Statistica Sinica **28** (2018), 319-338
doi:https://doi.org/10.5705/ss.202016-0468

A THRESHOLDING-BASED PREWHITENED LONG-RUN VARIANCE ESTIMATOR AND ITS DEPENDENCE-ORACLE PROPERTY

Ting Zhang

Boston University

Abstract: Statistical inference of time series data routinely relies on the estimation of long-run variances, defined as the sum of autocovariances of all orders. The current paper considers a new class of long-run variance estimators that first soaks up the dependence by a decision-based prewhitening filter, then regularizes autocorrelations of the resulting residual process by thresholding, and finally recolors back to obtain an estimator of the original process. Under mild regularity conditions, we prove that the proposed estimator (i) consistently estimates the long-run variance; (ii) achieves the parametric convergence rate when the underlying process has a sparse dependence structure as in finite-order moving average models; and (iii) enjoys the dependence-oracle property in the sense that it automatically reduces to the sample variance if the data are actually independent. Monte Carlo simulations were conducted to examine its finite-sample performance and make comparisons with existing estimators.

Key words and phrases: Long-run variance, prewhitening, thresholding.

1. Introduction

An important problem in time series analysis is to provide appropriate corrections to standard errors for data with serial correlation. In this case, the marginal variance should be replaced by the long-run variance, which involves autocovariances of all orders and serves as a key quantity in statistical inference with dependent data; see for example Ibragimov and Linnik (1971), Peligrad (1996), Maxwell and Woodroofe (2000), Wu (2005), and Bradley (2007), among others. Unlike the case of independent data where one can simply estimate the variance constant by its sample version, directly plugging sample autocovariances into the definition does not yield a consistent estimator for the long-run variance (Bratley, Fox and Schrage, 1987; Anderson, 1994). Therefore, a more sophisticated estimation procedure is needed for dependent data. The problem has attracted considerable attention in the literature, and popular candidates include the batch means estimator that achieves consistency by splitting the data into overlapping or nonoverlapping batches, and the lag-window estimator that achieves consistency by banding or tapering sample autocovariances. Their asymptotic properties have been studied by Song and Schmeiser (1995), Bühlmann (2002), Lahiri (2003), Jones et al. (2006), and Flegal and Jones (2010) for batch means estimators, and by Rosenblatt (1985), Newey and West (1987), Andrews (1991), Liu and Wu (2010), and Politis (2011) for lag-window estimators; see also the references therein. Other contributions can be found in Fishman (1971), Crane and Iglehart (1975), Schruben (1983), Goldsman, Meketon and Schruben (1990), Law and Kelton (2000), Aktaran-Kalaycı, Goldsman and Wilson (2007), Alexopoulos et al. (2007), Wu (2009), Xiao and Wu (2011), and Meterelliyoz, Alexopoulos and Goldsman (2012), among others. Nevertheless, due to direct or indirect estimation of a large number of autocovariances, existing methods usually do not achieve the parametric convergence rate. For example, given a sample of size n, the optimal rate of convergence of the overlapping batch means (OBM) estimator is $n^{1/3}$, as given in Flegal and Jones (2010).

Recently, Paparoditis and Politis (2012) considered another type of estimator by plugging thresholded sample autocovariances into the definition, and proved that the resulting estimator consistently estimates the spectral density and thus the long-run variance (by setting the frequency to zero). In particular, by seeking the connection between eigenvalues of an autocovariance matrix and the associated spectral density, they provided a stochastic error bound for their thresholded estimator based on a result of Xiao and Wu (2012) concerning thresholded autocovariance matrix estimation. As commented by Xiao and Wu (2012), compared with the conventional banded estimator, the thresholded estimator is desirable in that it can lead to better performance when there are a lot of zeros or very weak autocovariances. They also commented that, due to technical difficulties, their theoretical result was not able to reflect this advantage. Since Paparoditis and Politis (2012) relied on the result of Xiao and Wu (2012), the same difficulty arose. Due to the nonlinear nature of thresholding, studying the asymptotic properties of thresholding-based estimators can be nontrivial; see also Bickel and Levina (2008) for the case with independent data. The current paper aims to serve as the first step in solving this open problem by focusing on the relatively simpler quantity, the long-run variance, defined as the sum of autocovariances of all orders. In particular, we prove that the proposed thresholding-based long-run variance estimator (i) achieves the parametric convergence rate when the underlying process has a sparse dependence structure as in finite-order moving average models; and (ii) enjoys the dependence-oracle property in the sense that it automatically reduces to the sample variance when applied to independent data. For the general case where the underlying process does not necessarily have a sparse dependence structure, our stochastic error bound also improves over the one implied by Paparoditis and Politis (2012) in a polynomial of n for long-run variance estimation.

Another aspect of thresholding-based estimators that the current paper aims to improve is their performance under moderate to pronounced dependence strength. Although the technique of thresholding is intuitively desirable for situations with a lot of zero or very weak autocorrelations, as advocated by Xiao and Wu (2012), it may not perform well under pronounced autocorrelations, especially when the sample size is small. In particular, Paparoditis and Politis (2012) examined the finite-sample performance of their thresholded estimator for first-order autoregressive models via Monte Carlo simulations, and found that the performance can deteriorate quickly as one increases the autoregression coefficient, namely the dependence strength. Therefore, the standard class of plug-in thresholded long-run variance estimators may be too restrictive, and we consider a new class of thresholding-based estimators that first soaks up the dependence by prewhitening the observed time series; then regularizes autocorrelations of the resulting residual process by thresholding; and finally recolors back to obtain an estimator of the original process. Our simulation results in Section 4 show that the proposed estimator can bring significant improvements over the plug-in method of Paparoditis and Politis (2012), especially when the dependence strength is stronger than the m-dependence structure as in finite-order moving-average models. Although prewhitening (Press and Tukey, 1956) has its practical value of potentially reducing the dependence strength and flattening the spectral density function, it can cause additional difficulties in establishing asymptotic properties of the resulting estimator; see for example Andrews and Monahan (1992), but with a focus on lag-window estimators that are linear as functions of sample autocovariances. We shall here focus on thresholded estimators that are nonlinear functions of sample autocovariances. Note that prewhitening is not necessary and in fact not preferred if the observed sequence is already a white noise series, we propose to incorporate an additional decision rule into the prewhitening step to further improve its performance.

The remainder of the paper is organized as follows. Section 2 introduces the proposed thresholding-integrated prewhitening-sandwitched (TIPS) estimator for the long-run variance. Its asymptotic properties are studied in Section

3 with proofs provided in Section 6. Monte Carlo simulations are reported in Section 4; they examine the finite-sample performance of the proposed estimator and make comparisons with existing estimators.

