Journal of Multivariate Analysis 114 (2013) 40–53

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

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

Efficient inference for autoregressive coefficients in the presence of trends

D. Qiu [a], Q. Shao [b], L. Yang [a,c,*]

[a] Soochow University, Suzhou, 215006, China
[b] University of Toledo, Toledo, OH 43606, USA
[c] Michigan State University, East Lansing, MI 48824, USA

ARTICLE INFO

Article history:
Received 3 March 2012
Available online 21 July 2012

AMS 2000 subject classifications:
primary 62F12
62M10
secondary 62G08
62G20

Keywords:
Autoregressive time series
Local polynomial
Oracle efficiency
Yule–Walker estimator
Moving average

ABSTRACT

Time series often contain unknown trend functions and unobservable error terms. As is known, Yule–Walker estimators are asymptotically efficient for autoregressive time series. The focus of this article is the Yule–Walker estimators for time series with trends. A nonparametric detrending procedure is proposed. It is concluded that the asymptotic properties of the Yule–Walker estimators of autoregressive coefficients are not altered by the detrending procedure. The results of the simulation studies and real data application corroborate the asymptotic theory.

1. Introduction

Typically, the first step in time series analysis is to "separate" the deterministic trend and seasonality components from the stochastic noise component. One of the common approaches taught by textbooks is to apply a moving average filter to "remove" the slowly varying trend and the seasonality from the time series data, and then proceed to make inference based on the residual series which is used as a substitute of the unobserved time series without trends. Although much has been done for such residual based inference, little attention has been paid to the appropriateness of substituting the residual sequence for the unobservable time series except for trends with known parametric forms; see for example, [9] and Chapter 9 of [4].

There are two most relevant works pertaining to the asymptotic property of the Yule–Walker estimators for autoregressive coefficients of time series with nonparametric trends. In particular, Truong [8] tackled the issue and established the asymptotics of Yule–Walker estimators when the trend was estimated by moving average or kernel regression techniques under the restrictive assumption of Gaussian noise; Shao and Yang [6] showed the oracle efficiency of Yule–Walker estimators when the trend was estimated by B-splines under mild moment assumptions on the noise. By "oracle efficiency" we refer to the asymptotic equivalence of the autoregressive coefficient estimators based on the unobserved stationary noise sequence and the computed residual sequence. In other words, Yule–Walker estimators from

* Corresponding author at: Soochow University, Suzhou, 215006, China.
E-mail addresses: yang@stt.msu.edu, yanglijian@suda.edu.cn (L. Yang).

0047-259X/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2012.07.016

the residual sequence with the trend estimated and removed are as efficient as those from the time series with the trend known by "oracle" and deleted. Both of these two articles, however, did not provide data-driven smoothing parameter selection.

In this paper we will propose a nonparametric detrending procedure. This procedure constructed from local linear regression is a modified version of the well-known moving average filter. Compared with the classic moving average method, this new procedure has two major advantages: it not only automatically corrects boundary bias, but chooses the optimal moving average parameter based on the data as well. These two features make it more suitable for practitioners to deal with time series data with trends, as one sees from simulation results in Section 4, using the classic moving average filter can lead to erroneous estimation of the autoregressive coefficients. Moreover, under strict stationarity and similar moment conditions to those in [6], the oracle efficiency of the Yule–Walker estimators based on residuals is rigorously established without requiring Gaussianity of the noise sequence. It is worth mentioning that the major theoretical result of this paper is derived by bounding the moments of terms in the error decomposition, which is very different from the technique utilized by Shao and Yang [6] for *B*-spline smoothing. Their technique is closely related to Song and Yang [7].

The paper is organized as follows. Section 2 will describe the proposed local linear trend filter and its theoretical properties; Section 3 will outline the data-driven selection of the optimal moving average parameter; Section 4 will present the results of simulation studies and real data analysis. The proofs of all technical results will be provided in the Appendix.

2. Autoregressive coefficient estimation

The observed data $\{X_t\}_{t=1}^n$ we work with are of the form

$$X_t = m(t/n) + Y_t, \quad 1 \le t \le n$$

in which $m(\cdot)$ is the trend and the unobserved noise sequence $\{Y_t\}_{t=1}^n$ is a realization of an autoregressive time series with order p (i.e., AR(p)) that satisfies

$$Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + Z_t. \tag{1}$$

In (1) the innovation sequence $Z_t \sim \text{IID}\left(0, \sigma^2\right)$, i.e., the Z_t's are independent and identically distributed with $EZ_t = 0$, $EZ_t^2 = \sigma^2$, $-\infty < t < \infty$.

The parameters of interest are the autoregressive coefficients $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)^T$. According to Eq. (8.1.1) of [1], they satisfy

$$\boldsymbol{\phi} = \boldsymbol{\Gamma}_p^{-1}\boldsymbol{\gamma}_p, \qquad \boldsymbol{\Gamma}_p = \{\gamma(i-j)\}_{i,j=1}^p, \quad \boldsymbol{\gamma}_p = (\gamma(1), \ldots, \gamma(p))^T \tag{2}$$

in which $\gamma(l) \equiv E(Y_t Y_{t+l})$, $l = 0, \pm 1, \pm 2, \ldots$, denotes the autocovariance function of $\{Y_t\}_{t=-\infty}^{\infty}$. The classic Yule–Walker estimator of $\boldsymbol{\phi}$ is a method of moment estimator based on the noise sequence $\{Y_t\}_{t=1}^n$ and is calculated by

$$\tilde{\boldsymbol{\phi}} = \tilde{\boldsymbol{\Gamma}}_p^{-1}\tilde{\boldsymbol{\gamma}}_p, \qquad \tilde{\boldsymbol{\Gamma}}_p = \{\tilde{\gamma}(i-j)\}_{i,j=1}^p, \quad \tilde{\boldsymbol{\gamma}}_p = (\tilde{\gamma}(1), \ldots, \tilde{\gamma}(p))^T$$

in which the sample autocovariance function $\tilde{\gamma}(l)$ is calculated by

$$\tilde{\gamma}(l) = n^{-1}\sum_{t=1}^{n-l} Y_t Y_{t+l}, \quad 0 \le l \le n-1; \tag{3}$$

see Eq. (8.1.5) of [1]. Throughout this paper, we refer to $\tilde{\boldsymbol{\phi}}$ as the "infeasible" estimator of $\boldsymbol{\phi}$, as apparently it makes use of the unobserved sequence $\{Y_t\}_{t=1}^n$ and thus is not a proper statistic.

In Chapter 1 of [1] and other standard textbooks on time series analysis, various ad hoc estimators of the trend function $m(\cdot)$ are provided. Perhaps the most popular one is the moving average estimator defined by

$$\hat{m}(t/n) = N_{t,q}^{-1}\sum_{|i-t|\le q} X_i, \quad 1 \le t \le n,$$

in which q is the moving average lag and $N_{t,q} = \sum_{|i-t|\le q} 1$ is the number of indices i between $t-q$ and $t+q$. Assuming that $q \le (n-1)/2$, elementary algebra shows that

$$\hat{m}(t/n) = \begin{cases} (2q+1)^{-1}\sum_{i=t-q}^{t+q} X_i, & q+1 \le t \le n-q; \\ (q+t)^{-1}\sum_{i=1}^{t+q} X_i, & 1 \le t \le q; \\ (n-t+q+1)^{-1}\sum_{i=t-q}^{n} X_i, & n-q+1 \le t \le n. \end{cases} \tag{4}$$

The residual sequence of this estimator is used in the second step to compute the sample autocovariance function and the estimates of $\boldsymbol{\phi}$. Although this method is widely recommended, the only theoretical justification for this two-step approach was in [8], which, nonetheless, required Gaussianity of $\{Y_t\}_{t=-\infty}^{\infty}$ (and equivalently, of $\{Z_t\}_{t=-\infty}^{\infty}$). This last restriction rules out many interesting time series data, especially in economics and finance, that exhibit strong non-Gaussian features.

