et al.* [40] are the strongest to date in the i.i.d. setting. The main reference for a univariate strong invariance principle for dependent sequences is Philipp and Stout [48] who prove bounds similar to that of Komlós *et al.* [40] for a variety of weakly dependent processes including ϕ-mixing, regenerative and strongly mixing processes. Also, see Wu [63] for a univariate strong invariance principle for certain classes of dependent processes.



(a) $n = 10^3$ (b) $n = 10^4$

(c) $n = 10^5$

Figure 6. Kernel density of the maximum eigenvalue for the three MSVEs over 100 replications and increasing Monte Carlo sample sizes for Setting 1. The vertical line indicates the true eigenvalue of 2.683 calculated using (3.2).

Many of the univariate SIPs have been extended to the multivariate setting. For independent processes, Berkes and Philipp [5], Einmahl [16], and Zaitsev [64] extend the results of Komlós *et al.* [40]. For correlated processes, Eberlein [15] showed the existence of a strong invariance principle for Martingale sequences and Horvath [29] proved the KMT bound for multivariate extended renewal processes. For ϕ-mixing, strongly mixing, and absolutely regular processes, Kuelbs and Philipp [42] and Dehling and Philipp [12] extended the Philipp and Stout [48] results to the multivariate case.

Appendix: Strong consistency of MSVE

Before we begin the proof of Theorem 1 we note some useful properties of Brownian motion and lag windows which will be used often throughout the proof.

A.1. Brownian motion

Recall that $\{B(t)\}_{t\geq 0}$ denotes a p-dimensional standard Brownian motion and that $B^{(i)}$ denotes the ith component of $B(t)$.

Lemma 2 (Csörgő and Révész [9]). *Suppose Condition 2 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 $i = 1, \ldots, p$*

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

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

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

Let L be a lower triangular matrix and set $\Sigma = LL^T$. Define $C(t) := LB(t)$ and if $C^{(i)}(t)$ is the ith component of $C(t)$, define

$$\bar{C}_l^{(i)}(k) := \frac{1}{k}\left(C^{(i)}(l+k) - C^{(i)}(l)\right) \quad \text{and} \quad \bar{C}_n^{(i)} := \frac{1}{n}C^{(i)}(n).$$

Since $C^{(i)}(t) \sim N(0, t\Sigma_{ii})$, where Σ_{ii} is the ith diagonal of Σ, $C^{(i)}/\sqrt{\Sigma_{ii}}$ is a 1-dimensional standard Brownian motion. As a consequence, we have the following corollaries of Lemma 2.

Corollary 1. *Suppose Condition 2 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 $i = 1, \ldots, p$*

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

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

Corollary 2. *Suppose Condition 2 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 $i = 1, \ldots, p$*

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

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

A.2. Basic properties of lag windows

Recall that the lag window $w_n(\cdot)$ is such that it satisfies Condition 1. We will require the following results about the lag window $w_n(\cdot)$.

Lemma 3 (Damerdji [10]). *Under Condition 1,*

(i) $\Delta_1 w_n(s) = \sum_{k=s}^{b_n} \Delta_2 w_n(k)$,
(ii) $\sum_{k=s+1}^{b_n} \Delta_1 w_n(k) = w_n(s)$, *and*
(iii) $\sum_{k=1}^{b_n} \Delta_1 w_n(k) = 1$.

A.3. Proof of Theorem 1

Recall that

$$\widehat{\Sigma}_{w,n} = \frac{1}{n} \sum_{l=0}^{n-b_n} \sum_{k=1}^{b_n} k^2 \Delta_2 w_n(k) \big[\bar{Y}_l(k) - \bar{Y}_n\big]\big[\bar{Y}_l(k) - \bar{Y}_n\big]^T. \tag{A.3}$$

For $t = 1, 2, \ldots, n$, define $Z_t = Y_t - \bar{Y}_n$ and

$$\begin{aligned} d_n = \frac{1}{n} \Bigg\{ & \sum_{t=1}^{b_n} \Delta_1 w_n(t) \Bigg(\sum_{l=1}^{t-1} Z_l Z_l^T + \sum_{l=n-b_n+t+1}^{n} Z_l Z_l^T \Bigg) \\ & + \sum_{s=1}^{b_n-1} \Bigg[\sum_{t=1}^{b_n-s} \Delta_1 w_n(s+t) \\ & \times \Bigg(\sum_{l=1}^{t-1} \big(Z_l Z_{l+s}^T + Z_{l+s} Z_l^T \big) + \sum_{l=n-b_n+t+1}^{n-s} \big(Z_l Z_{l+s}^T + Z_{l+s} Z_l^T \big) \Bigg) \Bigg] \Bigg\}. \end{aligned} \tag{A.4}$$

Notice that in (A.4) we use the convention that empty sums are zero.

Lemma 4.
Under Condition 1, $\widehat{\Sigma}_{w,n} = \widehat{\Sigma}_S - d_n$.

Proof. For $i, j = 1, \ldots, p$, let $\widehat{\Sigma}_{w,ij}$ denote the (i, j)th entry of $\widehat{\Sigma}_{w,n}$. Then,

$$\begin{aligned} \widehat{\Sigma}_{w,ij} &= \frac{1}{n} \sum_{l=0}^{n-b_n} \sum_{k=1}^{b_n} k^2 \Delta_2 w_n(k) \big[\bar{Y}_l^{(i)}(k) - \bar{Y}_n^{(i)}\big]\big[\bar{Y}_l^{(j)}(k) - \bar{Y}_n^{(j)}\big] \\ &= \frac{1}{n} \sum_{l=0}^{n-b_n} \sum_{k=1}^{b_n} \Delta_2 w_n(k) \Bigg[\sum_{t=1}^{k} Y_{l+t}^{(i)} - k\bar{Y}_n^{(i)} \Bigg] \Bigg[\sum_{t=1}^{k} Y_{l+t}^{(j)} - k\bar{Y}_n^{(j)} \Bigg] \end{aligned} \tag{A.5}$$

$$= \frac{1}{n}\sum_{l=0}^{n-b_n}\sum_{k=1}^{b_n}\Delta_2 w_n(k)\left[\sum_{t=1}^{k} Z_{l+t}^{(i)}\right]\left[\sum_{t=1}^{k} Z_{l+t}^{(j)}\right]$$

$$= \frac{1}{n}\sum_{l=0}^{n-b_n}\sum_{k=1}^{b_n}\Delta_2 w_n(k)\left[\sum_{t=1}^{k} Z_{l+t}^{(i)} Z_{l+t}^{(j)} + \sum_{s=1}^{k-1}\sum_{t=1}^{k-s} Z_{l+t}^{(i)} Z_{l+t+s}^{(j)} + \sum_{s=1}^{k-1}\sum_{t=1}^{k-s} Z_{l+t}^{(j)} Z_{l+t+s}^{(i)}\right].$$

Notice that in (A.5), we use the convention that empty sums are zero. We will consider each term in (A.5) separately. For the first term, changing the order of summation and then using Lemma 3,

$$\begin{aligned}
&\frac{1}{n}\sum_{l=0}^{n-b_n}\sum_{k=1}^{b_n}\sum_{t=1}^{k}\Delta_2 w_n(k) Z_{l+t}^{(i)} Z_{l+t}^{(j)} \\
&\quad= \frac{1}{n}\sum_{l=0}^{n-b_n}\sum_{t=1}^{b_n}\sum_{k=t}^{b_n}\Delta_2 w_n(k) Z_{l+t}^{(i)} Z_{l+t}^{(j)} \\
&\quad= \frac{1}{n}\sum_{l=0}^{n-b_n}\sum_{t=1}^{b_n}\Delta_1 w_n(t) Z_{l+t}^{(i)} Z_{l+t}^{(j)} \\
&\quad= \frac{1}{n}\sum_{t=1}^{b_n}\Delta_1 w_n(t)\sum_{l=0}^{n-b_n} Z_{l+t}^{(i)} Z_{l+t}^{(j)} \qquad \text{(A.6)} \\
&\quad= \sum_{t=1}^{b_n}\Delta_1 w_n(t)\Big[\gamma_{n,ij}(0) - \frac{1}{n}\big(Z_1^{(i)} Z_1^{(j)} + \cdots + Z_{t-1}^{(i)} Z_{t-1}^{(j)} \\
&\qquad + Z_{n-b_n+t+1}^{(i)} Z_{n-b_n+t+1}^{(j)} + \cdots + Z_n^{(i)} Z_n^{(j)}\big)\Big] \\
&\quad= \gamma_{n,ij}(0) - \frac{1}{n}\sum_{t=1}^{b_n}\Delta_1 w_n(t)\left(\sum_{l=1}^{t-1} Z_l^{(i)} Z_l^{(j)} + \sum_{l=n-b_n+t+1}^{n} Z_l^{(i)} Z_l^{(j)}\right) \qquad \text{by Lemma 3.}
\end{aligned}$$

For the second term in (A.5), we change the order of summation from l, k, s, t to l, s, k, t to l, s, t, k to get

$$\begin{aligned}
&\frac{1}{n}\sum_{l=0}^{n-b_n}\sum_{k=1}^{b_n}\sum_{s=1}^{k-1}\sum_{t=1}^{k-s}\Delta_2 w_n(k) Z_{l+t}^{(i)} Z_{l+t+s}^{(j)} \\
&\quad= \frac{1}{n}\sum_{l=0}^{n-b_n}\sum_{s=1}^{b_n-1}\sum_{k=s+1}^{b_n}\sum_{t=1}^{k-s}\Delta_2 w_n(k) Z_{l+t}^{(i)} Z_{l+t+s}^{(j)} \\
&\quad= \frac{1}{n}\sum_{l=0}^{n-b_n}\sum_{s=1}^{b_n-1}\sum_{t=1}^{b_n-s}\sum_{k=t+s}^{b_n}\Delta_2 w_n(k) Z_{l+t}^{(i)} Z_{l+t+s}^{(j)}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{n}\sum_{l=0}^{n-b_n}\sum_{s=1}^{b_n-1}\sum_{t=1}^{b_n-s}\Delta_1 w_n(s+t)Z_{l+t}^{(i)}Z_{l+t+s}^{(j)} \qquad \text{by Lemma 3}\\
&= \frac{1}{n}\sum_{s=1}^{b_n-1}\sum_{l=0}^{n-b_n}\sum_{t=1}^{b_n-s}\Delta_1 w_n(s+t)Z_{l+t}^{(i)}Z_{l+t+s}^{(j)}\\
&= \frac{1}{n}\sum_{s=1}^{b_n-1}\sum_{t=1}^{b_n-s}\sum_{l=0}^{n-b_n}\Delta_1 w_n(s+t)Z_{l+t}^{(i)}Z_{l+t+s}^{(j)}\\
&= \frac{1}{n}\sum_{s=1}^{b_n-1}\sum_{t=1}^{b_n-s}\Delta_1 w_n(s+t)\sum_{l=0}^{n-b_n}Z_{l+t}^{(i)}Z_{l+t+s}^{(j)}\\
&= \sum_{s=1}^{b_n-1}\sum_{t=1}^{b_n-s}\Delta_1 w_n(s+t)\left[\gamma_{n,ij}(s)-\frac{1}{n}\sum_{l=1}^{t-1}Z_l^{(i)}Z_{l+s}^{(j)}-\frac{1}{n}\sum_{l=n-b_n+t+1}^{n-s}Z_l^{(i)}Z_{l+s}^{(j)}\right]\\
&= \sum_{s=1}^{b_n-1}w_n(s)\gamma_{n,ij}(s)\\
&\quad -\frac{1}{n}\sum_{s=1}^{b_n-1}\sum_{t=1}^{b_n-s}\Delta_1 w_n(s+t)\left[\sum_{l=1}^{t-1}Z_l^{(i)}Z_{l+s}^{(j)}+\sum_{l=n-b_n+t+1}^{n-s}Z_l^{(i)}Z_{l+s}^{(j)}\right] \qquad \text{by Lemma 3.}
\end{aligned}
\tag{A.7}
$$

Repeating the same steps as in the second term, we reduce the third term in (A.5) to

$$
\begin{aligned}
&\frac{1}{n}\sum_{l=0}^{n-b_n}\sum_{k=1}^{b_n}\sum_{s=1}^{k-1}\sum_{t=1}^{k-s}\Delta_2 w_n(k)Z_{l+t}^{(j)}Z_{l+t+s}^{(i)}\\
&\quad = \sum_{s=1}^{b_n-1}w_n(s)\gamma_{n,ji}(s)\\
&\qquad -\frac{1}{n}\sum_{s=1}^{b_n-1}\sum_{t=1}^{b_n-s}\Delta_1 w_n(s+t)\left[\sum_{l=1}^{t-1}Z_l^{(j)}Z_{l+s}^{(i)}+\sum_{l=n-b_n+t+1}^{n-s}Z_l^{(j)}Z_{l+s}^{(i)}\right]\\
&\quad = \sum_{s=1}^{b_n-1}w_n(-s)\gamma_{n,ij}(-s)\\
&\qquad -\frac{1}{n}\sum_{s=1}^{b_n-1}\sum_{t=1}^{b_n-s}\Delta_1 w_n(s+t)\left[\sum_{l=1}^{t-1}Z_l^{(j)}Z_{l+s}^{(i)}+\sum_{l=n-b_n+t+1}^{n-s}Z_l^{(j)}Z_{l+s}^{(i)}\right].
\end{aligned}
\tag{A.8}
$$

Using (A.6), (A.7), and (A.8) in (A.5)

$$\begin{aligned}\widehat{\Sigma}_{w,ij} &= \gamma_{n,ij}(0) + \sum_{s=1}^{b_n-1} w_n(s)\gamma_{n,ij}(s) + \sum_{s=-(b_n-1)}^{-1} w_n(s)\gamma_{n,ij}(s) \\ &\quad - \frac{1}{n}\sum_{t=1}^{b_n} \Delta_1 w_n(t)\left(\sum_{l=1}^{t-1} Z_l^{(i)} Z_l^{(j)} + \sum_{l=n-b_n+t+1}^{n} Z_l^{(i)} Z_l^{(j)}\right) \\ &\quad - \frac{1}{n}\sum_{s=1}^{b_n-1}\sum_{t=1}^{b_n-s} \Delta_1 w_n(s+t) \\ &\quad \times \left[\sum_{l=1}^{t-1}\left(Z_l^{(i)} Z_{l+s}^{(j)} + Z_{l+s}^{(i)} Z_l^{(j)}\right) + \sum_{l=n-b_n+t+1}^{n-s}\left(Z_l^{(i)} Z_{l+s}^{(j)} + Z_{l+s}^{(i)} Z_l^{(j)}\right)\right] \\ &= \sum_{s=-(b_n-1)}^{b_n-1} \gamma_{n,ij}(s) w_n(s) - d_{n,ij} = \widehat{\Sigma}_{S,ij} - d_{n,ij}.\end{aligned}$$