2. Long-Run Variance Estimation by TIPS

Let $(X_i)_{i\in\mathbb{Z}}$ be a stationary process with mean $\mu_X = E(X_0)$ and autocovariances $\gamma_{X,k} = \text{cov}(X_0, X_k)$, $k \in \mathbb{Z}$, then its long-run variance or time-average variance constant is defined as

$$g_X = \sum_{k\in\mathbb{Z}} \gamma_{X,k}, \tag{2.1}$$

which is typically unknown in practice and needs to be estimated from $X_1, \ldots, X_n$. We shall in the following suppress the subscript X for quantities related to the process (X_i), and use g and γ_k to denote g_X and $\gamma_{X,k}$, respectively. Let $\bar{X}_n = n^{-1}\sum_{i=1}^n X_i$ denote the sample mean, the major goal of the current paper is to propose and study a new class of long-run variance estimators, which can be obtained as follows. First, one estimates a first-order autoregressive model for (X_i) as

$$(X_i - \bar{X}_n) = \tilde{\varphi}(X_{i-1} - \bar{X}_n) + \tilde{V}_i, \quad i = 2, \ldots, n, \tag{2.2}$$

where $\tilde{\varphi}$ is the parameter estimator which could be obtained by solving least squares, and $(\tilde{V}_i)$ denotes the resulting residual process or the transformed process. The autoregressive model (2.2) is not meant to represent the belief in the underlying data generating mechanism, but is used as a tool to soak up the temporal dependence in the original process (X_i) so that the residual process $(\tilde{V}_i)$ can be closer to a white noise sequence. If (X_i) is already a white noise sequence, then the prewhitening (2.2) is not necessary and may introduce additional randomness into the final estimator by passing the randomness of $\tilde{\varphi}$; see for example Sul, Phillips and Choi (2005). To alleviate the problem, we propose to use the thresholding-integrated prewhitening

$$\hat{V}_i = (X_i - \bar{X}_n) - \hat{\varphi}(X_{i-1} - \bar{X}_n), \quad \hat{\varphi} = \tilde{\varphi}\mathbb{1}_{\{\tilde{\varphi}\geq\tau_n\}},$$

where τ_n is an appropriate threshold of significance. Second, one computes the sample autocovariances of the residual process $(\hat{V}_i)$ as

$$\hat{\gamma}_{\hat{V},k} = \frac{1}{n-1}\sum_{i=2}^{n-|k|}(\hat{V}_i - \bar{\hat{V}}_{n-1})(\hat{V}_{i+|k|} - \bar{\hat{V}}_{n-1}), \quad \bar{\hat{V}}_{n-1} = \frac{1}{n-1}\sum_{i=2}^{n}\hat{V}_i, \tag{2.3}$$

and forms the plug-in thresholded long-run variance estimator of $(\hat{V}_i)$ as

$$\hat{g}_{\hat{V}} = \sum_{|k|<n-1} \hat{\gamma}_{\hat{V},k} \mathbb{1}_{\{|\hat{\rho}_{\hat{V},k}| \geq \lambda_n\}}, \quad \hat{\rho}_{\hat{V},k} = \frac{\hat{\gamma}_{\hat{V},k}}{\hat{\gamma}_{\hat{V},0}}. \tag{2.4}$$

Since $\hat{V}_1 = X_1 - \bar{X}_n$ is also calculable if $\hat{\varphi} = 0$, a counterpart of (2.3) should be used in this case to make use of this additional residual. We also note that the plug-in thresholded estimator (2.4) has a different form from the one studied by Paparoditis and Politis (2012) in their asymptotic theory; see their equation (5). However, if one plugs the data-driven threshold suggested in Section 3.2 of Paparoditis and Politis (2012) into their estimator, then it becomes equivalent to the form given in (2.4). We shall here focus directly on the form (2.4) when developing our asymptotic theory for the proposed estimator. Given $\hat{g}_{\hat{V}}$ from the residual process $(\hat{V}_i)$, the third step is to recolor back to obtain the long-run variance estimator of the original process (X_i) as

$$\hat{g}_X = \frac{\hat{g}_{\hat{V}}}{(1-\hat{\varphi})^2}. \tag{2.5}$$

Since the technique of thresholding is used in obtaining both $\hat{\varphi}$ (before prewhitening) and $\hat{g}_{\hat{V}}$ (after prewhitening), we call $\hat{g}_X$ the thresholding-integrated prewhitening-sandwitched (TIPS) estimator. In the current paper, we choose to focus on first-order autoregressive models as the prewhitening filter because (i) autoregressive models have been found in the literature to yield reasonable approximations to a variety of time series processes; and (ii) first-order autoregressive models have the advantage of practical parsimony and computational simplicity and have been used by Andrews and Monahan (1992) and Rho and Shao (2013), among others, for prewhitening in their numerical work. If one has prior knowledge about the underlying data generating mechanism, a different prewhitening filter can be used to further improve the performance.

3. Asymptotic Theory

To study the asymptotic properties of the proposed estimator, we need to impose appropriate regularity conditions on the process (X_i). We shall here follow the framework of Wu (2005) and assume that

$$X_i = G(\mathcal{F}_i), \quad \mathcal{F}_i = (\ldots, \epsilon_{i-1}, \epsilon_i), \tag{3.1}$$

where ϵ_j, $j \in \mathbb{Z}$, are iid random variables, and G is a measurable function such that X_i is well defined. We interpret $\mathcal{F}_i$ and X_i as the input and output, and G as the transform that represents the underlying physical mechanism. The representation (3.1) covers a huge class of processes, including linear processes, bilinear processes, Volterra processes, and many other nonlinear processes; see for example Wiener (1958), Tong (1990), and Wu (2011).

For a random variable X, we write $X \in \mathcal{L}^q$, $q > 0$, if $\|X\|_q = \{E(|X|^q)\}^{1/q} < \infty$, and take $\|\cdot\| = \|\cdot\|_2$. Let $\epsilon_0^\star$ be identically distributed as ϵ_0 and independent of $(\epsilon_j)_{j\in\mathbb{Z}}$. Then $\mathcal{F}_i^\star = (\mathcal{F}_{-1}, \epsilon_0^\star, \epsilon_1, \ldots, \epsilon_i)$ represents the coupled shift process, and we define the functional dependence measure

$$\theta_{k,q} = \|G(\mathcal{F}_k) - G(\mathcal{F}_k^\star)\|_q. \tag{3.2}$$

The quantity $\theta_{k,q}$ measures the dependence of X_k on the innovation ϵ_0, and in particular, if $X_k = G(\mathcal{F}_k)$ does not depend on ϵ_0, then $\theta_{k,q} = 0$. The functional dependence measure (3.2) is easy to work with and is directly related to the underlying data generating mechanism. Let

$$\Theta_{m,q} = \sum_{k=m}^{\infty} \theta_{k,q}, \quad m \geq 0.$$

Throughout, we assume that $\gamma_0 = \text{var}(X_0) > 0$ and the short-range dependence condition $\Theta_{0,q} < \infty$ holds for some $q \geq 2$. In this case, the long-run variance $g < \infty$. Let the parameter estimator $\tilde{\varphi}$ in (2.2) be taken as the least squares estimator and $c_q = 6(q+4)\exp(q/4)\{(1+|\gamma_1/\gamma_0|)/(\gamma_0 - |\gamma_1|)\}\|X_0\|_4\Theta_{0,4}$.

Theorem 1. *Assume $X_0 \in \mathcal{L}^q$ for some $q > 4$, and $\theta_{k,q} = O(k^{-\delta})$ for some $\delta > 3/2$. If the thresholds $\lambda_n = \psi c_q\{(\log n)/n\}^{1/2}$ for some $\psi > 1$, $\tau_n \to 0$ and $n^{1/2}\tau_n \to \infty$, then (i)*

$$\hat{g}_X - g = O_p\left[\left\{\frac{(\log n)}{n}\right\}^{(\delta-1)/(2\delta-1)}\right];$$

(ii) if in addition $\theta_{k,2} = O(\phi^k)$ for some $0 < \phi < 1$, then

$$\hat{g}_X - g = O_p\{n^{-1/2}(\log n)^{3/2}\};$$

and (iii) if in addition there exists an $M < \infty$ such that $\theta_{k,2} = 0$ for all $k > M$, then

$$\hat{g}_X - g = O_p(n^{-1/2}).$$

Recently, Paparoditis and Politis (2012) considered an estimator of the spectral density, and thus the long-run variance, by directly plugging thresholded sample autocovariances into the definition. We denote it by $\tilde{g}_{\text{PP12}}$. By setting the frequency to zero, the stochastic error bound provided in their Theorem 1 becomes

$$\tilde{g}_{\text{PP12}} - g = O_p\left[\left\{\frac{(\log n)}{n}\right\}^{(\delta-1)/(2\delta)}\right].$$

Compared with their result, the stochastic error bound established in Theorem 1 (i) for the proposed TIPS estimator is generally better in a polynomial of n, and up to a scale of $(\log n)^{1/2}$ attains the rate conjectured in Remark 2 of Paparoditis and Politis (2012). In addition, we examine the stochastic error bound and the suggested convergence rate of the proposed TIPS estimator under different dependence strengths. In particular, Theorem 1 (ii) suggests that the convergence rate of the proposed TIPS estimator can be close to the parametric $n^{1/2}$ (up to a polynomial of $\log n$) if the dependence strength follows a geometric decay as in finite-order autoregressive models. On the other hand, Theorem 1 (iii) indicates that the proposed TIPS estimator enjoys the parametric $n^{1/2}$-convergence rate if the underlying process has a sparse dependence structure as in finite-order moving-average models.

