JOURNAL OF NONPARAMETRIC STATISTICS
2022, VOL. 34, NO. 2, 357–376
https://doi.org/10.1080/10485252.2022.2055756

Taylor & Francis
Taylor & Francis Group

Check for updates

Statistical inference for ARMA time series with moving average trend

Zening Song and Lijian Yang

Center for Statistical Science & Department of Industrial Engineering, Tsinghua University, Beijing, People's Republic of China

ABSTRACT

Maximum likelihood estimator (MLE) and Bayesian Information Criterion (BIC) order selection are examined for ARMA time series with slowly varying trend to validate the well-known detrending technique of moving average [Section 1.4, Brockwell, P.J., and Davis, R.A. (1991), *Time Series: Theory and Methods*, New York: Springer-Verlag]. In step one, a moving average equivalent to local linear regression is fitted to the raw data with a data-driven lag number, and subtracted from raw data to produce a sequence of residuals. The residuals are used in step two as substitutes of the latent ARMA series for MLE and BIC procedures. It is shown that with second order smooth trend and correctly chosen lag number, the two-step MLE is oracally efficient, i.e. it is asymptotically as efficient as the would-be MLE based on the unobserved ARMA series. At the same time, the two-step BIC consistently selects the orders as well. Simulation experiments corroborate the theoretical findings.

ARTICLE HISTORY
Received 9 February 2021
Accepted 6 March 2022

KEYWORDS
Asymptotic normality; local linear estimator; maximum likelihood; moving average; oracle efficiency

AMS SUBJECT CLASSIFICATIONS
62G05; 62G20; 62M10

1. Introduction

Much of the standard theory for time series analysis, such as the properties and inference for ARMA processes, is developed under the assumption of stationarity (see Chapters 3–5, 7–8, 10–11 of Brockwell and Davis 1991). On the other hand, all authors of time series textbooks are aware of the presence of nonstationarity, such as trend and seasonality, and present preliminary ad hoc steps to handle such phenomena (see Section 1.4 of Brockwell and Davis 1991, and Section 2.3 of Shumway and Stoffer 2017).

Consider an observed time series realisation $\mathbf{X} = (X_1, \ldots, X_n)^\mathsf{T}$ with slowly varying trend:

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

The data generating process in (1) can be alternatively written as

$$\mathbf{X} = \mathbf{m} + \mathbf{Y},$$

CONTACT Lijian Yang yanglijian@tsinghua.edu.cn Center for Statistical Science & Department of Industrial Engineering, Tsinghua University, Beijing 100084, People's Republic of China

© 2022 American Statistical Association and Taylor & Francis

with stochastic and deterministic components $\mathbf{Y} = (Y_1, \dots, Y_n)^{\mathsf{T}}$ and $\mathbf{m} = \{m(1/n), m(2/n) \dots, m(n/n)\}^{\mathsf{T}}$. The stochastic $\mathbf{Y}$ is the realisation of a stationary time series $\{Y_t\}_{t=-\infty}^{\infty}$, presumed in this work to be an autoregressive moving-average time series of orders p and q (ARMA(p, q)), satisfying

$$Y_t - \sum_{k=1}^{p} \phi_k Y_{t-k} = \epsilon_t + \sum_{k=1}^{q} \theta_k \epsilon_{t-k}, \quad t \in \mathbb{Z}, \tag{2}$$

where the innovations $\{\epsilon_t\}_{t=-\infty}^{\infty}$ is independent and identically distributed white noise with mean zero and variance σ^2 ($\epsilon_t \sim \mathrm{IID}(0, \sigma^2)$). The deterministic $\mathbf{m}$ consists of values of a smooth function $m(\cdot)$ at $t/n, 1 \le t \le n$, which is slowly varying in the sense that from time t to $t+1$ the increment in trend is of a negligible order $1/n$. For time series whose trend varies drastically, the deterministic component is typically represented as a parametric function $m(\cdot)$ of $t, 1 \le t \le n$, which is in general easier to estimate than the nonparametric trend in (1).

The most important parameters in model (2) consists of ARMA coefficients $\alpha = (\alpha_1, \dots, \alpha_{p+q})^{\mathsf{T}}$ with $\alpha_k = \phi_k$ for $1 \le k \le p$ and $\alpha_k = \theta_{k-p}$ for $p+1 \le k \le p+q$, and variance σ^2. Shao and Yang (2017) proposed two-step procedures to estimate $\boldsymbol{\alpha}$ and σ^2, and to determine ARMA orders p and q. The procedures are as follows: let $\hat{m}(\cdot)$ be some first-step estimator for the trend function $m(\cdot)$, $\hat{\mathbf{Y}}$ the vector of residuals obtained by subtracting the trend estimate $\hat{\mathbf{m}} = \{\hat{m}(1/n), \dots, \hat{m}(n/n)\}^{\mathsf{T}}$ from the observed $\mathbf{X}$

$$\hat{Y}_t = X_t - \hat{m}(t/n), \quad 1 \le t \le n, \tag{3}$$

and $\hat{\boldsymbol{\alpha}}$ the second-step maximum likelihood estimator (MLE) of $\boldsymbol{\alpha}$ based on $\hat{\mathbf{Y}}$. As a benchmark, let $\tilde{\boldsymbol{\alpha}}$ be the MLE of $\boldsymbol{\alpha}$ based on the unobservable $\mathbf{Y}$.

Oracle efficiency of estimation and consistency of order selection is established in Shao and Yang (2017) under generic causality/invertibility assumptions on $\{Y_t\}_{t=-\infty}^{\infty}$ and moment assumptions on the innovations $\{\epsilon_t\}_{t=-\infty}^{\infty}$ (assumptions (a,b) in Section 3). Additional assumption (c) in Section 3 is set forth for trend estimator $\hat{m}(\cdot)$ under which the difference between $\hat{\boldsymbol{\alpha}}$ and $\tilde{\boldsymbol{\alpha}}$ is of the asymptotically negligible order $o(n^{-1/2})$. In other words, the estimator based on $\hat{\mathbf{Y}}$ using estimated trend function $\hat{m}(\cdot)$ is asymptotically as efficient as the one based on $\mathbf{Y}$ using true trend function $m(\cdot)$ as if the unknown $m(\cdot)$ were known by 'oracle', thus the coined term 'oracle efficiency'. It was proved that the B-spline estimator for trend function $m(\cdot)$ meets assumption (c) if the trend function $m(\cdot)$ satisfies minimum smoothness Condition 1 and the B-spline smoothing parameter Condition 2.

While B-spline estimation is fast to compute, classic moving average estimation of trend (see Section 1.4 of Brockwell and Davis 1991 and Section 2.3 of Shumway and Stoffer 2017) is more widely used and intuitively understood by practitioners engaged in time series analysis. Therefore, it is interesting to investigate if moving average trend satisfies assumption (c) of Shao and Yang (2017), and if so, under what conditions on the moving average lag. This had been accomplished for the least-squares estimator (LSE) of the simpler AR(p) model by Qiu et al. (2013), the primary goal of the present paper is to establish the oracle efficiency of the MLE $\hat{\boldsymbol{\alpha}}$ obtained from moving average detrended residuals for the general ARMA model.

The organisation of the paper is as follows. The MLE and BIC procedures and their theoretical properties are described in Sections 2 and 3. Section 4 contains detailed steps to compute the crucial smoothing parameter, the moving average lag number. Section 5 summarises findings of Monte Carlo experiments, while Section 6 concludes. All technical proofs are in the Appendix.

2. Construction of estimators

In this section, we describe a data-driven step one moving average estimator of the trend in X_t that produces residuals $\hat{\mathbf{Y}}$. Also presented is step two MLE of ARMA parameters in $\mathbf{Y}$ based on the moving average residuals $\hat{\mathbf{Y}}$ from step one.

2.1. Moving average trend estimator

The methodological innovation that sets the current work apart from Shao and Yang (2017) is the choice of trend estimator $\hat{m}(\cdot)$ as the well-known moving average trend, which is described in detail next.

According to Section 1.4 of Brockwell and Davis (1991), the moving average trend estimator of lag l for $\mathbb{E}X_t = m_t$ is defined as

$$\hat{m}_t = (2l+1)^{-1} \sum_{i=t-l}^{t+l} X_i, \quad l+1 \le t \le n-l,$$

where the moving average lag l is a fixed positive integer, whose default value is 2. The moving average trend is appealing as $(2l+1)^{-1}\sum_{i=t-l}^{t+l} m_i$ approximates well m_t due to smoothness in trend and the variance of $(2l+1)^{-1}\sum_{i=t-l}^{t+l} Y_i$ is reduced to order $\mathcal{O}((2l+1)^{-1})$ due to stationarity in $\{Y_t\}_{t=-\infty}^{\infty}$ (Brockwell and Davis 1991, Theorem 7.1.1).

Qiu et al. (2013) had argued against this ad hoc choice of lag value, and shown that allowing a sequence of positive integers $l = l_n$ that tend to infinity at an appropriate rate produces an efficient follow-up estimator of AR(p) coefficients. To be more precise, one estimates $m(\cdot)$ by the local linear method,

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

where

$$N_{1t}^{-1} = \frac{4l^2 - 4lt + 6l + 4t^2 - 6t + 2}{(l+t)(l+t-1)(l+t+1)},$$
$$N_{2t}^{-1} = \frac{6(l-t+1)}{(l+t)(l+t-1)(l+t+1)}, \tag{5}$$

$$N_{3t}^{-1} = \frac{4(n-t)^2 + 4l^2 - 4l(n-t) + 2(n+l-t)}{(n+l-t+2)(n+l-t+1)(n+l-t)},$$
$$N_{4t}^{-1} = \frac{6(n-l-t)}{(n+l-t+2)(n+l-t+1)(n+l-t)}. \quad (6)$$

The above estimator can be alternatively written as a local linear estimator (Fan and Gijbels 1996)

$$\hat{m}_t = \hat{m}(t/n) = \hat{a},$$
$$(\hat{a}, \hat{b}) = \text{argmin}_{(a,b)\in\mathbb{R}^2} \sum_{i=1}^{n} \{X_i - a - b(i-t)/n\}^2 K_h\{(i-t)/n\}, \quad (7)$$

in which a uniform kernel $K(u) = \frac{1}{2} I_{[-1,1]}(u)$ and the bandwidth $h = h_n = l_n/n > 0$ are employed.

Notice that under the scheme of Qiu et al. (2013), not only the lag number l is sample size dependent, but the definition of trend estimator $\hat{m}_t$ is extended to $1 \le t \le l$ and $n - l + 1 \le t \le n$ as well. It will be shown in Section 3 that this first step data-driven moving average trend facilitates oracally efficient estimation of ARMA coefficients in the ensuing step.

