Statistics and Probability Letters 114 (2016) 30–37

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Recursive estimation of time-average variance constants through prewhitening

Wei Zheng [a,*], Yong Jin [b], Guoyi Zhang [c]

[a] Department of Mathematical Sciences, Indiana University-Purdue University, Indianapolis, IN, USA
[b] Department of Finance, Insurance and Real Estate, University of Florida, Gainesville, FL, USA
[c] Department of Mathematics and Statistics, University of New Mexico, Albuquerque, NM, USA

ARTICLE INFO

Article history:
Received 5 January 2016
Received in revised form 20 February 2016
Accepted 20 February 2016
Available online 10 March 2016

MSC:
primary 60F05
secondary 60F17

Keywords:
Central limit theorem
Consistency
Martingale
Prewhitening
Recursive estimation

ABSTRACT

The time-average variance constant (TAVC) has been an important component in the study of time series. Many real problems request a fast and recursive way in estimating TAVC. In this paper we apply $AR(1)$ prewhitening filter to the recursive algorithm by Wu (2009b), so that the memory complexity of order $O(1)$ is maintained and the accuracy of the estimate is improved. This is justified by both theoretical results and simulation studies.

1. Introduction

The time-average variance constant (TAVC) plays an important role in many time series inference problems such as unit root testing and statistical inference of the mean. Throughout the paper, we focus on a stationary time series $\{X_i\}_{i\in\mathbb{Z}}$ with mean $\mu = E(X_i)$ and finite variance. The TAVC typically has the representation of $\sigma^2 = \sum_{k\in\mathbb{Z}} \gamma(k)$, where $\gamma(k) = \mathrm{cov}(X_0, X_k)$ is the covariance function at lag k. The TAVC is proportional to the spectral density function evaluated at the origin by a constant and the estimation of the latter has been extensively studied in the literature. See for instance Stock (1994), Song and Schmeiser (1995), Phillips and Xiao (1998), Politis et al. (1999), Bühlmann (2002), Lahiri (2003), Alexopoulos and Goldsman (2004) and Jones et al. (2006).

In many applications, one has to sequentially update the estimate of σ^2 while the observations are being accumulated one by one. For the traditional methods of estimating σ^2, the computation time and the memory required will increase for each update of the estimate of σ as the sample size, say n, increases. This is prohibitive especially when multiple MCMC chains are run simultaneously. Wu (2009) modified the batch mean approach into a recursive version without sacrificing the convergence rate of $O(n^{-1/3})$. Meanwhile, the memory complexity of each update is reduced from the traditional $O(n)$ to $O(1)$.

* Correspondence to: Department of Mathematical Sciences, Indiana University-Purdue University, Indianapolis 402 N. Blackford, LD 270, Indianapolis, IN, 46202, USA. Tel.: +1 317 2780473; fax: +1 317 2743460.
E-mail addresses: weizheng@iupui.edu (W. Zheng), yong.jin@warrington.ufl.edu (Y. Jin), gzhang123@gmail.com (G. Zhang).

http://dx.doi.org/10.1016/j.spl.2016.02.014
0167-7152/

In this paper, we incorporate the prewhitening idea (Press and Tukey, 1956) into Wu (2009)'s approach in order to achieve reduced mean squared error (MSE) for the estimation of σ^2. The benefit of prewhitening is that the resulting residuals are less dependent than the original data and one can obtain more accurate estimate of the TAVC for the residuals due to larger effective degrees of freedom.[1]

We adopt the $AR(1)$ prefilter in constructing recursive estimate of the TAVC. By this small modification, we would achieve substantial improvements on estimation efficiency under some circumstances. Theoretical conditions are given for deciding if the prewhitening is needed or not. The rest of the paper is organized as follows. Section 2 provides the algorithms of recursive estimation of TAVC. In Section 3, we derive the asymptotic properties of the proposed estimators. In Section 4, we discuss efficiency comparisons between the proposed method and Wu's method in terms of both theoretical results and simulation studies. Section 5 concludes the paper. All the proofs are omitted here and could be found in a longer version of the paper.

2. Proposed algorithms

For notational convenience, we assume $\mu = 0$ throughout the paper without loss of generality. All results in the paper shall remain the same for any other values of μ. We first propose an algorithm for estimating σ^2 through prewhitening when μ is known. Then we generalize it to the unknown case. Here are some frequently used notations throughout the paper: $a_k = \lfloor ck^p \rfloor$; $t_i = \sum_{k\in\mathbb{N}} a_k I_{a_k \le i < a_{k+1}}$; $\rho = \gamma(1)/\gamma(0)$; $\hat{\gamma}_n(k) = n^{-1}\sum_{i=|k|+1}^{n} X_i X_{i-|k|}$; $\tilde{\gamma}_n(k) = n^{-1}\sum_{i=|k|+1}^{n}(X_i - \bar{X}_n)(X_{i-|k|} - \bar{X}_n)$; $\hat{\rho}_n = \hat{\gamma}_n(1)/\hat{\gamma}_n(0)$; $\tilde{\rho}_n = \tilde{\gamma}_n(1)/\tilde{\gamma}_n(0)$; and $v_n = \sum_{i=1}^{n} l_i$, where $l_i = i - t_i + 1$.

