Appl. Math. J. Chinese Univ.
2019, 34(4): 403-422

Asymptotics of estimators for nonparametric multivariate regression models with long memory

WANG Li-hong WANG Ming

Abstract. In this paper, a nonparametric multivariate regression model with long memory covariates and long memory errors is considered. We approximate the nonparametric multivariate regression function by the weighted additive one-dimensional functions. The local linear smoothing and least squares method are proposed for the one-dimensional regression estimation and the weight parameters estimation, respectively. The asymptotic behaviors of the proposed estimators are investigated.

§1 Introduction

In recent years long memory time series analysis has become an important tool for analyzing long range dependent data, see, e.g. [3], [9] and [30]. The recent monographs of [4] and [17] contain numerous additional references. In this article we will consider the estimation of a long memory nonparametric multivariate regression model.

Because of the poor convergence rate of the nonparametric multivariate regression estimation, known as "the curse of dimensionality", various dimension reduction methods using nonparametric or semiparametric analysis of time series, such as varying coefficient models, partially linear additive models, flexible semiparametric models, the popular LASSO type approach, have been pursued in the literature for independent or short range dependent processes, see, for example, [1], [7], [8], [12], [13], [14], [15], [22], [24], [25], [26], [35], and [36]. However, to the best of our knowledge, the dimension reduction methods are not yet sufficiently developed for long memory time series. This is the issue we intend to address in the current paper.

Specifically, we consider the stationary random process $(Y_t, X_t) \in \mathbf{R} \times \mathbf{R}^p$, $p \geq 1$, $t = 1, 2, \cdots$, and assume the availability of the data $\{Y_t, X_t, t = 1, \cdots, n\}$, for estimating the regression function $m(x) = E(Y_t|X_t = x)$. Without loss of generality we assume that $\mathrm{E}Y_t = 0$, otherwise we replace Y_t by $Y_t - \frac{1}{n}\sum_{t=1}^{n} Y_t$.

Received: 2016-12-22. Revised: 2019-05-27.
MR Subject Classification: 62M10.
Keywords: Additive model, local linear estimation, long memory time series.
Digital Object Identifier(DOI): https://doi.org/10.1007/s11766-019-3518-3.
Supported by the National Natural Science Foundation of China (No. 11671194 and No. 11501287) .

[22] proposed a semiparametric estimation method for estimating the regression function $m(x) = E(Y_t|X_t = x)$ under short range dependent assumption of the process (Y_t, X_t). They approximated the multivariate regression function $m(x)$ by an affine combination of one-dimensional marginal regression functions. The weight parameters involved in the approximation are estimated by least squares on the basis of the first-stage Nadaraya-Watson kernel estimates of the marginal regressions. It is shown that the convergence rate of the estimators does not depend on the dimension and thus the curse of dimensionality is avoided. The advantages of the method are also demonstrated by Monte Carlo simulation and real data examples. Due to its good performance, we shall follow this approach for the long memory processes in this paper, but apply the local linear smoothing method instead of Nadaraya-Watson kernel method for the first-stage nonparametric regression estimation. One motivation of using the local linear regression smoothers is that they repair the drawbacks of Nadaraya-Watson method. See [10] for additional discussion.

The local linear regression estimator has been studied extensively in the literature for time series models due to its superiority in function estimation. Asymptotic properties of the estimator under the conditions of independence as well as weak dependence are investigated, see, for example, [2], [5], [6], [10], [11], [20], [23], [27], [32] and [33]. [28] studied the nature of the asymptotic distributions for the local linear estimators of the regression models with independent designs and Gaussian-subordinated stationary long range dependent errors. [21] considered the local linear estimator of the conditional medians for stationary long memory linear time series models. [34] investigated the local linear estimation for a stationary long memory nonparametric spatio-temporal regression model.