2.2. Estimation of ARMA coefficients

Denote the true parameters of the ARMA process (2) by σ_0^2 and $\boldsymbol{\alpha}_0 = (\alpha_{01}, \ldots, \alpha_{0,p+q})^{\mathsf{T}}$ with $\alpha_{0k} = \phi_{0k}$ for $1 \le k \le p$ and $\alpha_{0k} = \theta_{0,k-p}$ for $p+1 \le k \le p+q$. Model (2) is rewritten as

$$\Phi(\boldsymbol{\alpha}_0, B) Y_t = \Theta(\boldsymbol{\alpha}_0, B)\epsilon_t, \quad (8)$$

where $\Phi(\boldsymbol{\alpha}_0, B) = 1 - \sum_{k=1}^{p} \phi_{0k} B^k$ and $\Theta(\boldsymbol{\alpha}_0, B) = 1 + \sum_{k=1}^{q} \theta_{0k} B^k$, where B denotes the backshift operator: $B^k Y_t = Y_{t-k}$. As in Shao and Yang (2017), it is assumed in this paper that the time series $\{Y_t\}_{t=-\infty}^{\infty}$ is causal and invertible, thus for some sequences of coefficients $\{\psi_{0j}\}_{j=0}^{\infty}$ and $\{\pi_{0j}\}_{j=0}^{\infty}$ such that $\sum_{j=0}^{\infty} |\psi_{0j}| < \infty, \sum_{j=0}^{\infty} |\pi_{0j}| < \infty$,

$$Y_t = \sum_{j=0}^{\infty} \psi_{0j}\epsilon_{t-j}, \epsilon_t = \sum_{j=0}^{\infty} \pi_{0j} Y_{t-j}, \quad \text{almost surely.}$$

According to equation (3.1.19) in Theorem 3.1.2 of Brockwell and Davis (1991), the constants $\{\pi_{0j}\}_{j=0}^{\infty}$ satisfy the following

$$\sum_{j=0}^{\infty} \pi_{0j} z^j = \Phi(\boldsymbol{\alpha}_0, z)/\Theta(\boldsymbol{\alpha}_0, z), \quad |z| \le 1. \quad (9)$$

If $\{\epsilon_t\}_{t=1}^{n}$ were observable and followed a normal distribution, the MLE would be calculated by minimising $n \log \sigma^2 + \sum_{t=1}^{n} \epsilon_t^2/\sigma^2$, which is proportional to the log-likelihood function $l(\boldsymbol{\alpha}, \sigma^2; \mathbf{Y})$. Therefore, the following objective function $Q_n(\boldsymbol{\alpha}; \mathbf{Y})$ could be used:

$$Q_n(\boldsymbol{\alpha}; \mathbf{Y}) = n^{-1} \sum_{t=p+1}^{n} \left(\sum_{j=0}^{t-1} \pi_j Y_{t-j} \right)^2$$

in which the $\{\pi_j\}_{j=0}^{\infty}$ are rational functions of candidate value $\boldsymbol{\alpha}$ defined by mimicking (9)

$$\sum_{j=0}^{\infty} \pi_j z^j = \sum_{j=0}^{\infty} \pi_j(\boldsymbol{\alpha}) z^j = \Phi(\boldsymbol{\alpha}, z)/\Theta(\boldsymbol{\alpha}, z). \tag{10}$$

Minimising $Q_n(\boldsymbol{\alpha}; \mathbf{Y})$ yields the following would-be estimators $\tilde{\boldsymbol{\alpha}}$ and $\tilde{\sigma}^2$:

$$\begin{aligned} \tilde{\boldsymbol{\alpha}} &= \operatorname{argmin} Q_n(\boldsymbol{\alpha}; \mathbf{Y}); \\ \tilde{\sigma}^2 &= Q_n(\tilde{\boldsymbol{\alpha}}; \mathbf{Y}). \end{aligned}$$

Pierce (1971) had argued that the above estimators have the same asymptotic properties as the MLE obtained from $l(\boldsymbol{\alpha}, \sigma^2; \mathbf{Y})$, and $(\tilde{\boldsymbol{\alpha}}^{\mathsf{T}}, \tilde{\sigma}^2)$ were named 'infeasible' MLE of $(\boldsymbol{\alpha}_0^{\mathsf{T}}, \sigma_0^2)$ in Shao and Yang (2017), as it relies on unobservable sequence $\{Y_t\}_{t=1}^n$, not the actual observations $\{X_t\}_{t=1}^n$.

The feasible replica of $(\tilde{\boldsymbol{\alpha}}^{\mathsf{T}}, \tilde{\sigma}^2)$ are constructed by substituting $\{Y_t\}_{t=1}^n$ with the residuals $\{\hat{Y}_t\}_{t=1}^n$ calculated by (3), that is,

$$\begin{aligned} \hat{\boldsymbol{\alpha}} &= \operatorname{argmin} Q_n(\boldsymbol{\alpha}; \hat{\mathbf{Y}}), \\ \hat{\sigma}^2 &= Q_n(\hat{\boldsymbol{\alpha}}; \hat{\mathbf{Y}}), \end{aligned} \tag{11}$$

where $Q_n(\boldsymbol{\alpha}; \hat{\mathbf{Y}}) = n^{-1} \sum_{t=p+1}^{n} (\sum_{j=0}^{t-1} \pi_j \hat{Y}_{t-j})^2$. The estimators $(\hat{\boldsymbol{\alpha}}^{\mathsf{T}}, \hat{\sigma}^2)$ are the two-step MLE of $(\boldsymbol{\alpha}_0^{\mathsf{T}}, \sigma_0^2)$ according to Shao and Yang (2017), which were shown to enjoy oracle efficiency under Assumptions (a–c) in the next section, namely, $\hat{\boldsymbol{\alpha}}^{\mathsf{T}}$ is asymptotically indistinguishable from $\tilde{\boldsymbol{\alpha}}$, while $\hat{\sigma}^2$ is consistent as $\tilde{\sigma}^2$.

The estimation of ARMA coefficient $\boldsymbol{\alpha}_0$ and variance σ_0^2 are outlined in Shao and Yang (2017):

- Step 1. Estimate the trend function $m(\cdot)$ and obtain the residuals $\{\hat{Y}_t\}_{t=1}^n$ by (3);
- Step 2. obtain the MLE $\hat{\boldsymbol{\alpha}}$ and $\hat{\sigma}^2$ according to (11).

3. Main results

The following assumptions are needed for the theoretical development.

(a) The parameter space Ξ is compact and consists of $\boldsymbol{\alpha}$ such that all roots of $\Phi(\boldsymbol{\alpha}, z) = 0$ and $\Theta(\boldsymbol{\alpha}, z) = 0$ are larger than one in absolute value, and they have no common roots. The true parameter value $\boldsymbol{\alpha}_0$ is in the interior of the parameter space Ξ.

(b) The innovations $\{\epsilon_t\}_{t=-\infty}^{\infty}$ are i.i.d. with $\mathbb{E}\epsilon_1^6 < \infty$.

(c) The trend estimator $\hat{m}(\cdot)$ satisfies the following constraints

$$\max_{1\le t\le n} \mathbb{E}\{m(t/n)-\hat{m}(t/n)\}^2 = o(n^{-1/2}), \tag{12}$$

$$\max_{1\le k\le p+q} n^{-1}\left|\sum_{t=p+1}^{n}\left(\sum_{j=0}^{t-1}\frac{\partial\pi_{0j}}{\partial\alpha_k}Y_{t-j}\right)\left[\sum_{j=0}^{t-1}\pi_{0j}\left\{\hat{m}\left(\frac{t-j}{n}\right)-m\left(\frac{t-j}{n}\right)\right\}\right]\right|$$
$$= o_p(n^{-1/2}), \tag{13}$$

$$\max_{1\le k\le p+q} n^{-1}\left|\sum_{t=p+1}^{n}\left(\sum_{j=0}^{t-1}\pi_{0j}Y_{t-j}\right)\left[\sum_{j=0}^{t-1}\frac{\partial\pi_{0j}}{\partial\alpha_k}\left\{\hat{m}\left(\frac{t-j}{n}\right)-m\left(\frac{t-j}{n}\right)\right\}\right]\right|$$
$$= o_p(n^{-1/2}), \tag{14}$$

in which $\frac{\partial\pi_{0j}}{\partial\alpha_k} = \frac{\partial\pi_j}{\partial\alpha_k}|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}$.

These assumptions above are essentially the same as in Shao and Yang (2017). Assumption (a) is standard for the ARMA process $\{Y_t\}_{t=-\infty}^{\infty}$ to be causal and invertible, Assumption (b) a mild requirement on moments of $\{\epsilon_t\}_{t=-\infty}^{\infty}$. What is of particular interest is Assumption (c) on the trend estimator $\hat{m}(\cdot)$, which is met by an appropriately defined B-spline estimator (Shao and Yang 2017, Theorem 4). In this work, it will be shown that Assumption (c) is also met by the moving average estimator $\hat{m}(t/n)$ in (4) with appropriate chosen moving average lag l_n.

According to Theorem 2 of Shao and Yang (2017), or equations (8.8.2)–(8.8.4) and Theorem 10.8.2 of Brockwell and Davis (1991), Assumptions (a) and (b) ensure that the infeasible MLE $\tilde{\boldsymbol{\alpha}}$ satisfies

$$\sqrt{n}(\tilde{\boldsymbol{\alpha}}-\boldsymbol{\alpha}_0) \xrightarrow{D} N_{p+q}(\mathbf{0},\mathbf{V}), \quad n\to\infty,$$

where $\mathbf{V}$ is the $(p+q)\times(p+q)$ covariance matrix defined as

$$\mathbf{V} = \sigma_0^2\begin{pmatrix}\mathbb{E}(\mathbf{u}_t\mathbf{u}_t^{\mathsf{T}}) & \mathbb{E}(\mathbf{u}_t\mathbf{v}_t^{\mathsf{T}})\\ \mathbb{E}(\mathbf{v}_t\mathbf{u}_t^{\mathsf{T}}) & \mathbb{E}(\mathbf{v}_t\mathbf{v}_t^{\mathsf{T}})\end{pmatrix}^{-1}$$

with $\mathbf{u}_t = (u_t,\ldots,u_{t+1-p})^{\mathsf{T}}$ and $\mathbf{v}_t = (v_t,\ldots,v_{t+1-q})^{\mathsf{T}}$, where the two autoregressive processes $\{u_t\}_{t=-\infty}^{\infty}$ and $\{v_t\}_{t=-\infty}^{\infty}$ are defined by $\Phi(\boldsymbol{\alpha}_0,B)u_t=\epsilon_t$ and $\Theta(\boldsymbol{\alpha}_0,B)v_t=\epsilon_t$, respectively.