□

Let $\widetilde{\gamma}_n(s)$, $\widetilde{\Sigma}_S$, $\widetilde{\Sigma}_{w,n}$ and $\tilde{d}_n$ be the Brownian motion analogs of (2.1), (2.2), (A.3), and (A.4). Specifically, for $t = 1, \ldots, n$, define Brownian motion increments $U_t = B(t) - B(t-1)$, so that $U_1, \ldots, U_n$ are $\overset{\text{i.i.d.}}{\sim} \mathrm{N}_p(0, I_p)$ where I_p is the $p \times p$ identity matrix. For $l = 0, \ldots, n - b_n$ and $k = 1, \ldots, b_n$ define $\bar{B}_l(k) = k^{-1}(B(l+k) - B(l))$, $\bar{B}_n = n^{-1}B(n)$, and $T_t = U_t - \bar{B}_n$. Then

$$\widetilde{\gamma}_n(s) = \frac{1}{n}\sum_{t\in I_s}(U_t - \bar{B}_n)(U_{t+s} - \bar{B}_n)^T = \frac{1}{n}\sum_{t\in I_s} T_t T_{t+s}^T, \tag{A.9}$$

$$\widetilde{\Sigma}_S = \sum_{s=-(b_n-1)}^{b_n-1} w_n(s)\widetilde{\gamma}_n(s), \tag{A.10}$$

$$\widetilde{\Sigma}_{w,n} = \frac{1}{n}\sum_{l=0}^{n-b_n}\sum_{k=1}^{b_n} k^2 \Delta_2 w_n(k)\big[\bar{B}_l(k) - \bar{B}_n\big]\big[\bar{B}_l(k) - \bar{B}_n\big]^T, \tag{A.11}$$

$$\begin{aligned}\tilde{d}_n &= \frac{1}{n}\left\{\sum_{t=1}^{b_n} \Delta_1 w_n(t)\left(\sum_{l=1}^{t-1} T_l T_l^T + \sum_{l=n-b_n+t+1}^{n} T_l T_l^T\right)\right. \\ &\quad + \sum_{s=1}^{b_n-1}\left[\sum_{t=1}^{b_n-s} \Delta_1 w_n(s+t)\right. \\ &\quad \left.\left.\times\left(\sum_{l=1}^{t-1}\left(T_l T_{l+s}^T + T_{l+s} T_l^T\right) + \sum_{l=n-b_n+t+1}^{n-s}\left(T_l T_{l+s}^T + T_{l+s} T_l^T\right)\right)\right]\right\}.\end{aligned} \tag{A.12}$$

Notice that in (A.12) we use the convention that empty sums are zero. Our goal is to show that $\widetilde{\Sigma}_{w,n} \to I_p$ as $n \to \infty$ with probability 1 in the following way. In Lemma 5, we show that $\widetilde{\Sigma}_{w,n} = \widetilde{\Sigma}_S - \tilde{d}_n$ and in Lemma 7 we show that the end term $\tilde{d}_n \to 0$ as $n \to \infty$ with probability 1. Lemma 12 shows that $\widetilde{\Sigma}_S \to I_p$ as $n \to \infty$ with probability 1, and hence $\widetilde{\Sigma}_{w,n} \to I_p$ as $n \to \infty$ with probability 1.

Lemma 5. *Under Condition 1,* $\widetilde{\Sigma}_{w,n} = \widetilde{\Sigma}_S - \tilde{d}_n$.

Proof. For $i, j = 1, \ldots, p$, let $\widetilde{\Sigma}_{w,ij}$ denote the (i, j)th entry of $\widetilde{\Sigma}_{w,n}$. Then,

$$\begin{aligned}
\widetilde{\Sigma}_{w,ij} &= \frac{1}{n}\sum_{l=0}^{n-b_n}\sum_{k=1}^{b_n} k^2\Delta_2 w_n(k)\big[\bar{B}_l^{(i)}(k) - \bar{B}_n^{(i)}\big]\big[\bar{B}_l^{(j)}(k) - \bar{B}_n^{(j)}\big] \\
&= \frac{1}{n}\sum_{l=0}^{n-b_n}\sum_{k=1}^{b_n} \Delta_2 w_n(k)\big[B^{(i)}(k+l) - B^{(i)}(l) - k\bar{B}_n^{(i)}\big] \\
&\quad \times \big[B^{(j)}(k+l) - B^{(j)}(l) - k\bar{B}_n^{(j)}\big] \\
&= \frac{1}{n}\sum_{l=0}^{n-b_n}\sum_{k=1}^{b_n} \Delta_2 w_n(k)\left[\sum_{t=1}^{k} U_{t+l}^{(i)} - k\bar{B}_n^{(i)}\right]\left[\sum_{t=1}^{k} U_{t+l}^{(j)} - k\bar{B}_n^{(j)}\right] \\
&= \frac{1}{n}\sum_{l=0}^{n-b_n}\sum_{k=1}^{b_n} \Delta_2 w_n(k)\left[\sum_{t=1}^{k} T_{t+l}^{(i)}\right]\left[\sum_{t=1}^{k} T_{t+l}^{(j)}\right] \\
&= \frac{1}{n}\sum_{l=0}^{n-b_n}\sum_{k=1}^{b_n} \Delta_2 w_n(k)\left[\sum_{t=1}^{k} T_{l+t}^{(i)}T_{l+t}^{(j)} + \sum_{s=1}^{k-1}\sum_{t=1}^{k-s} T_{l+t}^{(i)}T_{l+t+s}^{(j)} + \sum_{s=1}^{k-1}\sum_{t=1}^{k-s} T_{l+t}^{(j)}T_{l+t+s}^{(i)}\right].
\end{aligned} \tag{A.13}$$

In (A.13), we continue to use convention that empty sums are zero. We will look at each of the terms in (A.13) separately. For the first term, changing the order of summation and then using Lemma 3,

$$\begin{aligned}
&\frac{1}{n}\sum_{l=0}^{n-b_n}\sum_{k=1}^{b_n}\sum_{t=1}^{k} \Delta_2 w_n(k) T_{l+t}^{(i)}T_{l+t}^{(j)} \\
&\quad = \frac{1}{n}\sum_{l=0}^{n-b_n}\sum_{t=1}^{b_n}\sum_{k=t}^{b_n} \Delta_2 w_n(k) T_{l+t}^{(i)}T_{l+t}^{(j)} \\
&\quad = \frac{1}{n}\sum_{l=0}^{n-b_n}\sum_{t=1}^{b_n} T_{l+t}^{(i)}T_{l+t}^{(j)}\Delta_1 w_n(t)
\end{aligned} \tag{A.14}$$

$$= \frac{1}{n}\sum_{t=1}^{b_n}\Delta_1 w_n(t)\sum_{l=0}^{n-b_n} T_{l+t}^{(i)}T_{l+t}^{(j)}$$

$$= \widetilde{\gamma}_{n,ij}(0) - \frac{1}{n}\sum_{t=1}^{b_n}\Delta_1 w_n(t)\left(\sum_{l=1}^{t-1} T_l^{(i)}T_l^{(j)} + \sum_{l=n-b_n+t+1}^{n} T_l^{(i)}T_l^{(j)}\right).$$

For the second term in (A.13), we change the order of summation from l, k, s, t to l, s, k, t then to l, s, t, k and use Lemma 3 to get

$$\begin{aligned}
&\frac{1}{n}\sum_{l=0}^{n-b_n}\sum_{k=1}^{b_n}\sum_{s=1}^{k-1}\sum_{t=1}^{k-s}\Delta_2 w_n(k)T_{l+t}^{(i)}T_{l+t+s}^{(j)} \\
&= \frac{1}{n}\sum_{l=0}^{n-b_n}\sum_{s=1}^{b_n-1}\sum_{k=s+1}^{b_n}\sum_{t=1}^{k-s}\Delta_2 w_n(k)T_{l+t}^{(i)}T_{l+t+s}^{(j)} \\
&= \frac{1}{n}\sum_{l=0}^{n-b_n}\sum_{s=1}^{b_n-1}\sum_{t=1}^{b_n-s}\sum_{k=t+s}^{b_n}\Delta_2 w_n(k)T_{l+t}^{(i)}T_{l+t+s}^{(j)} \\
&= \frac{1}{n}\sum_{l=0}^{n-b_n}\sum_{s=1}^{b_n-1}\sum_{t=1}^{b_n-s}\Delta_1 w_n(s+t)T_{l+t}^{(i)}T_{l+t+s}^{(j)} \\
&= \frac{1}{n}\sum_{s=1}^{b_n-1}\sum_{l=0}^{n-b_n}\sum_{t=1}^{b_n-s}\Delta_1 w_n(s+t)T_{l+t}^{(i)}T_{l+t+s}^{(j)} \\
&= \frac{1}{n}\sum_{s=1}^{b_n-1}\sum_{t=1}^{b_n-s}\sum_{l=0}^{n-b_n}\Delta_1 w_n(s+t)T_{l+t}^{(i)}T_{l+t+s}^{(j)} \\
&= \frac{1}{n}\sum_{s=1}^{b_n-1}\sum_{t=1}^{b_n-s}\Delta_1 w_n(s+t)\sum_{l=0}^{n-b_n} T_{l+t}^{(i)}T_{l+t+s}^{(j)} \\
&= \frac{1}{n}\sum_{s=1}^{b_n-1}\sum_{t=1}^{b_n-s}\Delta_1 w_n(s+t)\sum_{l=t}^{n-b_n+t} T_l^{(i)}T_{l+s}^{(j)} \\
&= \sum_{s=1}^{b_n-1}\sum_{t=1}^{b_n-s}\Delta_1 w_n(s+t)\left[\tilde{\gamma}_{n,ij}(s) - \frac{1}{n}\sum_{l=1}^{t-1} T_l^{(i)}T_{l+s}^{(j)} - \frac{1}{n}\sum_{l=n-b_n+t+1}^{n-s} T_l^{(i)}T_{l+s}^{(j)}\right] \\
&= \sum_{s=1}^{b_n-1} w_n(s)\tilde{\gamma}_{n,ij}(s) - \frac{1}{n}\sum_{s=1}^{b_n-1}\sum_{t=1}^{b_n-s}\Delta_1 w_n(s+t)\left[\sum_{l=1}^{t-1} T_l^{(i)}T_{l+s}^{(j)} + \sum_{l=n-b_n+t+1}^{n-s} T_l^{(i)}T_{l+s}^{(j)}\right].
\end{aligned} \tag{A.15}$$

Repeating the same steps as in the second term, we reduce the third term in (A.13) to

$$\begin{aligned}
&\frac{1}{n}\sum_{l=0}^{n-b_n}\sum_{k=1}^{b_n}\sum_{s=1}^{k-1}\sum_{t=1}^{k-s}\Delta_2 w_n(k)T_{l+t}^{(j)}T_{l+t+s}^{(i)}\\
&\quad=\sum_{s=1}^{b_n-1}w_n(s)\tilde{\gamma}_{n,ji}(s)\\
&\qquad-\frac{1}{n}\sum_{s=1}^{b_n-1}\sum_{t=1}^{b_n-s}\Delta_1 w_n(s+t)\left[\sum_{l=1}^{t-1}T_l^{(j)}T_{l+s}^{(i)}+\sum_{l=n-b_n+t+1}^{n-s}T_l^{(j)}T_{l+s}^{(i)}\right] \qquad \text{(A.16)}\\
&\quad=\sum_{s=1}^{b_n-1}w_n(-s)\tilde{\gamma}_{n,ij}(-s)\\
&\qquad-\frac{1}{n}\sum_{s=1}^{b_n-1}\sum_{t=1}^{b_n-s}\Delta_1 w_n(s+t)\left[\sum_{l=1}^{t-1}T_l^{(j)}T_{l+s}^{(i)}+\sum_{l=n-b_n+t+1}^{n-s}T_l^{(j)}T_{l+s}^{(i)}\right].
\end{aligned}$$

Using (A.14), (A.15), and (A.16) in (A.13), we get

$$\begin{aligned}
\widetilde{\Sigma}_{w,ij}&=\tilde{\gamma}_{n,ij}(0)+\sum_{s=1}^{b_n-1}w_n(s)\tilde{\gamma}_{n,ij}(s)+\sum_{s=-(b_n-1)}^{-1}w_n(s)\tilde{\gamma}_{n,ij}(s)\\
&\quad-\frac{1}{n}\sum_{t=1}^{b_n}\Delta_1 w_n(t)\left(\sum_{l=1}^{t-1}T_l^{(i)}T_l^{(j)}+\sum_{l=n-b_n+t+1}^{n}T_l^{(i)}T_l^{(j)}\right)\\
&\quad-\frac{1}{n}\sum_{s=1}^{b_n-1}\sum_{t=1}^{b_n-s}\Delta_1 w_n(s+t)\\
&\quad\times\left[\sum_{l=1}^{t-1}\left(T_l^{(i)}T_{l+s}^{(j)}+T_{l+s}^{(i)}T_l^{(j)}\right)+\sum_{l=n-b_n+t+1}^{n-s}\left(T_l^{(i)}T_{l+s}^{(j)}+T_{l+s}^{(i)}T_l^{(j)}\right)\right]\\
&=\sum_{s=-(b_n-1)}^{b_n-1}\tilde{\gamma}_{n,ij}(s)w_n(s)-\tilde{d}_{n,ij}\\
&=\widetilde{\Sigma}_{S,ij}-\tilde{d}_{n,ij}.
\end{aligned}$$

□

Next, we show that as $n\to\infty$, $\tilde{d}_n\to 0$ with probability 1 implying $\widetilde{\Sigma}_{w,n}-\widetilde{\Sigma}_S\to 0$ with probability 1 as $n\to\infty$. To do so, we require a strong invariance principle for independent and identically distributed random variables.

Theorem 4 (Komlós *et al.* [40]). *Let $B(n)$ be a 1-dimensional standard Brownian motion. If $X_1, X_2, X_3 \ldots$ are independent and identically distributed univariate random variables with mean μ and standard deviation σ, such that $E[e^{|tX_1|}] < \infty$ in a neighborhood of $t = 0$, then as $n \to \infty$*

$$\sum_{i=1}^{n} X_i - n\mu - \sigma B(n) = O(\log n).$$

We begin with a technical lemma that will be used in a couple of places in the rest of the proof.