We will construct a modified moving average trend estimator. Under the assumptions on strict stationarity, the moments and causality of $\{Y_t\}_{t=-\infty}^{\infty}$ and proper conditions on the moving average lag q, this estimator leads to the same theoretical optimality as in [8]. Note that the optimality of the proposed trend estimator does not require Gaussianity of $\{Y_t\}_{t=-\infty}^{\infty}$.

To begin with, for any $1 \le t \le n$, the moving average trend estimator $\hat{m}(t/n)$ in (4) is rewritten as a Nadaraya–Watson estimator

$$\hat{m}(t/n) = \frac{\sum_{i=1}^{n} X_i K_h\{(i-t)/n\}}{\sum_{i=1}^{n} K_h\{(i-t)/n\}}, \tag{5}$$

where $K(u) = 0.5 I_{\{|u|\le 1\}}$ is the uniform kernel with I_A being the indicator function for the set A, $K_h(u) = K(u/h)/h$, and $h = q/n$ is called the bandwidth in the kernel smoothing literature. The above estimator $\hat{m}(t/n)$ is the solution to the following local constant least squares problem

$$\hat{m}(t/n) = \underset{a}{\operatorname{argmin}} \sum_{i=1}^{n} (X_i - a)^2 K_h\{(i-t)/n\}.$$

We propose to estimate $m(\cdot)$ by the local linear method. Specifically,

$$\hat{m}(t/n) = \hat{a}, \tag{6}$$

where $\hat{a}$ satisfies

$$\left(\hat{a}, \hat{b}\right) = \underset{(a,b)}{\operatorname{argmin}} \sum_{i=1}^{n} \{X_i - a - b(i-t)/n\}^2 K_h\{(i-t)/n\}.$$

The following is the explicit formula for the proposed trend estimator, the derivation of which is given in the Appendix:

$$\hat{m}(t/n) = \begin{cases} (2q+1)^{-1} \sum_{i=t-q}^{t+q} X_i, & q+1 \le t \le n-q; \\ N_{1t}^{-1} \sum_{i=1}^{t+q} X_i - N_{2t}^{-1} \sum_{i=1}^{t+q} (i-t) X_i, & 1 \le t \le q; \\ N_{3t}^{-1} \sum_{i=t-q}^{n} X_i - N_{4t}^{-1} \sum_{i=t-q}^{n} (i-t) X_i, & n-q+1 \le t \le n, \end{cases} \tag{7}$$

where

$$N_{1t}^{-1} = \frac{4q^2 - 4qt + 6q + 4t^2 - 6t + 2}{(q+t)(q+t-1)(q+t+1)},$$

$$N_{2t}^{-1} = \frac{6(q-t+1)}{(q+t)(q+t-1)(q+t+1)},$$

$$N_{3t}^{-1} = \frac{4(n-t)^2 + 4q^2 - 4q(n-t) + 2(n+q-t)}{(n+q-t+2)(n+q-t+1)(n+q-t)},$$

$$N_{4t}^{-1} = \frac{6(n-q-t)}{(n+q-t+2)(n+q-t+1)(n+q-t)}.$$

Fan and Gijbels [2] promoted local linear estimators over Nadaraya–Watson estimators in that it automatically corrects boundary bias. In other words, the local linear estimator in (6) satisfies

$$\max_{1\le t\le n} \left| E\hat{m}(t/n) - m(t/n) \right| = O\left(h^2\right),$$

whereas the local constant estimator in (4) or (5) has the rate of $O(h)$ at the boundary.

We denote the residuals by

$$\hat{Y}_t = X_t - \hat{m}(t/n), \quad 1 \le t \le n, \tag{8}$$

and the sample autocovariance function by

$$\hat{\gamma}(l) = n^{-1} \sum_{t=1}^{n-l} \hat{Y}_t \hat{Y}_{t+l}, \quad 0 \le l \le n-1. \tag{9}$$

The Yule–Walker estimator of $\boldsymbol{\phi}$ based on the residuals in (8) is defined by

$$\hat{\boldsymbol{\phi}} = \hat{\boldsymbol{\Gamma}}_p^{-1} \hat{\boldsymbol{\gamma}}_p, \qquad \hat{\boldsymbol{\Gamma}}_p = \left\{\hat{\gamma}(i-j)\right\}_{i,j=1}^{p}, \quad \hat{\boldsymbol{\gamma}}_p = \left(\hat{\gamma}(1), \ldots, \hat{\gamma}(p)\right)^T. \tag{10}$$

Before presenting the main results, we state the following assumptions.

(A1) *The trend function* $m(\cdot) \in C^2[0,1]$.
(A2) *The process* $\{Y_t\}_{t=-\infty}^{\infty}$ *is causal; i.e.,* $\inf_{|z|\le 1} \left|1 - \phi_1 z - \cdots - \phi_p z^p\right| > 0$.
(A3) $EY_t^6 < \infty$.
(A4) *The moving average lag* $q = q_n$ *satisfies* $n^{1/2}\log n \ll q_n \ll n^{5/6}$, *that is, as* $n \to \infty$, $q_n^{-1} n^{1/2} \log n \to 0$ *and* $q_n^{-1} n^{5/6} \to \infty$. *In particular, one can take* $q_n \sim n^{4/5}$.

Remark 1. Assumption (A2) and the condition that $Z_t \sim \mathrm{IID}\left(0, \sigma^2\right)$ imply that $\{Y_t\}_{t=-\infty}^{\infty}$ is a strictly stationary sequence with $\sum_{l=-\infty}^{\infty} |\gamma(l)| < \infty$. Also according to Assumption (A4), as $n \to \infty$, the bandwidth $h = h_n = q_n/n \ll n^{-1/8}$ and the number of $n' = n - 2q + 1$ satisfies $n'/n \to 1$.

One of the critical steps in implementing the detrending procedure is to find an appropriate q. We will provide a data-driven estimate of the optimal q in Section 3. The following result concerns the benchmark "infeasible" estimator $\tilde{\boldsymbol{\phi}}$.

Theorem 1 (*Theorem 8.1.1, Brockwell and Davis [1]*)**.** *Under Assumptions* (A2)–(A3), *as* $n \to \infty$

$$\sqrt{n}\left(\tilde{\boldsymbol{\phi}} - \boldsymbol{\phi}\right) \to_D N\left(0, \sigma^2 \boldsymbol{\Gamma}_p^{-1}\right),$$

where $\boldsymbol{\Gamma}_p$ *is the covariance matrix defined in* (2)*. Moreover,*

$$\tilde{\boldsymbol{\sigma}}^2 \to_P \boldsymbol{\sigma}^2,$$

where $\tilde{\sigma}^2 = \tilde{\gamma}(0) - \tilde{\boldsymbol{\phi}}^T \tilde{\boldsymbol{\gamma}}_p$.

With the smoothness of $m(\cdot)$ specified in Assumption (A1) and the order of q specified in Assumption (A4), the Yule–Walker estimator $\hat{\boldsymbol{\phi}}$ defined in (10) is oracally efficient as indicated by the following theorem. Its proof is given in the Appendix.

Theorem 2. *Under Assumptions* (A1)–(A4), *as* $n \to \infty$

$$\hat{\boldsymbol{\phi}} - \tilde{\boldsymbol{\phi}} = o_p\left(n^{-1/2}\right),$$

and hence

$$\sqrt{n}\left(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}\right) \to_D N\left(0, \sigma^2 \boldsymbol{\Gamma}_p^{-1}\right).$$

Let $\hat{\sigma}^2 = \hat{\gamma}(0) - \hat{\boldsymbol{\phi}}^T \hat{\boldsymbol{\gamma}}_p$ be the Yule–Walker estimators of σ^2 based on $\left\{\hat{Y}_t\right\}_{t=1}^{n}$. Theorem 2 implies that the two estimators $\hat{\sigma}^2$ and $\tilde{\sigma}^2$ are asymptotically closer than $n^{-1/2}$. On the other hand, Theorem 1 ensures consistency of $\tilde{\sigma}^2$ based on $\{Y_t\}_{t=1}^{n}$. The consistency of $\hat{\sigma}^2$ can be proved similarly to the counterpart of Theorem 3 in [6]. We summarize the conclusion as follows.