Recall also that the Bayesian Information Criterion (BIC) for an ARMA model is defined in Shumway and Stoffer (2017, p. 50) as

$$\mathrm{BIC}(p',q',\tilde{\boldsymbol{\alpha}}) = \log Q_n(\tilde{\boldsymbol{\alpha}};\mathbf{Y}) + \frac{p'+q'}{n}\log n,$$

and the orders $(\tilde{p},\tilde{q})$ selected from data $\mathbf{Y}$ minimise $\mathrm{BIC}(p',q',\tilde{\boldsymbol{\alpha}})$, that is,

$$(\tilde{p},\tilde{q}) = \operatorname*{argmin}_{(p',q')} \mathrm{BIC}(p',q',\tilde{\boldsymbol{\alpha}}). \tag{15}$$

It is established by Hannan (1980) that BIC is consistent in the sense that

$$\lim_{n\to\infty} \mathbb{P}(\tilde{p} = p; \tilde{q} = q) = 1.$$

Proposition 3.1 is a summary of Theorems 1–3 from Shao and Yang (2017) concerning the consistency, oracle efficiency of the MLE, and consistency of the BIC lag selection.

Proposition 3.1: *Under Assumptions* (a)–(c), *as* $n \to \infty$, $\hat{\boldsymbol{\alpha}} \xrightarrow{P} \boldsymbol{\alpha}_0, \hat{\sigma}^2 \xrightarrow{P} \sigma_0^2$, $\hat{\boldsymbol{\alpha}}$ *is oracally efficient in the sense that*

$$\hat{\boldsymbol{\alpha}} - \tilde{\boldsymbol{\alpha}} = o_p(n^{-1/2}), \quad \sqrt{n}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) \stackrel{D}{\Longrightarrow} N_{p+q}(\mathbf{0}, \mathbf{V}). \tag{16}$$

If $(\hat{p}, \hat{q}) = \operatorname{argmin}_{(p',q')} \mathrm{BIC}(p', q', \hat{\boldsymbol{\alpha}})$,

$$\lim_{n\to\infty} \mathbb{P}(\hat{p} = p; \hat{q} = q) = 1. \tag{17}$$

The main contribution of the current work is to establish that the data-driven moving average trend $\hat{m}(\cdot)$ defined in (4) satisfies the above high-level assumption (Assumption (c)) under the following elementary assumptions.

(c1) The trend function $m(\cdot) \in C^2[0, 1]$.
(c2) The moving average lag $l = l_n$ satisfies $n^{1/2} \ll l_n \ll n^{7/8}$ as $n \to \infty$.

Theorem 3.2: *Under Assumptions* (a)–(b), (c1), (c2), *Assumption* (c) *is fulfilled for the local linear estimator* $\hat{m}(\cdot)$ *defined in* (4), *hence Proposition* 3.1 *holds.*

Theorem 3.2, therefore, formally justifies the proposed two-step procedure using the moving average residual sequence to replace the unobserved stationary time series. The modelling procedure for $\mathbf{Y}$ is completely adapted to $\hat{\mathbf{Y}}$, including the selection of autoregressive order p and moving-average order q and estimation of $\boldsymbol{\alpha}_0$ and σ_0^2 according to Proposition 3.1.

It should be noted that our work has relaxed the lower bound on l_n from $n^{1/2} \log n$ in Qiu et al. (2013) to $n^{1/2}$. This new lower bound reflects the new approach of proving Theorem 3.2, which corresponds to the number of spline knots $N \ll n^{1/2}$ in Condition 2 to prove Theorem 4 in Shao and Yang (2017) resulting in the number of observations in each spline subinterval being $n/N \gg n^{1/2}$.

4. Implementation

The two-step procedure is determined by the moving average lag $l = l_n$, which is the data-driven rule-of-thumb (ROT) integer first proposed in Qiu et al. (2013)

$$\hat{l}_{\mathrm{ROT}} = \hat{l}_{n,\mathrm{ROT}} = \left[n^{4/5} (9/2)^{1/5} \{\hat{\gamma}(0)\}^{1/5} \left\{ \int_0^1 \hat{m}''(x)^2 \, \mathrm{d}x \right\}^{-1/5} \right]$$

in which for any $a \in \mathbb{R}$, $[a]$ denotes the largest integer less than or equal to a, $\hat{\gamma}(0) = n^{-1} \sum_{t=1}^{n} \widehat{Y}_t^2$ with $\widehat{Y}_t = X_t - \hat{m}(t/n), \hat{m}(\omega) = \hat{a} + \hat{b}\omega + \hat{c}\omega^2 + \hat{d}\omega^3, \hat{m}''(\omega) = 2\hat{c} + 6\hat{d}\omega,$

where $(\hat{a}, \hat{b}, \hat{c}, \hat{d})$ is the solution of

$$\underset{(a,b,c,d)\in\mathbb{R}^4}{\operatorname{argmin}} \sum_{i=1}^{n} (X_i - a - bi/n - ci^2/n^2 - di^3/n^3)^2.$$

The above is justified as a plug-in substitute of an asymptotically optimal bandwidth according to Fan and Gijbels (1996).

Thus, the moving average lag is computed from the data as

$$\hat{l}_{\mathrm{ROT}} = [n^{4/5}(9/2)^{1/5}\{\hat{\gamma}(0)(4\hat{c}^2 + 12\hat{c}\hat{d} + 12\hat{d}^2)^{-1}\}^{1/5}].$$

Note that this data-driven $\hat{l}_{n,\mathrm{ROT}}$ satisfies the order constraints spelled out in Assumption (c2).

5. Simulation

In this section, simulation experiments are carried out to illustrate the finite-sample behaviour of moving average trend estimators of ARMA coefficients based on the detrending time series.

For sample sizes $n = 100, 200, 400, 1000$, 100 sample paths of ARMA(1, 1) time series are generated. The white noise $\epsilon_t \sim N(0, 1)$ and the trend function is

$$m(\omega) = \sin(2\pi\omega), \quad \omega \in [0, 1t]. \tag{18}$$

The parameters of ARMA(1, 1) process are $(\phi_{01}, \theta_{01}) = (0.8, 0.8), (0.8, -0.6), (0.6, -0.4), (0.2, 0.8), (-0.8, 0.2), (0.4, -0.8), (-0.6, 0.4), (-0.2, -0.2), (-0.6, -0.6)$.

For each pair of (ϕ_{01}, θ_{01}) and sample size n, the sample mean and standard deviation of two-step MLE $(\hat{\phi}_1, \hat{\theta}_1)$ are tabulated in Table 1 to inspect the asymptotic behaviour. One

Table 1. Sample means and standard deviations of ARMA(1, 1) coefficients' estimates.

	Kernel trend estimates			
True coefficients	$n = 100$	$n = 200$	$n = 400$	$n = 1000$
$\phi_{01} = 0.8$	0.6516 ± 0.0914	0.7501 ± 0.0513	0.771 ± 0.037	0.7865 ± 0.0194
$\theta_{01} = 0.8$	0.8348 ± 0.0698	0.8099 ± 0.0455	0.807 ± 0.029	0.7997 ± 0.0259
$\phi_{01} = 0.8$	0.168 ± 0.471	0.447 ± 0.351	0.650 ± 0.127	0.7339 ± 0.0785
$\theta_{01} = -0.6$	-0.060 ± 0.475	-0.284 ± 0.345	-0.459 ± 0.138	-0.5424 ± 0.0948
$\phi_{01} = 0.6$	0.0925 ± 0.4584	0.258 ± 0.323	0.469 ± 0.197	0.523 ± 0.140
$\theta_{01} = -0.4$	0.0484 ± 0.4846	-0.0911 ± 0.3163	-0.272 ± 0.210	-0.327 ± 0.150
$\phi_{01} = 0.2$	0.111 ± 0.107	0.150 ± 0.078	0.1714 ± 0.0634	0.1868 ± 0.0364
$\theta_{01} = 0.8$	0.821 ± 0.075	0.8123 ± 0.0448	0.8100 ± 0.0354	0.8018 ± 0.0218
$\phi_{01} = -0.8$	-0.775 ± 0.105	-0.7840 ± 0.0712	-0.7970 ± 0.0445	-0.793 ± 0.035
$\theta_{01} = 0.2$	0.163 ± 0.168	0.164 ± 0.116	0.1890 ± 0.0702	0.1836 ± 0.0533
$\phi_{01} = 0.4$	0.365 ± 0.160	0.379 ± 0.112	0.3923 ± 0.0777	0.3944 ± 0.0509
$\theta_{01} = -0.8$	-0.7733 ± 0.0956	-0.7834 ± 0.0625	-0.7909 ± 0.0467	-0.7947 ± 0.0313
$\phi_{01} = -0.6$	-0.170 ± 0.504	-0.433 ± 0.336	-0.552 ± 0.176	-0.591 ± 0.108
$\theta_{01} = 0.4$	-0.104 ± 0.526	0.212 ± 0.359	0.337 ± 0.202	0.389 ± 0.124
$\phi_{01} = -0.2$	-0.0962 ± 0.3036	-0.137 ± 0.202	-0.158 ± 0.149	-0.1940 ± 0.0926
$\theta_{01} = -0.2$	-0.357 ± 0.313	-0.289 ± 0.202	-0.255 ± 0.153	-0.2182 ± 0.0949
$\phi_{01} = -0.6$	-0.626 ± 0.097	-0.6243 ± 0.0586	-0.6079 ± 0.0447	-0.612 ± 0.030
$\theta_{01} = -0.6$	-0.4349 ± 0.0778	-0.4814 ± 0.0513	-0.5298 ± 0.0319	-0.5585 ± 0.0259