2.1. When μ is known

To estimate σ^2, Wu (2009) proposed the statistic $v_n^{-1}\sum_{i=1}^{n}(\sum_{j=t_i}^{i} X_j)^2$, which allows the recursive algorithm with the memory complexity of $O(1)$. One way of improving the accuracy of the estimation is to prewhiten the original sequence so that the transformed sequence has weaker dependence and the corresponding estimation of TAVC would be more accurate. Here, we adopt the $AR(1)$ model as the prefilter to produce a new sequence, i.e. $e_i = X_i - \rho X_{i-1}$ with $\rho = \gamma(1)/\gamma(0)$. Due to the relationship $\sigma^2 = \sigma_e^2/(1-\rho)^2$, it is reasonable to estimate σ^2 by $V_n/(v_n(1-\rho)^2)$ when ρ is known, where

$$V_n = \sum_{i=1}^{n} W_i^2 \quad \text{and} \quad W_i = \sum_{j=t_i}^{i}(X_j - \rho X_{j-1}).$$

When ρ is unknown, one only need to plug in the estimate of ρ and hence σ^2 can be estimated by

$$\hat{\sigma}_n^2 = \hat{V}_n/(v_n(1-\hat{\rho}_n)^2), \tag{1}$$

where $\hat{V}_n = \sum_{i=1}^{n}\hat{W}_{i,n}^2$ and $\hat{W}_{i,n} = \sum_{j=t_i}^{i}(X_j - \hat{\rho}_n X_{j-1})$. It can be shown that

$$\hat{V}_n = \sum_{i=1}^{n}\left(\sum_{j=t_i}^{i} X_j\right)^2 + \hat{\rho}_n^2 \sum_{i=1}^{n}\left(\sum_{j=t_i}^{i} X_{j-1}\right)^2 - 2\hat{\rho}_n \sum_{i=1}^{n}\left(\sum_{j=t_i}^{i} X_j\right)\left(\sum_{j=t_i}^{i} X_{j-1}\right).$$

To simplify notations, let $S_{n,0} = \sum_{i=1}^{n} X_i^2$, $S_{n,1} = \sum_{i=2}^{n} X_i X_{i-1}$, $W_{i,0} = \sum_{j=t_i}^{i} X_j$, $W_{i,1} = \sum_{j=t_i}^{i} X_{j-1}$, $V_{n,0} = \sum_{i=1}^{n} W_{i,0}^2$, $V_{n,1} = \sum_{i=1}^{n} W_{i,1}^2$, and $V_{n,2} = \sum_{i=1}^{n} W_{i,0}W_{i,1}$. As a result, $\hat{\rho}_n = S_{n,1}/S_{n,0}$ and $\hat{V}_n = V_{n,0} + \hat{\rho}_n^2 V_{n,1} - 2\hat{\rho}_n V_{n,2}$. Now $\hat{\sigma}_n^2$ can be calculated recursively by the following Algorithm.

Algorithm 1. At stage n, we store $(k_n, v_n, X_n, S_{n,0}, S_{n,1}, \hat{\rho}_n, W_{n,0}, W_{n,1}, V_{n,0}, V_{n,1}, V_{n,2})$. Note that $t_n = a_{k_n}$. When $n = 1$, the vector is $(1, 1, X_1, X_1^2, 0, 0, X_1, 0, X_1^2, 0, 0)$. At stage $n+1$, we update the vector by

1. If $n + 1 = a_{1+k_n}$, let $W_{n+1,0} = X_{n+1}$, $W_{n+1,1} = X_n$ and $k_{n+1} = 1 + k_n$; Otherwise, let $W_{n+1,0} = W_{n,0} + X_{n+1}$, $W_{n+1,1} = W_{n,1} + X_n$ and $k_{n+1} = k_n$.
2. Let $S_{n+1,0} = S_{n,0} + X_{n+1}^2$, $S_{n+1,1} = S_{n,1} + X_{n+1}X_n$, $V_{n+1,0} = V_{n,0} + W_{n+1,0}^2$, $V_{n+1,1} = V_{n,1} + W_{n+1,1}^2$, $V_{n+1,2} = V_{n,2} + W_{n+1,0}W_{n+1,1}$, and $v_{n+1} = v_n + (n + 2 - a_{k_{n+1}})$.
3. Let $\hat{\rho}_{n+1} = S_{n+1,1}/S_{n+1,0}$.
4. Calculate $\hat{V}_{n+1} = V_{n+1,0} + \hat{\rho}_{n+1}^2 V_{n+1,1} - 2\hat{\rho}_{n+1}V_{n+1,2}$

Output: $\hat{\sigma}_{n+1}^2 = \hat{V}_{n+1}/(v_{n+1}(1-\hat{\rho}_{n+1})^2)$.

[1] For example, suppose the residuals $\{e_i\}_{i\in\mathbb{Z}}$ are generated from $\{X_i\}_{i\in\mathbb{Z}}$ by the mechanism of $a(L)X_i = b(L)e_i$, where L is the lag operate and $a(\cdot)$ and $b(\cdot)$ are polynomial functions so that their roots do not fall in the unit circle. It is well known that $\sigma^2 = a(1)^{-2}b(1)^2\sigma_e^2$, where σ_e^2 is the TAVC of $\{e_i\}_{i\in\mathbb{Z}}$. If $\{X_i\}_{i\in\mathbb{Z}}$ follows an ARMA model, $\{e_i\}_{i\in\mathbb{Z}}$ becomes the white noise sequence under the proper choice of $a(\cdot)$ and $b(\cdot)$. In the latter case, the estimation of σ_e^2 could be a lot more accurate than that of the original sequence.