Theorem 3. *Under Assumptions* (A1)–(A4), *as* $n \to \infty$, $\hat{\sigma}^2 - \tilde{\sigma}^2 = o_p\left(n^{-1/2}\right)$ *and hence* $\hat{\sigma}^2 - \sigma^2 = o_p(1)$.

Denote by $\chi^2_{p,1-\alpha}$ the $100(1-\alpha)$-th percentile of the chi-square distribution with p degrees of freedom. The next result follows directly from the above two theorems and Slutsky's Theorem.

Corollary 1. *Under Assumptions* (A1)–(A4), *for any* $\alpha \in (0,1)$, *a* $100(1-\alpha)\,\%$ *asymptotic confidence ellipsoid for* $\boldsymbol{\phi}$ *is*

$$n\hat{\sigma}^{-2}\left(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}\right)^T \hat{\boldsymbol{\Gamma}}_p \left(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}\right) \le \chi^2_{p,1-\alpha}.$$

In other words

$$\lim_{n\to\infty} P\left[n\hat{\sigma}^{-2}\left(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}\right)^T \hat{\boldsymbol{\Gamma}}_p \left(\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}\right) \le \chi^2_{p,1-\alpha}\right] = 1 - \alpha.$$

3. Implementation

In this section, we describe the procedure to implement the trend estimator. To this end, we first need to determine the moving average lag q for the purpose of trend estimation. According to Fan and Gijbels [2], with the design variable taking on values $i/n,\ 1 \le i \le n$ and $K(u) = 0.5I_{\{|u|\le 1\}}$, the optimal bandwidth h which asymptotically minimizes the mean integrated squared error $E\int_0^1 \left\{\hat{m}(x) - m(x)\right\}^2 dx$ is

$$h_{\text{opt}} = n^{-1/5}(9/2)^{1/5}\{\gamma(0)\}^{1/5}\left[\int_0^1 \left\{m''(x)\right\}^2 dx\right]^{-1/5}.$$

The corresponding optimal q is

$$q_{\text{opt}} = \left\lceil n^{4/5}(9/2)^{1/5}\{\gamma(0)\}^{1/5}\left[\int_0^1 \left\{m''(x)\right\}^2 dx\right]^{-1/5}\right\rceil,$$

where $\lceil a\rceil$ denotes the integer part of a real number a. This q_{opt} apparently satisfies Assumption (A4). However, it is not applicable, as it involves the second derivative of the unknown function $m(\cdot)$ and the unknown variance $\gamma(0)$.

We will provide a data driven practical method to determine the moving average lag q. Specifically, following Yang and Tschernig [10], we propose the rule-of-thumb (ROT) estimator defined as

$$\hat{q}_{\text{ROT}} = \hat{q}_{n,\text{ROT}} = \left\lceil n^{4/5}(9/2)^{1/5}\{\widehat{\gamma}(0)\}^{1/5}\left[\int_0^1 \left\{\widehat{m}''(x)\right\}^2 dx\right]^{-1/5}\right\rceil$$

where $\widehat{\gamma}(0) = n^{-1}\sum_{t=1}^{n}\widehat{Y}_t^2$ with $\widehat{Y}_t = X_t - \widehat{m}(t/n)$ and

$$\widehat{m}(x) = \hat{a} + \hat{b}x + \hat{c}x^2 + \hat{d}x^3, \qquad \widehat{m}''(x) = 2\hat{c} + 6\hat{d}x,$$

in which $\left(\hat{a}, \hat{b}, \hat{c}, \hat{d}\right)$ is the solution to the least squares problem

$$\underset{(a,b,c,d)}{\text{argmin}}\sum_{i=1}^{n}\left(X_i - a - bi/n - ci^2/n^2 - di^3/n^3\right)^2.$$

Simple algebra yields that

$$\int \left\{\widehat{m}''(x)\right\}^2 dx = 4\hat{c}^2 + 12\hat{c}\hat{d} + 12\hat{d}^2,$$

and thus

$$\hat{q}_{\text{ROT}} = \left\lceil n^{4/5}(9/2)^{1/5}\left\{\widehat{\gamma}(0)\left(4\hat{c}^2 + 12\hat{c}\hat{d} + 12\hat{d}^2\right)^{-1}\right\}^{1/5}\right\rceil.$$

Once $\hat{q}_{\text{ROT}}$ is obtained, one can compute the residuals $\hat{Y}_t$ from (8) and $\hat{\boldsymbol{\phi}}$ from (10).

4. Simulation studies and application

4.1. Simulation studies

In this section, we will carry out simulation experiments to illustrate the finite-sample behavior of Yule–Walker estimators of autoregressive coefficients based on the detrending time series. All calculations are implemented by the free software package R which is available from http://www.r-project.org.

We simulate 100 replicates of the time series in (1) with AR(p) errors for $p = 1, 2, 3$, respectively, and sample sizes $n = 100, 200, 400, 1000$. The white noise $Z_t \sim N(0, 1)$ and the trend function is defined as

$$m(x) = \sin(2\pi x), \quad x \in [0, 1].$$

We simulate the same autoregressive time series as those in [6], the parameters of which represent a wide range of autoregressive time series. Specifically, the autoregressive parameters are respectively $\phi_1 = -0.8, -0.4, -0.2, 0.2, 0.4, 0.8$ for AR(1), $(\boldsymbol{\phi}_1, \boldsymbol{\phi}_2) = (-0.8, -0.4), (0.6, 0.1), (0.2, 0.1)$ for AR(2), and $(\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \boldsymbol{\phi}_3) = (0.2, 0.64, -0.144), (1, -0.56, 0.08)$ for AR(3).

For each value of $\boldsymbol{\phi}$ and a sample size, there are two estimates and the corresponding sample standard deviations (SD) in Tables 1–3. The top one is the mean estimates of $\boldsymbol{\phi}$ and their sample standard deviations of the proposed method, and the bottom one is the estimates from (4) with $q = 2$, which is used in [1]. From the simulation studies in these tables, the estimates of $\boldsymbol{\phi}$ using the proposed method are much more accurate than their counterparts from the fixed $q = 2$. When the sample sizes are larger, the estimates of our method are closer to the true coefficient values. This fact is showed in Fig. 1, the boxplots of the 100 sample ratios $\hat{\phi}_1/\tilde{\phi}_1$ for the AR(1) process. The horizontal dashed line is $y = 1$. In addition, the coverage

Table 1
Estimates of AR(1) coefficients and standard errors.

ϕ_1	q	$n = 100$		$n = 200$		$n = 400$		$n = 1000$	
		$\hat{\phi}_1$	SD	$\hat{\phi}_1$	SD	$\hat{\phi}_1$	SD	$\hat{\phi}_1$	SD
−0.8	$\hat{q}_{\mathrm{ROT}}$	−0.789	0.049	−0.793	0.042	−0.796	0.032	−0.798	0.018
	2	−0.783	0.047	−0.786	0.040	−0.790	0.030	−0.791	0.017
−0.4	$\hat{q}_{\mathrm{ROT}}$	−0.416	0.092	−0.404	0.071	−0.403	0.051	−0.400	0.030
	2	−0.506	0.069	−0.502	0.048	−0.507	0.036	−0.507	0.020
−0.2	$\hat{q}_{\mathrm{ROT}}$	−0.234	0.099	−0.214	0.077	−0.209	0.053	−0.201	0.032
	2	−0.397	0.070	−0.391	0.048	−0.396	0.035	−0.396	0.020
0.2	$\hat{q}_{\mathrm{ROT}}$	0.128	0.106	0.162	0.080	0.178	0.050	0.193	0.030
	2	−0.214	0.073	−0.206	0.050	−0.210	0.035	−0.210	0.020
0.4	$\hat{q}_{\mathrm{ROT}}$	0.308	0.109	0.350	0.080	0.373	0.047	0.391	0.028
	2	−0.138	0.074	−0.131	0.053	−0.133	0.036	−0.132	0.021
0.8	$\hat{q}_{\mathrm{ROT}}$	0.637	0.102	0.724	0.060	0.762	0.035	0.786	0.019
	2	−0.022	0.070	−0.019	0.053	−0.020	0.036	−0.019	0.021

Table 2
Estimates of AR(2) coefficients and standard errors.

