

Autoregressive coefficient estimation in nonparametric analysis

Q. Shao^{a,*} and L. J. Yang^b

The article considers the Yule-Walker estimator of the autoregressive coefficient based on the observed time series that contains an unknown trend function and an autoregressive error term. The trend function is estimated by means of B-splines and then subtracted from the observations. The Yule-Walker estimator is obtained from the residual sequence. Asymptotic properties of this estimator are derived. The performance of the estimator is illustrated by simulation studies and real data analysis.

Keywords: Autoregressive time series, Yule-Walker estimator, B-splines.

JEL classification: C14.

1. INTRODUCTION

The dynamic system of many time series $\{x_t, t = 0, \pm 1, \pm 2, \dots\}$ in practice can be well described by the following model:

$$x_t = \sum_{k=1}^p \phi_k x_{t-k} + \epsilon_t, \quad (1)$$

where ϵ_t is white noise with mean 0 and variance σ^2 . The stochastic sequence that can be modelled by (1) is called an autoregressive time series with order p (AR(p)). Estimation of autoregressive coefficients $\phi = (\phi_1, \dots, \phi_p)'$ is one of the important components in analyzing such time series. There are many estimation methods based on the observations $\mathbf{x} = (x_1, \dots, x_n)'$. However, in practice, it happens often that \mathbf{x} is not observable. Instead, what can be observed are time series that contain trend functions.

In such a scenario, the classic approach in stationary time series analysis is that the trend is assumed to be a known function with some unknown parameters. After these parameters are estimated based on the observations and the model, the trend function is removed from the observations. The estimator of ϕ is then obtained according to the residual sequence. The advantage of this procedure is that the estimation of the parameters of the trend function does not affect the asymptotic properties of the estimator of ϕ . See, for example, chapter 9 of Fuller (1996) for details. The major drawback of this approach is that the assumption about the trend function is usually artificial.

In this article, we are interested in the asymptotic properties of the estimator of ϕ when the unknown trend function is not specified. In particular, the estimator of ϕ is obtained from observations $\mathbf{y} = (y_1, \dots, y_n)'$ of a time series that contains an unknown slowly varying trend function $g(\cdot)$. The time series \mathbf{y} satisfies the following model:

$$\mathbf{y} = \mathbf{g} + \mathbf{x}, \quad (2)$$

where $\mathbf{g} = (g(u_1), \dots, g(u_n))'$ with $u_i = \frac{i}{n}$ and the error term \mathbf{x} is an AR(p) time series defined by (1). Throughout this article, we will use bold lower-case letters to denote vectors, bold upper-case letters to denote matrices, and lower-case letters to denote both time series and their realizations.

When estimating ϕ from a time series with an unknown trend function, we will follow a commonly used procedure: estimate the trend function, subtract the estimated trend from the observed time series, and then estimate the model parameters ϕ . One of the critical steps in this procedure is to estimate the unknown trend function. It can be accomplished by non-parametric function estimation method, such as polynomial splines and local kernel smoothers. Fan and Gijbels (1996) provided a detailed discussion about local kernels, especially local polynomial smoothers. Truong (1991) and Altman (1993) discussed some asymptotic properties of the estimators of autoregressive parameters obtained by a residual sequence when the trend function is estimated by local kernel smoothers. However, there is little discussion about asymptotic properties of the estimators of autoregressive parameters when the trend function is estimated by polynomial splines.

Spline smoothing has been applied to time series analysis in recent years, for example, in Huang and Yang (2004). Compared to the local smoothing obtained by using a kernel, spline smoothing is global, i.e., only a single optimization is needed for the unknown

^aUniversity of Toledo

^bSoochow University and Michigan State University

*Correspondence to: Q. Shao Department of Mathematics, The University of Toledo, Toledo, OH 43606, USA.

[†]E-mail: qin.shao@utoledo.edu

[^]Corrections have been added after publication 24 Jan 2011.

function over an entire range, instead of optimization at every point in the range. There are two major advantages of spline over kernel smoothing: (i) computational expediency, see Xue and Yang (2006), Wang and Yang (2007), which show that spline can be thousands of times faster than kernel smoothing; and (ii) intuitive appeal, see Wang and Yang (2009) for an explicit formula of the spline confidence band.

We will focus on the asymptotic properties of the Yule-Walker estimator of ϕ obtained from the time series detrended by spline smoothing. We will show that after a time series is detrended by a uniformly consistent spline estimator of $g(\cdot)$, the Yule-Walker estimator of ϕ is consistent and normally distributed, which is similar to the conclusion obtained from a parametric model in Tsay and Tiao (1984). The article will be organized as follows: in Section 2, we will introduce spline trend estimation and Yule-Walker coefficient estimation; in Section 3, we will provide simulation studies of several models with autoregressive error terms and analyze annual global surface air temperatures; finally in Section 4, we will show that a Yule-Walker estimator obtained by polynomial splines is consistent and asymptotically normally distributed.

2. AUTOREGRESSIVE COEFFICIENT ESTIMATION

2.1. Polynomial Splines

Suppose that m is a positive integer. Consider a sequence of equally spaced points or knots $(-m + 1)h \leq \dots \leq 0 \leq h \leq 2h \dots \leq Nh \leq 1$. Notice that there are $N + 1$ subintervals that divide the interval $[0,1]$ into subintervals $J_j = [jh, (j + 1)h)$, $j = -m + 1, -m + 2, \dots, N - 1$ and $J_N = [Nh, 1]$, of width h . Let $G_N^{(m-2)} = G_N^{(m-2)}[0, 1]$ denote the space of functions that are polynomial of degree $m - 1$ on each J_j and have continuous $(m - 2)$ th derivatives. The B-spline basis of $G_N^{(m-2)}$ is $\{b_{j,m}(u), j = -m + 1, \dots, N\}$. For a realization of time series \mathbf{y} , define a vector $\mathbf{b}_j = (b_{j,m}(u_1), \dots, b_{j,m}(u_n))'$ and an $n \times (N + m)$ matrix

$$\mathbf{B} = (\mathbf{b}_{-m+1}, \dots, \mathbf{b}_N).$$

We will focus on the cases of $m = 1, 2$: $G_N^{(-1)}$ is the space of functions that are constant on each J_j and $G_N^{(0)}$ is the space of functions that are linear on each J_j and continuous on $[0,1]$.

The B-spline basis for $G_N^{(-1)}$ is $\{b_{j,1}(u), j = 0, \dots, N\}$, where $b_{j,1}(u)$ is the indicator function of J_j , i.e.

$$b_{j,1}(u) = \begin{cases} 1, & u \in J_j; \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

REMARK 1. For the basis of the space $G_N^{(-1)}$, each vector \mathbf{b}_j has at most $\lceil nh \rceil$ entries that have value one, the other $n - \lceil nh \rceil$ entries being zero, where $\lceil u \rceil$ is the smallest integer such that $u \leq \lceil u \rceil$.

REMARK 2. For the basis of the space $G_N^{(-1)}$, the column vectors of \mathbf{B} are orthogonal. In addition, there is only one nonzero entry in each row of \mathbf{B} . This nonzero entry is one.