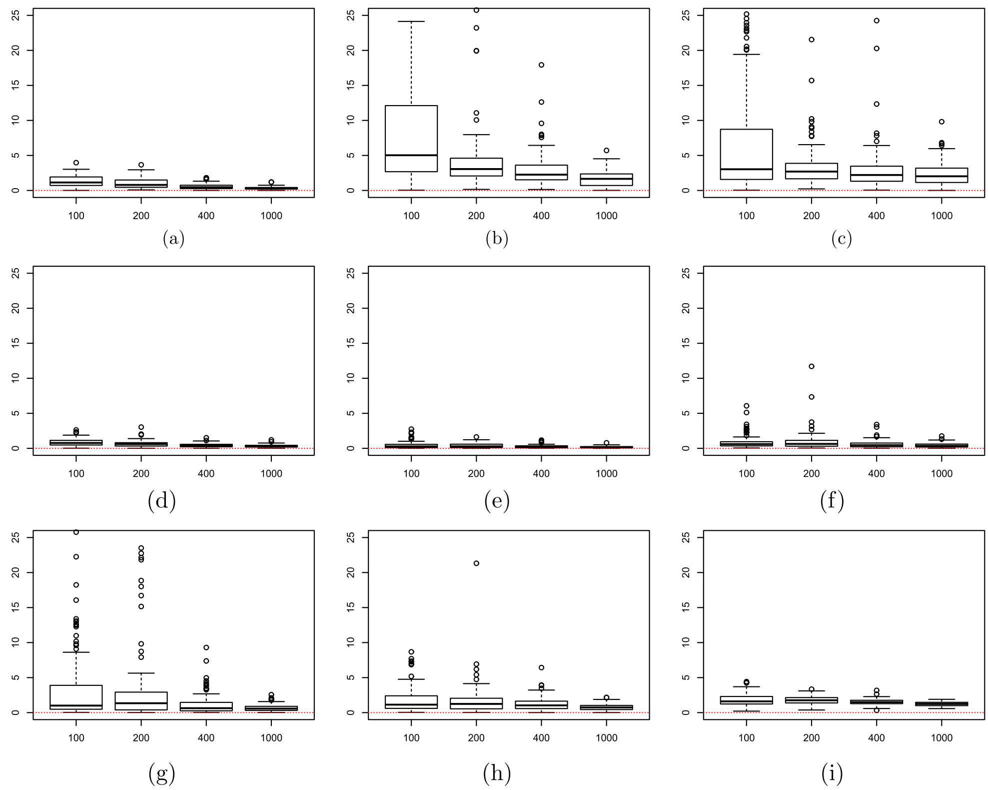

Figure 1. Boxplots of $\sqrt{n}\|\hat{\alpha} - \tilde{\alpha}\|$ for the ARMA(1, 1) process with trend m in Equation (18), $n = 100$, 200, 400, 1000: (a) $(\phi_{01}, \theta_{01}) = (0.8, 0.8)$; (b) $(\phi_{01}, \theta_{01}) = (0.8, -0.6)$; (c) $(\phi_{01}, \theta_{01}) = (0.6, -0.4)$; (d) $(\phi_{01}, \theta_{01}) = (0.2, 0.8)$; (e) $\phi_{01}, \theta_{01}) = (-0.8, 0.2)$; (f) $(\phi_{01}, \theta_{01}) = (0.4, -0.8)$; (g) $(\phi_{01}, \theta_{01}) = (-0.6, 0.4)$; (h) $(\phi_{01}, \theta_{01}) = (-0.2, -0.2)$; (i) $(\phi_{01}, \theta_{01}) = (-0.6, -0.6)$.

clearly sees that in all settings, $(\hat{\phi}_1, \hat{\theta}_1)$ converges in distribution to (ϕ_{01}, θ_{01}) as n increases, confirming the second equation of (16) in Proposition 3.1.

Meanwhile, to compare the two-step MLE $(\hat{\phi}_1, \hat{\theta}_1)$ and the infeasible MLE $(\tilde{\phi}_1, \tilde{\theta}_1)$, boxplots of $\sqrt{n}\|\widehat{\alpha} - \tilde{\alpha}\| = \sqrt{n}\sqrt{(\hat{\phi}_1 - \tilde{\phi}_1)^2 + (\hat{\theta}_1 - \tilde{\theta}_1)^2}$ are placed in Figure 1, each with a horizontal broken line at $y = 0$. These boxplots agree with the first equation of (16) in Proposition 3.1: for each setting the boxplot not only becomes narrower but also closer to 0, as the sample size increases from 100 to 1000.

The percentages of correctly selecting the ARMA(1, 1) models according to the BICs are summarised from the 100 replications. Table 2 includes both the percentages computed from the detrended residuals ('time series with trend') and the simulated ARMA(1, 1) sequences ('time series without trend') with sample size $n = 400$, 1000. We have also conducted simulation experiments with an MA(3) model. The trend function in (18) is used while the candidate MA parameters are $(\theta_{01}, \theta_{02}, \theta_{03}) = (0.2, 0.3, 0.2)$, $(0.2, 0.2, 0.3)$, $(0.3, -0.2, -0.3)$, $(0.2, 0.3, -0.2)$, $(-0.3, 0.2, 0.2)$, $(0.2, -0.2, -0.3)$. Percentages of correct lag selection for MA(3) model are listed in Table 3.

Table 2. Sample percentages of correct ARMA(1, 1) model selection by the BIC.

True coefficients	Time series with trend		Time series without trend	
(ϕ_{01}, θ_{01})	$n = 400$	$n = 1000$	$n = 400$	$n = 1000$
$(0.8, 0.8)$	0.87	0.89	0.82	0.92
$(0.8, 0.6)$	0.85	0.90	0.88	0.97
$(0.8, -0.6)$	0.30	0.86	0.67	0.91
$(0.6, 0.6)$	0.86	0.96	0.87	0.97
$(-0.8, -0.8)$	0.33	0.96	0.86	0.96
$(-0.8, -0.6)$	0.54	0.90	0.88	0.95
$(-0.8, 0.6)$	0.68	0.93	0.72	0.97
$(-0.6, -0.6)$	0.82	0.95	0.94	0.95

Table 3. Sample percentages of correct MA(3) model selection by the BIC.

True coefficients	Time series with trend		Time series without trend	
$(\theta_{01}, \theta_{02}, \theta_{03})$	$n = 400$	$n = 1000$	$n = 400$	$n = 1000$
$(0.2, 0.3, 0.2)$	0.54	0.76	0.86	0.94
$(0.2, 0.2, 0.3)$	0.89	0.92	0.90	1.00
$(0.3, -0.2, -0.3)$	0.81	0.84	0.96	0.96
$(0.2, 0.3, -0.2)$	0.77	0.78	0.89	0.94
$(-0.3, 0.2, 0.2)$	0.73	0.77	0.92	0.96
$(0.2, -0.2, -0.3)$	0.83	0.90	0.96	0.98

As one can see from Tables 2 and 3, for both the ARMA(1, 1) and the MA(3) models, the correct selection percentages increase with n, and come quite close to 1 at $n = 1000$. These results confirm again the last equation (Equation (17)) of Proposition 3.1.

6. Conclusions

This work illustrates that moving average trend with appropriately chosen lag can be used to remove a slowly varying trend in ARMA time series in order to identify the ARMA lags and compute MLE of ARMA coefficients, without loss of efficiency. This conclusion is justified with rigorous theory and confirmed in numerical experiments. It brings closure to a long-existing gap in time series literature about the legitimacy of the moving average detrending method taught in most textbooks.

Acknowledgements

Constructive and insightful comments from two Reviewers and the Associate Editor have led to substantial improvement of the paper.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work is part of the first author's doctoral dissertation under the supervision of the last author and has been supported exclusively by National Natural Science Foundation of China awards [NSFC12171269 and NSFC11771240].

ORCID

Lijian Yang http://orcid.org/0000-0003-3894-873X

References

Brockwell, P.J., and Davis, R.A. (1991), *Time Series: Theory and Methods*, New York: Springer-Verlag.
Fan, J., and Gijbels, I. (1996), *Local Polynomial Modelling and its Applications: Monographs on Statistics and Applied Probability* (Vol. 66), London: Chapman and Hall.
Fan, J., and Yao, Q. (2003), *Nonlinear Time Series: Nonparametric and Parametric Methods*, New York: Springer-Verlag.
Hannan, E.J. (1980), 'The Estimation of the Order of an ARMA Process', *The Annals of Statistics*, 8, 1071–1081.
Pierce, D.A. (1971), 'Least Squares Estimation in the Regression Model with Autoregressive-moving Average Errors', *Biometrika*, 58, 299–312.
Qiu, D., Shao, Q., and Yang, L. (2013), 'Efficient Inference for Autoregressive Coefficients in the Presence of Trends', *Journal of Multivariate Analysis*, 114, 40–53.
Shao, Q., and Yang, L. (2017), 'Oracally Efficient Estimation and Consistent Model Selection for Auto-regressive Moving Average Time Series with Trend', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79, 507–524.
Shumway, R.H., and Stoffer, D.S. (2017), *Time Series Analysis and Its Applications with R Examples*, Cham: Springer.
Yao, Q., and Brockwell, P.J. (2006), 'Gaussian Maximum Likelihood Estimation for ARMA Models. I. Time Series', *Journal of Time Series Analysis*, 27, 857–875.

Appendix

Theorem 3.2 is proved by verification of Equations (12)–(14) in Assumption (c). To begin with, the estimator $\hat{m}(\cdot)$ is decomposed as

$$\hat{m}(t/n) = \tilde{m}(t/n) + \tilde{Y}(t/n) \tag{A1}$$

with noise term $\tilde{Y}(t/n)$ and signal term $\tilde{m}(t/n)$ defined as

$$\tilde{m}(t/n) = \begin{cases} (2l+1)^{-1} \sum_{i=t-l}^{t+l} m(i/n), & l+1 \le t \le n-l, \\ \sum_{i=1}^{t+l} \{N_{1t}^{-1} - N_{2t}^{-1}(i-t)\} m(i/n), & 1 \le t \le l, \\ \sum_{i=t-l}^{n} \{N_{3t}^{-1} - N_{4t}^{-1}(i-t)\} m(i/n), & n-l+1 \le t \le n, \end{cases} \tag{A2}$$

$$\tilde{Y}(t/n) = \begin{cases} (2l+1)^{-1} \sum_{i=t-l}^{t+l} Y_i, & l+1 \le t \le n-l, \\ \sum_{i=1}^{t+l} \left\{ N_{1t}^{-1} - N_{2t}^{-1}(i-t) \right\} Y_i, & 1 \le t \le l, \\ \sum_{i=t-l}^{n} \left\{ N_{3t}^{-1} - N_{4t}^{-1}(i-t) \right\} Y_i, & n-l+1 \le t \le n. \end{cases} \tag{A3}$$

A.1 Proof of (12) in Assumption (c)

Fan and Gijbels (1996) proclaimed that one major advantage of local linear estimators over Nadaraya–Watson estimators is that it automatically corrects boundary bias. In our context, it results in the following basic lemma on estimation bias.

Lemma A.1: *Under Assumptions* (c1)–(c2), *the local linear estimator in* (7) *satisfies* $\max_{1\le t\le n} |m(t/n) - \tilde{m}(t/n)| = \mathcal{O}(h^2)$.