Lemma 6. *Let Conditions 1 and 2 hold. If, as $n \to \infty$, $b_n n^{-1} \sum_{k=1}^{b_n} k|\Delta_1 w_n(k)| \to 0$, then*

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

Proof.

$$\begin{aligned}
&\frac{b_n}{n}\left(\sum_{t=1}^{b_n}\left|\Delta_1 w_n(t)\right| + 2\sum_{s=1}^{b_n-1}\sum_{t=1}^{b_n-s}\left|\Delta_1 w_n(s+t)\right|\right)\\
&\quad= \frac{b_n}{n}\left(\sum_{t=1}^{b_n}\left|\Delta_1 w_n(t)\right| + 2\sum_{s=1}^{b_n-1}\sum_{k=s+1}^{b_n}\left|\Delta_1 w_n(k)\right|\right)\\
&\quad= \frac{b_n}{n}\left(\sum_{t=1}^{b_n}\left|\Delta_1 w_n(t)\right| + 2\sum_{k=2}^{b_n}\sum_{s=1}^{k-1}\left|\Delta_1 w_n(k)\right|\right)\\
&\quad= \frac{b_n}{n}\left(\sum_{t=1}^{b_n}\left|\Delta_1 w_n(t)\right| + 2\sum_{k=2}^{b_n}(k-1)\left|\Delta_1 w_n(k)\right|\right)\\
&\quad= \frac{b_n}{n}\left(\sum_{t=1}^{b_n}\left|\Delta_1 w_n(t)\right| + 2\sum_{k=1}^{b_n}(k-1)\left|\Delta_1 w_n(k)\right|\right)\\
&\quad\le \frac{b_n}{n}\left(2\sum_{k=1}^{b_n}k\left|\Delta_1 w_n(k)\right|\right)\\
&\quad\to 0 \qquad \text{by assumption.}
\end{aligned}$$

□

Lemma 7. *Let Conditions 1 and 2 hold and let $n > 2b_n$. If $b_n n^{-1} \sum_{k=1}^{b_n} k|\Delta_1 w_n(k)| \to 0$ and $b_n^{-1}\log n = O(1)$ as $n \to \infty$, then $\tilde{d}_n \to 0$ with probability 1 as $n \to \infty$.*

Proof. For $i, j = 1, \ldots, p$, we will show that as $n \to \infty$ with probability 1, $\tilde{d}_{n,ij} \to 0$. Recall

$$\begin{aligned}
\tilde{d}_{n,ij} &= \frac{1}{n}\sum_{t=1}^{b_n} \Delta_1 w_n(t)\left(\sum_{l=1}^{t-1} T_l^{(i)} T_l^{(j)} + \sum_{l=n-b_n+t+1}^{n} T_l^{(i)} T_l^{(j)}\right) \\
&\quad + \frac{1}{n}\sum_{s=1}^{b_n-1}\sum_{t=1}^{b_n-s} \Delta_1 w_n(s+t) \\
&\quad \times \left[\sum_{l=1}^{t-1}\left(T_l^{(i)} T_{l+s}^{(j)} + T_{l+s}^{(i)} T_l^{(j)}\right) + \sum_{l=n-b_n+t+1}^{n-s}\left(T_l^{(i)} T_{l+s}^{(j)} + T_{l+s}^{(i)} T_l^{(j)}\right)\right], \\
|\tilde{d}_{n,ij}| &\le \frac{1}{n}\sum_{t=1}^{b_n} \left|\Delta_1 w_n(t)\right|\left(\sum_{l=1}^{t-1} \left|T_l^{(i)} T_l^{(j)}\right| + \sum_{l=n-b_n+t+1}^{n} \left|T_l^{(i)} T_l^{(j)}\right|\right) \\
&\quad + \frac{1}{n}\sum_{s=1}^{b_n-1}\sum_{t=1}^{b_n-s} \left|\Delta_1 w_n(s+t)\right| \\
&\quad \times \left[\sum_{l=1}^{t-1}\left(\left|T_l^{(i)} T_{l+s}^{(j)}\right| + \left|T_{l+s}^{(i)} T_l^{(j)}\right|\right) + \sum_{l=n-b_n+t+1}^{n-s}\left(\left|T_l^{(i)} T_{l+s}^{(j)}\right| + \left|T_{l+s}^{(i)} T_l^{(j)}\right|\right)\right],
\end{aligned} \tag{A.17}$$

where we use the convention that empty sums are zero. Using the inequality $|ab| \le (a^2 + b^2)/2$ in the first and second terms in (A.17), we have for $t = 1, \ldots, b_n$

$$\sum_{l=1}^{t-1}\left|T_l^{(i)} T_l^{(j)}\right| \le \frac{1}{2}\sum_{l=1}^{t-1}\left(T_l^{(i)2} + T_l^{(j)2}\right) \le \frac{1}{2}\sum_{l=1}^{2b_n} T_l^{(i)2} + \frac{1}{2}\sum_{l=1}^{2b_n} T_l^{(j)2},$$

$$\sum_{l=n-b_n+t+1}^{n}\left|T_l^{(i)} T_l^{(j)}\right| \le \frac{1}{2}\sum_{l=n-b_n+t+1}^{n}\left(T_l^{(i)2} + T_l^{(j)2}\right) \le \frac{1}{2}\sum_{l=n-2b_n+1}^{n} T_l^{(i)2} + \frac{1}{2}\sum_{l=n-2b_n+1}^{n} T_l^{(j)2}.$$

Similarly, for the third and fourth terms in (A.17), for $t = 1, \ldots, b_n - 1$ and $s = 1, \ldots, b_n - 1$

$$\begin{aligned}
&\sum_{l=1}^{t-1}\left|T_l^{(i)} T_{l+s}^{(j)}\right| + \sum_{l=1}^{t-1}\left|T_{l+s}^{(i)} T_l^{(j)}\right| \\
&\quad \le \frac{1}{2}\sum_{l=1}^{2b_n} T_l^{(i)2} + \frac{1}{2}\sum_{l=1}^{2b_n} T_l^{(j)2} + \frac{1}{2}\sum_{l=1}^{b_n} T_{l+s}^{(j)2} + \frac{1}{2}\sum_{l=1}^{b_n} T_{l+s}^{(i)2} \\
&\quad \le \frac{1}{2}\sum_{l=1}^{2b_n} T_l^{(i)2} + \frac{1}{2}\sum_{l=1}^{2b_n} T_l^{(j)2} + \frac{1}{2}\sum_{l=1}^{2b_n} T_l^{(j)2} + \frac{1}{2}\sum_{l=1}^{2b_n} T_l^{(i)2}
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{l=1}^{2b_n} T_l^{(i)2} + \sum_{l=1}^{2b_n} T_l^{(j)2}, \\
&\sum_{l=n-b_n+t+1}^{n} \bigl|T_l^{(i)} T_{l+s}^{(j)}\bigr| + \sum_{l=n-b_n+t+1}^{n} \bigl|T_{l+s}^{(i)} T_l^{(j)}\bigr| \\
&\leq \frac{1}{2} \sum_{l=n-2b_n+1}^{n} \bigl(T_l^{(i)2} + T_l^{(j)2}\bigr) + \frac{1}{2} \sum_{l=n-b_n+1}^{n} \bigl(T_{l+s}^{(j)2} + T_{l+s}^{(i)2}\bigr) \\
&\leq \frac{1}{2} \sum_{l=n-2b_n+1}^{n} \bigl(T_l^{(i)2} + T_l^{(j)2}\bigr) + \frac{1}{2} \sum_{l=n-2b_n+1}^{n} \bigl(T_l^{(j)2} + T_l^{(i)2}\bigr) \\
&= \sum_{l=n-2b_n+1}^{n} T_l^{(i)2} + \sum_{l=n-2b_n+1}^{n} T_l^{(j)2}.
\end{aligned}
$$

Combining the above results in (A.17) we get,

$$
\begin{aligned}
|\tilde{d}_{n,ij}| &\leq \frac{1}{n} \left(\frac{1}{2} \sum_{l=1}^{2b_n} T_l^{(i)2} + \frac{1}{2} \sum_{l=1}^{2b_n} T_l^{(j)2} + \frac{1}{2} \sum_{l=n-2b_n+1}^{n} T_l^{(i)2} + \frac{1}{2} \sum_{l=n-2b_n+1}^{n} T_l^{(j)2} \right) \sum_{t=1}^{b_n} \bigl|\Delta_1 w_n(t)\bigr| \\
&\quad + \frac{1}{n} \sum_{s=1}^{b_n-1} \left[\left(\sum_{l=1}^{2b_n} T_l^{(i)2} + \sum_{l=1}^{2b_n} T_l^{(j)2} + \sum_{l=n-2b_n+1}^{n} T_l^{(i)2} + \sum_{l=n-2b_n+1}^{n} T_l^{(j)2} \right) \right. \\
&\quad \left. \times \sum_{t=1}^{b_n-s} \bigl|\Delta_1 w_n(s+t)\bigr| \right] \qquad \text{(A.18)} \\
&= \frac{1}{b_n} \left(\frac{1}{2} \sum_{l=1}^{2b_n} T_l^{(i)2} + \frac{1}{2} \sum_{l=1}^{2b_n} T_l^{(j)2} + \frac{1}{2} \sum_{l=n-2b_n+1}^{n} T_l^{(i)2} + \frac{1}{2} \sum_{l=n-2b_n+1}^{n} T_l^{(j)2} \right) \\
&\quad \times \frac{b_n}{n} \left(\sum_{t=1}^{b_n} \bigl|\Delta_1 w_n(t)\bigr| + 2 \sum_{s=1}^{b_n-1} \sum_{t=1}^{b_n-s} \bigl|\Delta_1 w_n(s+t)\bigr| \right).
\end{aligned}
$$

We will show that the first term in the product in (A.18) remains bounded with probability 1 as $n \to \infty$. Consider,

$$
\begin{aligned}
\frac{1}{2b_n} \sum_{l=1}^{2b_n} T_l^{(i)2} &= \frac{1}{2b_n} \sum_{l=1}^{2b_n} \bigl(U_l^{(i)} - \bar{B}_n^{(i)}\bigr)^2 \\
&= \frac{1}{2b_n} \sum_{l=1}^{2b_n} U_l^{(i)2} - 2\bar{B}_n^{(i)} \frac{1}{2b_n} \sum_{l=1}^{2b_n} U_l^{(i)} + \bigl(\bar{B}_n^{(i)}\bigr)^2
\end{aligned}
$$

$$
\begin{aligned}
&\leq \left| \frac{1}{2b_n} \sum_{l=1}^{2b_n} U_l^{(i)2} \right| + 2\left|\bar{B}_n^{(i)}\right| \left| \frac{1}{2b_n} \sum_{l=1}^{2b_n} U_l^{(i)} \right| + \left|\bar{B}_n^{(i)}\right|^2 \\
&< \left| \frac{1}{2b_n} \sum_{l=1}^{2b_n} U_l^{(i)2} \right| + \left| \frac{1}{2b_n} \sum_{l=1}^{2b_n} U_l^{(i)} \right| \left(\frac{2}{n}(1+\varepsilon)(2n\log\log n)^{1/2} \right) \\
&\quad + \left(\frac{1}{n}(1+\varepsilon)(2n\log\log n)^{1/2} \right)^2 \qquad \text{by Lemma 2} \\
&< \left| \frac{1}{2b_n} \sum_{l=1}^{2b_n} U_l^{(i)2} \right| + \left| \frac{1}{2b_n} \sum_{l=1}^{2b_n} U_l^{(i)} \right| O\big(\big(n^{-1}\log n\big)^{1/2}\big) + O\big(n^{-1}\log n\big).
\end{aligned}
$$

Since $U_l^{(i)}$ are Brownian motion increments, $U_l^{(i)} \overset{\text{i.i.d.}}{\sim} N(0,1)$ and by the classical strong law of large numbers, the above remains bounded with probability 1. Similarly $(2b_n)^{-1}\sum_{l=1}^{2b_n} T_l^{(j)2}$ remains bounded with probability 1 as $n \to \infty$. Next, consider $R_n = \sum_{l=1}^{n} U_l^{(i)2}$. Since $U_l^{(i)} \sim N(0,1)$, $R_n \sim \chi_n^2$. Thus, R_n has a moment generating function and an application of Theorem 4 implies there exists a finite random variable C_R such that, for sufficiently large n,

$$
\left| R_n - n - 2B^{(i)}(n) \right| < C_R \log n. \tag{A.19}
$$

Consider

$$
\begin{aligned}
|R_n - R_{n-2b_n}| &= \left| \big(R_n - n - 2B^{(i)}(n)\big) - \big(R_{n-2b_n} - (n-2b_n) - 2B^{(i)}(n-2b_n)\big) \right. \\
&\quad \left. - (n-2b_n) + n + 2B^{(i)}(n) - 2B^{(i)}(n-2b_n) \right| \\
&\leq \left| \big(R_n - n - 2B^{(i)}(n)\big) \right| + \left| \big(R_{n-2b_n} - (n-2b_n) - 2B^{(i)}(n-2b_n)\big) \right| \\
&\quad + \left| 2b_n + 2B^{(i)}(n) - 2B^{(i)}(n-2b_n) \right| \qquad \text{(A.20)} \\
&< C_R \log n + C_R \log(n-b_n) + 2b_n \\
&\quad + 2(1+\varepsilon)\left(2(2b_n)\left(\log\frac{n}{2b_n} + \log\log n \right) \right)^{1/2} \qquad \text{by (A.19) and Lemma 2} \\
&< 2C_R \log n + 2b_n + 4(1+\varepsilon)(2b_n \log n)^{1/2}.
\end{aligned}
$$

Finally,

$$
\begin{aligned}
&\frac{1}{2b_n} \sum_{l=n-2b_n+1}^{n} T_l^{(i)2} \\
&\quad = \frac{1}{2b_n} \sum_{l=n-2b_n+1}^{n} \big(U_l^{(i)} - \bar{B}_n^{(i)}\big)^2
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{2b_n} \sum_{l=n-2b_n+1}^{n} U_l^{(i)2} - \frac{2\bar{B}_n^{(i)}}{2b_n} \sum_{l=n-2b_n+1}^{n} U_l^{(i)} + \bigl(\bar{B}_n^{(i)}\bigr)^2 \\
&= \frac{1}{2b_n}(R_n - R_{n-2b_n}) - \frac{2}{n} B^{(i)}(n) \frac{1}{2b_n}\bigl(B^{(i)}(n) - B^{(i)}(n-2b_n)\bigr) + \left(\frac{1}{n} B^{(i)}(n)\right)^2 \\
&< \frac{1}{2b_n}|R_n - R_{n-2b_n}| + \frac{2}{n}\bigl|B^{(i)}(n)\bigr| \frac{1}{2b_n}\bigl|B^{(i)}(n) - B^{(i)}(n-2b_n)\bigr| + \left(\frac{1}{n}\bigl|B^{(i)}(n)\bigr|\right)^2 \\
&< \frac{1}{2b_n}\bigl(2C_R \log n + 2b_n + 4(1+\varepsilon)(2b_n \log n)^{1/2}\bigr) \\
&\quad + \frac{2}{n}(1+\varepsilon)(2n \log\log n)^{1/2} \frac{1}{2b_n}(1+\varepsilon)\left(2(2b_n)\left(\log \frac{n}{2b_n} + \log\log n\right)\right)^{1/2} \\
&\quad + \left(\frac{1}{n}(1+\varepsilon)(2n \log\log n)^{1/2}\right)^2 \qquad \text{by (A.20) and Lemma 2} \\
&< C_R b_n^{-1} \log n + 1 + \frac{4(1+\varepsilon)(2b_n \log n)^{1/2}}{2b_n} + \frac{1}{nb_n}(1+\varepsilon)^2 (2n \log n)^{1/2} (8b_n \log n)^{1/2} \\
&\quad + \frac{(1+\varepsilon)^2}{n}(2 \log n) \\
&< C_R b_n^{-1} \log n + 1 + 2(1+\varepsilon)\bigl(2b_n^{-1} \log n\bigr)^{1/2} \\
&\quad + 4(1+\varepsilon)^2 \left(\frac{\log n}{n}\right)^{1/2} \bigl(b_n^{-1} \log n\bigr)^{1/2} + 2(1+\varepsilon)^2 \frac{\log n}{n}.
\end{aligned}
$$

Since $b_n^{-1} \log n = O(1)$ as $n \to \infty$, the above term remains bounded with probability 1 as $n \to \infty$. Similarly, $(2b_n)^{-1} \sum_{l=n-2b_n+1}^{n} T_l^{(j)2}$ remains bounded with probability 1 as $n \to \infty$. The second term in the product in (A.18) converges to 0 by Lemma 6 and hence $\tilde{d}_{n,ij} \to 0$ with probability 1 as $n \to \infty$. □

Recall that $h(X_t) = Y_t^2$ for $t = 1, 2, 3, \ldots$, where the square is element-wise.