2.2. When μ is unknown

In this case, we have to centralize the data by the sample average and then apply Algorithm 1 to the centralized data. Hence one can estimate σ^2 by

$$\tilde{\sigma}_n^2 = \tilde{V}_n/(\upsilon_n(1-\tilde{\rho}_n)^2), \tag{2}$$

where

$$\tilde{V}_n = \sum_{i=1}^{n} \tilde{W}_{i,n}^2, \qquad \tilde{W}_{i,n} = \sum_{j=t_i}^{i}(X_j - \bar{X}_n - \tilde{\rho}_n(X_{j-1} - \bar{X}_n)). \tag{3}$$

Eq. (3) can be rewritten as

$$\tilde{V}_n = \tilde{V}_n^* + (1-\tilde{\rho}_n)^2\bar{X}_n^2 d_n - 2(1-\tilde{\rho}_n)\bar{X}_n U_n, \tag{4}$$

where $\tilde{\rho}_n = \sum_{i=2}^n (X_i - \bar{X}_n)(X_{i-1} - \bar{X}_n)/\sum_{i=1}^n (X_i - \bar{X}_n)^2$, $\tilde{V}_n^* = \sum_{i=1}^n (\sum_{j=t_i}^i (X_j - \tilde{\rho}_n X_{j-1}))^2$, $d_n = \sum_{i=1}^n l_i^2$ and $U_n = \sum_{i=1}^n l_i \sum_{j=t_i}^i (X_j - \tilde{\rho}_n X_{j-1})$. Note that $\tilde{V}_n^*$ is similar with $\hat{V}_n$ except that the $\hat{\rho}_n$ therein is replaced by $\tilde{\rho}_n$. Also note that $\bar{X}_{n+1} = (n\bar{X}_n + X_{n+1})/(n+1)$ and

$$\tilde{\rho}_n = \frac{\sum_{i=2}^{n} X_i X_{i-1} + \bar{X}_n(X_1 + X_n) - (n+1)\bar{X}_n^2}{\sum_{i=1}^{n} X_i^2 - n\bar{X}_n^2}.$$

We can update $\tilde{\sigma}_n^2$ recursively by the following Algorithm.

Algorithm 2. At stage n, we store $(k_n, \upsilon_n, d_n, X_1, X_n, \bar{X}_n, S_{n,0}, S_{n,1}, \tilde{\rho}_n, W_{n,0}, W_{n,1}, V_{n,0}, V_{n,1}, V_{n,2}, U_n)$. Note that $t_n = a_{k_n}$. When $n = 1$, the vector is $(1, 1, 1, X_1, X_1, X_1, X_1^2, 0, 0, X_1, 0, X_1^2, 0, 0, X_1)$. At stage $n+1$, we update the vector by

1. If $n+1 = a_{1+k_n}$, let $W_{n+1,0} = X_{n+1}$, $W_{n+1,1} = X_n$ and $k_{n+1} = 1 + k_n$; Otherwise, let $W_{n+1,0} = W_{n,0} + X_{n+1}$, $W_{n+1,1} = W_{n,1} + X_n$ and $k_{n+1} = k_n$.
2. Let $S_{n+1,0} = S_{n,0} + X_{n+1}^2$, $S_{n+1,1} = S_{n,1} + X_{n+1}X_n$, and $\bar{X}_{n+1} = (n\bar{X}_n + X_{n+1})/(n+1)$.
3. Let $\tilde{\rho}_{n+1} = (S_{n+1,1} + \bar{X}_{n+1}(X_1 + X_{n+1}) - (n+2)\bar{X}_{n+1}^2)/(S_{n+1,0} - (n+1)\bar{X}_{n+1}^2)$.
4. Let $V_{n+1,0} = V_{n,0} + W_{n+1,0}^2$, $V_{n+1,1} = V_{n,1} + W_{n+1,1}^2$, $V_{n+1,2} = V_{n,2} + W_{n+1,0}W_{n+1,1}$, and $\upsilon_{n+1} = \upsilon_n + (n+2-a_{k_{n+1}})$, $d_{n+1} = d_n + (n+2-a_{k_{n+1}})^2$.
5. Let $U_{n+1} = U_n + (n+2-a_{k_{n+1}})(W_{n+1,0} - \tilde{\rho}_{n+1}W_{n+1,1})$ and calculate $\tilde{V}_{n+1}^* = V_{n+1,0} + \tilde{\rho}_{n+1}^2 V_{n+1,1} - 2\tilde{\rho}_{n+1}V_{n+1,2}$.
6. Calculate $\tilde{V}_{n+1} = \tilde{V}_{n+1}^* + (1-\tilde{\rho}_{n+1})^2\bar{X}_{n+1}^2 d_{n+1} - 2(1-\tilde{\rho}_{n+1})\bar{X}_{n+1}U_{n+1}$

Output: $\tilde{\sigma}_{n+1}^2 = \tilde{V}_{n+1}/(\upsilon_{n+1}(1-\tilde{\rho}_{n+1})^2)$.