The B-spline basis for the piecewise linear spline space $G_N^{(0)}$ is $\{b_{j,2}(u), j = -1, \dots, N\}$, where $b_{j,2}(u)$ is defined as follows:

$$b_{j,2}(u) = K\left(\frac{u - (j + 1)h}{h}\right),$$

where $K(x) = (1 - |x|)_+$ with $(x)_+ = \max(x, 0)$; or equivalently

$$b_{j,2}(u) = \begin{cases} \frac{u}{h} - j, & u \in J_j; \\ j + 2 - \frac{u}{h}, & u \in J_{j+1}; \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

REMARK 3. For the basis of the space $G_N^{(0)}$, the column vectors of \mathbf{B} are not orthogonal. There are at most $2\lceil nh \rceil$ nonzero entries for each \mathbf{b}_j and at most two nonzero entries in each row of \mathbf{B} .

Denote the number of observations in J_j by n_j and define $j_i = \sum_{k=0}^{j-1} n_k + i$. Thus x_{j_i} is the i th observation in J_j . We have $\lceil nh \rceil - 1 \leq n_j \leq \lceil nh \rceil + 1$ and $\sum_{j=0}^N n_j = n$. ^

2.2. Yule-Walker estimation

Suppose that the autocovariance function of the time series $\{x_t; t = 0, \pm 1, \pm 2, \dots\}$ is $\gamma_k = E(x_t x_{t+k})$. If $\{x_t\}$ were observable, the sample autocovariance at lag k would be calculated by

$$\tilde{\gamma}_k = \frac{\sum_{t=1}^{n-k} X_t X_{t+k}}{n}, \tag{5}$$

and the Yule-Walker estimator of ϕ would be

$$\tilde{\phi} = \tilde{\Gamma}^{-1} \tilde{\gamma}, \tag{6}$$

where $\tilde{\gamma} = (\tilde{\gamma}_1, \tilde{\gamma}_2, \dots, \tilde{\gamma}_p)'$ and $\tilde{\Gamma}$ is the $p \times p$ estimated covariance matrix with (i,j) th entry $\tilde{\gamma}_{i-j}$.

To apply the Yule-Walker estimation procedure to the time series $\{y_t\}$ in eqn (2), we will estimate the trend function $g(u)$ by polynomial splines and remove it from the observations of the time series. Specifically, at any $u \in [0,1]$, the polynomial spline estimator of $g(u)$ is

$$\hat{g}(u) = \sum_{j=1-m}^N b_{j,m}(u) \hat{\beta}_j, \tag{7}$$

where $\hat{\beta} = (\hat{\beta}_{-m+1}, \dots, \hat{\beta}_N)'$ is obtained by

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} (\mathbf{y} - B\beta)'(\mathbf{y} - B\beta). \tag{8}$$

According to linear model theory, the estimator $\hat{\mathbf{g}} = (\hat{g}(u_1), \dots, \hat{g}(u_n))'$ is the projection of \mathbf{y} onto the space spanned by $\{\mathbf{b}_{-m+1}, \dots, \mathbf{b}_N\}$; that is,

$$\hat{\mathbf{g}} = \mathbf{B}(\mathbf{B}'\mathbf{B})^{-1} \mathbf{B}'\mathbf{y}. \tag{9}$$

We eliminate the trend by eqn (9) and obtain the residual sequence $\hat{\mathbf{x}}$ as follows:

$$\hat{\mathbf{x}} = \mathbf{y} - \hat{\mathbf{g}} = \{\mathbf{I} - \mathbf{B}(\mathbf{B}'\mathbf{B})^{-1} \mathbf{B}'\} \mathbf{y}.$$

Based on $\hat{\mathbf{x}}$, the Yule-Walker estimator of ϕ is

$$\hat{\phi} = \hat{\Gamma}^{-1} \hat{\gamma}, \tag{10}$$

where $\hat{\gamma}$ and $\hat{\Gamma}$ are calculated by eqn (5) with \mathbf{x} replaced by $\hat{\mathbf{x}}$.

The main results of this article indicate that $\tilde{\mathbf{x}}$ and $\hat{\mathbf{x}}$ are asymptotically equivalent under the following assumptions:

- [i] The trend function $g(\cdot) \in C^{(m)}[0,1]$, $m = 1,2$; that is, the trend function has m continuous derivatives.
- [ii] The subinterval length $h \sim n^{-1/(2m+1)}$; that is, the number of interior knots $N \sim n^{1/(2m+1)}$.
- [iii] The time series $\{x_t\}$ is causal; that is, there exists a sequence of constants $\{\psi_j\}$ such that $\sum_{j=0}^{\infty} |\psi_j| < \infty$ and

$$x_t = \sum_{j=0}^{\infty} \psi_j \epsilon_{t-j}.$$

- [iv] $E(\epsilon^4) < \infty$.

REMARK 4. Under assumption iii, $\sum_{|k|=0}^{\infty} |\gamma_k| < \infty$. Under assumptions iii and iv, $\tilde{\phi}$ defined by eqn (6) is a consistent estimator of ϕ . In addition,

$$\sqrt{n}(\tilde{\phi} - \phi) \xrightarrow{D} N(\mathbf{0}, \sigma^2 \mathbf{\Gamma}^{-1}) \text{ as } n \rightarrow \infty,$$

where ' \xrightarrow{D} ' denotes the convergence in distribution. See, for example, Theorem 8.1.1 of Brockwell and Davis (1991) for details.

The main results of the article are as follows:

THEOREM 1. Under assumptions i-iv, the Yule-Walker estimator $\hat{\phi}$ defined in eqn (10) and $\tilde{\phi}$ defined in eqn (6) satisfy

$$\hat{\phi} - \tilde{\phi} = o_p(n^{-1/2}).$$

We will provide the proof in Section 4. From Remark 4, the following theorem can be obtained immediately from Slutsky's Theorem.

THEOREM 2. Under assumptions i-iv, $\sqrt{n}(\hat{\phi} - \phi) \xrightarrow{D} N(\mathbf{0}, \sigma^2 \mathbf{\Gamma}^{-1})$ as $n \rightarrow \infty$, where $\mathbf{\Gamma}$ is the $p \times p$ covariance matrix with (i,j) th entry γ_{i-j} .

Theorem 2 implies that the non-parametric spline estimation of $g(\cdot)$ asymptotically does not affect the consistency of the Yule-Walker estimator of ϕ . An estimator of σ^2 based on the Yule-Walker estimator is

$$\hat{\sigma}^2 = \hat{\gamma}_0 - \hat{\gamma}' \hat{\phi}.$$

An interesting yet unresolved issue is the data driven selection of the autoregressive order p . It is well known that the Bayesian Information Criterion (BIC) consistently selects the correct order p of an autoregressive time series (see Brockwell and Davis 1991). We believe that the technical tools in our proofs can be used to show that the BIC computed from the residual sequence $\hat{\mathbf{x}}$ is asymptotically equivalent to that from the data \mathbf{x} . To prove this assertion, however, would require substantial further investigation.