Lemma 8. *Let a strong invariance principle for h hold as in (2.4). If Condition 2 holds, $b_n^{-1} \psi_h(n) \to 0$ and $b_n^{-1} \log n = O(1)$ as $n \to \infty$, then*

$$
\frac{1}{b_n} \sum_{k=1}^{b_n} h(X_k) \quad \text{and} \quad \frac{1}{b_n} \sum_{k=n-b_n+1}^{n} h(X_k),
$$

stay bounded with probability 1 as $n \to \infty$.

Proof. Equation (2.4) implies that $b_n^{-1} \sum_{k=1}^{b_n} h(X_k) \to \mathrm{E}_F h$ if $b_n^{-1} \psi_h(b_n) \to 0$ as $n \to \infty$. Since by assumption $b_n^{-1} \psi_h(n) \to 0$ as $n \to \infty$ and ψ_h is increasing, $b_n^{-1} \sum_{k=1}^{b_n} h(X_k)$ remains

bounded w.p. 1 as $n \to \infty$. Next, for all $\varepsilon > 0$ and sufficiently large $n(\varepsilon)$,

$$
\begin{aligned}
&\frac{1}{b_n}\Bigg\|\sum_{k=n-b_n+1}^{n} h(X_k)\Bigg\| \\
&\quad= \frac{1}{b_n}\Bigg\|\sum_{k=1}^{n} h(X_k) - \sum_{k=1}^{n-b_n} h(X_k)\Bigg\| \\
&\quad= \frac{1}{b_n}\Bigg\|\sum_{k=1}^{n} h(X_k) - n\mathrm{E}_F h + (n-b_n)\mathrm{E}_F h + b_n \mathrm{E}_F h - L_h B(n) + L_h B(n-b_n) \\
&\qquad + L_h\big(B(n) - B(n-b_n)\big) - \sum_{k=1}^{n-b_n} h(X_k)\Bigg\| \\
&\quad\le \frac{1}{b_n}\Bigg\|\sum_{k=1}^{n} h(X_k) - n\mathrm{E}_F h - L_h B(n)\Bigg\| + \frac{1}{b_n}\Bigg\|\sum_{k=1}^{n-b_n} h(X_k) - (n-b_n)\mathrm{E}_F h - L_h B(n-b_n)\Bigg\| \\
&\qquad + \frac{1}{b_n}\big\|L_h\big(B(n) - B(n-b_n)\big) + b_n \mathrm{E}_F h\big\| \\
&\quad< \frac{1}{b_n} D_h \psi_h(n) + \frac{1}{b_n} D_h \psi_h(n-b_n) + \frac{1}{b_n}\big\|L_h\big(B(n) - B(n-b_n)\big)\big\| + \|\mathrm{E}_F h\| \qquad \text{by (2.4)} \\
&\quad\le \frac{1}{b_n} D_h \psi_h(n) + \frac{1}{b_n} D_h \psi_h(n-b_n) + \frac{1}{b_n}\|L_h\|\left(\sum_{i=1}^{p}\big|B^{(i)}(n) - B^{(i)}(n-b_n)\big|^2\right)^{1/2} + \|\mathrm{E}_F h\| \\
&\quad\le \frac{1}{b_n} D_h \psi_h(n) + \frac{1}{b_n} D_h \psi_h(n-b_n) \\
&\qquad + \frac{1}{b_n}\|L_h\|\left(\sum_{i=1}^{p}\sup_{0\le s\le b_n}\big|B^{(i)}(n) - B^{(i)}(n-s)\big|^2\right)^{1/2} + \|\mathrm{E}_F h\| \\
&\quad< \frac{2}{b_n} D_h \psi_h(n) + \frac{p^{1/2}}{b_n}\|L_h\|(1+\varepsilon)\left(2b_n\left(\log\frac{n}{b_n} + \log\log n\right)\right)^{1/2} + \|\mathrm{E}_F h\| \quad \text{by Lemma 2} \\
&\quad< \|\mathrm{E}_F h\| + \frac{2}{b_n} D_h \psi_h(n) + O\big(\big(b_n^{-1}\log n\big)^{1/2}\big).
\end{aligned}
$$

Thus by the assumptions $b_n^{-1}\|\sum_{k=n-b_n+1}^{n} h(X_k)\|$ stays bounded w.p. 1 as $n \to \infty$. □

Lemma 9. *Suppose the strong invariance principles* (2.3) *and* (2.4) *hold. In addition, suppose Conditions* 1 *and* 2 *hold and* $n > 2b_n$, $b_n n^{-1}\sum_{k=1}^{b_n} k|\Delta_1 w_n(k)| \to 0$, $b_n^{-1}\psi(n) \to 0$, $b_n^{-1}\psi_h(n) \to 0$ *as* $n \to \infty$ *and* $b_n^{-1}\log n = O(1)$. *Then,* $d_n \to 0$ *with probability* 1 *as* $n \to \infty$.

Proof. For $i, j = 1, \ldots, p$, let $d_{n,ij}$ denote the (i, j)th element of the matrix d_n. We can follow the same steps as in Lemma 7 to obtain

$$|d_{n,ij}| \leq \frac{1}{b_n}\left(\frac{1}{2}\sum_{l=1}^{2b_n} Z_l^{(i)2} + \frac{1}{2}\sum_{l=1}^{2b_n} Z_l^{(j)2} + \frac{1}{2}\sum_{l=n-2b_n+1}^{n} Z_l^{(i)2} + \frac{1}{2}\sum_{l=n-2b_n+1}^{n} Z_l^{(j)2}\right) \times \frac{b_n}{n}\left(\sum_{t=1}^{b_n}\left|\Delta_1 w_n(t)\right| + 2\sum_{s=1}^{b_n-1}\sum_{t=1}^{b_n-s}\left|\Delta_1 w_n(s+t)\right|\right).$$

The second term in the product converges to 0 by Lemma 6. It remains to show that the following remains bounded with probability 1 as $n \to \infty$,

$$\frac{1}{b_n}\left(\frac{1}{2}\sum_{l=1}^{2b_n} Z_l^{(i)2} + \frac{1}{2}\sum_{l=1}^{2b_n} Z_l^{(j)2} + \frac{1}{2}\sum_{l=n-2b_n+1}^{n} Z_l^{(i)2} + \frac{1}{2}\sum_{l=n-2b_n+1}^{n} Z_l^{(j)2}\right).$$

We have,

$$\frac{1}{2b_n}\sum_{l=1}^{2b_n} Z_l^{(i)2} = \frac{1}{2b_n}\sum_{l=1}^{2b_n}\left(Y_l^{(i)} - \bar{Y}_n^{(i)}\right)^2 = \frac{1}{2b_n}\sum_{l=1}^{2b_n} Y_l^{(i)2} - 2\bar{Y}_{2b_n}^{(i)}\bar{Y}_n^{(i)} + \left(\bar{Y}_n^{(i)}\right)^2.$$

By the strong invariance principle for g, $\bar{Y}_n^{(i)} \to 0$, $\bar{Y}_{2b_n}^{(i)} \to 0$, and $(\bar{Y}_n^{(i)})^2 \to 0$ w.p. 1 as $n \to \infty$. By Lemma 8, $(2b_n)^{-1}\sum_{l=1}^{2b_n} Y_l^{(i)2}$ remains bounded w.p. 1 as $n \to \infty$. Thus, $(2b_n)^{-1}\sum_{l=1}^{2b_n} Z_l^{(i)2}$ remains bounded w.p. 1 as $n \to \infty$. Similarly $(2b_n)^{-1}\sum_{l=1}^{2b_n} Z_l^{(j)2}$ stay bounded w.p. 1 as $n \to \infty$. Now consider

$$\begin{aligned}\frac{1}{2b_n}\sum_{l=n-2b_n+1}^{n} Z_l^{(i)2} &= \frac{1}{2b_n}\sum_{l=n-2b_n+1}^{n}\left(Y_l^{(i)} - \bar{Y}_n^{(i)}\right)^2 \\ &= \frac{1}{2b_n}\sum_{l=n-2b_n+1}^{n} Y_l^{(i)2} - 2\bar{Y}_n^{(i)}\frac{1}{2b_n}\sum_{l=n-2b_n+1}^{n} Y_l^{(i)} + \left(\bar{Y}_n^{(i)}\right)^2.\end{aligned} \tag{A.21}$$

We will first show that $(2b_n)^{-1}\sum_{l=n-2b_n+1}^{n} Y_l^{(i)}$ remains bounded with probability 1. Let Σ_{ii} denote the ith diagonal entry of Σ, then

$$\frac{1}{2b_n}\sum_{l=n-2b_n+1}^{n} Y_l^{(i)} = \frac{1}{2b_n}\left(\sum_{l=1}^{n} Y_l^{(i)} - \sum_{l=1}^{n-2b_n} Y_l^{(i)}\right)$$

$$
\begin{aligned}
&= \frac{1}{2b_n}\left(\sum_{l=1}^{n} Y_l^{(i)} - \sqrt{\Sigma_{ii}}\,B^{(i)}(n)\right) - \frac{1}{2b_n}\left(\sum_{l=1}^{n-2b_n} Y_l^{(i)} - \sqrt{\Sigma_{ii}}\,B^{(i)}(n-2b_n)\right) \\
&\quad + \frac{1}{2b_n}\sqrt{\Sigma_{ii}}\big(B^{(i)}(n) - B^{(i)}(n-2b_n)\big) \\
&< \frac{1}{2b_n}\left|\sum_{l=1}^{n} Y_l^{(i)} - \sqrt{\Sigma_{ii}}\,B^{(i)}(n)\right| + \frac{1}{2b_n}\left|\sum_{l=1}^{n-2b_n} Y_l^{(i)} - \sqrt{\Sigma_{ii}}\,B^{(i)}(n-2b_n)\right| \\
&\quad + \frac{1}{2b_n}\left|\sqrt{\Sigma_{ii}}\big(B^{(i)}(n) - B^{(i)}(n-2b_n)\big)\right| \\
&< \frac{1}{2b_n}D\psi(n) + \frac{1}{2b_n}D\psi(n-2b_n) + \frac{1}{2b_n}\sqrt{\Sigma_{ii}}\sup_{0\le s\le 2b_n}\left|B^{(i)}(n) - B^{(i)}(n-s)\right| \qquad \text{by (2.3)} \\
&< \frac{2D}{2b_n}\psi(n) + \sqrt{\Sigma_{ii}}\,\frac{1}{2b_n}(1+\varepsilon)\left[2(2b_n)\left(\log\frac{n}{2b_n} + \log\log n\right)\right]^{1/2} \qquad \text{by Lemma 2} \\
&< Db_n^{-1}\psi(n) + \sqrt{\Sigma_{ii}}\,(1+\varepsilon)\big(2b_n^{-1}\log n\big)^{1/2} \\
&= O\big(b_n^{-1}\psi(n)\big) + O\big(\big(b_n^{-1}\log n\big)^{1/2}\big).
\end{aligned}
$$

By the strong invariance principle for g, $\bar{Y}_n^{(i)} \to 0$ and $(\bar{Y}_n^{(i)})^2 \to 0$ w.p. 1 as $n \to \infty$. By Lemma 8, $(2b_n)^{-1}\sum_{l=n-2b_n+1}^{n} Y_l^{(i)2}$ remains bounded w.p. 1 as $n \to \infty$. Combining these results in (A.21), $(2b_n)^{-1}\sum_{l=n-2b_n+1}^{n} Z_l^{(i)2}$ remains bounded w.p. 1 as $n \to \infty$. Similarly $(2b_n)^{-1}\sum_{l=n-2b_n+1}^{n} Z_l^{(j)2}$ remains bounded w.p. 1 as $n \to \infty$. □

Lemma 10 (Billingsley [6]). *For a family of random variables $\{X_n : n \ge 1\}$, if $E(|X_n|) \le s_n$ where s_n is a sequence such that $\sum_{n=1}^{\infty} s_n < \infty$, then $X_n \to 0$ w.p. 1 as $n \to \infty$.*

Lemma 11 (Whittle [62]). *Let $R_1, \ldots, R_n$ be i.i.d. standard normal variables and $A = \sum_{l=1}^{n}\sum_{k=1}^{n} a_{lk}R_lR_k$ where a_{lk} are real coefficients, then for $c \ge 1$ and for some constant K_c, we have*

$$
E\big[|A - EA|^{2c}\big] \le K_c\left(\sum_l\sum_k a_{lk}^2\right)^c.
$$

Lemma 12. *Let Conditions 1 and 2 hold and assume that*

(a) *there exists a constant $c \ge 1$ such that $\sum_n (b_n/n)^c < \infty$,*
(b) *$b_n n^{-1}\log n \to 0$ as $n \to \infty$,*

then $\widetilde{\Sigma}_S \to I_p$ w.p. 1 as $n \to \infty$.

Proof. Under the same conditions, Theorem 4.1 in Damerdji [10] shows $\widetilde{\Sigma}_{S,ii} \to 1$ as $n \to \infty$ w.p. 1. It is left to show that for all $i, j = 1, \ldots, p$, and $i \ne j$, $\widetilde{\Sigma}_{S,ij} \to 0$ w.p. 1 as $n \to \infty$. Recall

that

$$\begin{aligned}
\widetilde{\Sigma}_{S,ij} &= \sum_{s=-(b_n-1)}^{b_n-1} w_n(s)\tilde{\gamma}_{n,ij}(s) \\
&= \tilde{\gamma}_{n,ij}(0) + \frac{1}{n}\Bigg[\sum_{s=1}^{b_n-1} w_n(s)\sum_{t=1}^{n-s}\big(U_t^{(i)} - \bar{B}_n^{(i)}\big)\big(U_{t+s}^{(j)} - \bar{B}_n^{(j)}\big) \\
&\quad + \sum_{s=-(b_n-1)}^{-1} w_n(s)\sum_{t=1-s}^{n}\big(U_t^{(i)} - \bar{B}_n^{(i)}\big)\big(U_{t+s}^{(j)} - \bar{B}_n^{(j)}\big)\Bigg] \\
&= \tilde{\gamma}_{n,ij}(0) + \frac{1}{n}\Bigg[\sum_{s=1}^{b_n-1} w_n(s)\sum_{t=1}^{n-s}\big(U_t^{(i)} - \bar{B}_n^{(i)}\big)\big(U_{t+s}^{(j)} - \bar{B}_n^{(j)}\big) \\
&\quad + \sum_{s=1}^{b_n-1} w_n(s)\sum_{t=1+s}^{n}\big(U_t^{(i)} - \bar{B}_n^{(i)}\big)\big(U_{t-s}^{(j)} - \bar{B}_n^{(j)}\big)\Bigg] \\
&= \tilde{\gamma}_{n,ij}(0) + \sum_{s=1}^{b_n-1} w_n(s)\frac{1}{n}\Bigg[\sum_{t=1}^{n-s}\big(U_t^{(i)}U_{t+s}^{(j)} - \bar{B}_n^{(i)}U_{t+s}^{(j)} - \bar{B}_n^{(j)}U_t^{(i)} + \bar{B}_n^{(i)}\bar{B}_n^{(j)}\big) \\
&\quad + \sum_{t=1+s}^{n}\big(U_t^{(i)}U_{t-s}^{(j)} - \bar{B}_n^{(i)}U_{t-s}^{(j)} - \bar{B}_n^{(j)}U_t^{(i)} + \bar{B}_n^{(i)}\bar{B}_n^{(j)}\big)\Bigg].
\end{aligned}$$