To implement this algorithm, we have to decide the values of c and p in a_k. As shown later, the optimal choice of p is typically 3/2 and the value of c is determined by the dependence structure of $\{e_i\}_{i\in\mathbb{Z}}$. The details are deferred to Section 4.

3. Asymptotical properties of the estimators

In this section, we shall study the asymptotic properties of $\hat{\sigma}_n^2$ and $\tilde{\sigma}_n^2$ as defined by (1) and (2) respectively. Theorem 1 derives the asymptotic properties of V_n, which leads to Theorem 2 and Corollary 1 regarding the asymptotic properties of $\hat{\sigma}_n$ and $\tilde{\sigma}_n$.

Define $a \wedge b = \min(a, b)$. A random variable ξ is said to be in $\mathcal{L}^p (p > 0)$ if $\|\xi\|_p := [\mathbb{E}(|\xi|^p)]^{1/p} < \infty$, where $\|\xi\| = \|\xi\|_2$. For two sequences $\{a_n\}$ and $\{b_n\}$, write $a_n \sim b_n$, if $\lim_{n\to\infty} a_n/b_n = 1$; write $a_n = o(b_n)$ if $\lim_{n\to\infty} a_n/b_n = 0$; write $a_n = O(b_n)$ if $\limsup_{n\to\infty} a_n/b_n < \infty$; write $a_n = o_{a.s}(b_n)$ if $\lim_{n\to\infty} a_n/b_n = 0$ almost surely; write $a_n \asymp b_n$ or $a_n = O(b_n)$ if there exists a constant $c > 0$, such that $1/c \le |a_n/b_n| \le c$ for all large n.

Throughout the paper, we assume that $\{X_i\}$ is a stationary causal process of the form

$$X_i = g(\mathcal{F}_i), \quad \mathcal{F}_i = (\ldots, \varepsilon_{i-1}, \varepsilon_i),$$

and $\{\varepsilon_i\}$ are i.i.d. Let $X_i' = g(\mathcal{F}_i')$, where $\mathcal{F}_i' = (\mathcal{F}_{-1}, \varepsilon_0', \varepsilon_1, \ldots, \varepsilon_i)$ and ε_i' is an i.i.d. copy of ε_i. The quantity $\delta_\alpha(i) = \|X_i - X_i'\|_\alpha$ is called the physical dependence measure by Wu (2005). Recall $e_i = X_i - \rho X_{i-1}$ and define $e_i' = X_i' - \rho X_{i-1}'$. Let $\theta_\alpha(i) := \|e_i - e_i'\|_\alpha \le 2\delta_\alpha(i)$, it can be shown that the following two conditions are equivalent

$$\sum_{i=0}^{\infty} \delta_\alpha(i) < \infty, \tag{5}$$

$$\sum_{i=0}^{\infty} \theta_\alpha(i) < \infty \tag{6}$$

in view of $X_i = \sum_{k\geq 0} \rho^k e_{i-k}$. Particularly, for the $AR(1)$ model we have $\theta_\alpha(i) = \|\varepsilon_0 - \varepsilon_0'\|_\alpha 1_{i=0}$. Since e_i is adapted to $\mathcal{F}_i$ and $\mathbb{E}(e_i) = 0$, under mild condition of

$$\sum_{i=0}^{\infty} \|\mathcal{P}_0 e_i\| < \infty, \quad \text{where } \mathcal{P}_i \cdot = \mathbb{E}(\cdot|\mathcal{F}_i) - \mathbb{E}(\cdot|\mathcal{F}_{i-1}), \tag{7}$$

we have the functional central limit theorem

$$\frac{1}{\sqrt{n}} \left\{ \sum_{i\leq nt} e_i, 0 \leq t \leq 1 \right\} \xrightarrow{d} \{\sigma_e \mathbb{B}(t), 0 \leq t \leq 1\},$$

where $\mathbb{B}$ is the standard Brownian motion and $\sigma_e = \|\sum_{i=0}^{\infty} \mathcal{P}_0 e_i\| = \sum_{k\in\mathbb{Z}} \mathbb{E}(e_0 e_k) = (1-\rho)^2\sigma^2$. Note that (7) is weaker than (5). Now that σ_e^2 is the TAVC for the time series $\{e_i\}$. Wu (2009) proposed V_n/v_n to estimate σ_e^2. The properties of V_n are given in the following theorem, and its proof follows the same way as in Wu (2009).

Theorem 1. *Let $a_k = \lfloor ck^p \rfloor$, $k \geq 1$, where $c > 0$ and $p > 1$ are constants.*

(i) *Assume that $\mathbb{E}(X_i) = 0$, $X_i \in \mathcal{L}^\alpha$ and (6) holds for some $\alpha \in (2, 4]$, we have*

$$\|V_n - \mathbb{E}(V_n)\|_{\alpha/2} = O(n^\tau), \qquad \left\| \max_{n\leq N} |V_n - \mathbb{E}(V_n)| \right\|_{\alpha/2} = O(N^\tau \log N),$$

where $\tau = 2/\alpha + 3/2 - 3/(2p)$.