For independent data, the long-run variance equals the marginal variance and thus can be estimated by the sample variance $\hat{\gamma}_{X,0}$, arguably the best estimator under the setting, with a well-known parametric convergence rate. However, if one treats the data as dependent, a more sophisticated estimation procedure is needed and at a cost of a convergence rate that is usually inferior. For example, Flegal and Jones (2010) considered the class of batch means estimators and gave the corresponding optimal rate of convergence as $n^{1/3}$. Distinguishing between dependent and independent data can itself be a nontrivial task, and has been widely studied in the literature; see for example Box and Pierce (1970), Robinson (1991), Hong (1996), Escanciano and Lobato (2009), Shao (2011) and Xiao and Wu (2014) among others. Therefore, it would be desirable to have an estimator that consistently estimates the long-run variance for both dependent and independent data, and that reduces to the sample variance when the data are independent. We show that our proposed TIPS estimator enjoys both properties and thus solves the inconvenient dilemma for handling dependent and independent data.

Theorem 2. *Assume that (X_i) are iid random variables with $X_i \in \mathcal{L}^q$ for some $q > 4$. With the thresholds $\lambda_n = \psi c_q\{(\log n)/n\}^{1/2}$ for some $\psi > 1$, $\tau_n \to 0$ and $n^{1/2}\tau_n \to \infty$, we have $\lim_{n\to\infty} \mathrm{pr}(\hat{g}_X = \hat{\gamma}_{X,0}) = 1$.*

The proposed TIPS estimator thus enjoys the dependence-oracle property in the sense that it automatically reduces to the sample variance if the data are independent.

4. Monte Carlo Simulations

4.1. Simulation setting

We report here on some simulation studies to examine the finite-sample performance of the proposed TIPS estimator, and make comparisons with existing estimators. Let ϵ_k, $k \in \mathbb{Z}$, be iid random variables with $E(|\epsilon_0|^q) < \infty$ for some $q \geq 2$. We considered the linear process

$$\text{Model I}: X_i = \sum_{k=1}^{\infty} a_k \epsilon_{i-k+1} = a_1\epsilon_i + a_2\epsilon_{i-1} + a_3\epsilon_{i-2} + \cdots ;$$

and its nonlinear generalization

$$\text{Model II}: X_i = a_1\epsilon_i|\epsilon_i| + \sum_{k=2}^{\infty} a_k \epsilon_{i-k+1} = a_1\epsilon_i|\epsilon_i| + a_2\epsilon_{i-1} + a_3\epsilon_{i-2} + \cdots ,$$

and we allowed different choices of the coefficient sequence (a_k) to include processes with different dependence strengths. We shall here briefly mention the connection between the coefficient sequence (a_k) in Models I and II and the functional dependence measure $\theta_{k,q}$ defined in Section 3. For this, let $\epsilon_0^\star$ be identically distributed as ϵ_0 and independent of $(\epsilon_k)_{k\in\mathbb{Z}}$, then $\||\epsilon_0|-|\epsilon_0^\star|\|_q \leq \|\epsilon_0-\epsilon_0^\star\|_q \leq 2\|\epsilon_0\|_q$, and thus by (3.2),

$$\theta_{k,q} \leq |a_{k+1}| \cdot \|\epsilon_0 - \epsilon_0^\star\|_q, \quad k > 0,$$

holds for both Models I and II. Therefore, the conditions in Theorem 1 (i)–(iii) will be satisfied if (i) $a_k = O(k^{-\delta})$ for some $\delta > 3/2$ follows a polynomial decay[1]; (ii) $a_k = O(\phi^k)$ for some $0 < \phi < 1$ follows a geometric decay; and (iii) there exists an $M < \infty$ such that $a_k = 0$ for all $k > M$ respectively. To include processes with different strengths of dependence, we considered the following decay types of (a_k).

(a) Polynomial: $a_k = k^{-\delta}$, where $\delta \in \{2, 3, 5\}$;

(b) Exponential: $a_k = \phi^k$, where $\phi \in \{\pm 0.3, \pm 0.6\}$;

(c) Finite: $X_i = \epsilon_i + 0.4\epsilon_{i-1} + 0.3\epsilon_{i-2}$ and $X_i = \epsilon_i|\epsilon_i| + 0.4\epsilon_{i-1} + 0.3\epsilon_{i-2}$;

(d) Finite with season: $X_i = \epsilon_i + 0.5\epsilon_{i-s}$ and $X_i = \epsilon_i|\epsilon_i| + 0.5\epsilon_{i-s}$, where $s \in \{6, 12\}$.

For each configuration, given a realization $X_1, \ldots, X_n$, we considered seven types of long-run variance estimates: the proposed TIPS estimate; the plug-in thresh-

[1] This also implies that $\Theta_{m,q} = O\{m^{-(\delta-1)}\}$ for some $\delta - 1 > 1/2$, which is the condition required by Paparoditis and Politis (2012).

olded and hybrid estimates of Paparoditis and Politis (2012), denoted by PP12T and PP12H, respectively; the flat-top lag-window estimate of Politis (2001), denoted by P01; the prewhitened lag-window estimate of Andrews and Monahan (1992), denoted by AM92; the overlapping batch means (OBM) estimate; and the nonoverlapping batch means (NBM) estimates. For the innovation sequence (ϵ_k), we considered both standard normal innovations and Rademacher innovations.

4.2. Threshold and tuning parameter selection

We here discuss the choice of thresholds and tuning parameters for the long-run variance estimators considered. For the proposed TIPS estimator, by Theorem 1, one should choose $\lambda_n = \psi c_q \{(\log n)/n\}^{1/2}$ for some $\psi > 1$, where the constant c_q appears due to its appearance in the maximal deviation of sample autocovariances from their expected values; see for example Xiao and Wu (2012). Politis (2003) considered the problem of bandwidth selection for lag-window estimators, and their procedure depends on finding the smallest positive integer such that the sample autocorrelations are bounded in absolute value by $2\{(\log_{10} n)/n\}^{1/2}$ for a consecutive $K_n = o(\log n)$ lags. As commented in their Remark 2.3, the choice of $2\{(\log_{10} n)/n\}^{1/2}$ roughly corresponds to the construction of 95% simultaneous confidence interval for autocorrelations, which motivated us to consider using $2\{(\log_{10} n)/n\}^{1/2}$ as an estimate of $c_q\{(\log n)/n\}^{1/2}$ for finite-sample problems; see also Paparoditis and Politis (2012). The latter paper considered the class of plug-in thresholded estimators, and suggested the use of $\psi = 1.5$ in the thresholding rule, which also seems to be a reasonable choice for the proposed TIPS estimator, and is thus recommended. For an appropriate threshold of significance τ_n for the least squares estimator $\tilde{\varphi}$ used for prewhitening, we considered the choices $\tau_n = \varsigma n^{-1/2}$ for $\varsigma \in \{1.96, 2.58\}$. These correspond to hypothesis tests with 95% and 99% nominal levels, and the resulting estimates are denoted by $\mathrm{TIPS}_{.95}$ and $\mathrm{TIPS}_{.99}$, respectively. For the AM92 estimator, we used the data-driven bandwidth selector of Andrews and Monahan (1992); while the empirical bandwidth choice rule of Politis (2003) was used for the P01, PP12T, and PP12H estimators, following the suggestion of Paparoditis and Politis (2012). For the OBM and NBM estimators, we used the plug-in version of the optimal batch size given by Song and Schmeiser (1995).