(ϕ_1, ϕ_2)	q	$n = 100$		$n = 200$	
		$(\hat{\phi}_1, \hat{\phi}_2)$	SD	$(\hat{\phi}_1, \hat{\phi}_2)$	SD
(0.2, 0.1)	$\hat{q}_{\mathrm{ROT}}$	(0.129, 0.018)	(0.120, 0.106)	(0.166, 0.049)	(0.061, 0.050)
	2	(−0.361, −0.431)	(0.105, 0.080)	(−0.356, −0.439)	(0.075, 0.053)
(0.2, −0.1)	$\hat{q}_{\mathrm{ROT}}$	(0.146, −0.153)	(0.114, 0.102)	(0.174, −0.139)	(0.060, 0.048)
	2	(−0.266, −0.519)	(0.104, 0.076)	(−0.257, −0.528)	(0.073, 0.047)
(0.6, 0.1)	$\hat{q}_{\mathrm{ROT}}$	(0.507, 0.006)	(0.132, 0.094)	(0.561, 0.045)	(0.064, 0.050)
	2	(−0.164, −0.405)	(0.102, 0.084)	(−0.156, −0.414)	(0.075, 0.056)
(0.6, −0.1)	$\hat{q}_{\mathrm{ROT}}$	(0.528, −0.160)	(0.121, 0.101)	(0.568, −0.139)	(0.062, 0.049)
	2	(−0.046, −0.475)	(0.109, 0.077)	(−0.035, −0.485)	(0.078, 0.047)
(0.8, −0.4)	$\hat{q}_{\mathrm{ROT}}$	(0.788, −0.396)	(0.097, 0.089)	(0.787, −0.397)	(0.051, 0.045)
	2	(0.240, −0.563)	(0.103, 0.064)	(0.254, −0.573)	(0.075, 0.042)
(−0.8, −0.4)	$\hat{q}_{\mathrm{ROT}}$	(−0.753, −0.361)	(0.099, 0.091)	(−0.747, −0.357)	(0.054, 0.049)
	2	(−0.930, −0.652)	(0.087, 0.057)	(−0.927, −0.654)	(0.057, 0.038)
		$n = 400$		$n = 1000$	
		$(\hat{\phi}_1, \hat{\phi}_2)$	SD	$(\hat{\phi}_1, \hat{\phi}_2)$	SD
(0.2, 0.1)	$\hat{q}_{\mathrm{ROT}}$	(0.181, 0.074)	(0.027, 0.026)	(0.195, 0.088)	(0.010, 0.009)
	2	(−0.361, −0.436)	(0.051, 0.037)	(−0.361, −0.437)	(0.030, 0.022)
(0.2, −0.1)	$\hat{q}_{\mathrm{ROT}}$	(0.186, −0.119)	(0.026, 0.024)	(0.196, −0.109)	(0.010, 0.009)
	2	(−0.261, −0.525)	(0.049, 0.034)	(−0.261, −0.527)	(0.028, 0.021)
(0.6, 0.1)	$\hat{q}_{\mathrm{ROT}}$	(0.579, 0.072)	(0.028, 0.026)	(0.595, 0.087)	(0.010, 0.011)
	2	(−0.158, −0.409)	(0.052, 0.039)	(−0.158, −0.410)	(0.031, 0.023)
(0.6, −0.1)	$\hat{q}_{\mathrm{ROT}}$	(0.583, −0.119)	(0.027, 0.026)	(0.595, −0.108)	(0.010, 0.010)
	2	(−0.036, −0.480)	(0.051, 0.034)	(−0.035, −0.481)	(0.030, 0.022)
(0.8, −0.4)	$\hat{q}_{\mathrm{ROT}}$	(0.784, −0.406)	(0.024, 0.023)	(0.795, −0.402)	(0.009, 0.008)
	2	(0.258, −0.569)	(0.051, 0.031)	(0.260, −0.571)	(0.030, 0.019)
(−0.8, −0.4)	$\hat{q}_{\mathrm{ROT}}$	(−0.792, −0.398)	(0.024, 0.021)	(−0.794, −0.399)	(0.009, 0.009)
	2	(−0.934, −0.658)	(0.041, 0.028)	(−0.937, −0.660)	(0.025, 0.016)

frequencies from 100 replicates in Tables 4, 5 corroborate Corollary 1, that is the confidence ellipsoids based on $\hat{\phi}$ and $\tilde{\phi}$ are asymptotically the same at any level α for any AR(p).

4.2. Application

In this section, we analyze a real time series data set of the first differences of the annual global surface air temperatures in Celsius from 1880 through 1985. The data exhibit a remarkable nonlinear upward trend in Fig. 2. Hall and Keilegom [5] estimated the AR(1) coefficient by the observations directly before estimating the trend function. Their estimate was $\hat{\phi}_1 = 0.414$. Shao and Yang [6] estimated the AR(1) coefficient using a B-spline to detrend the data and their result was $\hat{\phi}_1 = 0.386$ with the standard error 0.090. We detrend the data by the modified moving average filter and analyze the residual sequence. It is straightforward to obtain the optimal bandwidth $h = q/n = 0.189$ and the coefficient estimate

Table 3
Estimates of AR(3) coefficients and standard errors.

(ϕ_1, ϕ_2, ϕ_3)	q	$n = 100$	
		$(\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3)$	SD
(0.2, 0.64, −0.144)	$\hat{q}_{ROT}$	(0.109, 0.502, −0.208)	(0.122, 0.077, 0.092)
	2	(−0.652, −0.320, −0.336)	(0.094, 0.106, 0.077)
(1, −0.56, 0.08)	$\hat{q}_{ROT}$	(0.897, −0.500, 0.007)	(0.121, 0.119, 0.099)
	2	(0.286, −0.545, −0.134)	(0.102, 0.068, 0.085)
		$n = 200$	
		$(\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3)$	SD
(0.2, 0.64, −0.144)	$\hat{q}_{ROT}$	(0.165, 0.570, −0.180)	(0.061, 0.041, 0.050)
	2	(−0.651, −0.322, −0.341)	(0.081, 0.081, 0.051)
(1, −0.56, 0.08)	$\hat{q}_{ROT}$	(0.955, −0.540, 0.051)	(0.062, 0.064, 0.048)
	2	(0.307, −0.560, −0.124)	(0.073, 0.046, 0.062)
		$n = 400$	
		$(\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3)$	SD
(0.2, 0.64, −0.144)	$\hat{q}_{ROT}$	(0.184, 0.605, −0.164)	(0.027, 0.021, 0.023)
	2	(−0.652, −0.316, −0.337)	(0.054, 0.048, 0.037)
(1, −0.56, 0.08)	$\hat{q}_{ROT}$	(0.975, −0.546, 0.062)	(0.026, 0.033, 0.024)
	2	(0.313, −0.558, −0.123)	(0.051, 0.034, 0.048)
		$n = 1000$	
		$(\hat{\phi}_1, \hat{\phi}_2, \hat{\phi}_3)$	SD
(0.2, 0.64, −0.144)	$\hat{q}_{ROT}$	(0.196, 0.627, −0.153)	(0.010, 0.008, 0.009)
	2	(−0.654, −0.315, −0.336)	(0.033, 0.031, 0.023)
(1, −0.56, 0.08)	$\hat{q}_{ROT}$	(0.992, −0.556, 0.076)	(0.010, 0.012, 0.009)
	2	(0.318, −0.560, −0.121)	(0.033, 0.021, 0.028)

Table 4
Coverage frequencies from 100 replicates in a confidence ellipsoid for ϕ in AR(1).