Proof: Matrix algebra leads to

$$\tilde{m}(t/n) - m(t/n) = e_0^{\mathsf{T}} \begin{pmatrix} \sum_{i=k_1}^{k_2} 1 & \sum_{i=k_1}^{k_2} (i-t) \\ \sum_{i=k_1}^{k_2} (i-t) & \sum_{i=k_1}^{k_2} (i-t)^2 \end{pmatrix}^{-1} \times \begin{pmatrix} \sum_{i=k_1}^{k_2} m(i/n) \\ \sum_{i=k_1}^{k_2} (i-t) m(i/n) \end{pmatrix} - m(t/n)$$

$$= e_0^{\mathsf{T}} \begin{pmatrix} \sum_{i=k_1}^{k_2} 1 & \sum_{i=k_1}^{k_2} (i-t) \\ \sum_{i=k_1}^{k_2} (i-t) & \sum_{i=k_1}^{k_2} (i-t)^2 \end{pmatrix}^{-1} \times \begin{pmatrix} \sum_{i=k_1}^{k_2} \{m(i/n) - m(t/n) - m'(t/n)(i-t)/n\} \\ \sum_{i=k_1}^{k_2} (i-t)\{m(i/n) - m(t/n) - m'(t/n)(i-t)/n\} \end{pmatrix}$$

in which

$$(k_1, k_2) = \begin{cases} (1, t+l) & 1 \le t \le l, \\ (t-l, t+l) & l+1 \le t \le n-l, \\ (t-l, n) & n-l+1 \le t \le n, \end{cases}$$

see (A.2) in Qiu et al. (2013). Hence $\tilde{m}(t/n) - m(t/n)$ is

$$\begin{cases} (2l+1)^{-1}\sum_{i=t-l}^{t+l}\{m(i/n)-m(t/n)-m'(t/n)(i-t)/n\}, & l+1\le t\le n-l, \\ \sum_{i=1}^{t+l}\{N_{1t}^{-1}-N_{2t}^{-1}(i-t)\}\{m(i/n)-m(t/n)-m'(t/n)(i-t)/n\}, & 1\le t\le l, \\ \sum_{i=t-l}^{n}\{N_{3t}^{-1}-N_{4t}^{-1}(i-t)\}\{m(i/n)-m(t/n)-m'(t/n)(i-t)/n\}, & n-l+1\le t\le n. \end{cases} \tag{A4}$$

When $|i-t|\le l = nh = o(n)$, one has

$$\max_{1\le i,t\le n, |i-t|\le l} |m(i/n)-m(t/n)-m'(t/n)(i-t)/n| \le C\|m''\|_\infty (l/n)^2 \le C\|m''\|_\infty h^2.$$

Note from the definition of N_{1t}^{-1}, N_{2t}^{-1}, N_{3t}^{-1} and N_{4t}^{-1} in (5) and (6) that there exists $C_N > 0$ such that

$$\begin{aligned} &0 < N_{1t}^{-1} \le C_N/l, \quad 0 < N_{2t}^{-1} \le C_N/l^2, \quad 1\le t\le l, \\ &0 < N_{3t}^{-1} \le C_N/l, \quad 0 < N_{4t}^{-1} \le C_N/l^2, \quad n-l+1\le t\le n, \end{aligned} \tag{A5}$$

one concludes from (A4) that $\max_{1\le t\le n} |m(t/n)-\tilde{m}(t/n)| = \mathcal{O}(h^2)$. ■

The next lemma concerns the estimation variance in $\tilde{Y}(t/n)$.

Lemma A.2: *Under Assumptions* (a)–(b), (c2), *as* $n\to\infty$,

$$\max_{1\le t\le n} \mathbb{E}\tilde{Y}^2(t/n) = \mathcal{O}(1/l) = \mathcal{O}(1/nh).$$

Proof: For $l+1\le t\le n-l$, $\tilde{Y}(t/n) = (2l+1)^{-1}\sum_{i=t-l}^{t+l} Y_i$, hence

$$\begin{aligned} \mathbb{E}\tilde{Y}^2(t/n) &= (2l+1)^{-2}\sum_{i,i'=t-l}^{t+l} \mathbb{E}Y_iY_{i'} = (2l+1)^{-2}\sum_{i,i'=t-l}^{t+l}\gamma(i-i') \\ &= (2l+1)^{-2}\left\{(2l+1)\gamma(0) + 2\sum_{j=1}^{2l}(2l+1-j)\gamma(j)\right\} \\ &\le (2l+1)^{-1}\left[|\gamma(0)| + 2\frac{2l}{2l+1}|\gamma(1)| + \cdots + 2\frac{1}{2l+1}|\gamma(2)|\right] \\ &\le (2l+1)^{-1}\left[|\gamma(0)| + 2|\gamma(1)| + \cdots + 2|\gamma(2)|\right] \le (2l+1)^{-1}\sum_{j=-\infty}^{\infty}|\gamma(j)|, \end{aligned}$$

in which $\gamma(j) \equiv \mathbb{E}(Y_tY_{t+j}), j = 0, \pm1, \pm2, \ldots$, denotes the autocovariance function of $\{Y_t\}_{t=-\infty}^{\infty}$. Since $\sum_{j=-\infty}^{\infty}|\gamma(j)| < \infty$, we have $\max_{l+1\le t\le n-l}\mathbb{E}\tilde{Y}^2(t/n) \le C/l$ for some constant $C > 0$.

For $1\le t\le l$, with N_{1t}^{-1}, N_{2t}^{-1} defined in (5), $\tilde{Y}(t/n) = N_{1t}^{-1}\sum_{i=1}^{t+l} Y_i - N_{2t}^{-1}\sum_{i=1}^{t+l}(i-t)Y_i$,

$$\begin{aligned} \mathbb{E}\tilde{Y}^2(t/n) &= \mathbb{E}\left(N_{1t}^{-1}\sum_{i=1}^{t+l} Y_i - N_{2t}^{-1}\sum_{i=1}^{t+l}(i-t)Y_i\right)^2 \\ &\le 2N_{1t}^{-2}\mathbb{E}\left(\sum_{i=1}^{t+l}Y_i\right)^2 + 2N_{2t}^{-2}\mathbb{E}\left(\sum_{i=1}^{t+l}(i-t)Y_i\right)^2 \end{aligned}$$

$$\leq 2N_{1t}^{-2}\sum_{i,i'=1}^{t+l}\mathbb{E}Y_iY_{i'} + 2N_{2t}^{-2}\sum_{i,i'=1}^{t+l}\mathbb{E}(i-t)(i'-t)Y_iY_{i'},$$

where

$$\begin{aligned}\left|\sum_{i,i'=1}^{t+l}\mathbb{E}Y_iY_{i'}\right| &= \left|\sum_{i,i'=1}^{t+l}\gamma(i-i')\right| = |(t+l)\gamma(0) + 2\sum_{j=1}^{t+l-1}(t+l-j)\gamma(j)| \\ &\leq (t+l)[|\gamma(0)| + 2|\gamma(1)| + \cdots + 2|\gamma(t+l-1)|] \\ &\leq 2l\sum_{j=-\infty}^{\infty}|\gamma(j)|, \end{aligned} \tag{A6}$$

$$\begin{aligned}\left|\sum_{i,i'=1}^{t+l}\mathbb{E}(i-t)(i'-t)Y_iY_{i'}\right| &= \left|\sum_{i,i'=1}^{t+l}(i-t)(i'-t)\gamma(i-i')\right| \\ &\leq l^2\left|\sum_{j=1-t-l}^{t+l-1}(t+l-j)\gamma(j)\right| \leq 2l^3\sum_{j=-\infty}^{\infty}|\gamma(j)|. \end{aligned} \tag{A7}$$

Hence (A6) and (A7) provide that

$$\max_{1\leq t\leq l}\mathbb{E}\tilde{Y}^2(t/n) \leq \max_{1\leq t\leq l}(4lN_{1t}^{-2} + 4l^3N_{2t}^{-2})\sum_{j=-\infty}^{\infty}|\gamma(j)| \leq C/l$$

for some constant $C > 0$.

Similarly, there exists a constant $C > 0$, such that $\max_{n-l+1\leq t\leq n}\mathbb{E}\tilde{Y}^2(t/n) \leq C/l$. The proof is complete. ■

To complete the proof of (12), note that

$$\begin{aligned}\{m(t/n) - \hat{m}(t/n)\}^2 &= \{m(t/n) - \tilde{m}(t/n) - \tilde{Y}(t/n)\}^2 \\ &\leq 2\{m(t/n) - \tilde{m}(t/n)\}^2 + 2\tilde{Y}^2(t/n),\end{aligned}$$

hence Lemmas A.1 and A.2 provide that

$$\begin{aligned}\max_{1\leq t\leq n}\mathbb{E}\{m(t/n) - \hat{m}(t/n)\}^2 &\leq 2\max_{1\leq t\leq n}\{m(t/n) - \tilde{m}(t/n)\}^2 + 2\max_{1\leq t\leq n}\mathbb{E}\tilde{Y}^2(t/n) \\ &\leq \mathcal{O}(h^4 + 1/l) = o(n^{-1/2}).\end{aligned}$$

The last inequality following from Assumption (c1) that $n^{1/2} \ll l \ll n^{7/8}$ implies that $h^4 \ll n^{-1/2}$ and $1/l \ll n^{-1/2}$.

A.2 Proof of (13), (14) in Assumption (c)

We begin with a technical lemma on mixed moments.

Lemma A.3: *Under Assumptions* (a) and (b), *there exist constants* $C_M > 0$, $\rho_M \in (0, 1)$ *such that for any set of indices* i_1, i_2, i_3, i_4 *with* $1 \leq i_1 \leq i_2 \leq i_3 \leq i_4 \leq n$

$$|\mathbb{E}Y_{i_1}Y_{i_2}Y_{i_3}Y_{i_4}| \leq C_M\rho_M^{\max(i_2-i_1, i_4-i_3)}. \tag{A8}$$

Proof: The terms in (A8) are bounded in four cases. Note that Assumptions (a) and (b) entail that for any $0 < r \leq 6$, the moment $E|Y_t|^r$ is finite and constant in t.

Case 1: $i_1 = i_2, i_3 = i_4$. Applying the Cauchy–Schwarz inequality,

$$|\mathbb{E}Y_{i_1}Y_{i_2}Y_{i_3}Y_{i_4}| \le \mathbb{E}|Y_{i_1}^2Y_{i_3}^2| \le (\mathbb{E}Y_{i_1}^4\mathbb{E}Y_{i_3}^4)^{1/2} < \infty.$$

Thus, there exist constant $C > 0$, such that $|\mathbb{E}Y_{i_1}Y_{i_2}Y_{i_3}Y_{i_4}| \le C$.

Case 2: $i_1 = i_2, i_3 \ne i_4$. Applying Proposition 2.5 of Fan and Yao (2003),

$$\begin{aligned}|\mathbb{E}Y_{i_1}^2Y_{i_3}Y_{i_4}| &= |\mathbb{E}Y_{i_1}^2Y_{i_3}Y_{i_4} - \mathbb{E}Y_{i_1}^2Y_{i_3}\mathbb{E}Y_{i_4}| \\ &\le 8\{\alpha(i_4 - i_3)\}^{1/6}\{\mathbb{E}Y_{i_1}^6\mathbb{E}Y_{i_3}^3\}^{1/3}\{\mathbb{E}|Y_{i_4}|^6\}^{1/6} \le C\{\alpha(i_4 - i_3)\}^{1/6}\end{aligned}$$

for some constant $C > 0$.