4.3. Simulation results and summary of findings

For the sample size $n \in \{250, 500\}$, the results are summarized in Tables 1–3, containing standardized mean squared errors for the estimates under different configurations. For each configuration, given a realization $X_1, \ldots, X_n$, we computed the standardized squared error $\{(\hat{g} - g)/g\}^2$ for each considered candidate estimate, and the averages based on 1000 realizations are reported in Tables 1–3. More details on the simulation procedure can be found in the Supplementary Material. Table 1 concerns linear processes with standard normal innovations, from which we observe the following. First, if the underlying process exhibits relatively strong dependence, for example the case with a polynomial decay with a relatively slow decay rate $\delta = 2$, then the proposed TIPS estimate can bring significant advantages over the class of plug-in thresholded estimates of Paparoditis and Politis (2012), as the integrated prewhitening of the proposed TIPS estimate takes into effect in soaking up the dependence. As expected, if we reduce the dependence strength, then the advantage can become less noticeable, and eventually disappear. Second, the proposed TIPS estimate can largely outperform the tapering-based estimate of Andrews and Monahan (1992) and Politis (2001) for situations where the autocovariances have an unordered sparse structure as in the two seasonal models considered; see also simulation results of Bickel and Levina (2008) and Xiao and Wu (2012) for similar findings in the context of covariance matrix estimation. In addition, the proposed TIPS estimate seems to outperform the batch means estimates for most of the cases considered, and the difference between $\text{TIPS}_{.95}$ and $\text{TIPS}_{.99}$ caused by choosing different τ_n in prewhitening seems to be subtle. If we increase the sample size, the performance of all the considered estimates improves, while the same pattern remains for their comparisons. We considered the situation for nonlinear processes in Table 2 and processes with non-Gaussian innovations in Table 3, and the above observations seem to hold there as well

5. Conclusion and Discussion

We consider a new class of long-run variance estimators, named as the TIPS, that first soaks up the dependence by a thresholding-integrated prewhitening filter, then regularizes autocorrelations of the resulting residual process by thresholding, and finally recolors back to obtain an estimator of the original process. Simulation results suggest that the proposed TIPS estimator can improve on the class of plug-in thresholded estimator of Paparoditis and Politis (2012) for processes with a relatively strong dependence strength, while it manages to deliver a similar performance for situations where direct thresholding as in Paparoditis and Politis (2012) is expectedly the most suitable. For the asymptotic theory,

Table 1. Standardized mean squared errors of different long-run variance estimates for Model I with standard normal innovations, under different dependence strengths. The results are based on 1,000 realizations for each configuration.

		Method							
n	*Decay Type*	$\text{TIPS}_{.95}$	$\text{TIPS}_{.99}$	PP12H	PP12T	P01	AM92	OBM	NBM
250	Polynomial (rate 2)	0.116	0.125	0.173	0.287	0.176	0.119	0.116	0.119
	Polynomial (rate 3)	0.058	0.077	0.086	0.094	0.087	0.033	0.069	0.074
	Polynomial (rate 5)	0.017	0.014	0.014	0.012	0.014	0.025	0.052	0.059
	Exponential (base 0.3)	0.041	0.044	0.055	0.131	0.058	0.042	0.072	0.079
	Exponential (base 0.6)	0.070	0.070	0.115	0.135	0.125	0.073	0.105	0.109
	Exponential (base −0.3)	0.017	0.019	0.135	0.393	0.122	0.016	0.073	0.099
	Exponential (base −0.6)	0.011	0.011	0.679	1.265	0.129	0.013	0.222	0.325
	Finite (no season)	0.062	0.062	0.051	0.073	0.061	0.059	0.075	0.080
	Finite (season 6)	0.032	0.028	0.201	0.025	0.209	0.213	0.117	0.118
	Finite (season 12)	0.038	0.034	0.209	0.031	0.210	0.211	0.247	0.239
500	Polynomial (rate 2)	0.104	0.104	0.129	0.183	0.126	0.109	0.086	0.090
	Polynomial (rate 3)	0.031	0.045	0.067	0.089	0.067	0.021	0.038	0.043
	Polynomial (rate 5)	0.011	0.010	0.010	0.009	0.011	0.013	0.033	0.037
	Exponential (base 0.3)	0.018	0.018	0.030	0.035	0.033	0.018	0.047	0.054
	Exponential (base 0.6)	0.036	0.036	0.057	0.076	0.060	0.037	0.061	0.066
	Exponential (base −0.3)	0.009	0.009	0.083	0.109	0.073	0.009	0.050	0.064
	Exponential (base −0.6)	0.006	0.006	0.394	0.820	0.061	0.007	0.120	0.159
	Finite (no season)	0.037	0.037	0.018	0.035	0.024	0.026	0.048	0.053
	Finite (season 6)	0.013	0.010	0.197	0.009	0.200	0.206	0.077	0.080
	Finite (season 12)	0.015	0.012	0.202	0.010	0.202	0.206	0.206	0.210

we utilized the functional dependence measure of Wu (2005), and our conditions are comparable to those required by Xiao and Wu (2012) and Paparoditis and Politis (2012). With a more dedicated focus on long-run variance estimation, the current stochastic error bound is superior to that of Paparoditis and Politis (2012) by a polynomial of n and, up to a scale of $(\log n)^{1/2}$, attains the rate conjectured in Remark 2 of their paper. As mentioned in Xiao and Wu (2012), compared with the conventional banded estimator, the thresholded estimator is desirable in that it can lead to better performance when there are sparse or very weak autocovariances. However, due to technical difficulties caused by the nonlinear nature of thresholding, they also commented that it was unfortunate that their theoretical result was not able to reflect this advantage. The same difficulty was observed by Paparoditis and Politis (2012) who relied on the result of Xiao and Wu (2012). The current paper aims to fill this gap in the setting of long-run variance estimation by proving that the proposed thresholding-based estimator (i) achieves the parametric convergence rate when the underlying process has a sparse dependence structure as in finite-order moving average models;

Table 2. Standardized mean squared errors of different long-run variance estimates for Model II with standard normal innovations, under different dependence strengths. The results are based on 1,000 realizations for each configuration.