We note that Theorem 2 does not contain the number of knots N in the asymptotics. This complicates the data driven selection of N as the usual bias-variance tradeoff in choosing a smoothing parameter (see, for example, Yang and Tschernig 1999) is unavailable. A more complex issue is the locations of the knots. Although Theorem 2 is proved under the assumption of knots being equally spaced, it is feasible to use other knot placement schemes, such as in Huang and Yang (2004), Xue and Yang (2006). We are in favour of the equally-spaced placement used in this article as it achieves the same asymptotic result and avoids the extra complication of determining the knot locations. In other words, it is as simple as possible to implement yet statistically efficient, as indicated in the simulation results of the next section.

3. SIMULATION STUDY AND DATA ANALYSIS

3.1. Simulation study

We simulate 100 samples of the time series in eqn (2) with AR(1), AR(2) and AR(3) errors respectively. The white noise variance $\sigma^2 = 1$. The sample sizes for each sample path are respectively $n = 100, 200, 400, 1000$. The trend function is defined as

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

The linear B-spline is applied and the number of knots used here is $N = \lceil cn^{1/5} \rceil$ with $c = 1, 2, 5$.

The autoregressive parameters ϕ_1 of AR(1) are chosen as $-0.8, -0.4, -0.2, 0.2, 0.4, 0.8$ so that the time series range from the relatively weakly to the relatively highly correlated. When choosing the autoregressive parameters of AR(2), we follow a referee's suggestion that the time series generated are positively correlated and/or do not have real roots except the AR(2) with $\phi = (-0.8, -0.4)'$. Among the AR(2), the time series with $\phi = (0.6, 0.1)'$ not only has strong positive correlations, but has complex roots. The coefficients of AR(3) are chosen such that $\phi = (0.2, 0.64, -0.144)'$ has three real roots with one positive and not too close to 1 and the others not far away from 1 and -1 , and $\phi = (1, -0.56, 0.08)'$ has two complex roots in the neighbourhood of the boundary of the unit circle.

The mean estimates of ϕ along with their sample standard deviations are summarized in Tables 1, 2 and 3, respectively. According to these simulations with three different numbers of knots, the number of knots has an impact on the coefficient estimates, especially when the sample size is small and/or the time series is positively correlated. The estimates are closer to the true values and more accurate for negatively correlated sequences. When the sample size is 1000, the estimates of three different knots are almost identical. The number of knots with $c = 1$ provides reasonable estimates for all the simulated models.

For the 100 sample paths of the AR(1) errors, we calculate the ratios $\hat{\phi}_1/\phi_1$ and obtain the boxplots in Figure 1 for $c = 1$. In Figure 1, the horizontal dashed line is $y = 1$. These boxplots coincide with Theorem 1: for each ϕ_1 the boxes not only become narrower and narrower but closer and closer to 1, as the sample size increases from 100 to 1000.

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

ϕ_1	c	$n = 100$	$n = 200$	$n = 400$	$n = 1000$
-0.8	1	-0.784 ± 0.053	-0.784 ± 0.045	-0.794 ± 0.033	-0.795 ± 0.018
	2	-0.796 ± 0.048	-0.795 ± 0.042	-0.799 ± 0.032	-0.799 ± 0.018
	5	-0.809 ± 0.046	-0.802 ± 0.041	-0.804 ± 0.031	-0.801 ± 0.018
-0.4	1	-0.395 ± 0.099	-0.382 ± 0.075	-0.395 ± 0.052	-0.391 ± 0.030
	2	-0.422 ± 0.092	-0.404 ± 0.072	-0.407 ± 0.051	-0.400 ± 0.030
	5	-0.462 ± 0.088	-0.424 ± 0.069	-0.420 ± 0.049	-0.406 ± 0.030
-0.2	1	-0.208 ± 0.112	-0.189 ± 0.080	-0.198 ± 0.054	-0.192 ± 0.032
	2	-0.241 ± 0.101	-0.212 ± 0.078	-0.212 ± 0.053	-0.202 ± 0.032
	5	-0.296 ± 0.095	-0.238 ± 0.074	-0.230 ± 0.052	-0.210 ± 0.033
0.2	1	0.154 ± 0.116	0.188 ± 0.081	0.190 ± 0.051	0.201 ± 0.030
	2	0.120 ± 0.108	0.165 ± 0.082	0.175 ± 0.051	0.193 ± 0.031
	5	0.032 ± 0.104	0.127 ± 0.078	0.148 ± 0.051	0.180 ± 0.032
0.4	1	0.332 ± 0.110	0.373 ± 0.079	0.384 ± 0.047	0.398 ± 0.028
	2	0.296 ± 0.105	0.352 ± 0.081	0.368 ± 0.048	0.390 ± 0.028
	5	0.190 ± 0.105	0.307 ± 0.077	0.335 ± 0.049	0.376 ± 0.030
0.8	1	0.676 ± 0.084	0.745 ± 0.055	0.769 ± 0.031	0.789 ± 0.019
	2	0.621 ± 0.099	0.719 ± 0.056	0.751 ± 0.035	0.782 ± 0.019
	5	0.447 ± 0.110	0.641 ± 0.065	0.694 ± 0.042	0.762 ± 0.021