ϕ_1	ϕ	$1-\alpha$	$n = 100$	$n = 200$	$n = 400$	$n = 1000$
−0.8	$\hat{\phi}$	0.95	0.97	0.96	0.94	0.95
	$\tilde{\phi}$		0.97	0.98	0.94	0.94
	$\hat{\phi}$	0.99	0.98	0.98	0.99	1.00
	$\tilde{\phi}$		0.99	0.98	0.99	1.00
−0.4	$\hat{\phi}$	0.95	0.91	0.92	0.91	0.93
	$\tilde{\phi}$		0.95	0.92	0.91	0.94
	$\hat{\phi}$	0.99	0.99	0.97	0.98	0.99
	$\tilde{\phi}$		0.97	0.97	1.00	0.99
−0.2	$\hat{\phi}$	0.95	0.91	0.93	0.89	0.93
	$\tilde{\phi}$		0.93	0.90	0.91	0.95
	$\hat{\phi}$	0.99	0.98	0.96	0.98	0.97
	$\tilde{\phi}$		0.98	0.98	1.00	0.98
0.2	$\hat{\phi}$	0.95	0.87	0.88	0.90	0.96
	$\tilde{\phi}$		0.92	0.91	0.96	0.97
	$\hat{\phi}$	0.99	0.96	0.92	0.98	0.99
	$\tilde{\phi}$		0.98	0.97	1.00	0.98
0.4	$\hat{\phi}$	0.95	0.79	0.85	0.87	0.96
	$\tilde{\phi}$		0.93	0.89	0.95	0.96
	$\hat{\phi}$	0.99	0.93	0.90	0.97	0.99
	$\tilde{\phi}$		0.98	0.96	1.00	1.00
0.8	$\hat{\phi}$	0.95	0.49	0.70	0.80	0.90
	$\tilde{\phi}$		0.87	0.94	0.95	0.97
	$\hat{\phi}$	0.99	0.68	0.84	0.88	0.98
	$\tilde{\phi}$		0.95	0.98	0.99	0.99

$\hat{\phi}_1 = 0.373$ with the standard error 0.119. The residual autocorrelations plotted in Fig. 3 are within the 95% confidence interval. Therefore, we conclude that the model is adequate. For the purpose of comparison, we also fit an AR(2) model to

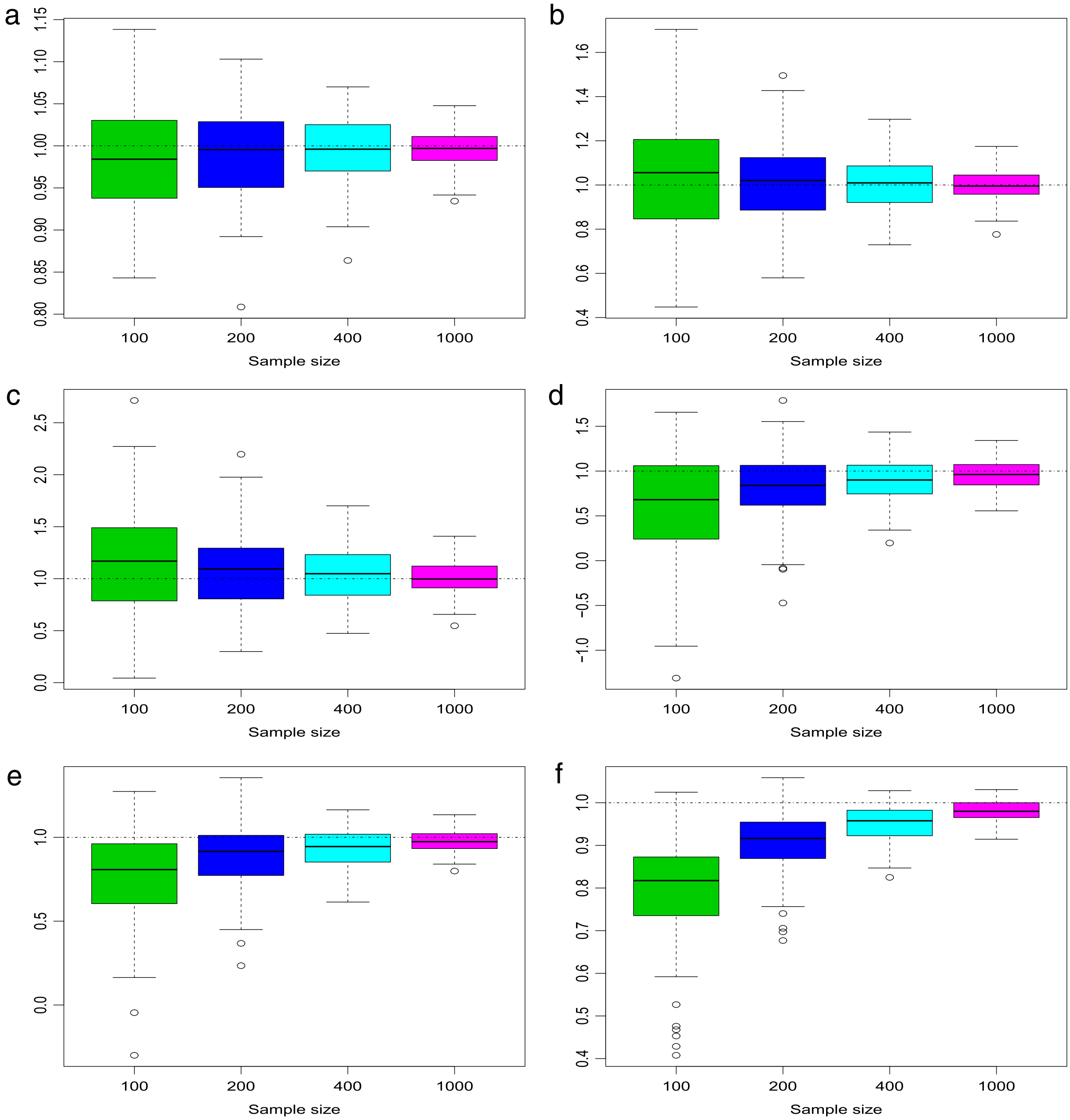

Fig. 1. (a) $\phi_1 = -0.8$, (b) $\phi_1 = -0.4$, (c) $\phi_1 = -0.2$, (d) $\phi_1 = 0.2$, (e) $\phi_1 = 0.4$, (f) $\phi_1 = 0.8$.

the data and the estimates are $(\hat{\phi}_1, \hat{\phi}_2) = (0.410, -0.099)$, which are fairly close to $(0.442, -0.068)$ obtained by Hall and Keilegom [5].

Acknowledgments

This work has been supported in part by NSF Award DMS 1007594, the Jiangsu Specially-Appointed Professor Program and the Jiangsu Key Discipline Program (Statistics), Jiangsu, China. The authors thank the anonymous referee for the critical comments and constructive suggestions which have led to substantial improvement of the paper.

Appendix

In this section, we derive (3) and prove Theorem 2 in Section 2. Hereafter, C is a constant and l is a fixed integer unless otherwise indicated. It is reasonable to assume that $q \geq l + 1$ from Assumption (A4).

Table 5
Coverage frequencies from 100 replicates in a confidence ellipsoid for ϕ in AR(2).