Case 3: $i_1 \ne i_2, i_3 = i_4$. Applying again Proposition 2.5 of Fan and Yao (2003),

$$\begin{aligned}|\mathbb{E}Y_{i_1}Y_{i_2}Y_{i_4}^2| &= |\mathbb{E}Y_{i_1}Y_{i_2}Y_{i_4}^2 - \mathbb{E}Y_{i_1}\mathbb{E}Y_{i_2}Y_{i_4}^2| \\ &\le 8\{\alpha(i_2 - i_1)\}^{1/6}\{\mathbb{E}|Y_{i_1}|^6\}^{1/6}\{\mathbb{E}Y_{i_2}^3\mathbb{E}Y_{i_4}^6\}^{1/3} \le C\{\alpha(i_2 - i_1)\}^{1/6}\end{aligned}$$

for some constant $C > 0$.

Case 4: $i_1 \ne i_2, i_3 \ne i_4$. Note that

$$\begin{aligned}|\mathbb{E}Y_{i_1}Y_{i_2}Y_{i_3}Y_{i_4}| &= |\mathbb{E}Y_{i_1}Y_{i_2}Y_{i_3}Y_{i_4} - \mathbb{E}Y_{i_1}\mathbb{E}Y_{i_2}Y_{i_3}Y_{i_4}| \\ &\le 8\{\alpha(i_2 - i_1)\}^{1/4} left\{\mathbb{E}|Y_{i_1}|^4\}^{1/4}\{\mathbb{E}|Y_{i_2}Y_{i_3}Y_{i_4}|^2\}^{1/2} \\ &\le C\{\alpha(i_2 - i_1)\}^{1/4}.\end{aligned}$$

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

$$\begin{aligned}|\mathbb{E}Y_{i_1}Y_{i_2}Y_{i_3}Y_{i_4}| &\le C\min\{\{\alpha(i_2 - i_1)\}^{1/4}, \{\alpha(i_4 - i_3)\}^{1/4}\} \\ &\le C\{\alpha(\min(i_2 - i_1, i_4 - i_3))\}^{1/4}.\end{aligned}$$

Since $\alpha(j) \le C_\alpha\rho_\alpha^j$ for some constants $C_\alpha > 0, \rho_\alpha \in (0, 1)$ according to Assumptions (a), (b), (A8) holds for Cases 1–4. ■

Clearly (13) and (14) follows from the next lemma, where $1 \le k \le p + q$ is fixed.

Lemma A.4: *Under Assumptions* (a)–(b), (c1), (c2), *as* $n \to \infty$,

$$T_{1n} = n^{-1}\left|\sum_{t=p+1}^{n}\left(\sum_{j=0}^{n}\frac{\partial\pi_{0j}}{\partial\alpha_k}Y_{t-j}\right)\left[\sum_{j=0}^{t-1}\pi_{0j}(m - \tilde{m})\left(\frac{t-j}{n}\right)\right]\right| = o_p(n^{-1/2}), \quad \text{(A9)}$$

$$T_{2n} = n^{-1}\left|\sum_{t=p+1}^{n}\left(\sum_{j=0}^{t-1}\pi_{0j}Y_{t-j}\right)\left[\sum_{j=0}^{t-1}\frac{\partial\pi_{0j}}{\partial\alpha_k}(m - \tilde{m})\left(\frac{t-j}{n}\right)\right]\right| = o_p(n^{-1/2}),$$

$$T_{3n} = n^{-1}\left|\sum_{t=p+1}^{n}\left(\sum_{j=0}^{t-1}\pi_{0j}Y_{t-j}\right)\left(\sum_{j=0}^{t-1}\frac{\partial\pi_{0j}}{\partial\alpha_k}\tilde{Y}\left(\frac{t-j}{n}\right)\right)\right| = o_p(n^{-1/2}),$$

$$T_{4n} = n^{-1}\left[\sum_{t=p+1}^{n}\left(\sum_{j=0}^{t-1}\frac{\partial\pi_{0j}}{\partial\alpha_k}Y_{t-j}\right)\left(\sum_{j=0}^{t-1}\pi_{0j}\tilde{Y}\left(\frac{t-j}{n}\right)\right)\right] = o_p(n^{-1/2}). \quad \text{(A10)}$$

To prove the above lemma, define $\Pi_j = \sup_{\boldsymbol{\alpha}\in\Xi}|\pi_j(\boldsymbol{\alpha})|, 0 \le j < \infty$. Note that under Assumption (a), π_j defined in (10) satisfies equation (3.3.6) of Brockwell and Davis (1991), which entails that there exist constants C_π and $0 < \rho_\pi < 1$ such that

$$\Pi_j \le C_\pi \rho_\pi^j, 0 \le j < \infty. \tag{A11}$$

Proofs for T_{1n}, T_{3n} in (A9) and (A10) are presented in the next two subsections, proofs for T_{2n}, T_{4n} are similar to that of T_{1n}, T_{3n} and thus omitted.

A.3 Proof of (A9)

Inequality (27) of Yao and Brockwell (2006) entails that $\{\partial\pi_{0j}/\partial\alpha_k\}_{j=0}^{n-1}$ is (C_s, s)-exponentially bounded for some $C_s > 0, s \in (0,1)$, while $\{\pi_{0j}\}_{j=0}^{n-1}$ is (C_π, ρ_π)-exponentially bounded by inequality (15) of Shao and Yang (2017). Now equation (3.3.10) of Brockwell and Davis (1991) ensures that $\mathbf{Y} = (Y_1, \ldots, Y_n)^\top$ is (C_y, ρ_y)-exponentially correlated for some $C_y > 0$, $\rho_y \in (0,1)$. Lemma 3 of Shao and Yang (2017) then implies that the sequence $\{\sum_{j=0}^{t-1}(\partial\pi_{0j}/\partial\alpha_k)Y_{t-j}\}_{t=p+1}^n$ is (C_ζ, ρ_ζ)-exponentially correlated for some $C_\zeta > 0$, $\rho_\zeta \in (0,1)$.

Next, Lemma 1 of Shao and Yang (2017) provides that

$$\begin{aligned}
&\max_{p+1\le t\le n}\left|\sum_{j=0}^{t-1}\pi_{0j}\left\{m\left(\frac{t-j}{n}\right)-\tilde{m}\left(\frac{t-j}{n}\right)\right\}\right| \\
&\le C_\pi(1-\rho_\pi)^{-1}\max_{1\le t\le n}\left|m\left(\frac{t-j}{n}\right)-\tilde{m}\left(\frac{t-j}{n}\right)\right| = \mathcal{O}\left(h^2\right),
\end{aligned}$$

where the last inequality uses Lemma A.1.

Applying Lemma 2 of Shao and Yang (2017) to the (C_ζ, ρ_ζ)-exponentially correlated $\{\sum_{j=0}^{t-1}(\partial\pi_{0j}/\partial\alpha_k)Y_{t-j}\}_{t=p+1}^n$ and $\{\sum_{j=0}^{t-1}\pi_{0j}\{m(\frac{t-j}{n})-\tilde{m}(\frac{t-j}{n})\}\}_{t=p+1}^n$, a deterministic sequence with uniform bound $\mathcal{O}(h^2)$, one obtains that

$$\begin{aligned}
T_{1n} &= n^{-1}\left|\sum_{t=p+1}^{n}\left(\sum_{j=0}^{t-1}\frac{\partial\pi_{0j}}{\partial\phi_k}Y_{t-j}\right)\left[\sum_{j=0}^{t-1}\pi_{0j}\left\{m\left(\frac{t-j}{n}\right)-\tilde{m}\left(\frac{t-j}{n}\right)\right\}\right]\right| \\
&= \mathcal{O}_p\left(n^{-1}\times n^{1/2}h^2\right) = \mathcal{O}_p\left(n^{-1/2}h^2\right) = o_p(n^{-1/2}).
\end{aligned}$$

Thus the proof of (A9) for term T_{1n} is completed.

A.4 Proof of (A10)

To prove the $o_p(n^{-1/2})$ bound for T_{3n}, note that

$$\begin{aligned}
\mathbb{E}T_{3n}^2 &= n^{-2}\mathbb{E}\left\{\sum_{t=p+1}^{n}\left(\sum_{j=0}^{t-1}\pi_{0j}Y_{t-j}\right)\left(\sum_{j=0}^{t-1}\frac{\partial\pi_{0j}}{\partial\alpha_k}\tilde{Y}\left(\frac{t-j}{n}\right)\right)\right\}^2 \\
&= n^{-2}\mathbb{E}\sum_{t_1,t_2=p+1}^{n}\sum_{j_1,j_1'=0}^{t_1-1}\sum_{j_2,j_2'=0}^{t_2-1}\pi_{0j_1}\frac{\partial\pi_{0j_1'}}{\partial\alpha_k}\pi_{0j_2}\frac{\partial\pi_{0j_2'}}{\partial\alpha_k}Y_{t_1-j_1}\tilde{Y}\left(\frac{t_1-j_1'}{n}\right)Y_{t_2-j_2}\tilde{Y}\left(\frac{t_2-j_2'}{n}\right) \\
&= n^{-2}\sum_{t_1,t_2=l+1}^{n-l}\sum_{j_1,j_1'=0}^{t_1-1}\sum_{j_2,j_2'=0}^{t_2-1}\pi_{0j_1}\frac{\partial\pi_{0j_1'}}{\partial\alpha_k}\pi_{0j_2}\frac{\partial\pi_{0j_2'}}{\partial\alpha_k}\mathbb{E}Y_{t_1-j_1}\tilde{Y}\left(\frac{t_1-j_1'}{n}\right)Y_{t_2-j_2}\tilde{Y}\left(\frac{t_2-j_2'}{n}\right) \\
&\quad + n^{-2}\sum_{t_1=l+1}^{n-l}\sum_{t_2=n-l+1}^{n}\sum_{j_1,j_1'=0}^{t_1-1}\sum_{j_2,j_2'=0}^{t_2-1}\pi_{0j_1}\frac{\partial\pi_{0j_1'}}{\partial\alpha_k}\pi_{0j_2}\frac{\partial\pi_{0j_2'}}{\partial\alpha_k}\mathbb{E}Y_{t_1-j_1}
\end{aligned}$$

$$\times \tilde{Y}\left(\frac{t_1-j_1'}{n}\right) Y_{t_2-j_2} \tilde{Y}\left(\frac{t_2-j_2'}{n}\right)$$

$$+ n^{-2} \sum_{t_1=l+1}^{n-l} \sum_{t_2=p+1}^{l} \sum_{j_1,j_1'=0}^{t_1-1} \sum_{j_2,j_2'=0}^{t_2-1} \pi_{0j_1} \frac{\partial \pi_{0j_1'}}{\partial \alpha_k} \pi_{0j_2} \frac{\partial \pi_{0j_2'}}{\partial \alpha_k} \mathbb{E} Y_{t_1-j_1}$$

$$\times \tilde{Y}\left(\frac{t_1-j_1'}{n}\right) Y_{t_2-j_2} \tilde{Y}\left(\frac{t_2-j_2'}{n}\right)$$