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

ϕ	c	$n = 100$		$n = 200$	
(0.2, 0.1)	1	(0.152 ± 0.124, 0.041 ± 0.102)		(0.188 ± 0.062, 0.069 ± 0.050)	
	2	(0.120 ± 0.121, 0.012 ± 0.103)		(0.168 ± 0.063, 0.051 ± 0.050)	
	5	(0.021 ± 0.123, -0.070 ± 0.109)		(0.128 ± 0.061, 0.013 ± 0.050)	
(0.2, -0.1)	1	(0.166 ± 0.116, -0.134 ± 0.100)		(0.195 ± 0.061, -0.119 ± 0.047)	
	2	(0.137 ± 0.113, -0.160 ± 0.100)		(0.176 ± 0.061, -0.138 ± 0.047)	
	5	(0.062 ± 0.118, -0.224 ± 0.097)		(0.145 ± 0.060, -0.167 ± 0.047)	
(0.6, 0.1)	1	(0.535 ± 0.122, 0.030 ± 0.097)		(0.578 ± 0.059, 0.060 ± 0.050)	
	2	(0.495 ± 0.132, 0.002 ± 0.094)		(0.559 ± 0.063, 0.044 ± 0.049)	
	5	(0.345 ± 0.135, -0.088 ± 0.101)		(0.503 ± 0.066, 0.001 ± 0.051)	
(0.6, -0.1)	1	(0.542 ± 0.120, -0.141 ± 0.099)		(0.581 ± 0.061, -0.125 ± 0.048)	
	2	(0.515 ± 0.118, -0.163 ± 0.101)		(0.567 ± 0.062, -0.138 ± 0.048)	
	5	(0.418 ± 0.124, -0.229 ± 0.096)		(0.533 ± 0.061, -0.168 ± 0.050)	
(0.8, -0.4)	1	(0.751 ± 0.102, -0.405 ± 0.086)		(0.782 ± 0.055, -0.405 ± 0.042)	
	2	(0.732 ± 0.108, -0.421 ± 0.081)		(0.771 ± 0.055, -0.414 ± 0.041)	
	5	(0.671 ± 0.112, -0.460 ± 0.079)		(0.749 ± 0.056, -0.434 ± 0.041)	
(-0.8, -0.4)	1	(-0.753 ± 0.099, -0.361 ± 0.091)		(-0.747 ± 0.054, -0.357 ± 0.049)	
	2	(-0.804 ± 0.101, -0.411 ± 0.090)		(-0.792 ± 0.052, -0.402 ± 0.046)	
	5	(-0.850 ± 0.095, -0.456 ± 0.082)		(-0.816 ± 0.050, -0.426 ± 0.044)	
ϕ	c	$n = 400$		$n = 1000$	
(0.2, 0.1)	1	(0.191 ± 0.027, 0.084 ± 0.026)		(0.202 ± 0.010, 0.095 ± 0.010)	
	2	(0.177 ± 0.028, 0.070 ± 0.026)		(0.195 ± 0.010, 0.088 ± 0.009)	
	5	(0.149 ± 0.028, 0.043 ± 0.027)		(0.183 ± 0.011, 0.076 ± 0.009)	
(0.2, -0.1)	1	(0.195 ± 0.027, -0.110 ± 0.024)		(0.203 ± 0.010, -0.102 ± 0.009)	
	2	(0.182 ± 0.027, -0.123 ± 0.024)		(0.196 ± 0.010, -0.110 ± 0.009)	
	5	(0.161 ± 0.026, -0.144 ± 0.025)		(0.186 ± 0.010, -0.119 ± 0.009)	
(0.6, 0.1)	1	(0.585 ± 0.027, 0.078 ± 0.026)		(0.598 ± 0.010, 0.090 ± 0.011)	
	2	(0.573 ± 0.028, 0.067 ± 0.026)		(0.593 ± 0.011, 0.086 ± 0.010)	
	5	(0.536 ± 0.030, 0.036 ± 0.027)		(0.581 ± 0.011, 0.074 ± 0.010)	
(0.6, -0.1)	1	(0.589 ± 0.026, -0.113 ± 0.025)		(0.599 ± 0.010, -0.105 ± 0.010)	
	2	(0.578 ± 0.027, -0.123 ± 0.026)		(0.595 ± 0.010, -0.109 ± 0.009)	
	5	(0.555 ± 0.027, -0.145 ± 0.027)		(0.585 ± 0.011, -0.119 ± 0.009)	
(0.8, -0.4)	1	(0.788 ± 0.024, -0.403 ± 0.023)		(0.797 ± 0.009, -0.399 ± 0.009)	
	2	(0.781 ± 0.024, -0.409 ± 0.023)		(0.794 ± 0.009, -0.403 ± 0.008)	
	5	(0.765 ± 0.024, -0.423 ± 0.023)		(0.788 ± 0.010, -0.409 ± 0.008)	
(-0.8, -0.4)	1	(-0.780 ± 0.026, -0.386 ± 0.023)		(-0.779 ± 0.009, -0.384 ± 0.009)	
	2	(-0.800 ± 0.025, -0.406 ± 0.022)		(-0.797 ± 0.010, -0.402 ± 0.008)	
	5	(-0.815 ± 0.024, -0.421 ± 0.021)		(-0.804 ± 0.010, -0.409 ± 0.008)	

3.2. Application

Figure 2 is the scatter plot of the first differences of annual global surface air temperatures in Celsius from 1880 through 1985. It has a pronounced nonlinear upward trend. Hall and Keilegom (2003) estimated the AR(1) coefficient by the observations directly before estimating the trend function. Their estimate was $\hat{\phi}_1 = 0.414$. We detrend the data using the linear B-spline and then analyze the residual sequence. The number of knots used is $N = \lceil n^{1/5} \rceil$ according to the simulation. After removing the trend, the residuals are fitted by an AR(1) model. The coefficient estimate is $\hat{\phi}_1 = 0.386$ with standard error 0.090. To check whether the model is adequate, we plot the residual autocorrelation function in Figure 3. All autocorrelations in Figure 3 are within the 95% confidence interval, which indicates that the model is adequate.

4. PROOFS

We will prove Theorem 1 by first providing three lemmas. Hereafter, $U(\cdot)$ denotes the uniform boundedness of a matrix and $u(\cdot)$ denotes the uniform boundedness of a scalar.

According to eqn (9), notice

$$\hat{\mathbf{g}} = \mathbf{B}(\mathbf{B}\mathbf{B})^{-1}\mathbf{B}'\{\mathbf{g} + \mathbf{x}\} = \tilde{\mathbf{g}} + \tilde{\mathbf{x}}. \tag{11}$$

In eqn (11), $\tilde{\mathbf{g}} = \mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\mathbf{g}$ is the projection of \mathbf{g} and $\tilde{\mathbf{x}} = \mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\mathbf{x}$ is the projection of \mathbf{x} . Then we obtain

$$\hat{\mathbf{x}} = (\mathbf{g} - \tilde{\mathbf{g}}) + (\mathbf{x} - \tilde{\mathbf{x}}). \tag{12}$$

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

ϕ	c	$n = 100$
(0.2, 0.64, -0.144)	1	(0.134 ± 0.114, 0.532 ± 0.068, -0.190 ± 0.090)
	2	(0.097 ± 0.116, 0.489 ± 0.078, -0.212 ± 0.089)
	5	(-0.071 ± 0.134, 0.314 ± 0.110, -0.271 ± 0.080)
(1, -0.56, 0.08)	1	(0.913 ± 0.113, -0.504 ± 0.117, 0.022 ± 0.099)
	2	(0.882 ± 0.117, -0.495 ± 0.115, -0.004 ± 0.096)
	5	(0.775 ± 0.116, -0.466 ± 0.105, -0.085 ± 0.097)
ϕ	c	$n = 200$
(0.2, 0.64, -0.144)	1	(0.180 ± 0.058, 0.590 ± 0.041, -0.167 ± 0.050)
	2	(0.163 ± 0.057, 0.569 ± 0.040, -0.179 ± 0.051)
	5	(0.109 ± 0.063, 0.505 ± 0.045, -0.210 ± 0.049)
(1, -0.56, 0.08)	1	(0.969 ± 0.061, -0.546 ± 0.065, 0.065 ± 0.048)
	2	(0.952 ± 0.062, -0.540 ± 0.064, 0.049 ± 0.048)
	5	(0.910 ± 0.061, -0.523 ± 0.060, 0.011 ± 0.045)
ϕ	c	$n = 400$
(0.2, 0.64, -0.144)	1	(0.189 ± 0.026, 0.613 ± 0.020, -0.159 ± 0.023)
	2	(0.179 ± 0.027, 0.600 ± 0.020, -0.166 ± 0.023)
	5	(0.144 ± 0.028, 0.554 ± 0.024, -0.192 ± 0.022)
(1, -0.56, 0.08)	1	(0.980 ± 0.027, -0.549 ± 0.034, 0.067 ± 0.024)
	2	(0.968 ± 0.027, -0.544 ± 0.033, 0.056 ± 0.024)
	5	(0.938 ± 0.027, -0.532 ± 0.032, 0.028 ± 0.025)
ϕ	c	$n = 1000$
(0.2, 0.64, -0.144)	1	(0.198 ± 0.010, 0.631 ± 0.008, -0.150 ± 0.009)
	2	(0.195 ± 0.010, 0.625 ± 0.008, -0.154 ± 0.009)
	5	(0.184 ± 0.010, 0.610 ± 0.008, -0.163 ± 0.009)
(1, -0.56, 0.08)	1	(0.995 ± 0.010, -0.558 ± 0.012, 0.079 ± 0.009)
	2	(0.990 ± 0.010, -0.555 ± 0.012, 0.074 ± 0.009)
	5	(0.979 ± 0.010, -0.550 ± 0.012, 0.063 ± 0.009)