		Method							
n	*Decay Type*	$\text{TIPS}_{.95}$	$\text{TIPS}_{.99}$	PP12H	PP12T	P01	AM92	OBM	NBM
250	Polynomial (rate 2)	0.124	0.149	0.175	0.201	0.177	0.096	0.105	0.108
	Polynomial (rate 3)	0.063	0.064	0.065	0.061	0.067	0.054	0.090	0.090
	Polynomial (rate 5)	0.042	0.039	0.038	0.037	0.037	0.055	0.074	0.086
	Exponential (base 0.3)	0.069	0.090	0.103	0.123	0.106	0.054	0.084	0.085
	Exponential (base 0.6)	0.093	0.093	0.129	0.198	0.138	0.103	0.103	0.108
	Exponential (base −0.3)	0.094	0.128	0.192	0.195	0.185	0.064	0.105	0.119
	Exponential (base −0.6)	0.067	0.067	0.257	0.383	0.152	0.060	0.121	0.142
	Finite (no season)	0.060	0.072	0.099	0.174	0.102	0.055	0.087	0.093
	Finite (season 6)	0.117	0.115	0.129	0.114	0.137	0.138	0.119	0.124
	Finite (season 12)	0.121	0.120	0.135	0.119	0.135	0.140	0.185	0.178
500	Polynomial (rate 2)	0.089	0.104	0.140	0.192	0.140	0.082	0.074	0.076
	Polynomial (rate 3)	0.039	0.043	0.045	0.046	0.045	0.028	0.043	0.048
	Polynomial (rate 5)	0.023	0.021	0.022	0.021	0.022	0.028	0.052	0.057
	Exponential (base 0.3)	0.030	0.037	0.054	0.104	0.054	0.028	0.050	0.054
	Exponential (base 0.6)	0.077	0.077	0.072	0.121	0.067	0.091	0.063	0.069
	Exponential (base −0.3)	0.037	0.050	0.095	0.149	0.094	0.032	0.056	0.065
	Exponential (base −0.6)	0.046	0.046	0.151	0.200	0.070	0.038	0.071	0.084
	Finite (no season)	0.033	0.034	0.044	0.099	0.047	0.036	0.054	0.061
	Finite (season 6)	0.050	0.049	0.117	0.048	0.118	0.122	0.068	0.070
	Finite (season 12)	0.056	0.054	0.120	0.053	0.121	0.124	0.138	0.137

and (ii) enjoys the dependence-oracle property in that it automatically reduces to the sample variance when applied to independent data. This is convenient, as determining the existence of serial correlation can itself be a highly nontrivial problem; see for example Box and Pierce (1970), Robinson (1991), Hong (1996), Escanciano and Lobato (2009), Shao (2011), Xiao and Wu (2014), and references therein.

Although first-order autoregressive filters have been popular choices in practice for prewhitening due to their practical parsimony and simplicity (Andrews and Monahan, 1992; Rho and Shao, 2013), our results can be generalized to allow higher-order autoregressive prewhitening filters. In this case, instead of using the first-order filter as in (2.2), one filters (X_i) by the autoregressive model

$$(X_i - \bar{X}_n) = \sum_{r=1}^{p} \hat{\varphi}_r (X_{i-r} - \bar{X}_n) + \hat{V}_i, \quad i = p+1, \ldots, n. \tag{5.1}$$

In respect to this, the factor $(1-\hat{\varphi})^2$ in (2.5) should be replaced by $(1-\sum_{r=1}^{p} \hat{\varphi}_r)^2$ in the recoloring step. Corollary 1 provides the asymptotic property of the resulting prewhitened long-run variance estimator under the following additional

Table 3. Standardized mean squared errors of different long-run variance estimates for Model I with Rademacher innovations, under different dependence strengths. The results are based on 1,000 realizations for each configuration.

		Method							
n	*Decay Type*	$\text{TIPS}_{.95}$	$\text{TIPS}_{.99}$	PP12H	PP12T	P01	AM92	OBM	NBM
250	Polynomial (rate 2)	0.118	0.127	0.180	0.292	0.181	0.120	0.119	0.122
	Polynomial (rate 3)	0.056	0.074	0.083	0.090	0.083	0.028	0.056	0.063
	Polynomial (rate 5)	0.011	0.009	0.007	0.006	0.007	0.018	0.052	0.061
	Exponential (base 0.3)	0.031	0.034	0.049	0.125	0.055	0.031	0.071	0.074
	Exponential (base 0.6)	0.067	0.067	0.110	0.133	0.127	0.069	0.098	0.104
	Exponential (base −0.3)	0.009	0.013	0.131	0.391	0.119	0.009	0.070	0.095
	Exponential (base −0.6)	0.004	0.004	0.715	1.231	0.129	0.007	0.175	0.270
	Finite (no season)	0.051	0.051	0.044	0.067	0.057	0.047	0.073	0.078
	Finite (season 6)	0.026	0.022	0.196	0.018	0.203	0.212	0.112	0.108
	Finite (season 12)	0.033	0.028	0.205	0.024	0.206	0.212	0.250	0.244
500	Polynomial (rate 2)	0.103	0.103	0.131	0.180	0.128	0.109	0.082	0.088
	Polynomial (rate 3)	0.026	0.042	0.066	0.088	0.067	0.017	0.039	0.042
	Polynomial (rate 5)	0.008	0.007	0.006	0.005	0.006	0.009	0.028	0.032
	Exponential (base 0.3)	0.016	0.016	0.029	0.037	0.032	0.016	0.043	0.047
	Exponential (base 0.6)	0.033	0.033	0.064	0.076	0.068	0.035	0.066	0.071
	Exponential (base −0.3)	0.004	0.004	0.087	0.112	0.077	0.005	0.039	0.055
	Exponential (base −0.6)	0.002	0.002	0.425	0.797	0.061	0.003	0.132	0.187
	Finite (no season)	0.031	0.031	0.014	0.032	0.020	0.023	0.046	0.053
	Finite (season 6)	0.010	0.007	0.194	0.005	0.198	0.204	0.072	0.076
	Finite (season 12)	0.010	0.008	0.201	0.005	0.201	0.202	0.207	0.209

assumption.

(PW) $\hat{\varphi}_r - \varphi_r = O_p(n^{-1/2})$ for some $\varphi_r \in \mathbb{R}$, $r = 1, \ldots, p$, satisfying $1 - \sum_{r=1}^{p} \varphi_r \neq 0$.

Corollary 1. *The results of Theorem* 1 *continue to hold for finite-order autoregressive prewhitening filters under the additional assumption (PW).*

Assumption (PW) states that estimators from the autoregressive filter (5.1) have the usual parametric $n^{1/2}$-convergence rate, and is adopted from Andrews and Monahan (1992); see Assumption D of their paper. Note that $1 - \sum_{r=1}^{p} \varphi_r \neq 0$ is implied by the assumption that all roots of the polynomial $\Upsilon(B) = 1 - \sum_{r=1}^{p} \varphi_r B^r$ lie outside of the unit circle, which is typically required for the autoregressive process to be stationary. When one has prior knowledge of the underlying data generating mechanism, a different prewhitening filter can be used for potential performance improvement, which however is beyond the scope of the current paper. We leave this as a possible future research topic.

6. Useful Lemmas and Proofs of Main Results

Recall that $\tilde{V}_i = (X_i - \bar{X}_n) - \tilde{\varphi}(X_{i-1} - \bar{X}_n)$, $i = 2, \ldots, n$, from (2.2), where

$$\tilde{\varphi} = \left\{\sum_{i=2}^{n}(X_{i-1} - \bar{X}_n)^2\right\}^{-1}\left\{\sum_{i=2}^{n}(X_{i-1} - \bar{X}_n)(X_i - \bar{X}_n)\right\}$$

is the least squares estimator. The following lemma states that there exists a $\varphi \in (-1, 1)$ such that $\tilde{\varphi} - \varphi = O_p(n^{-1/2})$ regardless of whether the first-order autoregressive model (2.2) used for prewhitening represents the true underlying data generating mechanism. Recall from Section 3 that $\gamma_0 = \mathrm{var}(X_0) > 0$ is assumed throughout the paper. Proofs of Lemmas 1–6 are provided in the Supplementary Material.