$$+ n^{-2} \sum_{t_1=n-l+1}^{n} \sum_{t_2=l+1}^{n-l} \sum_{j_1,j_1'=0}^{t_1-1} \sum_{j_2,j_2'=0}^{t_2-1} \pi_{0j_1} \frac{\partial \pi_{0j_1'}}{\partial \alpha_k} \pi_{0j_2} \frac{\partial \pi_{0j_2'}}{\partial \alpha_k} \mathbb{E} Y_{t_1-j_1}$$

$$\times \tilde{Y}\left(\frac{t_1-j_1'}{n}\right) Y_{t_2-j_2} \tilde{Y}\left(\frac{t_2-j_2'}{n}\right)$$

$$+ n^{-2} \sum_{t_1=p+1}^{l} \sum_{t_2=l+1}^{n-l} \sum_{j_1,j_1'=0}^{t_1-1} \sum_{j_2,j_2'=0}^{t_2-1} \pi_{0j_1} \frac{\partial \pi_{0j_1'}}{\partial \alpha_k} \pi_{0j_2} \frac{\partial \pi_{0j_2'}}{\partial \alpha_k} \mathbb{E} Y_{t_1-j_1}$$

$$\times \tilde{Y}\left(\frac{t_1-j_1'}{n}\right) Y_{t_2-j_2} \tilde{Y}\left(\frac{t_2-j_2'}{n}\right)$$

$$+ n^{-2} \sum_{t_1,t_2=p+1}^{l} \sum_{j_1,j_1'=0}^{t_1-1} \sum_{j_2,j_2'=0}^{t_2-1} \pi_{0j_1} \frac{\partial \pi_{0j_1'}}{\partial \alpha_k} \pi_{0j_2} \frac{\partial \pi_{0j_2'}}{\partial \alpha_k} \mathbb{E} Y_{t_1-j_1} \tilde{Y}\left(\frac{t_1-j_1'}{n}\right) Y_{t_2-j_2} \tilde{Y}\left(\frac{t_2-j_2'}{n}\right)$$

$$+ n^{-2} \sum_{t_1=p+1}^{l} \sum_{t_2=n-l+1}^{n} \sum_{j_1,j_1'=0}^{t_1-1} \sum_{j_2,j_2'=0}^{t_2-1} \pi_{0j_1} \frac{\partial \pi_{0j_1'}}{\partial \alpha_k} \pi_{0j_2} \frac{\partial \pi_{0j_2'}}{\partial \alpha_k} \mathbb{E} Y_{t_1-j_1}$$

$$\times \tilde{Y}\left(\frac{t_1-j_1'}{n}\right) Y_{t_2-j_2} \tilde{Y}\left(\frac{t_2-j_2'}{n}\right)$$

$$+ n^{-2} \sum_{t_1=n-l+1}^{n} \sum_{t_2=p+1}^{l} \sum_{j_1,j_1'=0}^{t_1-1} \sum_{j_2,j_2'=0}^{t_2-1} \pi_{0j_1} \frac{\partial \pi_{0j_1'}}{\partial \alpha_k} \pi_{0j_2} \frac{\partial \pi_{0j_2'}}{\partial \alpha_k} \mathbb{E} Y_{t_1-j_1}$$

$$\times \tilde{Y}\left(\frac{t_1-j_1'}{n}\right) Y_{t_2-j_2} \tilde{Y}\left(\frac{t_2-j_2'}{n}\right)$$

$$+ n^{-2} \sum_{t_1,t_2=n-l+1}^{n} \sum_{j_1,j_1'=0}^{t_1-1} \sum_{j_2,j_2'=0}^{t_2-1} \pi_{0j_1} \frac{\partial \pi_{0j_1'}}{\partial \alpha_k} \pi_{0j_2} \frac{\partial \pi_{0j_2'}}{\partial \alpha_k} \mathbb{E} Y_{t_1-j_1} \tilde{Y}\left(\frac{t_1-j_1'}{n}\right) Y_{t_2-j_2} \tilde{Y}\left(\frac{t_2-j_2'}{n}\right).$$

One needs to show that each of the nine terms in the above sum is of order $o(n^{-1})$. To save space, we illustrate only with the first term

$$n^{-2} \sum_{t_1,t_2=l+1}^{n-l} \sum_{j_1,j_1'=0}^{t_1-1} \sum_{j_2,j_2'=0}^{t_2-1} \pi_{0j_1} \frac{\partial \pi_{0j_1'}}{\partial \alpha_k} \pi_{0j_2} \frac{\partial \pi_{0j_2'}}{\partial \alpha_k} \mathbb{E} Y_{t_1-j_1} \tilde{Y}\left(\frac{t_1-j_1'}{n}\right) Y_{t_2-j_2} \tilde{Y}\left(\frac{t_2-j_2'}{n}\right)$$

$$= n^{-2} \sum_{t_1,t_2=l+1}^{n-l} \sum_{j_1=0}^{t_1-1} \sum_{j_2=0}^{t_2-1} \sum_{j_1'=0}^{t_1-l-1} \sum_{j_2'=0}^{t_2-l-1} \pi_{0j_1} \frac{\partial \pi_{0j_1'}}{\partial \alpha_k} \pi_{0j_2} \frac{\partial \pi_{0j_2'}}{\partial \alpha_k} \mathbb{E} Y_{t_1-j_1}$$

$$
\times \sum_{i_1=t_1-j_1'-l}^{t_1-j_1'+l} \frac{Y_{i_1}}{2l+1} Y_{t_2-j_2} \sum_{i_2=t_2-j_2'-l}^{t_2-j_2'+l} \frac{Y_{i_2}}{2l+1}
$$

$$
+ n^{-2} \sum_{t_1,t_2=l+1}^{n-l} \sum_{j_1=0}^{t_1-1} \sum_{j_2=0}^{t_2-1} \sum_{j_1'=0}^{t_1-l-1} \sum_{j_2'=t_2-l}^{t_2-1} \pi_{0j_1} \frac{\partial \pi_{0j_1'}}{\partial \alpha_k} \pi_{0j_2} \frac{\partial \pi_{0j_2'}}{\partial \alpha_k} \mathbb{E} Y_{t_1-j_1} \sum_{i_1=t_1-j_1'-l}^{t_1-j_1'+l} \frac{Y_{i_1}}{2l+1}
$$

$$
\times Y_{t_2-j_2} \left\{ N_{1,t_2-j_2'}^{-1} \sum_{i_2=1}^{t_2-j_2'+l} Y_{i_2} - N_{2,t_2-j_2'}^{-1} \sum_{i_2=1}^{t_2-j_2'+l} (i_2 - t_2 + j_2') Y_{i_2} \right\}
$$

$$
+ n^{-2} \sum_{t_1,t_2=l+1}^{n-l} \sum_{j_1=0}^{t_1-1} \sum_{j_2=0}^{t_2-1} \sum_{j_1'=t_1-l}^{t_1-1} \sum_{j_2'=0}^{t_2-l-1} \pi_{0j_1} \frac{\partial \pi_{0j_1'}}{\partial \alpha_k} \pi_{0j_2} \frac{\partial \pi_{0j_2'}}{\partial \alpha_k} \mathbb{E} Y_{t_2-j_2} \sum_{i_2=t_2-j_2'-l}^{t_2-j_2'+l} \frac{Y_{i_2}}{2l+1}
$$

$$
\times Y_{t_1-j_1} \left\{ N_{1,t_1-j_1'}^{-1} \sum_{i_1=1}^{t_1-j_1'+l} Y_{i_1} - N_{2,t_1-j_1'}^{-1} \sum_{i_1=1}^{t_1-j_1'+l} (i_1 - t_1 + j_1') Y_{i_1} \right\}
$$

$$
+ n^{-2} \sum_{t_1,t_2=l+1}^{n-l} \sum_{j_1=0}^{t_1-1} \sum_{j_2=0}^{t_2-1} \sum_{j_1'=t_1-l}^{t_1-1} \sum_{j_2'=t_2-l}^{t_2-1} \pi_{0j_1} \frac{\partial \pi_{0j_1'}}{\partial \alpha_k} \pi_{0j_2} \frac{\partial \pi_{0j_2'}}{\partial \alpha_k} \mathbb{E} Y_{t_1-j_1}
$$

$$
\times \left\{ N_{1,t_1-j_1'}^{-1} \sum_{i_1=1}^{t_1-j_1'+l} Y_{i_1} - N_{2,t_1-j_1'}^{-1} \sum_{i_1=1}^{t_1-j_1'+l} (i_1 - t_1 + j_1') Y_{i_1} \right\}
$$

$$
\times Y_{t_2-j_2} \left\{ N_{1,t_2-j_2'}^{-1} \sum_{i_2=1}^{t_2-j_2'+l} Y_{i_2} - N_{2,t_2-j_2'}^{-1} \sum_{i_2=1}^{t_2-j_2'+l} (i_2 - t_2 + j_2') Y_{i_2} \right\}.
$$

The first of the above four terms is bounded by

$$
n^{-2}(2l+1)^{-2} \sum_{t_1,t_2=l+1}^{n-l} \sum_{j_1=0}^{t_1-1} \sum_{j_2=0}^{t_2-1} \sum_{j_1'=0}^{t_1-l-1} \sum_{j_2'=0}^{t_2-l-1} \sum_{i_1=t_1-j_1'-l}^{t_1-j_1'+l} \sum_{i_2=t_2-j_2'-l}^{t_2-j_2'+l}
$$

$$
\times \left| \pi_{0j_1} \frac{\partial \pi_{0j_1'}}{\partial \alpha_k} \pi_{0j_2} \frac{\partial \pi_{0j_2'}}{\partial \alpha_k} \mathbb{E} Y_{t_1-j_1} Y_{i_1} Y_{t_2-j_2} Y_{i_2} \right|
$$

$$
\leq C n^{-2}(2l+1)^{-2} \sum_{t_1,t_2=l+1}^{n-l} \sum_{j_1=0}^{t_1-1} \sum_{j_2=0}^{t_2-1} \sum_{j_1'=0}^{t_1-l-1} \sum_{j_2'=0}^{t_2-l-1} \sum_{i_1=t_1-j_1'-l}^{t_1-j_1'+l} \sum_{i_2=t_2-j_2'-l}^{t_2-j_2'+l}
$$

$$
\times \rho^{j_1+j_2+j_1'+j_2'} |\mathbb{E} Y_{t_1-j_1} Y_{i_1} Y_{t_2-j_2} Y_{i_2}|.
$$

We will establish next that the complicated sum above is $\mathcal{O}(l^{-2}) = o(n^{-1})$. In other words,

$$
\sum_{t_1,t_2=l+1}^{n-l} \sum_{j_1=0}^{t_1-1} \sum_{j_2=0}^{t_2-1} \sum_{j_1'=0}^{t_1-l-1} \sum_{j_2'=0}^{t_2-l-1} \sum_{i_1=t_1-j_1'-l}^{t_1-j_1'+l} \sum_{i_2=t_2-j_2'-l}^{t_2-j_2'+l} \rho^{j_1+j_2+j_1'+j_2'} |\mathbb{E} Y_{t_1-j_1} Y_{i_1} Y_{t_2-j_2} Y_{i_2}| = \mathcal{O}(n^2). \tag{A12}
$$

Define $\#(t_1 - j_1, i_1, t_2 - j_2, i_2)$ = the number of distinct indices in $t_1 - j_1, i_1, t_2 - j_2, i_2$. Obviously $\#(t_1 - j_1, i_1, t_2 - j_2, i_2) \in \{1, 2, 3, 4\}$. We will bound the LHS of (A12) according to the index $\#(t_1 - j_1, i_1, t_2 - j_2, i_2)$.