(ii) *Assume that $\mathbb{E}(X_i) = 0$, $X_i \in \mathcal{L}^\alpha$ and (6) holds for some $\alpha > 4$, we have*

$$\|V_n - \mathbb{E}(V_n)\|_{\alpha/2} = O(n^{2-3/(2p)}), \qquad \left\| \max_{n\leq N} |V_n - \mathbb{E}(V_n)| \right\|_{\alpha/2} = O(N^{2-3/(2p)}).$$

(iii) *Assume that $\mathbb{E}(X_i) = 0$, and*

$$\sum_{i=0}^{\infty} \|\mathcal{P}_0 e_i\|_\alpha < \infty, \tag{8}$$

for $\alpha > 4$, we have

$$\lim_{n\to\infty} \frac{\|V_n - \mathbb{E}(V_n)\|}{n^{2-3/(2p)}} = \frac{\sigma_e^2 p^2 c^{3/(2p)}}{\sqrt{12p - 9}}.$$

(iv) *Assume that $\mathbb{E}X_i = 0$, and for some $q \in (0, 1]$ either* (a)

$$\sum_{k=0}^{\infty} k^q \|\mathcal{P}_0 e_k\| < \infty \tag{9}$$

or (b) $\sum_{k=0}^{\infty} \|\mathcal{P}_0 X_k\| < \infty$ *and*

$$\sum_{k=0}^{\infty} k^q \|\mathcal{P}_0(e_k - \rho e_{k+1})\| < \infty, \tag{10}$$

are satisfied, we have

$$\mathbb{E}(V_n) - v_n(1-\rho)^2\sigma^2 = O(n^{1+(1-q)(1-1/p)}).$$

Lemma 1 indicates that the magnitude of $\hat{V}_n - V_n$ and $\tilde{V}_n - V_n$ are both negligible compared to that of $V_n - v_n(1-\rho)^2\sigma^2$. Hence the almost sure bounds of $\hat{\sigma}^2 - \sigma^2$ and $\tilde{\sigma}^2 - \sigma^2$ could be resorted to Theorem 1.

Lemma 1. *Assume $p > 1$, $\mathbb{E}(X_i) = 0$, $X_i \in \mathcal{L}^\alpha$, and (5) holds with $\alpha \geq 4$.* (i) *For $1 < \beta < \alpha$, let $\beta' = 2 \wedge \beta$, then we have*

$$\left\| \max_{k\leq n} |\hat{V}_k - V_k| \right\|_{\beta/4} = O(n^{3/2-1/p} n^{(1-1/p)(2/\beta'-1)}).$$

(ii) *If $4/3 \leq \beta < \alpha$ and the conditions of Theorem 1*(iv) *hold, we have*

$$\|\hat{V}_n/v_n - (1-\rho)^2\sigma^2\|_{\beta/4} = O(n^{-\phi(p)}), \tag{11}$$

where $\phi(p) = (2p)^{-1} \wedge q(1-1/p)$.

As will be shown later, (11) plays the critical role in determining the deviation of $\hat{\sigma}_n^2$ and $\tilde{\sigma}_n^2$ from σ^2. It can be shown easily that $\phi(p)$ reaches its maximum of $q/(2q+1)$ at $p = 1 + (2q)^{-1}$. By imposing the latter condition, we derive the following results regarding asymptotic properties of $\hat{\sigma}_n$ and $\tilde{\sigma}_n$.

Theorem 2. *Suppose* $\mathbb{E}(X_i) = 0$, (9) *holds with* $q \in (0, 1]$, *and* $p = 1 + (2q)^{-1}$.

(i) *Assume* $X_i \in \mathcal{L}^\alpha$ *and* (5) *holds with* $\alpha > 4$, *then we have*

$$\hat{\sigma}_n^2 - \sigma^2 = o_{a.s.}(n^{-q/(2q+1)} \log n). \tag{12}$$

(ii) *Assume* $X_i \in \mathcal{L}^4$ *and* (5) *holds with* $\alpha = 4$, *then we have*

$$\hat{\sigma}_n^2 - \sigma^2 = o_{a.s.}(n^{-q/(2q+1)} \log^{2+\epsilon} n) \tag{13}$$

for any small $\epsilon > 0$.

Corollary 1. *Suppose* $\mathbb{E}(X_i) = 0$, (9) *holds with* $q \in (0, 1]$, *and* $p = 1 + (2q)^{-1}$.

(i) *Assume* $X_i \in \mathcal{L}^\alpha$ *and* (5) *holds with* $\alpha > 4$, *then we have*

$$\tilde{\sigma}_n^2 - \sigma^2 = o_{a.s.}(n^{-q/(2q+1)} \log n). \tag{14}$$

(ii) *Assume* $X_i \in \mathcal{L}^4$ *and* (5) *holds with* $\alpha = 4$, *then we have*

$$\tilde{\sigma}_n^2 - \sigma^2 = o_{a.s.}(n^{-q/(2q+1)} \log^{2+\epsilon} n) \tag{15}$$

for any small $\epsilon > 0$.

(iii) *Assume* $X_i \in \mathcal{L}^\alpha$ *and* (5) *holds with* $\alpha \geq 4$, *then we have*

$$\|\tilde{V}_n/v_n - (1-\rho)^2\sigma^2\|_{\beta/4} = O(n^{-q/(2q+1)}) \tag{16}$$

when $4/3 \leq \beta < \alpha$.