Since

$$\sum_{t=1}^{n-s} U_{t+s}^{(j)} = B^{(j)}(n) - B^{(j)}(s), \qquad \sum_{t=1}^{n-s} U_t^{(i)} = B^{(i)}(n-s),$$

$$\sum_{t=1+s}^{n} U_{t-s}^{(j)} = B^{(j)}(n-s) \quad \text{and} \quad \sum_{t=1+s}^{n} U_t^{(i)} = B^{(i)}(n) - B^{(i)}(s),$$

we get $\widetilde{\Sigma}_{S,ij} =$

$$\begin{aligned}
&= \tilde{\gamma}_{n,ij}(0) + \sum_{s=1}^{b_n-1} w_n(s)\Bigg[\frac{1}{n}\sum_{t=1}^{n-s} U_t^{(i)}U_{t+s}^{(j)} - \frac{1}{n}\bar{B}_n^{(i)}\big(B^{(j)}(n) - B^{(j)}(s)\big) - \frac{1}{n}\bar{B}_n^{(j)}B^{(i)}(n-s) \\
&\quad + \left(\frac{n-s}{n}\right)\bar{B}_n^{(i)}\bar{B}_n^{(j)} + \frac{1}{n}\sum_{t=1+s}^{n} U_t^{(i)}U_{t-s}^{(j)} - \frac{1}{n}\bar{B}_n^{(i)}B^{(j)}(n-s) - \frac{1}{n}\bar{B}_n^{(j)}\big(B^{(i)}(n) - B^{(i)}(s)\big) \\
&\quad + \left(\frac{n-s}{n}\right)\bar{B}_n^{(i)}\bar{B}_n^{(j)}\Bigg]
\end{aligned}$$

$$
\begin{aligned}
&= \tilde{\gamma}_{n,ij}(0) + \sum_{s=1}^{b_n-1} w_n(s)\Bigg[\frac{1}{n}\sum_{t=1}^{n-s} U_t^{(i)}U_{t+s}^{(j)} + \frac{1}{n}\sum_{t=1+s}^{n} U_t^{(i)}U_{t-s}^{(j)} \\
&\quad + 2\left(1-\frac{s}{n}\right)\bar{B}_n^{(i)}\bar{B}_n^{(j)} - 2\bar{B}_n^{(i)}\bar{B}_n^{(j)} \\
&\quad + \frac{1}{n}\bar{B}_n^{(i)}B^{(j)}(s) - \frac{1}{n}\bar{B}_n^{(j)}B^{(i)}(n-s) - \frac{1}{n}\bar{B}_n^{(i)}B^{(j)}(n-s) + \frac{1}{n}\bar{B}_n^{(j)}B^{(i)}(s)\Bigg] \\
&= \tilde{\gamma}_{n,ij}(0) + \sum_{s=1}^{b_n-1} w_n(s)\Bigg[\frac{1}{n}\sum_{t=1}^{n-s} U_t^{(i)}U_{t+s}^{(j)} + \frac{1}{n}\sum_{t=1+s}^{n} U_t^{(i)}U_{t-s}^{(j)} - 2\left(1+\frac{s}{n}\right)\bar{B}_n^{(i)}\bar{B}_n^{(j)} \\
&\quad + \bar{B}_n^{(i)}\bar{B}_n^{(j)} - \frac{1}{n}\bar{B}_n^{(j)}B^{(i)}(n-s) + \bar{B}_n^{(i)}\bar{B}_n^{(j)} - \frac{1}{n}\bar{B}_n^{(i)}B^{(j)}(n-s) \\
&\quad + \frac{1}{n}\bar{B}_n^{(i)}B^{(j)}(s) + \frac{1}{n}\bar{B}_n^{(j)}B^{(i)}(s)\Bigg] \\
&= \tilde{\gamma}_{n,ij}(0) + \sum_{s=1}^{b_n-1} w_n(s)\Bigg[\frac{1}{n}\sum_{t=1}^{n-s} U_t^{(i)}U_{t+s}^{(j)} + \frac{1}{n}\sum_{t=1+s}^{n} U_t^{(i)}U_{t-s}^{(j)} - 2\left(1+\frac{s}{n}\right)\bar{B}_n^{(i)}\bar{B}_n^{(j)} \\
&\quad + \frac{1}{n}\bar{B}_n^{(j)}\big(B^{(i)}(n) - B^{(i)}(n-s)\big) + \frac{1}{n}\bar{B}_n^{(i)}\big(B^{(j)}(n) - B^{(j)}(n-s)\big) + \frac{1}{n}\bar{B}_n^{(i)}B^{(j)}(s) \\
&\quad + \frac{1}{n}B^{(i)}(s)\bar{B}_n^{(j)}\Bigg].
\end{aligned}
\tag{A.22}
$$

We will show that each of the terms goes to 0 with probability 1 as $n \to \infty$.

1.

$$
\begin{aligned}
\tilde{\gamma}_{n,ij}(0) &= \frac{1}{n}\sum_{t=1}^{n} T_t^{(i)}T_t^{(j)} \\
&= \frac{1}{n}\sum_{t=1}^{n}\big(U_t^{(i)} - \bar{B}_n^{(i)}\big)\big(U_t^{(j)} - \bar{B}_n^{(j)}\big) \\
&= \frac{1}{n}\sum_{t=1}^{n} U_t^{(i)}U_t^{(j)} - \bar{B}_n^{(j)}\frac{1}{n}\sum_{t=1}^{n} U_t^{(i)} - \bar{B}_n^{(i)}\frac{1}{n}\sum_{t=1}^{n} U_t^{(j)} + \bar{B}_n^{(i)}\bar{B}_n^{(j)}.
\end{aligned}
\tag{A.23}
$$

We will show that each of the terms in (A.23) goes to 0 with probability 1, as $n \to \infty$. First, we will use Lemma 11 to show that $n^{-1}\sum_{t=1}^{n} U_t^{(i)}U_t^{(j)} \to 0$ with probability 1 as $n \to \infty$. Define

$$
R_1 = U_1^{(i)}, R_2 = U_2^{(i)}, \ldots, R_n = U_n^{(i)}, R_{n+1} = U_1^{(j)}, \ldots, R_{2n} = U_n^{(j)}.
$$

Thus, $\{R_i : 1 \le i \le 2n\}$ is an i.i.d. sequence of normally distributed random variables. Define for $1 \le l, k, \le 2n$,

$$a_{lk} = \begin{cases} \frac{1}{n}, & \text{if } 1 \le l \le n \text{ and } k = l + n, \\ 0 & \text{otherwise.} \end{cases}$$

Then,

$$A := \sum_{l=1}^{2n} \sum_{k=1}^{2n} a_{lk} R_l R_k = \sum_{t=1}^{n} \frac{1}{n} U_t^{(i)} U_t^{(j)}.$$

By Lemma 11, for all $c \ge 1$ there exists K_c such that

$$\mathrm{E}\big[|A - \mathrm{E}A|^{2c}\big] \le K_c \bigg(\sum_l \sum_k a_{lk}^2 \bigg)^c.$$

Since $i \ne j$, $\mathrm{E}[A] = 0$,

$$\mathrm{E}\left(\left| \frac{1}{n} \sum_{t=1}^{n} U_t^{(i)} U_t^{(j)} \right|^{2c} \right) \le K_c \left(\sum_{l=1}^{2n} \sum_{k=1}^{2n} a_{lk}^2 \right)^c = K_c \left(\sum_{t=1}^{n} \frac{1}{n^2} \right)^c = K_c n^{-c}.$$

Note that $\sum_{n=0}^{\infty} n^{-c} < \infty$ for all $c > 1$, hence by Lemma 10, $n^{-1} \sum_{t=1}^{n} U_t^{(i)} U_t^{(j)} \to 0$ with probability 1 as $n \to \infty$. Next in (A.23),

$$\begin{aligned} \bar{B}_n^{(j)} \frac{1}{n} \sum_{t=1}^{n} U_t^{(i)} &\le \frac{1}{n} \big| B^{(j)}(n) \big| \left| \frac{1}{n} \sum_{t=1}^{n} U_t^{(i)} \right| \\ &< \frac{1}{n} (1 + \varepsilon) \sqrt{2n \log\log n} \left| \frac{1}{n} \sum_{t=1}^{n} U_t^{(i)} \right| \qquad \text{by Lemma 2} \\ &< \sqrt{2} (1 + \varepsilon) \left(\frac{\log n}{n} \right)^{1/2} \left| \frac{1}{n} \sum_{t=1}^{n} U_t^{(i)} \right|. \end{aligned}$$

By the classical SLLN

$$\left| \frac{1}{n} \sum_{t=1}^{n} U_t^{(i)} \right| \to 0 \qquad \text{w.p. 1 as } n \to \infty.$$

Similarly,

$$\bar{B}_n^{(i)} \frac{1}{n} \sum_{t=1}^{n} U_t^{(j)} \to 0 \qquad \text{w.p. 1 as } n \to \infty.$$

Finally,

$$\begin{aligned}\bar{B}_n^{(i)}\bar{B}_n^{(j)} &\le \frac{1}{n^2}\big|B^{(i)}(n)\big|\big|B^{(j)}(n)\big| \\ &< \frac{1}{n^2}(1+\varepsilon)^2(2n\log\log n) \qquad \text{by Lemma 2} \\ &< 2(1+\varepsilon)^2\left(\frac{\log n}{n}\right) \\ &\to 0 \qquad \text{as } n\to\infty.\end{aligned}$$

Thus, $\tilde{\gamma}_{n,ij}(0)\to 0$ with probability 1 as $n\to\infty$.

2. Now consider the term $\sum_{s=1}^{b_n-1} w_n(s)n^{-1}\sum_{t=1}^{n-s} U_t^{(i)}U_{t+s}^{(j)}$. Define

$$\begin{aligned}R_1 = U_1^{(i)}, R_2 = U_2^{(i)}, \ldots, R_n = U_n^{(i)}, \\ R_{(n+1)} = U_1^{(j)}, \ldots, R_{2n} = U_n^{(j)}.\end{aligned}$$

Thus, $\{R_i : 1\le i\le 2n\}$ is an i.i.d. sequence of normally distributed random variables. Next, define for $1\le l,k\le 2n$

$$a_{lk} = \begin{cases} \dfrac{1}{n}w_n\big(k-(n+l)\big), \\ \qquad \text{if } 1\le l\le n-1, n+2\le k\le 2n, \text{ and } 1\le k-(n+l)\le b_n-1, \\ 0, \qquad \text{otherwise.}\end{cases}$$

Then,

$$\begin{aligned}A &:= \sum_{l=1}^{2n}\sum_{k=1}^{2n} a_{lk}R_lR_k \\ &= \sum_{l=1}^{n-1}\sum_{k=n+2}^{2n} I\big\{1\le k-(n+l)\le b_n-1\big\}\frac{1}{n}w_n\big(k-(n+l)\big)R_lR_k \\ &= \sum_{l=1}^{n-1}\sum_{s=2-l}^{n-l} I\{1\le s\le b_n-1\}\frac{1}{n}w_n(s)R_lR_{n+l+s} \qquad \text{letting } k-(n+l)=s \\ &= \sum_{s=1}^{n-1}\sum_{l=1}^{n-s} I\{1\le s\le b_n-1\}\frac{1}{n}w_n(s)R_lR_{n+l+s} \\ &\quad + \sum_{s=(3-n)}^{0}\sum_{l=(2-s)}^{n-1} I\{1\le s\le b_n-1\}\frac{1}{n}w_n(s)R_lR_{n+l+s}\end{aligned}$$

$$= \sum_{s=1}^{b_n-1} \sum_{l=1}^{n-s} \frac{1}{n} w_n(s) U_l^{(i)} U_{l+s}^{(j)} \qquad \text{since } n > 2b_n \geq 2$$

$$= \sum_{s=1}^{b_n-1} \sum_{l=1}^{n-s} \frac{1}{n} w_n(s) U_l^{(i)} U_{l+s}^{(j)}.$$

Using Lemma 11, for $c \geq 1$ and some constant K_c,

$$\mathrm{E}\left[\left(\sum_{s=1}^{b_n-1} w_n(s) \frac{1}{n} \sum_{t=1}^{n-s} U_t^{(i)} U_{t+s}^{(j)}\right)^{2c}\right] \leq K_c \left(\sum_l \sum_k a_{lk}^2\right)^c,$$

where

$$\sum_l \sum_k a_{lk}^2 = \sum_{s=1}^{b_n-1} \sum_{t=1}^{n-s} \frac{1}{n^2} w_n^2(s) = \frac{1}{n^2} \sum_{s=1}^{b_n-1} (n-s) w_n^2(s) \leq \frac{n}{n^2} \sum_{s=1}^{b_n-1} 1 \leq \frac{b_n}{n}.$$

Thus, by Assumption (a) and Lemma 10,

$$\sum_{s=1}^{b_n-1} w_n(s) \frac{1}{n} \sum_{t=1}^{n-s} U_t^{(i)} U_{t+s}^{(j)} \to 0 \qquad \text{w.p. 1 as } n \to \infty.$$

3. By letting $t - s = l$,

$$\sum_{s=1}^{b_n-1} w_n(s) \frac{1}{n} \sum_{t=1+s}^{n} U_t^{(i)} U_{t-s}^{(j)} = \sum_{s=1}^{b_n-1} w_n(s) \frac{1}{n} \sum_{l=1}^{n-s} U_{l+s}^{(i)} U_l^{(j)}.$$

This is similar to the previous part with just the i and j components interchanged. A similar argument will lead to $\sum_{s=1}^{b_n-1} w_n(s) n^{-1} \sum_{t=1+s}^{n} U_t^{(i)} U_{t-s}^{(j)} \to 0$ with probability 1 as $n \to \infty$.