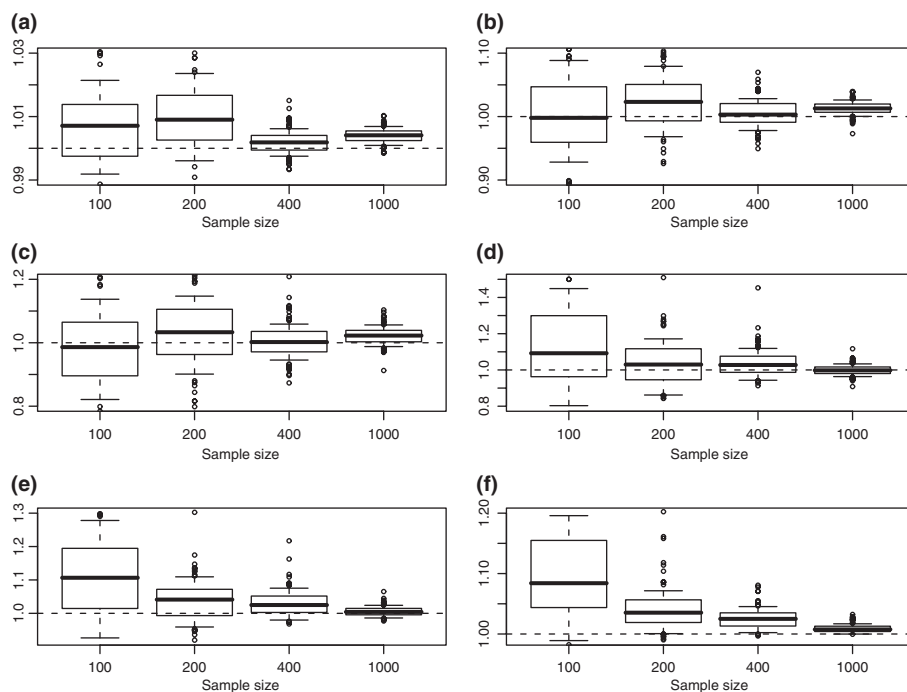


Figure 1. (a) $\phi_1 = -0.8$, (b) $\phi_1 = -0.4$, (c) $\phi_1 = -0.2$, (d) $\phi_1 = 0.2$, (e) $\phi_1 = 0.4$, (f) $\phi_1 = 0.8$

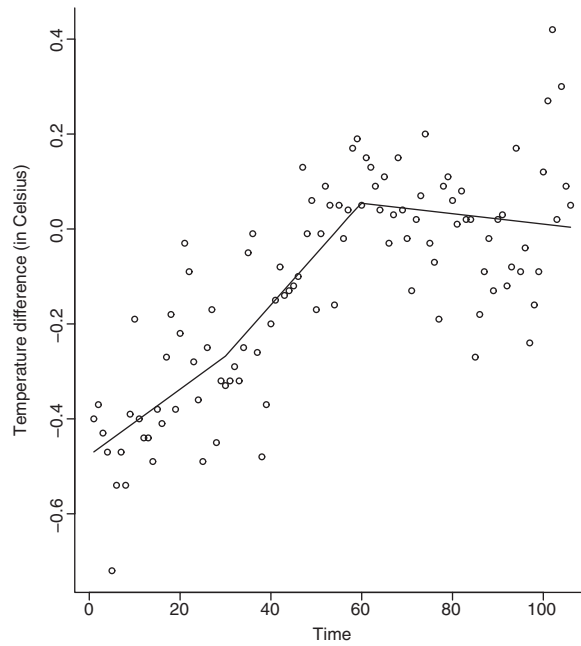


Figure 2. First differences of annual global surface air temperatures in Celsius from 1880 through 1985

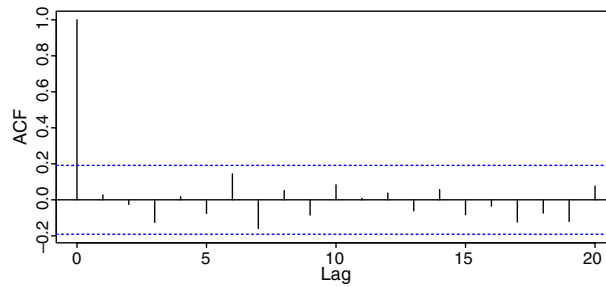


Figure 3. Residual autocorrelations

REMARK 5. Under assumptions i and ii, according to Theorem 5.1 of Huang (2003),

$$\|g - \tilde{g}\|_\infty = \sup_{u \in [0,1]} |g(u) - \tilde{g}(u)| = O(h^m), \tag{13}$$

where $\tilde{g}(u)$ is defined similarly to $\hat{g}(u)$ in eqn (7). The only difference is that $\hat{\beta}$ is calculated by replacing \mathbf{y} in eqn (8) by \mathbf{g} .

Define $\mathbf{x}_{t,n-k} = (x_t, \dots, x_{n-k})'$, $\tilde{\mathbf{x}}_{t,n-k} = (\tilde{x}_t, \dots, \tilde{x}_{n-k})'$, $\mathbf{g}_{t,n-k} = (g_t, \dots, g_{n-k})'$, and $\tilde{\mathbf{g}}_{t,n-k} = (\tilde{g}_t, \dots, \tilde{g}_{n-k})'$. From eqn (13), it is obvious that for any non-negative integers $1 \leq t_1, t_2 \leq p-1$ and $1 \leq k \leq p$,

$$\frac{1}{n} (\mathbf{g}_{t_1,n-k} - \tilde{\mathbf{g}}_{t_1,n-k})' (\mathbf{g}_{k+t_1,n} - \tilde{\mathbf{g}}_{k+t_1,n}) = O(n^{-2m/(2m+1)}), \tag{14}$$

and

$$\frac{1}{n} (\mathbf{g}_{t_1,n-k} - \tilde{\mathbf{g}}_{t_1,n-k})' \mathbf{x}_{t_2,n-t_1-k+t_2} = O_p(n^{-(4m+1)/(4m+2)}). \tag{15}$$

LEMMA 1. If $g(\cdot) \in C^{(1)}[0,1]$ is estimated using the basis of the space $G_N^{(-1)}$ defined in eqn (3), the following hold for $1 \leq t \leq \max\{1, p-1\}$ and $1 \leq k \leq p$:

- [i] $\|\tilde{\mathbf{x}}_{t,n-k}\| = O_p(h^{-1/2})$, where $\|\cdot\|$ is the Euclidean norm;
- [ii] $\hat{\gamma}_k \xrightarrow{P} \gamma_k$;
- [iii] $\hat{\gamma}_k - \tilde{\gamma}_k = O_p(h^2) = O_p(n^{-2/3})$.