4. The comparison of efficiencies

Assume that (8) holds with $\alpha > 8$ and (9) holds with $q = 1$. Theorem 2 suggests optimal $p = 3/2$ in the sequence $a_k = \lfloor ck^p \rfloor$. Let $S_n = \sum_{i=1}^n e_i$, $\eta = 2\sum_{k=1}^\infty k\gamma(k)$, and $\eta_e = 2\sum_{k=1}^\infty k\mathbb{E}(e_0 e_k)$. As $l \to \infty$, we have

$$l\sigma_e^2 - \mathbb{E}(S_l^2) = \eta_e + o(1).$$

Following the proof of Lemma 1 and the arguments in Section 4 of Wu (2009), we have

$$\|\tilde{V}_n/v_n - \sigma_e^2\|^2 \sim \|V_n/v_n - \sigma_e^2\|^2 \sim \left(\sigma_e^4 \frac{16c^{2/3}}{9} + \eta^2 \frac{256}{81c^{4/3}}\right) n^{-2/3}.$$

In minimizing the MSE, we choose $c = 4\sqrt{2}|\eta_e|/(3\sigma_e^2)$ so that

$$\|\tilde{V}_n/v_n - \sigma_e^2\|^2 \sim \frac{2^{14/3}}{3^{5/3}} \eta_e^{2/3} \sigma_e^{8/3} n^{-2/3}. \tag{17}$$

By direct calculations we have

$$\tilde{\sigma}_n^2 - \sigma^2 = \frac{\Delta_n}{(1-\tilde{\rho}_n)^2},$$

where $\Delta_n = \tilde{V}_n/v_n - \sigma_e^2 + (2 - \rho - \tilde{\rho}_n)(\tilde{\rho}_n - \rho)\sigma^2$. By Proposition 1 and (17) we have

$$\|\Delta_n\|^2 \sim \frac{2^{14/3}}{3^{5/3}} \eta_e^{2/3} \sigma_e^{8/3} n^{-2/3}. \tag{18}$$

Let $\breve{\sigma}^2$ be the Wu's estimator of σ^2.

Theorem 3. *Suppose* $\mathbb{E}(X_i) = 0$, (9) *holds with* $q \in (0, 1]$, *and* $p = 1 + 1/2q$.

(i) *Assume* $X_i \in \mathcal{L}^\alpha$, *and* (5) *holds with* $\alpha > 40$.

$$\|\tilde{\sigma}_n^2 - \sigma^2\|_{\beta/20} - \frac{\|\Delta_n\|_{\beta/20}}{(1-\rho)^2} = o(n^{-1/3}) \tag{19}$$

for $40 \leq \beta < \alpha$.

(ii) *Assume $X_i \in \mathcal{L}^\alpha$, and (5) holds with $\alpha > 40$. We have*

$$\frac{\|\tilde{\sigma}_n^2 - \sigma^2\|^2}{\|\check{\sigma}_n^2 - \sigma^2\|^2} \to \left(1 - \frac{2\gamma(1)}{(1-\rho)^2\eta}\right)^{2/3} := \lambda. \tag{20}$$

To estimate c, we can refer to Algorithm 3 in Wu (2009) except that $\hat{\gamma}(k)$ therein should be replaced by an estimate of $\mathbb{E}(e_0 e_k) = (1+\rho^2)\gamma(k) - \rho(\gamma(k+1) + \gamma(k-1))$. Therefore, we could substitute ρ and $\gamma(k)$ by their estimates based on a small portion of the initial data. Since $\lambda < 1$ if and only if $2\gamma(1)/\eta > 0$, the proposed algorithm here would outperform Wu's algorithm whenever $\gamma(1) \neq 0$ and

$$\sum_{k\geq 2} k\gamma(k)/\gamma(1) > -1.$$

4.1. ARMA model

Consider $ARMA(\tilde{p}, q)$ model

$$a(L)X_i = b(L)\varepsilon_i, \tag{21}$$

where $a(\cdot)$ and $b(\cdot)$ are polynomials of orders $\tilde{p}$ and q respectively. Suppose $a(z) = \prod_{i=1}^{p_0}(1 - r_i z)^{m_i}$ with $\sum_{i=1}^{p_0} m_i = \tilde{p}$ and $|r_i| < 1$. In the sequel, we will discuss the calculation of λ for the case of $\tilde{p} > q$. No extra difficulty is needed to deal with the case of $\tilde{p} \leq q$. From Brockwell and Davis (1991), the autocovariance function has the representation of

$$\gamma(k) = \sum_{i=1}^{p_0}\sum_{j=0}^{m_i-1} d_{ij} k^j r_i^k, \quad k \geq 0, \tag{22}$$

where d_{ij} could be found by solving the system of linear equations (22) for $k = 0, 1, \ldots, p-1$. See Brockwell and Davis (1991) or Karanasos (1998) for details. Define $\Xi_i(r) = \sum_{k=1}^{\infty} k^i r^k$, then $\Xi_0(r) = r/(1-r)$ and $\Xi_i(r),\ i \geq 1$ could be derived recursively by the relation $\Xi_{i+1}(r) = r\Xi_i'(r), i \geq 0$. For example, we have $\Xi_1(r) = r/(1-r)^2$ and $\Xi_2(r) = (r^2+r)/(1-r)^3$. For model (21) we have

$$\frac{2\gamma(1)}{(1-\rho)^2\eta} = \frac{\sum_{i=1}^{p_0}\sum_{j=0}^{m_i-1} d_{ij} r_i/(1-\rho)^2}{\sum_{i=1}^{p_0}\sum_{j=0}^{m_i-1} d_{ij}\Xi_{j+1}(r_i)} \quad \text{and} \quad \rho = \frac{\sum_{i=1}^{p_0}\sum_{j=0}^{m_i-1} d_{ij} r_i}{\sum_{i=1}^{p_0} d_{i0}}.$$