Lemma 1. *Assume $\Theta_{0,2} < \infty$. Let $\varphi = \gamma_1/\gamma_0$ be the first-order autocorrelation of (X_i), then $\varphi \in (-1, 1)$ and $\tilde{\varphi} - \varphi = O_p(n^{-1/2})$.*

Let $U_i = X_i - \varphi X_{i-1}$, $i = 2, \ldots, n$, and

$$\hat{\gamma}_{U,k} = \frac{1}{n-1}\sum_{i=2}^{n-|k|}(U_i - \bar{U}_{n-1})(U_{i+|k|} - \bar{U}_{n-1}), \quad \bar{U}_{n-1} = \frac{1}{n-1}\sum_{i=2}^{n}U_i.$$

The following lemma states that sample autocovariances of $(\tilde{V}_i)$ can be well approximated by those of (U_i) in a uniform manner, and the leading term of the approximation error will depend on

$$\Gamma_{n,k,1} = \frac{1}{n-1}\sum_{i=2}^{n-|k|}(X_{i-1} - \mu)\{U_{i+|k|} - (1-\varphi)\mu\};$$

$$\Gamma_{n,k,2} = \frac{1}{n-1}\sum_{i=2}^{n-|k|}(X_{i+|k|-1} - \mu)\{U_i - (1-\varphi)\mu\};$$

$$\Gamma_{n,k,3} = \frac{1}{n-1}\sum_{i=2}^{n-|k|}(X_{i-1} - \mu)(X_{i+|k|-1} - \mu).$$

Lemma 2. *Assume $\Theta_{0,2} < \infty$. Then*

$$\max_{|k|<n-1}|\hat{\gamma}_{\tilde{V},k} - \hat{\gamma}_{U,k} + (\tilde{\varphi} - \varphi)(\Gamma_{n,k,1} + \Gamma_{n,k,2}) - (\tilde{\varphi} - \varphi)^2\Gamma_{n,k,3}| = O_p(n^{-3/2}).$$

Let $\hat{\rho}_{U,k} = \hat{\gamma}_{U,k}/\hat{\gamma}_{U,0}$ be sample autocorrelations of (U_i), whose population counterpart is denoted by $\rho_{U,k} = \mathrm{cor}(U_0, U_k)$. Recall that $c_q = 6(q + 4)\exp(q/4)\{(1 + |\gamma_1/\gamma_0|)/(\gamma_0 - |\gamma_1|)\}\|X_0\|_4\Theta_{0,4}$. Let $c_q^\star = 6(q + 4)\exp(q/4)(1 + |\gamma_1/\gamma_0|)^2\|X_0\|_4\Theta_{0,4}$, the following lemma provides a uniform bound for $\hat{\rho}_{U,k}$, $|k| < n - 1$, and is useful in proving Lemma 4.

Lemma 3. *Assume* $X_0 \in \mathcal{L}^q$ *for some* $q > 4$, *and* $\theta_{k,q} = O(k^{-\delta})$ *for some* $\delta > 3/2$. *Then for any* $\xi > 1$,

$$\lim_{n\to\infty} \mathrm{pr}\left\{ \max_{|k|<n-1} \left| \hat{\rho}_{U,k} - \left(1 - \frac{|k|}{n-1}\right)\rho_{U,k} \right| \le c_q \xi \left(\frac{\log n}{n}\right)^{1/2} \right\} = 1.$$

The following lemma states that the sum of thresholded sample autocorrelations of the generally unobservable process (U_i) can be well approximated by a truncated sum that only involves estimators for nonzero autocorrelations.

Lemma 4. *Assume* $X_0 \in \mathcal{L}^q$ *for some* $q > 4$, *and* $\theta_{k,q} = O(k^{-\delta})$ *for some* $\delta > 3/2$. *If the threshold* $\lambda_n = \psi c_q\{(\log n)/n\}^{1/2}$ *for some* $\psi > 1$, *then (i) for any sequence* $l_n \to \infty$ *and* $l_n/n \to 0$,

$$\sum_{|k|<n-1} \hat{\rho}_{U,k} \mathbb{1}_{\{|\hat{\rho}_{U,k}|\ge\lambda_n\}} = \sum_{|k|\le l_n} \hat{\rho}_{U,k} \mathbb{1}_{\{\rho_{U,k}\ne 0\}} + O_p\left(\sum_{|k|>l_n} |\rho_{U,k}| + \lambda_n l_n\right);$$

and (ii) if in addition there exists an $M < \infty$ *such that* $\theta_{k,2} = 0$ *for all* $k > M$, *then the bound in (i) can be improved to*

$$\sum_{|k|<n-1} \hat{\rho}_{U,k} \mathbb{1}_{\{|\hat{\rho}_{U,k}|\ge\lambda_n\}} = \sum_{|k|\le l_n} \hat{\rho}_{U,k} \mathbb{1}_{\{\rho_{U,k}\ne 0\}} + O_p\left(\sum_{|k|>l_n} |\rho_{U,k}|\right).$$

Let $\gamma_{U,k} = \mathrm{cov}(U_0, U_k)$, $k \in \mathbb{Z}$, denote sample autocovariances of the process (U_i). The following lemma concerns the sum of thresholded autocovariances of $(\tilde{V}_i)$, and is useful in proving Lemma 6.

Lemma 5. *Assume* $X_0 \in \mathcal{L}^q$ *for some* $q > 4$, *and* $\theta_{k,q} = O(k^{-\delta})$ *for some* $\delta > 3/2$. *If the threshold* $\lambda_n = \psi c_q\{(\log n)/n\}^{1/2}$ *for some* $\psi > 1$, *then (i) for any sequence* $l_n \to \infty$ *and* $l_n/n \to 0$,

$$\sum_{|k|<n-1} \hat{\gamma}_{\tilde{V},k} \mathbb{1}_{\{|\hat{\rho}_{\tilde{V},k}|\ge\lambda_n\}} = \sum_{|k|\le l_n} \hat{\gamma}_{U,k} \mathbb{1}_{\{\gamma_{U,k}\ne 0\}} + O_p\left(\sum_{|k|>l_n} |\gamma_{U,k}| + \lambda_n l_n\right);$$

and (ii) if in addition there exists an $M < \infty$ *such that* $\theta_{k,2} = 0$ *for all* $k > M$, *then the bound in (i) can be improved to*

$$\sum_{|k|<n-1} \hat{\gamma}_{\tilde{V},k} \mathbb{1}_{\{|\hat{\rho}_{\tilde{V},k}|\ge\lambda_n\}} = \sum_{|k|\le l_n} \hat{\gamma}_{U,k} \mathbb{1}_{\{\gamma_{U,k}\ne 0\}} + O_p\left(\sum_{|k|>l_n} |\gamma_{U,k}| + n^{-1/2} + \frac{l_n}{n}\right).$$

The following lemma states that the residual process $(\hat{V}_i)$ can be used to estimate

$$g_U = \sum_{k\in\mathbb{Z}} \gamma_{U,k},$$

the long-run variance of the generally unobserved process (U_i). An explicit stochastic error bound is given, which is useful in proving Theorem 1.