PROOF: i. Consider a partition of \mathbf{B} as follows:

$$\mathbf{B} = \begin{pmatrix} \mathbf{1B}_{t-1} \\ \mathbf{tB}_{n-k} \\ \mathbf{(n-k+1)B}_n \end{pmatrix},$$

where the $(j-i+1) \times (N+m)$ matrix $\mathbf{;B}_j$ is the submatrix consisting of the entries from the i th row to the j th row of \mathbf{B} .

Notice that we have the following conclusion regarding the Euclidean norm of \mathbf{b}_j ,

$$\sup_{0 \leq j \leq N} \|\mathbf{b}_j\|^2 - nh = o(nh). \quad \wedge$$

Under assumption ii, according to eqn (3), the (j, k) th entry of $(\mathbf{B}'\mathbf{B})^{-1}$ satisfies

$$(\mathbf{B}'\mathbf{B})_{j,k}^{-1} = \begin{cases} u(n^{-1}h^{-1}), & j = k, \\ 0, & j \neq k, \end{cases}$$

or $(\mathbf{B}'\mathbf{B})^{-1} = U(n^{-1}h^{-1})$. On the other hand,

$$\begin{aligned} \frac{1}{n} E\|\mathbf{B}'\mathbf{x}\|^2 &= \frac{1}{n} \sum_{j=0}^N E \left(\sum_{i=1}^n b_{j,1}(u_i)x_i \right)^2 = \frac{1}{n} \sum_{j=0}^N E \left(\sum_{i=1}^{n_j} x_{ji} \right)^2 \\ &= \sum_{j=0}^N \frac{n_j}{n} \sum_{|k| < n_j} \left(1 - \frac{|k|}{n_j} \right) \gamma_k \rightarrow \sum_{k=-\infty}^{\infty} \gamma_k. \end{aligned} \quad (16)$$

The last step above is obtained from the fact that $1/n_j = 1/(nh) + u(n^{-1}h^{-1})$; that is $1/n_j$ is uniformly bounded for all j . From Remark 4, $\sum_{k=-\infty}^{\infty} \gamma_k < \infty$. Therefore, $\|\mathbf{B}'\mathbf{x}\| = O_p(n^{1/2})$ and thus

$$\|(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\mathbf{x}\| = O_p(n^{-1/2}h^{-1}).$$

According to eqn (11), notice that $\tilde{\mathbf{x}}_{t,n-k}$ can be rewritten as follows:

$$\tilde{\mathbf{x}}_{t,n-k} = \mathbf{tB}_{n-k}(\mathbf{B}'\mathbf{B})^{-1}(\mathbf{B}'\mathbf{x}), \quad (17)$$

and thus

$$\|\tilde{\mathbf{x}}_{t,n-k}\|^2 \leq [nh]O_p(n^{-1}h^{-2}).$$

The proof is complete. □

ii. According to eqn (12), we obtained

$$\begin{aligned} \hat{\gamma}_k &= \frac{1}{n} \tilde{\mathbf{x}}'_{1,n-k} \tilde{\mathbf{x}}_{k+1,n} = \frac{1}{n} (\mathbf{g}_{1,n-k} - \tilde{\mathbf{g}}_{1,n-k})' (\mathbf{g}_{k+1,n} - \tilde{\mathbf{g}}_{k+1,n}) + \frac{1}{n} (\mathbf{g}_{k+1,n} - \tilde{\mathbf{g}}_{k+1,n})' \mathbf{x}_{1,n-k} \\ &\quad - \frac{1}{n} (\mathbf{g}_{k+1,n} - \tilde{\mathbf{g}}_{k+1,n})' \tilde{\mathbf{x}}_{1,n-k} + \frac{1}{n} (\mathbf{g}_{1,n-k} - \tilde{\mathbf{g}}_{1,n-k})' \mathbf{x}_{k+1,n} \\ &\quad - \frac{1}{n} (\mathbf{g}_{1,n-k} - \tilde{\mathbf{g}}_{1,n-k})' \tilde{\mathbf{x}}_{k+1,n} + \frac{1}{n} \mathbf{x}'_{1,n-k} \mathbf{x}_{k+1,n} - \frac{1}{n} \mathbf{x}'_{1+k,n} \tilde{\mathbf{x}}_{1,n-k} \\ &\quad - \frac{1}{n} \tilde{\mathbf{x}}'_{1+k,n} \mathbf{x}_{1,n-k} + \frac{1}{n} \tilde{\mathbf{x}}'_{1,n-k} \tilde{\mathbf{x}}_{k+1,n}. \end{aligned} \quad (18) \wedge$$

From (i), it is obvious that for any non-negative integers $1 \leq t_1, t_2 \leq \max\{1, p-1\}$ and $1 \leq k \leq p$,

$$\frac{1}{n} (\mathbf{g}_{t_1,n-k} - \tilde{\mathbf{g}}_{t_1,n-k})' \tilde{\mathbf{x}}_{t_2,n-t_1-k+t_2} = O_p(h^m). \quad (19)$$

From (i), eqns (14), (15) and (19), it can be concluded that the order of the dominant term of (18) is

$$\frac{1}{n} \mathbf{x}'_{1,n-k} \mathbf{x}_{k+1,n} = \tilde{\gamma}_k = \overset{p}{\rightarrow} \gamma_k. \quad \wedge$$

The proof is complete. □

iii. Notice that

$$\begin{aligned} \hat{\gamma}_k - \tilde{\gamma}_k &= \frac{1}{n} (\tilde{\mathbf{x}}'_{1,n-k} \tilde{\mathbf{x}}_{k+1,n} - \mathbf{x}'_{1,n-k} \mathbf{x}_{k+1,n}) = \frac{1}{n} (\mathbf{g}_{1,n-k} - \tilde{\mathbf{g}}_{1,n-k})' (\mathbf{g}_{k+1,n} - \tilde{\mathbf{g}}_{k+1,n}) + \frac{1}{n} (\mathbf{g}_{k+1,n} - \tilde{\mathbf{g}}_{k+1,n})' \mathbf{x}_{1,n-k} \\ &\quad - \frac{1}{n} (\mathbf{g}_{k+1,n} - \tilde{\mathbf{g}}_{k+1,n})' \tilde{\mathbf{x}}_{1,n-k} + \frac{1}{n} (\mathbf{g}_{1,n-k} - \tilde{\mathbf{g}}_{1,n-k})' \mathbf{x}_{k+1,n} \\ &\quad - \frac{1}{n} (\mathbf{g}_{1,n-k} - \tilde{\mathbf{g}}_{1,n-k})' \tilde{\mathbf{x}}_{k+1,n} - \frac{1}{n} \mathbf{x}'_{1+k,n} \tilde{\mathbf{x}}_{1,n-k} - \frac{1}{n} \tilde{\mathbf{x}}'_{1+k,n} \mathbf{x}_{1,n-k} \\ &\quad + \frac{1}{n} \tilde{\mathbf{x}}'_{1,n-k} \tilde{\mathbf{x}}_{k+1,n}. \end{aligned} \quad (20)$$