If all the r_i's are distinct, these two equations reduce to

$$\frac{2\gamma(1)}{(1-\rho)^2\eta} = \frac{\sum_{i=1}^{\tilde{p}} d_i r_i/(1-\rho)^2}{\sum_{i=1}^{\tilde{p}} d_i r_i/(1-r_i)^2} \quad \text{and} \quad \rho = \frac{\sum_{i=1}^{\tilde{p}} d_i r_i}{\sum_{i=1}^{\tilde{p}} d_i} \tag{23}$$

with the convention of $d_i = d_{i0}$.

Simulation studies are carried out for three different examples of Model (21) and the results are displayed in Figs. 1–3. In all these three examples, we adopted $\sigma_\varepsilon^2 = 1$, and the value of λ is calculated by (23). The black line therein is associated with Wu's estimator and the red line is associated with the proposed estimator. Figs. 1 and 2 verify that when $\lambda < 1$ the proposed method outperforms Wu's. The main reason of the smaller MSE is due to the reduced bias by introducing the prewhitening. On the other hand, Fig. 3 shows that when $\lambda > 1$ the prewhitening is not recommended.

5. Conclusions

This paper studies a prewhitened version of Wu's algorithm for estimating the TAVC in a recursive way. Asymptotic properties of the estimate is derived for a general causal stationary process and simulation studies are carried out for both linear and nonlinear time series models. Here we adopt the simplest prefilter, i.e. the $AR(1)$ model, which achieves significant improvements on the estimation efficiency whenever $\gamma(1) \neq 0$ and $\sum_{k\geq 2} k\gamma(k)/\gamma(1) > -1$. In practice, one may keep two lines of updating algorithms simultaneously in the initial stage and then choose one from them according to the latter condition. The simple form of $AR(1)$ facilitates the construction and implementation of the recursive estimation of TAVC with the memory complexity of $O(1)$.

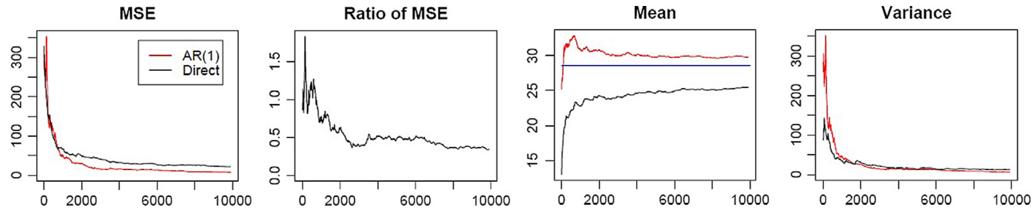

Fig. 1. $\lambda \approx 0.4150$ for ARMA Model with $a(z) = (1-0.8z)(1+0.05z)(1-0.02z)$ and $b(z) = 1+0.1z$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

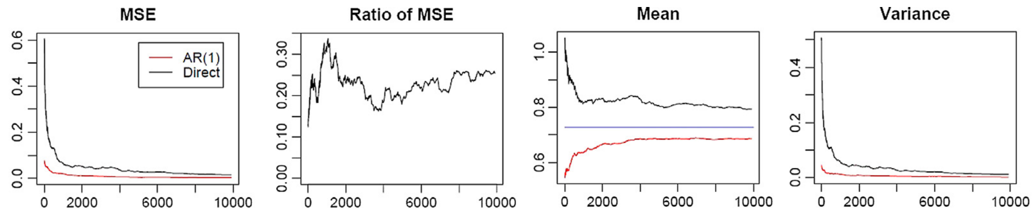

Fig. 2. $\lambda \approx 0.2220$ for ARMA Model with $a(z) = (1+0.9z)(1+0.8z)(1+0.5z)(1-0.2z)$ and $b(z) = 1+0.9z+0.8z^2+0.8z^3$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

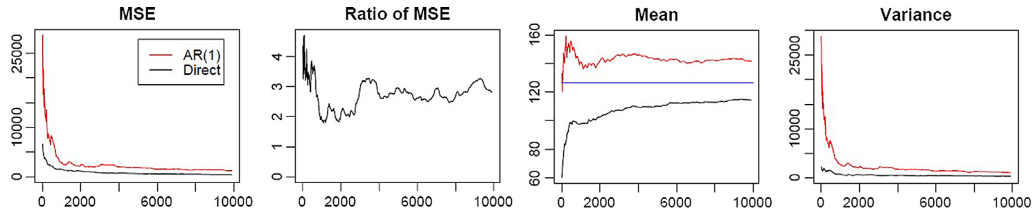

Fig. 3. $\lambda \approx 2.9947$ for ARMA Model with $a(z) = (1-0.8z)(1+0.5z)(1-0.2z)$ and $b(z) = 1+0.9z+0.8z^2$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Appendix