Lemma 6. *Assume $X_0 \in \mathcal{L}^q$ for some $q > 4$, and $\theta_{k,q} = O(k^{-\delta})$ for some $\delta > 3/2$. If the threshold $\lambda_n = \psi c_q\{(\log n)/n\}^{1/2}$ for some $\psi > 1$, then (i) for any sequence $l_n \to \infty$ and $l_n/n \to 0$,*

$$\hat{g}_{\hat{V}} = g_U + O_p\left\{(\frac{l_n}{n})^{1/2} + n^{-1}\sum_{|k|\le l_n}|k\gamma_{U,k}| + \sum_{|k|>l_n}|\gamma_{U,k}| + \lambda_n l_n\right\};$$

and (ii) if in addition there exists an $M < \infty$ such that $\theta_{k,2} = 0$ for all $k > M$, then the bound in (i) can be improved to

$$\hat{g}_{\hat{V}} = g_U + O_p\left\{n^{-1/2} + n^{-1}\sum_{|k|\le l_n}|k\gamma_{U,k}| + \sum_{|k|>l_n}|\gamma_{U,k}|\right\}.$$

Proof. (Theorem 1) Since $\theta_{k,2} \le \theta_{k,q} = O(k^{-\delta})$ for some $\delta > 3/2$ under the stated conditions, we have $|\gamma_k| \le \sum_{i=0}^{\infty}\theta_i\theta_{i+|k|} = O(|k|^{1-2\delta})$ and $\sum_{k=0}^{\infty}|k\gamma_k| < \infty$. Then by Lemma 6 (i),

$$\hat{g}_{\hat{V}} - g_U = O_p\left\{\left(\frac{l_n}{n}\right)^{1/2} + n^{-1} + l_n^{2(1-\delta)} + \lambda_n l_n\right\}$$

holds for any sequence $l_n \to \infty$ and $l_n/n \to 0$. By choosing $l_n = (n/\log n)^{1/(4\delta-2)}$ in the above error bound, we obtain that

$$\hat{g}_{\hat{V}} - g_U = O_p\left[\left\{\frac{(\log n)}{n}\right\}^{(\delta-1)/(2\delta-1)}\right].$$

By Lemma 1, the parameter estimator $\tilde{\varphi} = \varphi + O_p(n^{-1/2})$, and thus under the stated conditions on τ_n we have $\hat{\varphi} = \varphi + O_p(n^{-1/2})$. Therefore,

$$\hat{g} = \frac{g_U}{(1-\varphi)^2} + O_p\left[n^{-1/2} + \left\{\frac{(\log n)}{n}\right\}^{(\delta-1)/(2\delta-1)}\right].$$

As $U_i = X_i - \varphi X_{i-1}$,

$$\gamma_{U,k} = \mathrm{cov}(X_i - \varphi X_{i-1}, X_{i+k} - \varphi X_{i+k-1}) = (1+\varphi^2)\gamma_k - \varphi(\gamma_{k-1} + \gamma_{k+1}),$$

and thus

$$g_U = \sum_{k\in\mathbb{Z}}\gamma_{U,k} = (1-\varphi)^2 g.$$

Since $(\delta-1)/(2\delta-1) < 1/2$, (i) follows. If $\theta_{k,2} = O(\phi^k)$ for some $0 < \phi < 1$, as in (ii), then $|\gamma_k| \le \sum_{i=0}^{\infty}\theta_i\theta_{i+|k|} = O(\phi^{|k|})$ and by Lemma 6 (i),

$$\hat{g}_{\hat{V}} - g_U = O_p\left\{\left(\frac{l_n}{n}\right)^{1/2} + n^{-1} + \phi^{l_n} + \lambda_n l_n\right\}.$$

By letting $l_n = (\log\log n - \log n)/(2\log\phi)$, (ii) follows by a similar argument. On the other hand, if there exists an $M < \infty$ such that $\theta_{k,2} = 0$ for all $k > M$, then $\sum_{|k|>l_n} |\gamma_{U,k}| = 0$ for all sufficiently large l_n, and (iii) follows by choosing any $l_n \to \infty$ in Lemma 6 (ii).

Proof. (Theorem 2) If the underlying process (X_i) does follow a stationary first-order autoregressive model

$$(X_i - \mu) = \varphi(X_{i-1} - \mu) + \epsilon_i,$$

where $\varphi \in (-1, 1)$ is the autoregressive coefficient and (ϵ_i) are iid random variables, then by Lemma 1 the parameter estimator $\tilde{\varphi} = \varphi + O_p(n^{-1/2})$. In this case, $U_i = \epsilon_i + (1-\varphi)\mu$ and thus $\rho_{U,k} = 0$ if $k \neq 0$. With ν_n from the proof of Lemma 4, by Lemma 3 we have

$$\lim_{n\to\infty} \text{pr}\left\{\max_{0<|k|<n-1} |\hat{\rho}_{U,k}| \leq \frac{(\psi+1)\nu_n}{2}\right\} = 1,$$

and thus

$$\lim_{n\to\infty} \text{pr}\left(\sum_{|k|<n} \hat{\gamma}_{U,k} \mathbb{1}_{\{|\hat{\rho}_{U,k}|\geq\lambda_n\}} = \hat{\gamma}_{U,0}\right) = 1. \tag{6.1}$$

As $\tilde{V}_i = (X_i - \bar{X}_n) - \tilde{\varphi}(X_{i-1} - \bar{X}_n)$, by the proof of Lemma 5 we have

$$\lim_{n\to\infty} \text{pr}\left(\sum_{|k|<n} \hat{\gamma}_{\tilde{V},k} \mathbb{1}_{\{|\hat{\rho}_{\tilde{V},k}|\geq\lambda_n\}} = \hat{\gamma}_{\tilde{V},0}\right) = 1.$$

If $\varphi \neq 0$, then by Lemma 1 and the condition that $\tau_n \to 0$, we have $\tilde{\varphi} \geq \tau_n$ and thus $\hat{V}_i = \tilde{V}_i$ with probability tending to one as $n \to \infty$. As a result, the proposed TIPS estimator has the property that

$$\lim_{n\to\infty} \text{pr}\left\{\hat{g}_X = \frac{\hat{\gamma}_{\tilde{V},0}}{(1-\tilde{\varphi})^2}\right\} = 1.$$

If $\varphi = 0$, then by Lemma 1 and the condition that $n^{1/2}\tau_n \to \infty$, we have $\tilde{\varphi} < \tau_n$ and thus $\hat{V}_i = X_i - \bar{X}_n = U_i - \bar{X}_n$ with probability tending to one as $n \to \infty$. Hence, by (6.1) and the construction of $\hat{g}_X$, we have

$$\lim_{n\to\infty} \text{pr}\{\hat{g}_X = \hat{\gamma}_{X,0}\} = 1,$$

entailing the dependence-oracle property.

Proof. (Corollary 1) The proof follows similarly to that of Theorem 1, and we shall here only outline the key differences. With a little abuse of notation, let

$$U_i = X_i - \sum_{r=1}^{p} \varphi_r X_{i-r}, \quad i = p+1, \ldots, n,$$

$$\hat{\gamma}_{U,k} = \frac{1}{n-p} \sum_{i=p+1}^{n-|k|} (U_i - \bar{U}_{n-p})(U_{i+|k|} - \bar{U}_{n-p}), \quad \bar{U}_{n-p} = \frac{1}{n-p} \sum_{i=p+1}^{n} U_i.$$

Similarly, define

$$\hat{U}_i = X_i - \sum_{r=1}^{p} \hat{\varphi}_r X_{i-r}, \quad i = p+1, \ldots, n,$$

$$\hat{\gamma}_{\hat{U},k} = \frac{1}{n-p} \sum_{i=p+1}^{n-|k|} (\hat{U}_i - \bar{\hat{U}}_{n-p})(\hat{U}_{i+|k|} - \bar{\hat{U}}_{n-p}), \quad \bar{\hat{U}}_{n-p} = \frac{1}{n-p} \sum_{i=p+1}^{n} \hat{U}_i.$$

Then by (5.1) we have $\hat{V}_i = \hat{U}_i - (1 - \sum_{r=1}^{p} \hat{\varphi}_r)\bar{X}_n$, $i = p+1, \ldots, n$, and thus $\hat{\gamma}_{\hat{V},k} = \hat{\gamma}_{\hat{U},k}$, $|k| < n-p$. Let

$$\hat{D}_i = (\hat{U}_i - \bar{\hat{U}}_{n-p}) - (U_i - \bar{U}_{n-p}) = -\sum_{r=1}^{p} (\hat{\varphi}_r - \varphi_r) \left(X_{i-r} - \frac{1}{n-p} \sum_{j=p+1}^{n} X_{j-r} \right)$$

denote the centered difference, $i = p+1, \ldots, n$, then as $p < \infty$ is fixed, the proof of Theorem 1 continues to follow.