(ϕ_1, ϕ_2)	ϕ	$1-\alpha$	$n = 100$	$n = 200$	$n = 400$	$n = 1000$
(0.2, 0.1)	$\hat{\phi}$	0.95	0.78	0.80	0.88	0.94
	$\tilde{\phi}$		0.88	0.89	0.91	0.95
	$\hat{\phi}$	0.99	0.90	0.92	0.96	1.00
	$\tilde{\phi}$		0.96	0.97	1.00	0.99
(0.2, −0.1)	$\hat{\phi}$	0.95	0.83	0.82	0.90	0.94
	$\tilde{\phi}$		0.91	0.91	0.93	0.96
	$\hat{\phi}$	0.99	0.94	0.93	0.99	1.00
	$\tilde{\phi}$		0.97	0.96	1.00	1.00
(0.6, 0.1)	$\hat{\phi}$	0.95	0.61	0.75	0.80	0.91
	$\tilde{\phi}$		0.87	0.91	0.92	0.94
	$\hat{\phi}$	0.99	0.82	0.86	0.93	0.99
	$\tilde{\phi}$		0.97	0.97	0.98	1.00
(0.6, −0.1)	$\hat{\phi}$	0.95	0.78	0.82	0.88	0.96
	$\tilde{\phi}$		0.89	0.89	0.94	0.96
	$\hat{\phi}$	0.99	0.88	0.90	0.96	1.00
	$\tilde{\phi}$		0.95	0.95	1.00	1.00
(0.8, 0.4)	$\hat{\phi}$	0.95	0.84	0.86	0.89	0.97
	$\tilde{\phi}$		0.91	0.92	0.94	0.98
	$\hat{\phi}$	0.99	0.91	0.92	0.98	0.99
	$\tilde{\phi}$		0.96	0.96	0.99	1.00
(0.8, −0.4)	$\hat{\phi}$	0.95	0.92	0.92	0.96	0.93
	$\tilde{\phi}$		0.92	0.92	0.96	0.92
	$\hat{\phi}$	0.99	0.97	0.99	0.98	0.99
	$\tilde{\phi}$		0.96	0.99	0.98	0.99

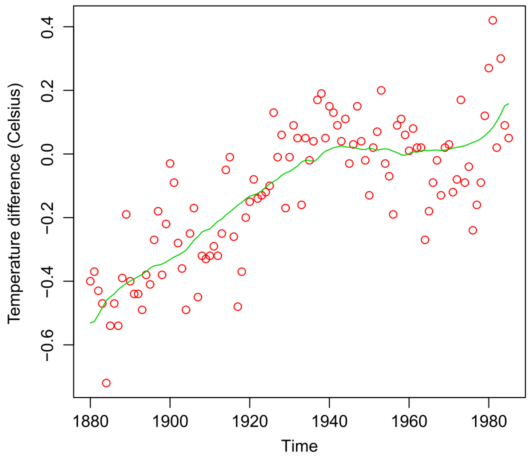

Fig. 2. First differences of annual global surface air temperatures in Celsius from 1880 through 1985.

Consider the objective function

$$\sum_{i=1}^{n} \{X_i - a - b\,(i-t)\,/n\}^2\, K_h\,\{(i-t)\,/n\}\,. \tag{A.1}$$

Differentiating (A.1) with respect to a and b, and setting these partial derivatives as 0, we obtain the equations as follows:

$$\begin{pmatrix} \sum_{i=k_1}^{k_2} 1 & \sum_{i=k_1}^{k_2} (i-t) \\ \sum_{i=k_1}^{k_2} (i-t) & \sum_{i=k_1}^{k_2} (i-t)^2 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \sum_{i=k_1}^{k_2} X_i \\ \sum_{i=k_1}^{k_2} (i-t) X_i \end{pmatrix}, \tag{A.2}$$

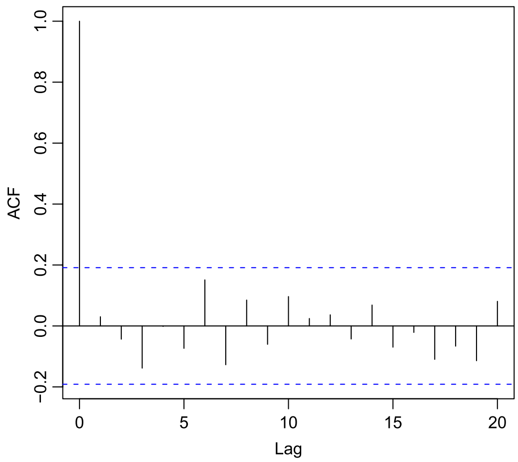

Fig. 3. Residual autocorrelations.

where $k_1 = t - q$ and $k_2 = t + q$ for $q + 1 \le t \le n - q$; $k_1 = 1$ and $k_2 = t + q$ for $1 \le t \le q$; $k_1 = t - q$ and $k_2 = n$ for $n - q + 1 \le t \le n$. It is straightforward to get (7) by solving Eqs. (A.2).

To prove Theorem 2, note from (3) and (9),

$$\hat{\gamma}(l) - \tilde{\gamma}(l) = n^{-1} \sum_{t=1}^{n-l} \left(\hat{Y}_t \hat{Y}_{t+l} - Y_t Y_{t+l} \right)$$
$$= n^{-1} \sum_{t=1}^{q} \left\{ \left(\hat{Y}_t - Y_t \right) Y_{t+l} + \left(\hat{Y}_{t+l} - Y_{t+l} \right) Y_t \right\} - n^{-1} \sum_{t=q+1}^{n-q} \left\{ \left(\hat{Y}_t - Y_t \right) Y_{t+l} + \left(\hat{Y}_{t+l} - Y_{t+l} \right) Y_t \right\}$$
$$- n^{-1} \sum_{t=n-q+1}^{n-l} \left\{ \left(\hat{Y}_t - Y_t \right) Y_{t+l} + \left(\hat{Y}_{t+l} - Y_{t+l} \right) Y_t \right\} + n^{-1} \sum_{t=1}^{n-l} \left(\hat{Y}_t - Y_t \right) \left(\hat{Y}_{t+l} - Y_{t+l} \right). \quad (A.3)$$

We will discuss the property of each term of (A.3) in several lemmas. The following proposition concerns the uniform convergence rate of the difference between Y_t and $\hat{Y}_t$. Its proof is similar to Theorem 6.5 of [3] and therefore omitted.

Proposition A.1. *Under Assumptions* (A1)–(A4), *as* $n \to \infty$,

$$\max_{1 \le t \le n} \left| Y_t - \hat{Y}_t \right| = \sup_{0 \le x \le 1} \left| \hat{m}(x) - m(x) \right| = O_p \left\{ (q/n)^2 + q^{-1/2} (\log n)^{1/2} \right\}.$$

Lemma A.1. *Under Assumptions* (A1)–(A4), *as* $n \to \infty$,

$$n^{-1} \sum_{t=1}^{n-l} \left(\hat{Y}_t - Y_t \right) \left(\hat{Y}_{t+l} - Y_{t+l} \right) = O_p \left\{ (q/n)^4 + q^{-1} \log n \right\}.$$

Proof. Proposition A.1 implies

$$n^{-1} \left| \sum_{t=1}^{n-l} \left(\hat{Y}_t - Y_t \right) \left(\hat{Y}_{t+l} - Y_{t+l} \right) \right| \le n^{-1} (n - l) \left(\max_{1 \le t \le n} \left| Y_t - \hat{Y}_t \right| \right)^2$$
$$= O_p \left\{ (q/n)^4 + q^{-1} \log n \right\}.$$

The proof is complete. □

Next, we will consider the sums of the first and last q observations in (A.3).