The dominant terms of eqn (20) are $O_p(n^{-2/3})$ from (i), eqns (14), (15) and (19). The proof is complete. □

The extension of the above lemma to $G_N^{(0)}$ is not trivial due to the fact that the vectors of the B-spline basis of $G_N^{(0)}$ are not orthogonal. The following lemma is needed to prove results similar to Lemma 4.1 if $g(\cdot) \in C^{(2)}[0,1]$ is estimated by a basis of $G_N^{(0)}$.

LEMMA 2. For the basis of $G_N^{(0)} \{b_j, j = -1, \dots, N\}$, $B'B = U(nh)$.

PROOF: According to $u_j = j/n$ and the definition of $\{b_j, j = -1, \dots, N\}$ in eqn (4), for $0 \leq j = k \leq N-1$,

$$\begin{aligned} \|b_j\|^2 &= \sum_{i=1}^n b_{j,2}^2(u_i) = \sum_{i=1}^{n_j} b_{j,2}^2(u_i) + \sum_{i=1}^{n_{j+1}} b_{j,2}^2(u_{(j+1)_i}) \\ &= \sum_{i=1}^{n_j} \left(\frac{j_i}{nh} - j\right)^2 + \sum_{i=1}^{n_{j+1}} \left(j + 2 - \frac{(j+1)_i}{nh}\right)^2. \end{aligned}$$

For $1 \leq i \leq n_j$,

$$\frac{i-1}{nh} \leq \frac{j_i}{nh} - j \leq \frac{i}{nh}, \tag{21}$$

and

$$\frac{n_j - i}{nh} \leq (j+1) - \frac{j_i}{nh} \leq \frac{n_j - i + 1}{nh}. \tag{22}$$

Thus it can be shown that

$$\frac{1}{6n^2h^2} (n_j - 1)n_j(2n_j - 1) \leq \sum_{i=1}^{n_j} \left(\frac{j_i}{nh} - j\right)^2 \leq \frac{1}{6n^2h^2} n_j(n_j + 1)(2n_j + 1).$$

Therefore, $\|b_j\|^2 = \frac{2}{3}nh + u(nh)$ for $0 \leq j \leq N-1$. Similarly we can show that $\|b_j\|^2 = \frac{1}{3}nh + u(nh)$ for $j = -1$ and N . For $-1 \leq j \leq N-1$, from eqns (21) and (22),

$$\begin{aligned} b'_j b_{j+1} &= \sum_{i=1}^{n_{j+1}} \left(j + 2 - \frac{(j+1)_i}{nh}\right) \left(\frac{(j+1)_i}{nh} - j - 1\right) \\ &= \frac{1}{6}nh + u(nh). \end{aligned}$$

It is obvious that $b'_j b_k = 0$ for any $|j-k| > 1$. The proof is complete. □

LEMMA 3. If $g(\cdot) \in C^{(2)}[0,1]$ is estimated by the basis of the space $G_N^{(0)}$ defined in eqn (4), the following hold for $1 \leq t \leq \max\{1, p-1\}$ and $1 \leq k \leq p$: ^

i. $\|\tilde{x}_{t,n-k}\| = O_p(h^{-1/2})$;

ii. $\hat{\gamma}_k \xrightarrow{p} \gamma_k$;

iii. $\hat{\gamma}_k - \tilde{\gamma}_k = O_p(h^4) = O_p(n^{-4/5})$. ^

PROOF: i. Lemma 2 implies that $(B'B)^{-1} = U(n^{-1}h^{-1})$. It suffices to show similar results to eqn (16) for a basis of $G_N^{(0)}$.

$$\begin{aligned} \frac{1}{n} E\|B'x\|^2 &= \frac{1}{n} E(b'_{-1}x)^2 + \frac{1}{n} E(b'_N x)^2 + \frac{1}{n} \sum_{j=0}^{N-1} E\left\{ \sum_{i=1}^{n_j} \left(\frac{j_i}{nh} - j\right) x_{j_i} + \sum_{i=1}^{n_{j+1}} \left(j + 2 - \frac{(j+1)_i}{nh}\right) x_{(j+1)_i} \right\}^2 \\ &= \frac{1}{n} E(b'_{-1}x)^2 + \frac{1}{n} E(b'_N x)^2 + R_{1n} + R_{2n} + R_{3n}, \end{aligned} \tag{23} \text{^}$$

where

$$\begin{aligned} R_{1n} &= \frac{1}{n} \sum_{j=0}^{N-1} E\left\{ \sum_{i=1}^{n_j} \left(\frac{j_i}{nh} - j\right) x_{j_i} \right\}^2, \\ R_{2n} &= \frac{1}{n} \sum_{j=0}^{N-1} E\left\{ \sum_{i=1}^{n_{j+1}} \left(j + 2 - \frac{(j+1)_i}{nh}\right) x_{(j+1)_i} \right\}^2, \\ R_{3n} &= \frac{2}{n} \sum_{j=0}^{N-1} E\left\{ \sum_{i=1}^{n_j} \sum_{k=1}^{n_{j+1}} \left(\frac{j_i}{nh} - j\right) \left(j + 2 - \frac{(j+1)_k}{nh}\right) x_{j_i} x_{(j+1)_k} \right\}. \end{aligned} \text{^}$$

Since terms R_{1n} and R_{2n} are very similar, we only discuss R_{1n} . From $n_j = nh + u(nh)$, eqns (21) and (22), the dominant term of R_{1n} is

$$\begin{aligned} \frac{1}{n} \sum_{j=0}^{N-1} E \left\{ \sum_{i=1}^{n_j} \frac{i}{nh} x_{ji} \right\}^2 &= \frac{1}{n^3 h^2} \sum_{j=0}^{N-1} \sum_{i=0}^{n_j} i^2 \gamma_0 + \frac{1}{n^3 h^2} \sum_{j=0}^{N-1} \sum_{i=1}^{n_j} \sum_{k=1, i \neq k}^{n_j} ik \gamma_{i-k} \\ &= \frac{1}{n^3 h^2} \sum_{j=0}^{N-1} \sum_{|k| < n_j} \left(\sum_{i=|k|+1}^{n_j} i^2 \right) \gamma_k - \frac{2}{n^3 h^2} \sum_{j=0}^{N-1} \sum_{k=1}^{n_j-1} \sum_{i=k+1}^{n_j} ik \gamma_k \\ &= \frac{1}{n^3 h^2} \sum_{j=0}^{N-1} \sum_{|k| < n_j} \left(\sum_{i=1}^{n_j} i^2 - \sum_{i=1}^{|k|} i^2 \right) \gamma_k - \frac{1}{n^3 h^2} \sum_{j=0}^{N-1} \sum_{k=1}^{n_j-1} k(n_j - k)(n_j + k + 1) \gamma_k. \end{aligned} \quad \wedge$$