Proposition 1. (i) *Assume* $\mathbb{E}(X_i) = 0$, $X_i \in \mathcal{L}^\alpha$ *and* (5) *holds with* $\alpha > 2$. *For* $0 < \beta < \alpha$, *the following sequences are all uniformly integrable:*

$$\{[n^{1/2}(\hat{\gamma}_n(k) - \gamma(k))]^{\beta/2}, n \geq 1\}$$
$$\{[n^{1/2}(\hat{\rho}_n(k) - \rho(k))]^{\beta/2}, n \geq 1\}$$
$$\{[n^{1/2}(\tilde{\gamma}_n(k) - \gamma(k))]^{\beta/2}, n \geq 1\}$$
$$\{[n^{1/2}(\tilde{\rho}_n(k) - \rho(k))]^{\beta/2}, n \geq 1\}.$$

(ii) *Assume* $\mathbb{E}(X_i) = 0$, $X_i \in \mathcal{L}^4$ *and* (5) *holds with* $\alpha = 4$, *we have*

$$\sqrt{n}(\hat{\rho}_n - \rho) \xrightarrow{d} N(0, \sigma_\rho^2),$$

$$\sqrt{n}(\tilde{\rho}_n - \rho) \xrightarrow{d} N(0, \sigma_\rho^2),$$

where $\sigma_\rho^2 = \|\sum_{i=0}^{\infty} \mathcal{P}_0 X_i \tilde{e}_i\|^2/\gamma(0)^2$ *and* $\tilde{e}_i = X_{i-1} - \rho X_i$.

(iii) *Assume* $\mathbb{E}(X_i) = 0$, $X_i \in \mathcal{L}^\alpha$ *and* (5) *holds with* $\alpha \geq 4$. *Then for* $0 < \beta < \alpha$

$$\|\sqrt{n}(\hat{\rho}_n - \rho)\|_{\beta/2} \to \|\sigma_\rho \mathcal{G}\|_{\beta/2},$$

$$\|\sqrt{n}(\tilde{\rho}_n - \rho)\|_{\beta/2} \to \|\sigma_\rho \mathcal{G}\|_{\beta/2},$$

where σ_ρ *is as defined in* (ii) *and* $\mathcal{G}$ *stands for a standard normal distribution.*

Proposition 2. *If* $\mathbb{E}(X_i) = 0$, $X_i \in \mathcal{L}^4$ *and* (5) *holds with* $\alpha = 4$, *we have*

$$\hat{\gamma}_n(k) - \gamma(k) = o_{a.s.}(n^{-1/2}\log^2 n),$$
$$\hat{\rho}_n(k) - \rho(k) = o_{a.s.}(n^{-1/2}\log^2 n),$$

$$\tilde{\gamma}_n(k) - \gamma(k) = o_{a.s.}(n^{-1/2}\log^2 n), \tag{24}$$

$$\tilde{\rho}_n(k) - \rho(k) = o_{a.s.}(n^{-1/2}\log^2 n). \tag{25}$$

References

Alexopoulos, C., Goldsman, D., 2004. To batch or not to batch? ACM Trans. Model. Comput. Simul. 14, 76–114.
Bühlmann, P., 2002. Bootstraps for time series. Statist. Sci. 17, 52–72.
Brockwell, P.J., Davis, R.A., 1991. Time Series: Theory and Methods, seconnd ed. Springer, New York.
Jones, G.L., Haran, M., Caffo, B.S., Neath, R., 2006. Fixed-width output analysis for Markov chain Monte Carlo. J. Amer. Statist. Assoc. 101, 1537–1547.
Karanasos, M., 1998. A new method for obtaining the autocovariance of an AMRA model: An exact form solution. Econometric Theory 14, 622–640.
Lahiri, S.N., 2003. Resampling Methods for Dependent Data. Springer, New York.
Phillips, P.C.B., Xiao, Z., 1998. A primer on unit root testing. J. Econ. Surv. 12, 423–469.
Politis, D.N., Romano, J.P., Wolf, M., 1999. Subsampling. Springer, New York.
Press, H., Tukey, J.W., 1956. Power Spectral Methods of Analysis and their Application to Problems in Airplane Dynamic. In: Flight Test Manual, NATO, Advisory Group for Aeronautical Research and Development, vol. IV-C. pp. 1–41. Reprinted as Bell System Monograph No. 2606.
Song, W.M., Schmeiser, B.W., 1995. Optimal mean-squared-error batch sizes. Manage. Sci. 41, 110–123.
Stock, J.H., 1994. Unit roots, structural breaks and trends. In: Handbook of Econometrics, Vol. IV. North-Holland, Amsterdam, pp. 2739–2841.
Wu, W.B., 2005. Nonlinear system theory: Another look at dependence. Proc. Natl. Acad. Sci. USA 102, 14150–14154.
Wu, W.B., 2009. Recursive estimation of time-average variance constants. Ann. Appl. Probab. 19, 1529–1552.