Lemma A.2. *Under Assumptions* (A1)–(A4), *as* $n \to \infty$,

$$n^{-1} \left\{ \sum_{t=n-q+1}^{n-l} \left(\hat{Y}_t - Y_t \right) Y_{t+l} + \sum_{t=1}^{q} \left(\hat{Y}_t - Y_t \right) Y_{t+l} \right\} = O_p \left\{ (q/n)^3 + q^{1/2} n^{-1} (\log n)^{1/2} \right\} = o_p \left(n^{-1/2} \right),$$
$$n^{-1} \left\{ \sum_{t=n-q+1}^{n-l} \left(\hat{Y}_{t+l} - Y_{t+l} \right) Y_t + \sum_{t=1}^{q} \left(\hat{Y}_{t+l} - Y_{t+l} \right) Y_t \right\} = O_p \left\{ (q/n)^3 + q^{1/2} n^{-1} (\log n)^{1/2} \right\} = o_p \left(n^{-1/2} \right).$$

Proof. Without loss of generality, we set $l = 0$. By the strong law of large numbers (SLLN) for an α-mixing sequence (Proposition 2.8 of [3]), $q^{-1}\sum_{t=1}^{q}|Y_t| < \infty$. Hence

$$\left|n^{-1}\left\{\sum_{t=n-q+1}^{n}\left(\hat{Y}_t - Y_t\right)Y_t + \sum_{t=1}^{q}\left(\hat{Y}_t - Y_t\right)Y_t\right\}\right| \le qn^{-1}\max_{1\le t\le n}\left|\hat{Y}_t - Y_t\right|\left(q^{-1}\sum_{t=1}^{q}|Y_t| + q^{-1}\sum_{t=n-q+1}^{n}|Y_t|\right)$$
$$= qn^{-1}O_p\left\{(q/n)^2 + q^{-1/2}(\log n)^{1/2}\right\}.$$

The proof is complete. □

In the following, we will consider the terms that involve the $n - 2q$ terms in the middle in (A.3). Note that we have the following decomposition:

$$n^{-1}\sum_{t=q+1}^{n-q}\left(\hat{Y}_t - Y_t\right)Y_{t+l} = n^{-1}\sum_{t=q+1}^{n-q}\left\{m(t/n) - \hat{m}(t/n)\right\}Y_{t+l} = T_{1n} + T_{2n}$$

where

$$\begin{aligned} T_{1n} &= -n^{-1}\sum_{t=q+1}^{n-q}\frac{\sum_{i=1}^{n}Y_iK_h\{(i-t)/n\}}{\sum_{i=1}^{n}K_h\{(i-t)/n\}}Y_{t+l} \\ &= -n^{-1}(2q+1)^{-1}\sum_{t=q+1}^{n-q}\sum_{i=1}^{n}Y_iY_{t+l}I_{\{|i-t|\le q\}}, \\ T_{2n} &= n^{-1}\sum_{t=q+1}^{n-q}\frac{\sum_{i=1}^{n}\{m(t/n) - m(i/n)\}K_h\{(i-t)/n\}}{\sum_{i=1}^{n}K_h\{(i-t)/n\}}Y_{t+l} \\ &= n^{-1}(2q+1)^{-1}\sum_{t=q+1}^{n-q}\sum_{i=1}^{n}\{m(t/n) - m(i/n)\}I_{\{|i-t|\le q\}}Y_{t+l}. \end{aligned}$$

We introduce two propositions as follows regarding the convergence rates of T_{1n} and T_{2n}.

Proposition A.2. *Under Assumptions* (A1)–(A2), *as* $n \to \infty$, $T_{2n} = O_p\left(n^{-3/2}\right)$.

Proof. By the definition of T_{2n},

$$\begin{aligned} ET_{2n}^2 &= E\left[n^{-1}(2q+1)^{-1}\sum_{t=q+1}^{n-q}\sum_{i=1}^{n}\{m(t/n) - m(i/n)\}I_{\{|i-t|\le q\}}Y_{t+l}\right]^2 \\ &\le \frac{|\gamma(0)| + 2\sum_{j=1}^{n-2q+1}|\gamma(j)|}{n(2q+1)^2}\max_{q+1\le t\le n-q}\left[\sum_{i=1}^{n}\{m(t/n) - m(i/n)\}I_{\{|i-t|\le q\}}\right]^2 \\ &\le \frac{\sum_{j=-\infty}^{\infty}|\gamma(j)|}{n(2q+1)^2}\left(\left\|m'\right\|_\infty q/n\right)^2 = \frac{q^2\sum_{j=-\infty}^{\infty}|\gamma(j)|}{n^3(2q+1)^2}\left\|m'\right\|_\infty^2 \\ &\le \frac{\sum_{j=-\infty}^{\infty}|\gamma(j)|}{4n^3}\left\|m'\right\|_\infty^2, \end{aligned}$$

where $\left\|m'\right\|_\infty$ is the supremum norm of m', that is $\left\|m'\right\|_\infty = \sup_x\left|m'(x)\right|$. The proof is complete by the Markov inequality. □

For a strictly stationary time series $\{Y_t\}_{t=-\infty}^{\infty}$, define the α-mixing coefficient

$$\alpha(n) = \sup_{A\in\mathcal{F}_{-\infty}^{0}, B\in\mathcal{F}_{n}^{\infty}}|P(A)P(B) - P(AB)|,$$

where $\mathcal{F}_i^j$ denotes the σ-algebra generated by $\{Y_t\}_{t=i+1}^{j-1}$. In particular, if an autoregressive moving average time series $\{Y_t\}_{t=-\infty}^{\infty}$ is causal, then $\alpha(n)$ decreases at an exponential rate, and thus

$$\sum_{n=1}^{\infty}\alpha(n) < \infty.$$

Applying Proposition 2.5 of [3], for $j \geq i$, it holds that

$$\left|\text{cov}\left(Y_i, Y_j\right)\right| \leq 8\left\{\alpha(j-i)\right\}^{1/r}\left\{E\left|Y_i\right|^p\right\}^{1/p}\left\{E\left|Y_j\right|^q\right\}^{1/q},$$

where p, q, r are constants satisfying $p, q, r > 1$, $1/p + 1/q + 1/r = 1$.

Proposition A.3. *Under Assumptions* (A1)–(A4), *as* $n \to \infty$,

$$ET_{1n}^2 = O\left(n^{-1}q^{-2} + q^{-2} + n^{-1}q^{-1}\right), \tag{A.4}$$

$$T_{1n} = O_p\left(n^{-1/2}q^{-1} + q^{-1} + n^{-1/2}q^{-1/2}\right) = O_p\left(q^{-1}\right) = o_p\left(n^{-1/2}\right). \tag{A.5}$$

Proof. Without loss of generality, we set $l = 0$. Define

$$R_{i,i',t,t'} = Y_iY_{i'}Y_tY_{t'}I_{\{|i-t|\leq q,\,|i'-t'|\leq q\}}.$$

Note that

$$ET_{1n}^2 = n^{-2}(2q+1)^{-2}\sum_{t,t'=q+1}^{n-q}\sum_{i,i'=1}^{n}ER_{i,i',t,t'}. \tag{A.6}$$

Define $\#\left(i, i', t, t'\right)$ = the number of distinct indices in i, i', t, t'. Obviously $\#\left(i, i', t, t'\right) \in \{1, 2, 3, 4\}$. We will consider how to bound the terms in (A.6) for four different cases.

Case 1: $\#\left(i, i', t, t'\right) = 1$. The four indices are equal, that is $i = i' = t = t'$. We have

$$\left|n^{-2}(2q+1)^{-2}\sum_{t,t'=q+1}^{n-q}\sum_{i,i'=1}^{n}ER_{i,i',t,t'}I_{\{\#(i,i',t,t')=1\}}\right| \leq n^{-2}(2q+1)^{-2}\sum_{t=1}^{n}\left|E\left(Y_t^4\right)\right| \leq Cn^{-1}q^{-2}.$$

Case 2: $\#\left(i, i', t, t'\right) = 2$. Given the range of i, i', t, t', there are at most $2n^2 + n$ such quadruples. First we consider that there are only two different indices. Without loss of generality, we assume $i = i' = 1$ and $t = t' = 2$. Apply the Cauchy–Schwarz inequality,

$$\left|ER_{i,i',t,t'}I_{\{\#(i,i',t,t')=2\}}\right| \leq E\left|Y_1^2Y_2^2\right| \leq \left(EY_1^4EY_2^4\right)^{1/2} < \infty.$$

Next we consider that there is only one different index and the other three are identical. Without loss of generality, we assume $i = i' = t = 1$, $t' = 2$, using the same method we obtain

$$\begin{aligned}\left|ER_{i,i',t,t'}I_{\{\#(i,i',t,t')=2\}}\right| &\leq E\left|Y_1^3Y_2\right| \leq \left\{EY_1^4E\left(Y_1Y_2\right)^2\right\}^{1/2}\\ &\leq \left\{EY_1^4\left(EY_1^4EY_2^4\right)^{1/2}\right\}^{1/2} < \infty.\end{aligned}$$