From Kronecker's Lemma (See Fuller (1996) page 126 for details), we have \wedge

$$\frac{1}{n^3 h^2} \sum_{j=0}^{N-1} \sum_{|k| < n_j} \left(\frac{\sum_{i=1}^{|k|} i^2}{\sum_{i=1}^{n_j} i^2} \right) \gamma_k \rightarrow 0 \quad \text{as } n \rightarrow \infty,$$

and

$$\frac{1}{n^3 h^2} \sum_{j=0}^{N-1} \frac{1}{n_j^3} \sum_{k=1}^{n_j-1} k(n_j - k)(n_j + k + 1) \gamma_k \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Therefore, $R_{1n} \rightarrow \frac{1}{3} \sum_{k=0}^{\infty} \gamma_k < \infty$. The absolute value of the dominant term of R_{3n} is

$$\begin{aligned} &\left| \frac{2}{n^3 h^2} \sum_{j=0}^{N-1} \left| \sum_{i=1}^{n_j} \sum_{k=1}^{n_{j+1}} i(n_{j+1} - k) \gamma_{(j+1)k-j} \right| \right| \\ &\leq \frac{2}{n^3 h^2} \sum_{j=0}^{N-1} \left\{ n_j(n_{j+1} - 1) \sum_{i=1}^{n_j} \sum_{k=1}^{n_{j+1}} \left| \gamma_{(j+1)k-j} \right| \right\} \\ &\leq \frac{2}{n} \sum_{j=0}^{N-1} \left(n_j |\gamma_{n_j}| + \sum_{k=1}^{n_j} k |\gamma_k| + \sum_{k=1}^{n_{j+1}} k |\gamma_{2n_j-k}| \right). \end{aligned} \quad \wedge$$

By following a similar discussion to that of R_{1n} and applying Kronecker's lemma, we can show that $R_{3n} \rightarrow 0$ as $n \rightarrow \infty$. It is straightforward that the other two terms in (23) are $O_p(1)$ from the above discussion. Therefore, $\|\mathbf{B}'\mathbf{x}\| = O_p(n^{1/2})$.

According to eqn (17),

$$\|\tilde{\mathbf{x}}_{t,n-k}\|^2 \leq [nh] \left(\frac{i}{nh} + \frac{n_{j+1} - i}{nh} \right)^2 O_p(n^{-1}h^{-2}) = O_p(h^{-1}). \quad \wedge$$

This completes the proof of (i). The proofs of (ii) and (iii) are very similar to those of Lemma 1. \square

PROOF OF THEOREM 1: Theorem 1 is obtained immediately according to the Lemmas 1 and 3 and

$$\hat{\phi} - \tilde{\phi} = \hat{\Gamma}^{-1}(\hat{\gamma} - \tilde{\gamma}) + \hat{\Gamma}^{-1}(\tilde{\Gamma} - \hat{\Gamma})\tilde{\Gamma}^{-1}\tilde{\gamma}. \quad \wedge$$

THEOREM 3. Under assumptions i-iv, $\hat{\sigma}^2 - \sigma^2 = o_p(1)$.

PROOF: Define $\tilde{\sigma}^2$ in a similar way to $\hat{\sigma}^2$ using $\tilde{\gamma}$ and $\tilde{\phi}$. It is known that $\tilde{\sigma}^2 - \sigma^2 = o_p(1)$. It suffices to show that $\hat{\sigma}^2 - \tilde{\sigma}^2 = o_p(1)$. Notice that

$$\begin{aligned} \hat{\sigma}^2 - \tilde{\sigma}^2 &= (\hat{\gamma}_0 - \hat{\gamma}'\hat{\phi}) - (\tilde{\gamma}_0 - \tilde{\gamma}'\tilde{\phi}) \\ &= (\hat{\gamma}_0 - \tilde{\gamma}_0) - \hat{\gamma}'(\hat{\phi} - \tilde{\phi}) - (\hat{\gamma}' - \tilde{\gamma}')\tilde{\phi} \\ &= o_p(1). \end{aligned} \quad (24)$$

Theorem 1 and (ii) of Lemmas 1 and 3 are used to establish eqn (24). \square

In applying Theorem 2 to real data analysis, the covariance matrix Γ^{-1} can be estimated by $\hat{\Gamma}^{-1}$ and the white noise variance σ^2 can be replaced by its consistent estimate $\hat{\sigma}^2$.

Acknowledgements

Lijian Yang's research partially supported by NSF awards DMS 0706518 and 1007594 and a grant from the Risk Management Institute, National University of Singapore.

REFERENCES

- Altman, N. (1993) Estimating error correlation in non-parametric regression. *Statistics & Probability Letters* **18**, 213–18.
- Brockwell, P. J. and Davis, R. A. (1991) *Time Series: Theory and Methods, 2nd ed.* New York: Springer.
- Fan, J. and Gijbels, I. (1996) *Local Polynomial Modeling and Its Applications.* London: Chapman & Hall.
- Fuller, W. A. (1996) *Introduction to Statistical Time Series.* New York: Wiley.
- Hall, P. and Keilegom, I. V. (2003) Using difference-based methods for inference in non-parametric regression with time series errors. *Journal of the Royal Statistical Society Series B* **65**, 443–56.
- Huang, J. Z. (2003) Local asymptotics for polynomial spline regression. *Annals of Statistics* **31**, 1600–35.
- Huang, J. and Yang, L. (2004) Identification of nonlinear additive autoregressive models. *Journal of the Royal Statistical Society Series B* **66**, 463–77.
- Truong, Y. K. (1991) Nonparametric curve estimation with time series errors. *Journal of Statistical Planning and Inference* **28**, 167–83.
- Tsay, R. S. and Tiao, G. C. (1984) Consistent estimates of autoregressive parameters and extended sample autocorrelation function for stationary and nonstationary ARMA models. *Journal of the American Statistical Association* **79**, 84–96.
- Wang, L. and Yang, L. (2007) Spline-backfitted kernel smoothing of nonlinear additive autoregression model. *Annals of Statistics* **35**, 2474–503.
- Wang, J. and Yang, L. (2009) Polynomial spline confidence bands for regression curves. *Statistica Sinica* **19**, 325–42.
- Xue, L. and Yang, L. (2006) Additive coefficient modelling via polynomial spline. *Statistica Sinica* **16**, 1423–46.
- Yang, L. and Tschernig, R. (1999) Multivariate bandwidth selection for local linear regression. *Journal of the Royal Statistical Society Series B* **61**, 793–815.