4.

$$\begin{aligned}
&\sum_{s=1}^{b_n-1} 2 w_n(s) \left(1 + \frac{s}{n}\right) \bar{B}_n^{(i)} \bar{B}_n^{(j)} \\
&\quad \leq \left| \sum_{s=1}^{b_n-1} 2 w_n(s) \left(1 + \frac{s}{n}\right) \bar{B}_n^{(i)} \bar{B}_n^{(j)} \right| \\
&\quad \leq \sum_{s=1}^{b_n-1} 2 \left|w_n(s)\right| \left(1 + \frac{s}{n}\right) \left|\bar{B}_n^{(i)}\right| \left|\bar{B}_n^{(j)}\right| \\
&\quad \leq \frac{2}{n^2} \sum_{s=1}^{b_n-1} \left(1 + \frac{s}{n}\right) \left|B^{(i)}(n)\right| \left|B^{(j)}(n)\right| \qquad \text{since } \left|w_n(s)\right| \leq 1
\end{aligned}$$

$$
\begin{aligned}
&< \frac{2}{n^2}(1+\varepsilon)^2 2n \log\log n \sum_{s=1}^{b_n-1}\left(1+\frac{s}{n}\right) \qquad \text{by Lemma 2}\\
&< 4(1+\varepsilon)^2 n^{-1}\log n \sum_{s=1}^{b_n-1} 2\\
&\le 8(1+\varepsilon)^2 b_n n^{-1}\log n\\
&\to 0.
\end{aligned}
$$

5. Next,

$$
\begin{aligned}
&\sum_{s=1}^{b_n-1} w_n(s)\frac{1}{n}\bar{B}_n^{(j)}\bigl(B^{(i)}(n)-B^{(i)}(n-s)\bigr)\\
&\quad\le \left|\sum_{s=1}^{b_n-1} w_n(s)\frac{1}{n}\bar{B}_n^{(j)}\bigl(B^{(i)}(n)-B^{(i)}(n-s)\bigr)\right|\\
&\quad\le \sum_{s=1}^{b_n-1}\frac{1}{n^2}\bigl|B^{(j)}(n)\bigr|\bigl|B^{(i)}(n)-B^{(i)}(n-s)\bigr| \qquad \text{since } \bigl|w_n(s)\bigr|\le 1\\
&\quad\le \frac{1}{n^2}\bigl|B^{(j)}(n)\bigr|\sum_{s=1}^{b_n-1}\sup_{0\le m\le b_n}\bigl|B^{(i)}(n)-B^{(i)}(n-m)\bigr|\\
&\quad< \frac{1}{n^2}\bigl((1+\varepsilon)(2n\log\log n)^{1/2}\bigr)\\
&\qquad\times(1+\varepsilon)\left(2b_n\left(\log\frac{n}{b_n}+\log\log n\right)\right)^{1/2}\sum_{s=1}^{b_n-1}1 \qquad \text{by Lemma 2}\\
&\quad< 2^{1/2}(1+\varepsilon)^2\frac{1}{n^2}(n\log n)^{1/2}(4b_n\log n)^{1/2}b_n\\
&\quad< 2^{3/2}(1+\varepsilon)^2\left(\frac{b_n}{n}\right)^{1/2}n^{-1}b_n\log n\\
&\quad\to 0.
\end{aligned}
$$

6. Similar to the previous term, but exchanging the i and j indices,

$$
\sum_{s=1}^{b_n-1} w_n(s)\frac{1}{n}\bar{B}_n^{(i)}\bigl(B^{(j)}(n)-B^{(j)}(n-s)\bigr)\to 0 \qquad \text{with probability 1 as } n\to\infty.
$$

7.

$$
\begin{aligned}
&\sum_{s=1}^{b_n-1} w_n(s)\frac{1}{n}\bar{B}_n^{(i)}B^{(j)}(s)\\
&\quad\leq \left|\sum_{s=1}^{b_n-1} w_n(s)\frac{1}{n}\bar{B}_n^{(i)}B^{(j)}(s)\right|\\
&\quad\leq \sum_{s=1}^{b_n-1} \bigl|w_n(s)\bigr|\frac{1}{n}\bigl|\bar{B}_n^{(i)}\bigr|\bigl|B^{(j)}(s)\bigr|\\
&\quad\leq \frac{1}{n^2}\bigl|B^{(i)}(n)\bigr|\sum_{s=1}^{b_n-1}\bigl|B^{(j)}(s)\bigr| \qquad \text{since } \bigl|w_n(s)\bigr|\leq 1\\
&\quad< \frac{1}{n^2}(1+\varepsilon)(2n\log\log n)^{1/2}\sum_{s=1}^{b_n-1}\sup_{1\leq m\leq b_n}\bigl|B^{(j)}(m)\bigr| \qquad \text{by Lemma 2}\\
&\quad< \frac{1}{n^2}(1+\varepsilon)(2n\log\log n)^{1/2}\sup_{1\leq m\leq b_n}\bigl|B^{(j)}(m+0)-B^{(j)}(0)\bigr|\sum_{s=1}^{b_n-1}1\\
&\quad< \frac{b_n}{n^2}(1+\varepsilon)(2n\log\log n)^{1/2}\sup_{0\leq t\leq n-b_n}\sup_{0\leq m\leq b_n}\bigl|B^{(j)}(t+m)-B^{(j)}(t)\bigr|\\
&\quad< \frac{b_n}{n^2}(1+\varepsilon)(2n\log\log n)^{1/2}(1+\varepsilon)\left(2b_n\left(\log\frac{n}{b_n}+\log\log n\right)\right)^{1/2}\\
&\quad= 2^{3/2}(1+\varepsilon)^2\frac{b_n^{1/2}}{n^{1/2}}b_n n^{-1}\log n\\
&\quad\to 0.
\end{aligned}
$$

8. Similar to the previous term, by exchanging the i and j index,

$$\sum_{s=1}^{b_n-1} w_n(s)\frac{1}{n}\bar{B}_n^{(j)}B^{(i)}(s)\to 0 \qquad \text{w.p. 1 as } n\to\infty.$$

Since each term in (A.22) goes to 0, we get that

$$\widetilde{\Sigma}_{S,ij}\to 0 \qquad \text{w.p. 1 as } n\to\infty. \qquad \square$$

Lemma 13. *Let Conditions 1 and 2 hold. In addition, suppose there exists a constant $c\geq 1$ such that $\sum_n (b_n/n)^c<\infty$, $n>2b_n$, $b_n n^{-1}\log n\to 0$, and $b_n n^{-1}\sum_{k=1}^{b_n} k|\Delta_1 w_n(k)|\to 0$, then $\widetilde{\Sigma}_{w,n}\to I_p$ w.p. 1 as $n\to\infty$ where I_p is the $p\times p$ identity matrix.*

Proof. The result follows from Lemmas 5, 7 and 12. □

The following corollary is an immediate consequence of the previous lemma.

Corollary 3. *Under the conditions of Lemma 13,* $L\widetilde{\Sigma}_{w,n}L^T \to LL^T = \Sigma$ *w.p.* 1 *as* $n \to \infty$.

Lemma 14. *Suppose* (2.3) *holds and Conditions* 1 *and* 2 *hold. If as* $n \to \infty$,

$$b_n \psi(n)^2 \log n \left(\sum_{k=1}^{b_n} \left| \Delta_2 w_n(k) \right| \right)^2 \to 0 \quad \textit{and} \tag{A.24}$$

$$\psi(n)^2 \sum_{k=1}^{b_n} \left| \Delta_2 w_n(k) \right| \to 0, \tag{A.25}$$

then $\widehat{\Sigma}_{w,n} \to \Sigma$ *w.p.* 1.

Proof. For $i, j = 1, \ldots, p$, let Σ_{ij} and $\widehat{\Sigma}_{w,ij}$ denote the (i, j)th element of Σ and $\widehat{\Sigma}_{w,n}$, respectively. Recall

$$\widehat{\Sigma}_{w,ij} = \frac{1}{n} \sum_{l=0}^{n-b_n} \sum_{k=1}^{b_n} k^2 \Delta_2 w_n(k) \big[\bar{Y}_l^{(i)}(k) - \bar{Y}_n^{(i)} \big] \big[\bar{Y}_l^{(j)}(k) - \bar{Y}_n^{(j)} \big].$$

We have

$$\begin{aligned}
&|\widehat{\Sigma}_{w,ij} - \Sigma_{ij}| \\
&\quad = \left| \frac{1}{n} \sum_{l=0}^{n-b_n} \sum_{k=1}^{b_n} k^2 \Delta_2 w_n(k) \big[\bar{Y}_l^{(i)}(k) - \bar{Y}_n^{(i)} \big] \big[\bar{Y}_l^{(j)}(k) - \bar{Y}_n^{(j)} \big] - \Sigma_{ij} \right| \\
&\quad = \left| \frac{1}{n} \sum_{l=0}^{n-b_n} \sum_{k=1}^{b_n} k^2 \Delta_2 w_n(k) \big[\bar{Y}_l^{(i)}(k) - \bar{Y}_n^{(i)} \pm \bar{C}_l^{(i)}(k) \pm \bar{C}_n^{(i)} \big] \right. \\
&\qquad \left. \times \big[\bar{Y}_l^{(j)}(k) - \bar{Y}_n^{(j)} \pm \bar{C}_l^{(j)}(k) \pm \bar{C}_n^{(j)} \big] - \Sigma_{ij} \right| \\
&\quad = \left| \frac{1}{n} \sum_{l=0}^{n-b_n} \sum_{k=1}^{b_n} k^2 \Delta_2 w_n(k) \big[\big(\bar{Y}_l^{(i)}(k) - \bar{C}_l^{(i)}(k) \big) + \big(\bar{C}_l^{(i)}(k) - \bar{C}_n^{(i)} \big) - \big(\bar{Y}_n^{(i)} - \bar{C}_n^{(i)} \big) \big] \right. \\
&\qquad \left. \times \big[\big(\bar{Y}_l^{(j)}(k) - \bar{C}_l^{(j)}(k) \big) + \big(\bar{C}_l^{(j)}(k) - \bar{C}_n^{(j)} \big) - \big(\bar{Y}_n^{(j)} - \bar{C}_n^{(j)} \big) \big] - \Sigma_{ij} \right|
\end{aligned} \tag{A.26}$$

$$
\begin{aligned}
&\leq \left| \frac{1}{n} \sum_{l=0}^{n-b_n} \sum_{k=1}^{b_n} k^2 \Delta_2 w_n(k) \big(\bar{C}_l^{(i)}(k) - \bar{C}_n^{(i)}\big)\big(\bar{C}_l^{(j)}(k) - \bar{C}_n^{(j)}\big) - \Sigma_{ij} \right| \\
&\quad + \frac{1}{n} \sum_{l=0}^{n-b_n} \sum_{k=1}^{b_n} k^2 \big|\Delta_2 w_n(k)\big| \big[\big| \big(\bar{Y}_l^{(i)}(k) - \bar{C}_l^{(i)}(k)\big)\big(\bar{Y}_l^{(j)}(k) - \bar{C}_l^{(j)}(k)\big) \big| \\
&\quad + \big| \big(\bar{Y}_l^{(i)}(k) - \bar{C}_l^{(i)}(k)\big)\big(\bar{Y}_n^{(j)} - \bar{C}_n^{(j)}\big) \big| \\
&\quad + \big| \big(\bar{Y}_l^{(i)}(k) - \bar{C}_l^{(i)}(k)\big)\big(\bar{C}_l^{(j)}(k) - \bar{C}_n^{(j)}\big) \big| + \big| \big(\bar{Y}_n^{(i)} - \bar{C}_n^{(i)}\big)\big(\bar{Y}_n^{(j)} - \bar{C}_n^{(j)}\big) \big| \\
&\quad + \big| \big(\bar{Y}_n^{(i)} - \bar{C}_n^{(i)}\big)\big(\bar{Y}_l^{(j)}(k) - \bar{C}_l^{(j)}(k)\big) \big| + \big| \big(\bar{Y}_n^{(i)} - \bar{C}_n^{(i)}\big)\big(\bar{C}_l^{(j)}(k) - \bar{C}_n^{(j)}\big) \big| \\
&\quad + \big| \big(\bar{C}_l^{(i)}(k) - \bar{C}_n^{(i)}\big)\big(\bar{Y}_l^{(j)}(k) - \bar{C}_l^{(j)}(k)\big) \big| + \big| \big(\bar{C}_l^{(i)}(k) - \bar{C}_n^{(i)}\big)\big(\bar{Y}_n^{(j)} - \bar{C}_n^{(j)}\big) \big| \big].
\end{aligned}
$$

We will show that each of the nine terms in (A.26) goes to 0 with probability 1 as $n \to \infty$. To do that, let us first establish a useful inequality. From (2.3), for any component i, and sufficiently large n,

$$
\left| \sum_{t=1}^{n} Y_t^{(i)} - C^{(i)}(n) \right| < D\psi(n). \tag{A.27}
$$

1. $|\frac{1}{n} \sum_{l=0}^{n-b_n} \sum_{k=1}^{b_n} k^2 \Delta_2 w_n(k) (\bar{C}_l^{(i)}(k) - \bar{C}_n^{(i)})(\bar{C}_l^{(j)}(k) - \bar{C}_n^{(j)}) - \Sigma_{ij}|$.
Notice that

$$
\frac{1}{n} \sum_{l=0}^{n-b_n} \sum_{k=1}^{b_n} k^2 \Delta_2 w_n(k) \big(\bar{C}_l^{(i)}(k) - \bar{C}_n^{(i)}\big)\big(\bar{C}_l^{(j)}(k) - \bar{C}_n^{(j)}\big),
$$

is equivalent to the ijth entry in $L \widetilde{\Sigma}_{w,n} L^T$. Then by Corollary 3

$$
\left| \frac{1}{n} \sum_{l=0}^{n-b_n} \sum_{k=1}^{b_n} k^2 \Delta_2 w_n(k) \big(\bar{C}_l^{(i)}(k) - \bar{C}_n^{(i)}\big)\big(\bar{C}_l^{(j)}(k) - \bar{C}_n^{(j)}\big) - \Sigma_{ij} \right| \to 0 \qquad \text{as } n \to \infty \text{ w.p. 1.}
$$

2. $\frac{1}{n} \sum_{l=0}^{n-b_n} \sum_{k=1}^{b_n} k^2 |\Delta_2 w_n(k)| |(\bar{Y}_l^{(i)}(k) - \bar{C}_l^{(i)}(k))(\bar{Y}_l^{(j)}(k) - \bar{C}_l^{(j)}(k))|$.
Note that for any component i,

$$
\begin{aligned}
\big|k\big(\bar{Y}_l^{(i)}(k) - \bar{C}_l^{(i)}(k)\big)\big| &= \left| \sum_{t=1}^{k} Y_{l+t}^{(i)} - C^{(i)}(k+l) + C^{(i)}(l) \right| \\
&= \left| \sum_{t=1}^{k+l} Y_t^{(i)} - \sum_{t=1}^{l} Y_t^{(i)} - C^{(i)}(k+l) + C^{(i)}(l) \right| \\
&< \left| \sum_{t=1}^{l+k} Y_t^{(i)} - C^{(i)}(k+l) \right| + \left| \sum_{t=1}^{l} Y_t^{(i)} - C^{(i)}(l) \right|
\end{aligned} \tag{A.28}
$$

$$\begin{aligned}&\le D\psi(l+k)+D\psi(l) \qquad \text{by (A.27)}\\ &\le 2D\psi(n) \qquad \text{since } l+k\le n.\end{aligned}$$

By (A.28),

$$\begin{aligned}&\frac{1}{n}\sum_{l=0}^{n-b_n}\sum_{k=1}^{b_n}k^2\bigl|\Delta_2 w_n(k)\bigr|\bigl|\bigl(\bar{Y}_l^{(i)}(k)-\bar{C}_l^{(i)}(k)\bigr)\bigl(\bar{Y}_l^{(j)}(k)-\bar{C}_l^{(j)}(k)\bigr)\bigr|\\ &<\frac{1}{n}\sum_{l=0}^{n-b_n}\sum_{k=1}^{b_n}\bigl|\Delta_2 w_n(k)\bigr|\bigl(2D\psi(n)\bigr)^2\\ &=4D^2\biggl(\frac{n-b_n+1}{n}\biggr)\psi(n)^2\sum_{k=1}^{b_n}\bigl|\Delta_2 w_n(k)\bigr|\\ &\to 0 \qquad \text{as } n\to\infty \text{ with probability 1.}\end{aligned}$$

3. $\frac{1}{n}\sum_{l=0}^{n-b_n}\sum_{k=1}^{b_n}k^2|\Delta_2 w_n(k)||(\bar{Y}_l^{(i)}(k)-\bar{C}_l^{(i)}(k))(\bar{Y}_n^{(j)}-\bar{C}_n^{(j)})|$.
Note that for any component i, using (A.27),

$$\bigl|\bar{Y}_n^{(i)}-\bar{C}_n^{(i)}\bigr|=\frac{1}{n}\biggl|\sum_{t=1}^{n}Y_t^{(i)}-C^{(i)}(n)\biggr|<\frac{1}{n}D\psi(n). \tag{A.29}$$