Thus

$$\left|n^{-2}(2q+1)^{-2}\sum_{t,t'=q+1}^{n-q}\sum_{i,i'=1}^{n}ER_{i,i',t,t'}I_{\{\#(i,i',t,t')=2\}}\right| \leq n^{-2}(2q+1)^{-2}\left(2n^2+n\right)C \leq Cq^{-2}.$$

Case 3: $\#\left(i, i', t, t'\right) = 3$. We need to bound three different forms of mixed moments: $E\left|Y_i^2Y_jY_k\right|$, $E\left|Y_iY_j^2Y_k\right|$, $E\left|Y_iY_jY_k^2\right|$, $1 \leq i < j < k \leq n$, each having at most n^2q triples. These three kinds of terms can be bounded in a similar manner. Therefore, we only discuss the details about $E\left|Y_iY_j^2Y_k\right|$. We apply Proposition 2.5 of [3] again

$$\begin{aligned}\left|EY_iY_j^2Y_k\right| &= \left|EY_iY_j^2Y_k - EY_iEY_j^2Y_k\right|\\ &\leq 8\left\{\alpha(j-i)\right\}^{1/6}\left\{E\left|Y_i\right|^6\right\}^{1/6}\left\{EY_j^6EY_k^3\right\}^{1/3} \leq C\left\{\alpha(j-i)\right\}^{1/6}.\end{aligned}$$

Hence

$$\left| n^{-2}(2q+1)^{-2}\sum_{t,t'=q+1}^{n-q}\sum_{i,i'=1}^{n} ER_{i,i',t,t'}I_{\{\#(i,i',t,t')=3\}}\right| \le Cn^{-2}(2q+1)^{-2}q\sum_{1\le i<j\le n}^{n}\{\alpha(j-i)\}^{1/6}$$

$$\le Cn^{-2}(2q+1)^{-2}nq\sum_{l=1}^{n}\{\alpha(l)\}^{1/6}\le Cn^{-1}q^{-1}.$$

Case 4: $\#(i,i',t,t') = 4$. The four indices are different from each other. Without loss of generality, we assume $1\le i<i'<t<t'\le n$. Note that we have

$$|EY_iY_{i'}Y_tY_{t'} - EY_iEY_{i'}Y_tY_{t'}| \le 8\left\{\alpha\left(i'-i\right)\right\}^{1/4}\left\{E|Y_i|^4\right\}^{1/4}\left\{E|Y_{i'}Y_tY_{t'}|^2\right\}^{1/2},$$

and hence

$$|EY_iY_{i'}Y_tY_{t'}| = |EY_iY_{i'}Y_tY_{t'} - EY_iEY_{i'}Y_tY_{t'}| \le C\left\{\alpha\left(i'-i\right)\right\}^{1/4}.$$

Note that stationarity of $\{Y_t\}_{t=-\infty}^{\infty}$ implies

$$|EY_iY_{i'}Y_tY_{t'}| \le C\min\left\{\left\{\alpha\left(t'-t\right)\right\}^{1/4},\left\{\alpha\left(i'-i\right)\right\}^{1/4}\right\}$$

$$\le C\left\{\alpha\left(\max\left(i'-i,t'-t\right)\right)\right\}^{1/4}.$$

Since $|i-t|\le q$ and $\left|i'-t'\right|\le q$, we conclude that $\max\left(i'-i,t'-t\right)\le q$. Thus

$$\left| n^{-2}(2q+1)^{-2}\sum_{t,t'=q+1}^{n-q}\sum_{i,i'=1}^{n} ER_{i,i',t,t'}I_{\{\#(i,i',t,t')=4\}}\right| \le n^{-2}(2q+1)^{-2}\sum_{\substack{1\le i<i'<t<t'\le n\\ \max(i'-i,t'-t)\le q}}|EY_iY_{i'}Y_tY_{t'}|$$

$$\le n^{-2}(2q+1)^{-2}2C\sum_{l=0}^{q}\{\alpha(l)\}^{1/4}n^2 l$$

$$\le Cq^{-2}\sum_{l=0}^{q}\{\alpha(l)\}^{1/4}l\le Cq^{-2}. \tag{A.7}$$

The inequality (A.7) is obtained based on that $\alpha(n)$ decreases at an exponential rate.

From Cases 1–4, (A.4) holds, and thus (A.5) is true. The proof is complete. □

Lemma A.3. *Under Assumptions* (A1)–(A4), *as* $n\to\infty$,

$$n^{-1}\sum_{t=q+1}^{n-q}(\hat{Y}_t - Y_t)Y_{t+l} = O_p\left(q^{-1}\right) = o_p\left(n^{-1/2}\right),$$

$$n^{-1}\sum_{t=q+1}^{n-q}(\hat{Y}_{t+l} - Y_{t+l})Y_t = O_p\left(q^{-1}\right) = o_p\left(n^{-1/2}\right).$$

Proof. According to Propositions A.2 and A.3, we have

$$n^{-1}\sum_{t=q+1}^{n-q}(\hat{Y}_t - Y_t)Y_{t+l} = T_{1n}+T_{2n}$$

$$= o_p\left(n^{-1/2}\right) + O_p\left(n^{-3/2}\right) = o_p\left(n^{-1/2}\right),$$

and

$$n^{-1}\sum_{t=q+1}^{n-q}(\hat{Y}_{t+l} - Y_{t+l})Y_t = o_p\left(n^{-1/2}\right).$$

The proof is complete. □

Lemma A.4. *Under Assumptions* (A1)–(A4), *as* $n\to\infty$

$$\left|\hat{\boldsymbol{\Gamma}}_p - \tilde{\boldsymbol{\Gamma}}_p\right| + \left|\hat{\boldsymbol{\gamma}}_p - \tilde{\boldsymbol{\gamma}}_p\right| = O_p(c_n) = o_p\left(n^{-1/2}\right),$$

where $c_n = (q/n)^3 + q^{1/2}n^{-1}(\log n)^{1/2} + q^{-1}\log n$ *and* $|\mathbf{A}|$ *denotes the maximal absolute value of the entries of a matrix or a vector* $\mathbf{A}$.

Proof. Lemmas A.1–A.3 imply

$$\left|\hat{\boldsymbol{\gamma}}_p - \tilde{\boldsymbol{\gamma}}_p\right| = O_p\left(q^4 n^{-4} + q^{-1}\log n\right) + O_p\left(q^{-1}\right) = o_p\left(n^{-1/2}\right).$$

The proof is complete. □

Theorem 2 can be obtained immediately by Lemma A.4, Slutsky's Theorem and Theorem 1.

References

[1] P.J. Brockwell, R.A. Davis, Time Series: Theory and Methods, second ed., Springer, New York, 1991.
[2] J. Fan, I. Gijbels, Local Polynomial Modeling and its Applications, Chapman & Hall, London, 1996.
[3] J. Fan, Q. Yao, Nonlinear Time Series: Nonparametric and Parametric Methods, Springer, New York, 2003.
[4] W.A. Fuller, Introduction to Statistical Time Series, Wiley, New York, 1996.
[5] P. Hall, I.V. Keilegom, Using difference-based methods for inference in non-parametric regression with time series errors, Journal of the Royal Statistical Society. Series B. Statistical Methodology 65 (2003) 443–456.
[6] Q. Shao, L. Yang, Autogressive coefficient estimation in nonparametric analysis, Journal of Time Series Analysis 32 (2011) 587–597.
[7] Q. Song, L. Yang, Spline confidence bands for variance function, Journal of Nonparametric Statistics 21 (2009) 589–609.
[8] Y.K. Truong, Nonparametric curve estimation with time series errors, Journal of Statistical Planning and Inference 28 (1991) 167–183.
[9] R.S. Tsay, G.C. Tiao, Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary ARMA models, Journal of the American Statistical Association 79 (1984) 84–96.
[10] L. Yang, R. Tscherning, Multivariate bandwidth selection for local linear regression, Journal of the Royal Statistical Society. Series B. Statistical Methodology 61 (1999) 793–815.