Case 1: $\#(t_1-j_1,i_1,t_2-j_2,i_2)=1$. The four indices are equal, that is $t_1-j_1=i_1=t_2-j_2=i_2$, then $j_2=t_2-t_1+j_1$. Therefore,

$$\sum_{t_1,t_2=l+1}^{n-l}\sum_{j_1=0}^{t_1-1}\sum_{j_2=0}^{t_2-1}\sum_{j_1'=0}^{t_1-l-1}\sum_{j_2'=0}^{t_2-l-1}\sum_{i_1=t_1-j_1'-l}^{t_1-j_1'+l}\sum_{i_2=t_2-j_2'-l}^{t_2-j_2'+l}\rho^{j_1+j_2+j_1'+j_2'}$$
$$\times|\mathbb{E}Y_{t_1-j_1}Y_{i_1}Y_{t_2-j_2}Y_{i_2}I_{\{\#(t_1-j_1,i_1,t_2-j_2,i_2)=1\}}|$$
$$\leq C\sum_{t_1=l+1}^{n-l}\sum_{j_1=0}^{t_1-1}\sum_{j_1'=\max\{0,j_1-l\}}^{\min\{t_1-l-1,j_1+l\}}\sum_{t_2=\max\{l+1,t_1-j_1\}}^{n-l}\sum_{j_2'=\max\{0,t_2-t_1+j_1-l\}}^{\min\{t_2-l-1,t_2-t_1+j_1+l\}}\rho^{t_2-t_1+2j_1+j_1'+j_2'}\leq C'n.$$

Case 2: $\#(t_1-j_1,i_1,t_2-j_2,i_2)=2$. First we consider the possibility that the four indices t_1-j_1,i_1,t_2-j_2,i_2 take two different values, each twice. For instance,

$$\sum_{t_1,t_2=l+1}^{n-l}\sum_{j_1=0}^{t_1-1}\sum_{j_2=0}^{t_2-1}\sum_{j_1'=0}^{t_1-l-1}\sum_{j_2'=0}^{t_2-l-1}\sum_{i_1=t_1-j_1'-l}^{t_1-j_1'+l}\sum_{i_2=t_2-j_2'-l}^{t_2-j_2'+l}\rho^{j_1+j_2+j_1'+j_2'}$$
$$\times|\mathbb{E}Y_{t_1-j_1}Y_{i_1}Y_{t_2-j_2}Y_{i_2}I_{\{t_1-j_1=i_1<t_2-j_2=i_2\}}|$$
$$\leq C'\sum_{t_1=l+1}^{n-l}\sum_{j_1=0}^{t_1-1}\sum_{j_1'=\max\{0,j_1-l\}}^{\min\{t_1-l-1,j_1+l\}}\sum_{t_2=\max\{l+1,t_1-j_1\}}^{n-l}\sum_{j_2=0}^{t_2-t_1+j_1}\sum_{j_2'=\max\{0,j_2-l\}}^{\min\{t_2-l-1,j_2+l\}}\rho^{j_1+j_2+j_1'+j_2'}\leq C''n^2.$$

Next we consider the case that the four indices t_1-j_1,i_1,t_2-j_2,i_2 take two different values, one thrice, the other once. For instance,

$$\sum_{t_1,t_2=l+1}^{n-l}\sum_{j_1=0}^{t_1-1}\sum_{j_2=0}^{t_2-1}\sum_{j_1'=0}^{t_1-l-1}\sum_{j_2'=0}^{t_2-l-1}\sum_{i_1=t_1-j_1'-l}^{t_1-j_1'+l}\sum_{i_2=t_2-j_2'-l}^{t_2-j_2'+l}\rho^{j_1+j_2+j_1'+j_2'}$$
$$\times|\mathbb{E}Y_{t_1-j_1}Y_{i_1}Y_{t_2-j_2}Y_{i_2}I_{\{t_1-j_1=i_1=t_2-j_2<i_2\}}|$$
$$\leq C\sum_{t_1=l+1}^{n-l}\sum_{j_1=0}^{t_1-1}\sum_{j_1'=\max\{0,j_1-l\}}^{\min\{t_1-l-1,j_1+l\}}\sum_{t_2=\max\{l+1,t_1-j_1\}}^{n-l}\sum_{j_2'=0}^{t_2-t_1+j_1+l}\sum_{i_2=t_2-j_2'-l}^{t_2-j_2'+l}\rho^{t_2-t_1+2j_1+j_1'+j_2'}\leq C'nl.$$

Case 3: $\#(t_1-j_1,i_1,t_2-j_2,i_2)=3$. For instance, let $\tilde{\rho}=\max\{\rho,\rho_M\}\in(0,1)$,

$$\sum_{t_1,t_2=l+1}^{n-l}\sum_{j_1=0}^{t_1-1}\sum_{j_2=0}^{t_2-1}\sum_{j_1'=0}^{t_1-l-1}\sum_{j_2'=0}^{t_2-l-1}\sum_{i_1=t_1-j_1'-l}^{t_1-j_1'+l}\sum_{i_2=t_2-j_2'-l}^{t_2-j_2'+l}\rho^{j_1+j_2+j_1'+j_2'}$$
$$\times|\mathbb{E}Y_{t_1-j_1}Y_{i_1}Y_{t_2-j_2}Y_{i_2}I_{\{t_1-j_1<i_1<t_2-j_2=i_2\}}|$$
$$\leq C\sum_{t_1=l+1}^{n-l}\sum_{j_1=0}^{t_1-1}\sum_{j_1'=0}^{\min\{t_1-l-1,j_1+l\}}\sum_{i_1=t_1-j_1}^{t_1-j_1'+l}\sum_{t_2=\max l+1,i_1}^{n-l}\sum_{j_2=0}^{t_2-i_1}\sum_{j_2'=\max\{0,j_2-l\}}^{\min\{t_2-l-1,j_2+l\}}$$
$$\times\rho^{j_1+j_2+j_1'+j_2'}\{\alpha(i_1-t_1+j_1)\}^{1/6}$$
$$\leq C'\sum_{t_1=l+1}^{n-l}\sum_{j_1=0}^{t_1-1}\sum_{j_1'=0}^{\min\{t_1-l-1,j_1+l\}}\sum_{i_1=t_1-j_1}^{t_1-j_1'+l}\sum_{t_2=\max\{l+1,i_1\}}^{n-l}\sum_{j_2=0}^{t_2-i_1}\sum_{j_2'=\max\{0,j_2-l\}}^{\min\{t_2-l-1,j_2+l\}}\tilde{\rho}^{j_1+j_2+j_1'+j_2'+(i_1-t_1+j_1)/6}$$
$$\leq C''n^2.$$

Case 4: $\#(t_1-j_1,i_1,t_2-j_2,i_2)=4$. The four indices t_1-j_1,i_1,t_2-j_2,i_2 are all distinct. Let $\tilde{\rho}=\max\{\rho,\rho_M\}\in(0,1)$. One such scenario is

$$\sum_{t_1,t_2=l+1}^{n-l}\sum_{j_1=0}^{t_1-1}\sum_{j_2=0}^{t_2-1}\sum_{j_1'=0}^{t_1-l-1}\sum_{j_2'=0}^{t_2-l-1}\sum_{i_1=t_1-j_1'-l}^{t_1-j_1'+l}\sum_{i_2=t_2-j_2'-l}^{t_2-j_2'+l}\rho^{j_1+j_2+j_1'+j_2'}$$
$$\times|\mathbb{E}Y_{t_1-j_1}Y_{i_1}Y_{t_2-j_2}Y_{i_2}I_{\{t_1-j_1<i_1<t_2-j_2<i_2,i_1-t_1+j_1<i_2-t_2+j_2\}}|$$
$$\le C\sum_{t_1=l+1}^{n-l}\sum_{j_1=0}^{t_1-1}\sum_{j_1'=0}^{\min\{t_1-l-1,j_1+l-1\}}\sum_{i_1=\max\{t_1-j_1+1,t_1-j_1'-l\}}^{t_1-j_1'+l}\sum_{t_2=\max\{l+1,i_1+1\}}^{n-l}\sum_{j_2=0}^{t_2-i_1-1}$$
$$\times\sum_{j_2'=0}^{\min\{t_2-l-1,j_2+l-1,t_1-j_1-i_1+j_2+l\}}\sum_{i_2=\max\{t_2-j_2+1,t_2-j_2'-l,i_1-t_1+j_1+t_2-j_2\}}^{t_2-j_2'+l}\tilde{\rho}^{j_1+j_1'-t_2+2j_2+j_2'+i_2}$$
$$\le C'n^2.$$

Another possible scenario is

$$\sum_{t_1,t_2=l+1}^{n-l}\sum_{j_1=0}^{t_1-1}\sum_{j_2=0}^{t_2-1}\sum_{j_1'=0}^{t_1-l-1}\sum_{j_2'=0}^{t_2-l-1}\sum_{i_1=t_1-j_1'-l}^{t_1-j_1'+l}\sum_{i_2=t_2-j_2'-l}^{t_2-j_2'+l}\rho^{j_1+j_2+j_1'+j_2'}$$
$$\times|\mathbb{E}Y_{t_1-j_1}Y_{i_1}Y_{t_2-j_2}Y_{i_2}I_{\{i_1<i_2<t_1-j_1<t_2-j_2,i_2-i_1<t_2-j_2-t_1+j_1\}}|$$
$$\le C\sum_{j_1'=0}^{n-2l-1}\sum_{i_1=1}^{n}\sum_{j_2'=0}^{\min\{n-i_1,n-2l-1\}}\sum_{i_2=\max\{1,i_1+1\}}^{n}\sum_{t_1=\max\{l+1,i_2+1,j_1'+l+1\}}^{n-l}\sum_{j_1=0}^{t_1-i_2}$$
$$\times\sum_{j_2=0}^{n-l-1}\sum_{t_2=\max\{l+1,j_2'+l+1,t_1-j_1+j_2+1,i_2-i_1+j_2+t_1-j_1\}}^{n-l}\tilde{\rho}^{j_1'+j_2'-t_1+2j_1+t_2}$$
$$\le C'n^2.$$

Summarising the above, (A12) is proved and so is (A10).