Let $m_j(x_j) = E(Y_t|X_{tj} = x_j)$, $j = 1, \cdots, p$. We approximate $m(x) = E(Y_t|X_t = x)$ by

$$m_w(x) = \sum_{j=1}^{p} w_j m_j(x_j)$$

with some weights w_j, $j = 1, \cdots, p$, where $x = (x_1, \cdots, x_p)^T$. To allow for the conditional heteroscedasticity, define the errors as $\sigma(X_t)\varepsilon_t := Y_t - m_w(X_t)$ and $\sigma_j(X_t)\eta_t := Y_t - m_j(X_{tj})$, $j = 1, \cdots, p$, $t = 1, 2, \cdots$. That is,

$$Y_t = m_w(X_t) + \sigma(X_t)\varepsilon_t, \quad t = 1, 2, \cdots, \tag{1}$$

and

$$Y_t = m_j(X_{tj}) + \sigma_j(X_t)\eta_t, \quad j = 1, \cdots, p, \quad t = 1, 2, \cdots \tag{2}$$

We assume that the processes $\{\varepsilon_t,\ t = 1, 2, \cdots\}$ and $\{\eta_t,\ t = 1, 2, \cdots\}$ are mutually independent long memory processes with zero mean and variance one, and the conditional heteroscedasticity is permitted. We also assume that $\{\varepsilon_t,\ t = 1, 2, \cdots\}$ and $\{\eta_t,\ t = 1, 2, \cdots\}$ are independent of $\{X_t,\ t = 1, 2, \cdots\}$. The long memory property of the processes $\{\varepsilon_t\}$, $\{\eta_t\}$ and $\{X_t\}$ will be discussed in details in the next section.

The main idea is that we first use the local linear smoother, $\hat{m}_j(x_j)$, to estimate the one-dimensional function $m_j(x_j)$, and then least squares estimators of the weights w_j, $\hat{w}_j$, can be obtained. Finally, we can estimate the multivariate regression function $m(x)$ by

$$\hat{m}(x) = \sum_{j=1}^{p} \hat{w}_j \hat{m}_j(x_j).$$

It is worth mentioning that $\hat{m}(x)$ can only be taken as the approximated value of $m(x)$ as $\sum_{j=1}^{p} w_j m_j(x_j)$ may not equal to $m(x)$ except the full linear regression case. As stated in [22], because the closed form for the parametric estimator of w_j can be obtained and no iterative algorithm is involved, this method is fast to be computed especially when p is large, and thus avoids the curse of dimensionality. We shall show that, in the long memory case, the method can also avoid the curse of dimensionality.

Now we estimate $m_j(x_j)$ by using local linear fitting method. That is, we approximate $m_j(\cdot)$ in a neighborhood of x_j as

$$m_j(\tilde{x}) \approx m_j(x_j) + (\tilde{x} - x_j) m_j'(x_j),$$

where $m_j'(\cdot)$ is the first derivative of the function $m_j(\cdot)$. This suggests the following estimator of $m_j(x_j)$:

$$\hat{m}_j(x_j) = (1, 0) U_n^{-1} V_n, \tag{3}$$

where

$$U_n := U_n(x_j) = \begin{pmatrix} U_0 & U_1 \\ U_1 & U_2 \end{pmatrix}, \quad V_n := V_n(x_j) = \begin{pmatrix} V_0 \\ V_1 \end{pmatrix},$$

where

$$U_l := U_l(x_j) = \frac{1}{n} \sum_{t=1}^{n} \left(\frac{X_{tj} - x_j}{h} \right)^l K_h(X_{tj} - x_j), \quad l = 0, 1, 2,$$

and

$$V_l := V_l(x_j) = \frac{1}{n} \sum_{t=1}^{n} \left(\frac{X_{tj} - x_j}{h} \right)^l K_h(X_{tj} - x_j) Y_t, \quad l = 0, 1,$$

and $h = h_n$ is a sequence of bandwidths tending to zero at an appropriate rate as n tends to infinity, K is a kernel satisfying the conditions given in the next section, and $K_h(t) = K(t/h)/h$. As suggested in [29], since the performance of the estimator $\hat{m}_j(x_j)$ will be poor for large values of x_j, we shall assume that X_{tj} falls into an interval $[a_n, q_n]$, where $a_n < 0 < q_n$, and $-a_n$ and q_n tend to infinity slowly.