Supplementary Materials

The Supplementary Material contains proofs of Lemmas 1–6, and additional details on the simulation procedure.

Acknowledgment

I am grateful to the co-Editor, an associate editor, and two anonymous referees for their helpful comments and suggestions. I also thank Dimitris N. Politis for his helpful comments. The research was Supported by NSF Grant DMS-1461796.

References

Aktaran-Kalaycı, T., Goldsman, D. and Wilson, J. R. (2007). Linear combinations of overlapping variance estimators for simulation. *Operations Research Letters* **35**, 439–447.

Alexopoulos, C., Argon, N. T., Goldsman, D., Tokol, G. and Wilson, J. R. (2007). Overlapping variance estimators for simulation. *Operations Research* **55**, 1090–1103.

Anderson, T. W. (1994). *The Statistical Analysis of Time Series*. New York: Wiley.

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59**, 817–858.

Andrews, D. W. K. and Monahan, J. C. (1992). An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica* **60**, 953–966.

Bickel, P. J. and Levina, E. (2008). Covariance regularization by thresholding. *The Annals of Statistics* **36**, 2577–2604.

Box, G. E. P. and Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association* **65**, 1509–1526.

Bradley, R. C. (2007). *Introduction to Strong Mixing Conditions*. Utah: Kendrick Press.

Bratley, P., Fox, B. L. and Schrage, L. E. (1987). *A Guide to Simulation*. New York: Springer.

Bühlmann, P. (2002). Bootstraps for time series. *Statistical Science* **17**, 52–72.

Crane, M. A. and Iglehart, D. L. (1975). Simulating stable stochastic systems: III. Regenerative processes and discrete-event simulations. *Operations Research* **23**, 33–45.

Escanciano, J. C. and Lobato, I. N. (2009). An automatic Portmanteau test for serial correlation. *Journal of Econometrics* **151**, 140–149.

Fishman, G. S. (1971). Estimating sample size in computing simulation experiments. *Management Science* **18**, 21–38.

Flegal, J. M. and Jones, G. L. (2010). Batch means and spectral variance estimators in markov chain monte carlo. *The Annals of Statistics* **38**, 1034–1070.

Goldsman, D., Meketon, M. and Schruben, L. (1990). Properties of standardized time series weighted area variance estimators. *Management Science* **36**, 602–612.

Hong, Y. (1996). Consistent testing for serial correlation of unknown form. *Econometrica* **64**, 837–864.

Ibragimov, I. A. and Linnik, I. V. (1971). *Independent and Stationary Sequences of Random Variables*. Groningen: Wolters-Noordhoff.

Jones, G. L., Haran, M., Caffo, B. S. and Neath, R. (2006). Fixed-width output analysis for markov chain monte carlo. *Journal of the American Statistical Association* **101**, 1537–1547.

Lahiri, S. (2003). *Resampling Methods for Dependent Data*. New York: Springer.

Law, A. M. and Kelton, W. D. (2000). *Simulation Modeling and Analysis*. New York: McGraw-Hill, 3rd ed.

Liu, W. and Wu, W. B. (2010). Asymptotics of spectral density estimates. *Econometric Theory* **26**, 1218–1245.

Maxwell, M. and Woodroofe, M. (2000). Central limit theorems for additive functionals of markov chains. *The Annals of Probability* **28**, 713–724.

Meterelliyoz, M., Alexopoulos, C. and Goldsman, D. (2012). Folded overlapping variance estimators for simulation. *European Journal of Operational Research* **220**, 135–146.

Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* **55**, 703–708.

Paparoditis, E. and Politis, D. N. (2012). Nonlinear spectral density estimation: thresholding the correlogram. *Journal of Time Series Analysis* **33**, 386–397.

Peligrad, M. (1996). On the asymptotic normality of sequences of weak dependent random variables. *Journal of Theoretical Probability* **9**, 703–715.

Politis, D. N. (2001). On nonparametric function estimation with infinite-order flat-top kernels. In *Probability and Statistical Models with Applications*, C. A. Charalambides, M. V.

Koutras and N. Balakrishnan, eds. Boca Raton: Chapman and Hall/CRC.

Politis, D. N. (2003). Adaptive bandwidth choice. *Journal of Nonparametric Statistics* **15**, 517–533.

Politis, D. N. (2011). Higher-order accurate, positive semi-definite estimation of large-sample covariance and spectral density matrices. *Econometric Theory* **27**, 703–744.

Press, H. and Tukey, J. W. (1956). Power spectral methods of analysis and their application to problems in airplane dynamics. In *Flight Test Manual, NATO Advisory Group for Aeronautical Research and Development*, vol. **IV-C**. Reprinted as Bell System Monograph No. 2606.

Rho, Y. and Shao, X. (2013). Improving the bandwidth-free inference methods by prewhitening. *Journal of Statistical Planning and Inference* **143**, 1912–1922.

Robinson, P. M. (1991). Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. *Journal of Econometrics* **47**, 67–84.

Rosenblatt, M. (1985). *Stationary Sequences and Random Fields.* Boston: Birkhäuser.

Schruben, L. (1983). Confidence interval estimation using standardized time series. *Operations Research* **31**, 1090–1108.

Shao, X. (2011). Testing for white noise under unknown dependence and its applications to diagnostic checking for time series models. *Econometric Theory* **27**, 312–343.

Song, W. T. and Schmeiser, B. W. (1995). Optimal mean-squared-error batch sizes. *Management Science* **41**, 110–123.

Sul, D., Phillips, P. C. B. and Choi, C.-Y. (2005). Prewhitening bias in hac estimation. *Oxford Bulletin of Economics and Statistics* **67**, 517–546.

Tong, H. (1990). *Non-Linear Time Series: A Dynamical System Approach (Oxford Statistical Science Series, 6).* U.K.: Oxford University Press.

Wiener, N. (1958). *Nonlinear Problems In Random Theory (Technology Press Research Monographs).* Cambridge: The MIT Press.

Wu, W. B. (2005). Nonlinear system theory: another look at dependence. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 14150–14154.

Wu, W. B. (2009). Recursive estimation of time-average variance constants. *The Annals of Applied Probability* **19**, 1529–1552.

Wu, W. B. (2011). Asymptotic theory for stationary processes. *Statistics and its Interface* **4**, 207–226.

Xiao, H. and Wu, W. B. (2011). A single-pass algorithm for spectrum estimation with fast convergence. *IEEE Transactions on Information Theory* **57**, 4720–4731.

Xiao, H. and Wu, W. B. (2012). Covariance matrix estimation for stationary time series. *The Annals of Statistics* **40**, 466–493.

Xiao, H. and Wu, W. B. (2014). Portmanteau test and simultaneous inference for serial covariances. *Statistica Sinica* **24**, 577–600.

Department of Mathematics and Statistics, Boston University, 111 Cummington Mall , Boston, Massachusetts 02215, USA.

E-mail: tingz@bu.edu

(Received October 2016; accepted December 2016)