By (A.28) and (A.29),

$$\begin{aligned}&\frac{1}{n}\sum_{l=0}^{n-b_n}\sum_{k=1}^{b_n}k^2\bigl|\Delta_2 w_n(k)\bigr|\bigl|\bigl(\bar{Y}_l^{(i)}(k)-\bar{C}_l^{(i)}(k)\bigr)\bigl(\bar{Y}_n^{(j)}-\bar{C}_n^{(j)}\bigr)\bigr|\\ &<\frac{1}{n}\sum_{l=0}^{n-b_n}\sum_{k=1}^{b_n}k\bigl|\Delta_2 w_n(k)\bigr|\bigl(2D\psi(n)\bigr)\biggl(\frac{1}{n}D\psi(n)\biggr)\\ &=2D^2\psi(n)^2\biggl(\frac{n-b_n+1}{n}\biggr)\frac{1}{n}\sum_{k=1}^{b_n}k\bigl|\Delta_2 w_n(k)\bigr|\\ &\le 2D^2\psi(n)^2\biggl(\frac{n-b_n+1}{n}\biggr)\frac{b_n}{n}\sum_{k=1}^{b_n}\bigl|\Delta_2 w_n(k)\bigr|\\ &\to 0 \qquad \text{as } n\to\infty \text{ with probability 1.}\end{aligned}$$

4. Now

$$\frac{1}{n}\sum_{l=0}^{n-b_n}\sum_{k=1}^{b_n}k^2\bigl|\Delta_2 w_n(k)\bigr|\bigl|\bigl(\bar{Y}_l^{(i)}(k)-\bar{C}_l^{(i)}(k)\bigr)\bigl(\bar{C}_l^{(j)}(k)-\bar{C}_n^{(j)}\bigr)\bigr|$$

$$\leq \frac{1}{n}\sum_{l=0}^{n-b_n}\sum_{k=1}^{b_n} k^2\big|\Delta_2 w_n(k)\big|\big|\big(\bar{Y}_l^{(i)}(k)-\bar{C}_l^{(i)}(k)\big)\bar{C}_l^{(j)}(k)\big| + \frac{1}{n}\sum_{l=0}^{n-b_n}\sum_{k=1}^{b_n} k^2\big|\Delta_2 w_n(k)\big|\big|\big(\bar{Y}_l^{(i)}(k)-\bar{C}_l^{(i)}(k)\big)\bar{C}_n^{(j)}\big|.$$

We will show that both parts of the sum converge to 0 with probability 1 as $n \to \infty$. Consider the first sum.

$$\begin{aligned}
&\frac{1}{n}\sum_{l=0}^{n-b_n}\sum_{k=1}^{b_n} k^2\big|\Delta_2 w_n(k)\big|\big|\big(\bar{Y}_l^{(i)}(k)-\bar{C}_l^{(i)}(k)\big)\bar{C}_l^{(j)}(k)\big| \\
&\quad\leq \frac{1}{n}\sum_{l=0}^{n-b_n}\sum_{k=1}^{b_n} k^2\big|\Delta_2 w_n(k)\big|\big|\bar{Y}_l^{(i)}(k)-\bar{C}_l^{(i)}(k)\big|\big|\bar{C}_l^{(j)}(k)\big| \\
&\quad\leq \frac{1}{n}\sum_{l=0}^{n-b_n}\sum_{k=1}^{b_n} k\big|\Delta_2 w_n(k)\big|\big(2D\psi(n)\big)\left(2(1+\varepsilon)\sqrt{b_n\Sigma_{ii}}\frac{1}{k}(\log n)^{1/2}\right) \qquad \text{by (A.2) and (A.28)} \\
&\quad= \left(\frac{n-b_n+1}{n}\right)4D(1+\varepsilon)\sqrt{\Sigma_{ii}b_n\log n}\,\psi(n)\sum_{k=1}^{b_n}\big|\Delta_2 w_n(k)\big| \\
&\quad\to 0 \qquad \text{by Condition 2 and (A.24).}
\end{aligned}$$

The second part is

$$\begin{aligned}
&\frac{1}{n}\sum_{l=0}^{n-b_n}\sum_{k=1}^{b_n} k^2\big|\Delta_2 w_n(k)\big|\big|\big(\bar{Y}_l^{(i)}(k)-\bar{C}_l^{(i)}(k)\big)\bar{C}_n^{(j)}\big| \\
&\quad= \frac{1}{n}\sum_{l=0}^{n-b_n}\sum_{k=1}^{b_n} k^2\big|\Delta_2 w_n(k)\big|\big|\bar{Y}_l^{(i)}(k)-\bar{C}_l^{(i)}(k)\big|\big|\bar{C}_n^{(j)}\big| \\
&\quad\leq \frac{1}{n}\sum_{l=0}^{n-b_n}\sum_{k=1}^{b_n} k\big|\Delta_2 w_n(k)\big|\big(2D\psi(n)\big)\left(\frac{1}{n}(1+\varepsilon)[2n\Sigma_{ii}\log\log n]^{1/2}\right) \qquad \text{by (A.28) and (A.1)} \\
&\quad< \left(\frac{n-b_n+1}{n}\right)2\sqrt{2\Sigma_{ii}}D(1+\varepsilon)\psi(n)\frac{(n\log n)^{1/2}}{n}\sum_{k=1}^{b_n} k\big|\Delta_2 w_n(k)\big| \\
&\quad< \left(\frac{n-b_n+1}{n}\right)2\sqrt{2\Sigma_{ii}}D(1+\varepsilon)\psi(n)\frac{(\log n)^{1/2}}{n^{1/2}}b_n\sum_{k=1}^{b_n}\big|\Delta_2 w_n(k)\big|
\end{aligned}$$

$$< \left(\frac{n-b_n+1}{n}\right) 2\sqrt{2\Sigma_{ii}} D(1+\varepsilon)\psi(n)(b_n \log n)^{1/2} \frac{b_n^{1/2}}{n^{1/2}} \sum_{k=1}^{b_n} \bigl|\Delta_2 w_n(k)\bigr|$$

$\to 0$ by Condition 2 and (A.24).

5. Next,

$$\frac{1}{n} \sum_{l=0}^{n-b_n} \sum_{k=1}^{b_n} k^2 \bigl|\Delta_2 w_n(k)\bigr| \bigl|\bigl(\bar{Y}_n^{(i)} - \bar{C}_n^{(i)}\bigr)\bigl(\bar{Y}_n^{(j)} - \bar{C}_n^{(j)}\bigr)\bigr|$$

$$< \frac{1}{n} \sum_{l=0}^{n-b_n} \sum_{k=1}^{b_n} k^2 \bigl|\Delta_2 w_n(k)\bigr| \frac{1}{n^2} D^2 \psi(n)^2 \qquad \text{by (A.29)}$$

$$\leq \left(\frac{n-b_n+1}{n}\right) D^2 \frac{b_n^2}{n^2} \psi(n)^2 \sum_{k=1}^{b_n} \bigl|\Delta_2 w_n(k)\bigr|$$

$\to 0$ by Condition 2 and (A.25).

6.

$$\frac{1}{n} \sum_{l=0}^{n-b_n} \sum_{k=1}^{b_n} k^2 \bigl|\Delta_2 w_n(k)\bigr| \bigl|\bigl(\bar{Y}_n^{(i)} - \bar{C}_n^{(i)}\bigr)\bigl(\bar{Y}_l^{(j)}(k) - \bar{C}_l^{(j)}(k)\bigr)\bigr|$$

$$< \frac{1}{n} \sum_{l=0}^{n-b_n} \sum_{k=1}^{b_n} k \bigl|\Delta_2 w_n(k)\bigr| \left(\frac{1}{n} D\psi(n)\right)\bigl(2D\psi(n)\bigr)$$

$$< \left(\frac{n-b_n+1}{n}\right) 2D^2 \psi(n)^2 \frac{1}{n} \sum_{k=1}^{b_n} k \bigl|\Delta_2 w_n(k)\bigr|$$

$$< \left(\frac{n-b_n+1}{n}\right) 2D^2 \psi(n)^2 \frac{b_n}{n} \sum_{k=1}^{b_n} \bigl|\Delta_2 w_n(k)\bigr|$$

$\to 0$ by Condition 2 and (A.25).

7.

$$\frac{1}{n} \sum_{l=0}^{n-b_n} \sum_{k=1}^{b_n} k^2 \bigl|\Delta_2 w_n(k)\bigr| \bigl|\bigl(\bar{Y}_n^{(i)} - \bar{C}_n^{(i)}\bigr)\bigl(\bar{C}_l^{(j)}(k) - \bar{C}_n^{(j)}\bigr)\bigr|$$

$$\leq \frac{1}{n} \sum_{l=0}^{n-b_n} \sum_{k=1}^{b_n} k^2 \bigl|\Delta_2 w_n(k)\bigr| \bigl|\bigl(\bar{Y}_n^{(i)} - \bar{C}_n^{(i)}\bigr)\bar{C}_l^{(j)}(k)\bigr|$$

$$+ \frac{1}{n} \sum_{l=0}^{n-b_n} \sum_{k=1}^{b_n} k^2 \bigl|\Delta_2 w_n(k)\bigr| \bigl|\bigl(\bar{Y}_n^{(i)} - \bar{C}_n^{(i)}\bigr)\bar{C}_n^{(j)}\bigr|.$$

We will show that each of the two terms goes to 0 with probability 1 as $n \to \infty$.

$$\begin{aligned}
&\frac{1}{n}\sum_{l=0}^{n-b_n}\sum_{k=1}^{b_n} k^2 \big|\Delta_2 w_n(k)\big|\big|\big(\bar{Y}_n^{(i)} - \bar{C}_n^{(i)}\big)\bar{C}_l^{(j)}(k)\big| \\
&\quad < \frac{1}{n}\sum_{l=0}^{n-b_n}\sum_{k=1}^{b_n} k^2 \big|\Delta_2 w_n(k)\big|\left(\frac{1}{n}D\psi(n)\right)\left(2(1+\varepsilon)\sqrt{b_n \Sigma_{ii}}\frac{1}{k}(\log n)^{1/2}\right) \\
&\qquad\qquad \text{by (A.2) and (A.29)} \\
&\quad < \left(\frac{n-b_n+1}{n}\right)2D(1+\varepsilon)\sqrt{\Sigma_{ii}}\psi(n)\frac{\sqrt{b_n \log n}}{n}\sum_{k=1}^{b_n} k\big|\Delta_2 w_n(k)\big| \\
&\quad \le \left(\frac{n-b_n+1}{n}\right)2D(1+\varepsilon)\sqrt{\Sigma_{ii}}\frac{b_n}{n}\sqrt{b_n \log n}\psi(n)\sum_{k=1}^{b_n}\big|\Delta_2 w_n(k)\big| \\
&\quad \to 0 \qquad \text{by Condition 2 and (A.24).}
\end{aligned}$$

For the second term,

$$\begin{aligned}
&\frac{1}{n}\sum_{l=0}^{n-b_n}\sum_{k=1}^{b_n} k^2 \big|\Delta_2 w_n(k)\big|\big|\big(\bar{Y}_n^{(i)} - \bar{C}_n^{(i)}\big)\bar{C}_n^{(j)}\big| \\
&\quad = \frac{1}{n}\sum_{l=0}^{n-b_n}\sum_{k=1}^{b_n} k^2 \big|\Delta_2 w_n(k)\big|\big|\big(\bar{Y}_n^{(i)} - \bar{C}_n^{(i)}\big)\big|\big|\bar{C}_n^{(j)}\big| \\
&\quad < \frac{1}{n}\sum_{l=0}^{n-b_n}\sum_{k=1}^{b_n} k^2 \big|\Delta_2 w_n(k)\big|\left(\frac{1}{n}D\psi(n)\right)\left(\frac{1}{n}(1+\varepsilon)[2n\Sigma_{ii}\log\log n]^{1/2}\right) \\
&\qquad\qquad \text{by (A.29) and (A.1)} \\
&\quad \le \left(\frac{n-b_n+1}{n}\right)\sqrt{2\Sigma_{ii}}D(1+\varepsilon)\frac{\sqrt{n\log n}\psi(n)}{n^2}\sum_{k=1}^{b_n} k^2\big|\Delta_2 w_n(k)\big| \\
&\quad \le \left(\frac{n-b_n+1}{n}\right)\sqrt{2\Sigma_{ii}}D(1+\varepsilon)\frac{b_n^2\sqrt{n\log n}\psi(n)}{n^2}\sum_{k=1}^{b_n}\big|\Delta_2 w_n(k)\big| \\
&\quad \le \left(\frac{n-b_n+1}{n}\right)\sqrt{2\Sigma_{ii}}D(1+\varepsilon)\frac{b_n}{n}\frac{b_n^{1/2}}{n^{1/2}}(b_n\log n)^{1/2}\psi(n)\sum_{k=1}^{b_n}\big|\Delta_2 w_n(k)\big| \\
&\quad \to 0 \qquad \text{by Condition 2 and (A.24).}
\end{aligned}$$

8.

$$\frac{1}{n}\sum_{l=0}^{n-b_n}\sum_{k=1}^{b_n} k^2\big|\Delta_2 w_n(k)\big|\big|\big(\bar{C}_l^{(i)}(k)-\bar{C}_n^{(i)}\big)\big(\bar{Y}_l^{(i)}(k)-\bar{C}_l^{(j)}(k)\big)\big|.$$

This term is the same as term 4 except for a change of components. Thus, the same argument can be used to show that it converges to 0 with probability 1 as $n\to\infty$.

9.

$$\frac{1}{n}\sum_{l=0}^{n-b_n}\sum_{k=1}^{b_n} k^2\big|\Delta_2 w_n(k)\big|\big|\big(\bar{C}_l^{(i)}(k)-\bar{C}_n^{(i)}\big)\big(\bar{Y}_n^{(j)}-\bar{C}_n^{(j)}\big)\big|.$$

This term is the same as term 7 except for a change of components. Thus the same argument can be used to show that it converges to 0 w.p. 1 as $n\to\infty$.

Since each of the nine terms converges to 0 with probability 1, $|\widehat{\Sigma}_{ij}-\Sigma_{ij}|\to 0$ as $n\to\infty$ with probability 1. □

Since we proved that $\widehat{\Sigma}_S=\widehat{\Sigma}_{w,n}+d_n\to\Sigma+0$ as $n\to\infty$ with probability 1, we have the desired result for Theorem 1.

A.4. Proof of Theorem 2

Let $S=\{S_t\}_{t\ge1}$ be a strictly stationary stochastic process on a probability space $(\Omega,\mathcal{F},P)$ and set $\mathcal{F}_k^l=\sigma(S_k,\dots,S_l)$. Define the α-mixing coefficients for $n=1,2,3,\dots$ as

$$\alpha(n)=\sup_{k\ge1}\sup_{A\in\mathcal{F}_1^k,B\in\mathcal{F}_{k+n}^\infty}\big|P(A\cap B)-P(A)P(B)\big|.$$

The process S is said to be *strongly mixing* if $\alpha(n)\to0$ as $n\to\infty$. It is easy to see that Harris ergodic Markov chains are strongly mixing; see, for example, Jones [35].

Theorem 5 (Kuelbs and Philipp [42]). *Let $f(S_1), f(S_2),\dots$ be an $\mathbb{R}^p$-valued stationary process such that $E_F\|f\|^{2+\delta}<\infty$ for some $0<\delta\le1$. Let $\alpha_f(n)$ be the mixing coefficients of the process $\{f(S_t)\}_{t\ge1}$ and suppose, as $n\to\infty$,*

$$\alpha_f(n)=O\big(n^{-(1+\varepsilon)(1+2/\delta)}\big)\qquad \text{for } \varepsilon>0.$$

Then there exists a p-vector θ_f, a $p\times p$ lower triangular matrix L_f, and a finite random variable D_f, such that, with probability 1,

$$\left\|\sum_{t=1}^{n} f(X_t)-n\theta_f-L_fB(n)\right\|<D_fn^{1/2-\lambda_f} \tag{A.30}$$

for some $\lambda_f>0$ depending on ε, δ, and p only.

Corollary 4. *Let $E_F\|f\|^{2+\delta} < \infty$ for some $\delta > 0$. If X is a polynomially ergodic Markov chain of order $\xi \geq (1+\varepsilon)(1+2/\delta)$ for some $\varepsilon > 0$, then (A.30) holds for any initial distribution.*

Proof. Let α be the mixing coefficient for the Markov chain $X = \{X_t\}_{t\geq 1}$ and α_f be the mixing coefficient for the mapped process $\{f(X_t)\}_{t\geq 1}$. Then the elementary properties of sigma-algebras [cf. Chow and Teicher [8], p. 16] shows that $\alpha_f(n) \leq \alpha(n)$ for all n. Since X is polynomially ergodic of order ξ we also have that $\alpha(n) \leq \mathrm{E}_F M n^{-\xi}$ for all n and hence if $\xi \geq (1+\varepsilon)(1+2/\delta)$, then $\alpha_f(n) \leq \mathrm{E}_F M n^{-\xi} = O(n^{-(1+\varepsilon)(1+2/\delta)})$. The result follows from Theorem 5 and thus the strong invariance principle as stated, holds at stationarity. A standard Markov chain argument (see, e.g., Proposition 17.1.6 in Meyn and Tweedie [45]) shows that if the result holds for any initial distribution, then it holds for every initial distribution. □

Proof of Theorem 2. Since $\mathrm{E}_F\|g\|^{4+\delta} < \infty$ implies $\mathrm{E}_F\|g\|^{2+\delta} < \infty$ and X is a polynomially ergodic Markov chain of order $\xi \geq (1+\varepsilon)(1+2/\delta)$ we have from Corollary 4 that an SIP holds such that

$$\left\| \sum_{t=1}^{n} g(X_t) - n\theta - LB(n) \right\| < Dn^{1/2-\lambda_g}$$

for some $\lambda_g > 0$ depending on ε, δ, and p only.

Since $\mathrm{E}_F\|g\|^{4+\delta} < \infty$ implies $\mathrm{E}_F\|h\|^{2+\delta} < \infty$ and X is a polynomially ergodic Markov chain of order $\xi \geq (1+\varepsilon)(1+2/\delta)$ we have from Corollary 4 that an SIP holds such that

$$\left\| \sum_{t=1}^{n} h(X_t) - n\theta_h - L_h B(n) \right\| < D_h n^{1/2-\lambda_h}$$

for some $\lambda_h > 0$ depending on ε, δ, and p only.

Setting $\lambda = \min\{\lambda_g, \lambda_h\}$ shows that (2.3) and (2.4) hold with

$$\psi(n) = \psi_h(n) = n^{1/2-\lambda}.$$

The rest now follows easily from Theorem 1. □

Acknowledgments

The authors thank Tiefeng Jiang and Gongjun Xu for helpful discussions. The second author's work is partially supported by NSF DMS-13-08270. The third author's work is partially supported by NSF DMS-13-10096 and NIH NIBIB R01 EB012547.

References

[1] Acosta, F., Huber, M.L. and Jones, G.L. (2015). Markov chain Monte Carlo with linchpin variables. Preprint.

[2] Anderson, T.W. (1971). *The Statistical Analysis of Time Series*. New York–London–Sydney: Wiley. MR0283939
[3] Andrews, D.W.K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59** 817–858. MR1106513
[4] Atchadé, Y.F. (2011). Kernel estimators of asymptotic variance for adaptive Markov chain Monte Carlo. *Ann. Statist.* **39** 990–1011. MR2816345
[5] Berkes, I. and Philipp, W. (1979). Approximation theorems for independent and weakly dependent random vectors. *Ann. Probab.* **7** 29–54. MR0515811
[6] Billingsley, P. (2008). *Probability and Measure*. New York: Wiley.
[7] Choi, H.M. and Hobert, J.P. (2013). The Polya-Gamma Gibbs sampler for Bayesian logistic regression is uniformly ergodic. *Electron. J. Stat.* **7** 2054–2064. MR3091616
[8] Chow, Y.S. and Teicher, H. (1978). *Probability Theory*: *Independence*, *Interchangeability*, *Martingales*. New York–Heidelberg: Springer. MR0513230
[9] Csörgő, M. and Révész, P. (1981). *Strong Approximations in Probability and Statistics*. *Probability and Mathematical Statistics*. New York–London: Academic Press. MR0666546
[10] Damerdji, H. (1991). Strong consistency and other properties of the spectral variance estimator. *Manage. Sci.* **37** 1424–1440.
[11] de Jong, R.M. (2000). A strong consistency proof for heteroskedasticity and autocorrelation consistent covariance matrix estimators. *Econometric Theory* **16** 262–268. MR1763435
[12] Dehling, H. and Philipp, W. (1982). Almost sure invariance principles for weakly dependent vector-valued random variables. *Ann. Probab.* **10** 689–701. MR0659538
[13] Doss, C.R., Flegal, J.M., Jones, G.L. and Neath, R.C. (2014). Markov chain Monte Carlo estimation of quantiles. *Electron. J. Stat.* **8** 2448–2478. MR3285872
[14] Doss, H. and Hobert, J.P. (2010). Estimation of Bayes factors in a class of hierarchical random effects models using a geometrically ergodic MCMC algorithm. *J. Comput. Graph. Statist.* **19** 295–312. MR2675092
[15] Eberlein, E. (1986). On strong invariance principles under dependence assumptions. *Ann. Probab.* **14** 260–270. MR0815969
[16] Einmahl, U. (1989). Extensions of results of Komlós, Major, and Tusnády to the multivariate case. *J. Multivariate Anal.* **28** 20–68. MR0996984
[17] Flegal, J.M. and Gong, L. (2015). Relative fixed-width stopping rules for Markov chain Monte Carlo simulations. *Statist. Sinica* **25** 655–675. MR3379093
[18] Flegal, J.M., Haran, M. and Jones, G.L. (2008). Markov chain Monte Carlo: Can we trust the third significant figure? *Statist. Sci.* **23** 250–260. MR2516823
[19] Flegal, J.M. and Jones, G.L. (2010). Batch means and spectral variance estimators in Markov chain Monte Carlo. *Ann. Statist.* **38** 1034–1070. MR2604704
[20] Flegal, J.M. and Jones, G.L. (2011). Implementing MCMC: Estimating with confidence. In *Handbook of Markov Chain Monte Carlo* (S. Brooks, A. Gelman, X.L. Meng and G.L. Jones, eds.). Boca Raton, FL: Chapman & Hall.
[21] Fort, G. and Moulines, E. (2003). Polynomial ergodicity of Markov transition kernels. *Stochastic Process. Appl.* **103** 57–99. MR1947960
[22] Franklin, J.N. (2012). *Matrix Theory*. Courier Corporation.
[23] Geyer, C.J. (1992). Practical Markov chain Monte Carlo (with discussion). *Statist. Sci.* **7** 473–511.
[24] Geyer, C.J. (2011). Introduction to Markov chain Monte Carlo. In *Handbook of Markov Chain Monte Carlo* (S. Brooks, A. Gelman, G.L. Jones and X.L. Meng, eds.). Boca Raton, FL: Chapman & Hall.
[25] Glynn, P.W. and Whitt, W. (1992). The asymptotic validity of sequential stopping rules for stochastic simulations. *Ann. Appl. Probab.* **2** 180–198. MR1143399

[26] Gong, L. and Flegal, J.M. (2016). A practical sequential stopping rule for high-dimensional Markov chain Monte Carlo. *J. Comput. Graph. Statist.* **25** 684–700. MR3533633
[27] Hobert, J.P. and Geyer, C.J. (1998). Geometric ergodicity of Gibbs and block Gibbs samplers for a hierarchical random effects model. *J. Multivariate Anal.* **67** 414–430. MR1659196
[28] 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* **89** 731–743. MR1946508
[29] Horváth, L. (1984). Strong approximation of extended renewal processes. *Ann. Probab.* **12** 1149–1166. MR0757773
[30] Jarner, S.F. and Hansen, E. (2000). Geometric ergodicity of Metropolis algorithms. *Stochastic Process. Appl.* **85** 341–361. MR1731030
[31] Jarner, S.F. and Roberts, G.O. (2002). Polynomial convergence rates of Markov chains. *Ann. Appl. Probab.* **12** 224–247. MR1890063
[32] Jarner, S.F. and Roberts, G.O. (2007). Convergence of heavy-tailed Monte Carlo Markov chain algorithms. *Scand. J. Stat.* **34** 781–815. MR2396939
[33] Johnson, A.A. and Jones, G.L. (2015). Geometric ergodicity of random scan Gibbs samplers for hierarchical one-way random effects models. *J. Multivariate Anal.* **140** 325–342. MR3372572
[34] Johnson, L.T. and Geyer, C.J. (2012). Variable transformation to obtain geometric ergodicity in the random-walk Metropolis algorithm. *Ann. Statist.* **40** 3050–3076. MR3097969
[35] Jones, G.L. (2004). On the Markov chain central limit theorem. *Probab. Surv.* **1** 299–320. MR2068475
[36] Jones, G.L., Haran, M., Caffo, B.S. and Neath, R. (2006). Fixed-width output analysis for Markov chain Monte Carlo. *J. Amer. Statist. Assoc.* **101** 1537–1547. MR2279478
[37] Jones, G.L. and Hobert, J.P. (2001). Honest exploration of intractable probability distributions via Markov chain Monte Carlo. *Statist. Sci.* **16** 312–334. MR1888447
[38] Jones, G.L., Roberts, G.O. and Rosenthal, J.S. (2014). Convergence of conditional Metropolis–Hastings samplers. *Adv. in Appl. Probab.* **46** 422–445. MR3215540
[39] Kass, R.E., Carlin, B.P., Gelman, A. and Neal, R.M. (1998). Markov chain Monte Carlo in practice: A roundtable discussion. *Amer. Statist.* **52** 93–100. MR1628427
[40] Komlós, J., Major, P. and Tusnády, G. (1975). An approximation of partial sums of independent RV's and the sample DF. I. *Z. Wahrsch. Verw. Gebiete* **32** 111–131. MR0375412
[41] Kosorok, M.R. (2000). Monte Carlo error estimation for multivariate Markov chains. *Statist. Probab. Lett.* **46** 85–93. MR1731349
[42] Kuelbs, J. and Philipp, W. (1980). Almost sure invariance principles for partial sums of mixing B-valued random variables. *Ann. Probab.* **8** 1003–1036. MR0602377
[43] Liu, J.S. (2008). *Monte Carlo Strategies in Scientific Computing*. *Springer Series in Statistics*. New York: Springer. MR2401592
[44] Marchev, D. and Hobert, J.P. (2004). Geometric ergodicity of van Dyk and Meng's algorithm for the multivariate Student's t model. *J. Amer. Statist. Assoc.* **99** 228–238. MR2054301
[45] Meyn, S. and Tweedie, R.L. (2009). *Markov Chains and Stochastic Stability*, 2nd ed. Cambridge: Cambridge Univ. Press. MR2509253
[46] Mykland, P., Tierney, L. and Yu, B. (1995). Regeneration in Markov chain samplers. *J. Amer. Statist. Assoc.* **90** 233–241. MR1325131
[47] Petrone, S., Roberts, G.O. and Rosenthal, J.S. (1999). A note on convergence rates of Gibbs sampling for nonparametric mixtures. *Far East J. Theor. Stat.* **3** 213–225. MR1764864
[48] Philipp, W. and Stout, W.F. (1975). *Almost Sure Invariance Principles for Partial Sums of Weakly Dependent Random Variables* **161**. Providence, RI: AMS.
[49] Polson, N.G., Scott, J.G. and Windle, J. (2013). Bayesian inference for logistic models using Pólya–Gamma latent variables. *J. Amer. Statist. Assoc.* **108** 1339–1349. MR3174712

[50] Robert, C.P. and Casella, G. (2004). *Monte Carlo Statistical Methods*, 2nd ed. *Springer Texts in Statistics*. New York: Springer. MR2080278
[51] Roberts, G.O. and Rosenthal, J.S. (1999). Convergence of slice sampler Markov chains. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **61** 643–660. MR1707866
[52] Roberts, G.O. and Rosenthal, J.S. (2004). General state space Markov chains and MCMC algorithms. *Probab. Surv.* **1** 20–71. MR2095565
[53] Roberts, G.O. and Tweedie, R.L. (1996). Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms. *Biometrika* **83** 95–110. MR1399158
[54] Rosenthal, J.S. (1996). Analysis of the Gibbs sampler for a model related to James–Stein estimators. *Stat. Comput.* **6** 269–275.
[55] Roy, V. and Hobert, J.P. (2007). Convergence rates and asymptotic standard errors for Markov chain Monte Carlo algorithms for Bayesian probit regression. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **69** 607–623. MR2370071
[56] Strassen, V. (1964). An invariance principle for the law of the iterated logarithm. *Z. Wahrsch. Verw. Gebiete* **3** 211–226. MR0175194
[57] Tan, A. and Hobert, J.P. (2009). Block Gibbs sampling for Bayesian random effects models with improper priors: Convergence and regeneration. *J. Comput. Graph. Statist.* **18** 861–878. MR2598033
[58] Tan, A., Jones, G.L. and Hobert, J.P. (2013). On the geometric ergodicity of two-variable Gibbs samplers. In *Advances in Modern Statistical Theory and Applications*: *A Festschrift in Honor of Morris L. Eaton* (G.L. Jones and X. Shen, eds.) 25–42. Beachwood, OH: IMS.
[59] Tierney, L. (1994). Markov chains for exploring posterior distributions. *Ann. Statist.* **22** 1701–1762. MR1329166
[60] Tjøstheim, D. (1990). Non-linear time series and Markov chains. *Adv. in Appl. Probab.* **22** 587–611.
[61] Vats, D., Flegal, J.M. and Jones, G.L. (2016). Multivariate output analysis for Markov chain Monte Carlo. Preprint. Available at arXiv:1512.07713.
[62] Whittle, P. (1960). Bounds for the moments of linear and quadratic forms in independent variables. *Theory Probab. Appl.* **5** 302–305.
[63] Wu, W.B. (2007). Strong invariance principles for dependent random variables. *Ann. Probab.* **35** 2294–2320. MR2353389
[64] Zaitsev, A.Y. (1998). Multidimensional version of the results of Komlós, Major and Tusnády for vectors with finite exponential moments. *ESAIM Probab. Stat.* **2** 41–108 (electronic). MR1616527

Received January 2016 and revised June 2016