Define

$$\hat{M} = \begin{pmatrix} \hat{m}_1(X_{11}) & \cdots & \hat{m}_p(X_{1p}) \\ \vdots & \vdots & \vdots \\ \hat{m}_1(X_{n1}) & \cdots & \hat{m}_p(X_{np}) \end{pmatrix},$$

and $y = (Y_1, \cdots, Y_n)^T$. Let $w = (w_1, \cdots, w_p)^T$. The least squares estimators for the weights are

$$\hat{w} = (\hat{M}^T \hat{M})^{-1} \hat{M}^T y.$$

In the following section, we will study the asymptotic properties of the estimators $\hat{w}$ and $\hat{m}(x)$ for long memory processes. It is shown that the convergence rates of the estimators do not depend on the dimension p. This means that, for the long memory nonparametric multivariate regression models, this dimension reduction estimation method not only incorporates fast computation when p is large but also solves the curse of dimensionality problem. Section 3 illustrates the estimation method with a small real data analysis. The proofs of the theorems are in Section 4.

Throughout the paper, all limits are taken as $n \to \infty$, unless specified otherwise, $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution, and $\xrightarrow{\mathcal{P}}$ denotes convergence in probability. For any two real sequences $\{a_n\}$ and $\{b_n\}$, $a_n \sim b_n$ means that there are constants $c > 0$ and $C < \infty$ such that $c \le a_n/b_n \le C$ for all sufficiently large n.

§2 Asymptotic properties of the estimators

Throughout this paper, we assume the following conditions:

Assumption(A)

(A1) $\{\varepsilon_t,\ t = 1, 2, \cdots\}$ is a zero mean stationary process with the autocovariance function satisfying

$$\gamma_\varepsilon(k) = \mathrm{Cov}(\varepsilon_t, \varepsilon_{t+k}) \sim G_\varepsilon |k|^{2d_\varepsilon - 1}$$

for large k, where $0 < d_\varepsilon < \frac{1}{2}$ and G_ε is a positive constant.

(A2) $\{\eta_t,\ t = 1, 2, \cdots\}$ is a zero mean stationary process with the autocovariance function satisfying

$$\gamma_\eta(k) = \mathrm{Cov}(\eta_t, \eta_{t+k}) \sim G_\eta |k|^{2d_\eta - 1}$$

for large k, where $0 < d_\eta < \frac{1}{2}$ and G_η is a positive constant.

(A3) The marginal density function of X_{tj}, $f_j(\cdot)$, is positive, bounded and Lipschitz continuous. The joint density function of X_t, $f_X(\cdot)$, is positive, bounded and differentiable with the derivative f'_X satisfying $\int \|f'_X(x)\|dx < \infty$, $\int \|f'_X(x_j, x')\|^2/f_j(x_j)dx'dx_j < \infty$, where $\|\cdot\|$ denotes the Euclidean norm and $x' = (x_1 \cdots, x_{j-1}, x_{j+1}, \cdots, x_p)^T$.

(A4) The joint density function of the random vectors X_t and X_s, $f_{X,t,s}$, exists and satisfies that

$$D_{t,s}(x, y) \le G_X |s - t|^{2d_X - 1} \|f'_X(x)\| \|f'_X(y)\|$$

for any $x, y \in \mathbf{R^p}$ as $|s - t| \to \infty$, where $D_{t,s}(x, y) = |f_{X,t,s}(x, y) - f_X(x) f_X(y)|$, $0 < d_X < \frac{1}{2}$ and G_X is a positive constant. Moreover, the joint density function of the random vectors X_t, X_s and X_k, $f_{X,t,s,k}$, exists and satisfies that

$$D_{t,s,k}(x, y, z) \le G'_X \Big(|k - t|^{2d_X - 1} \|f'_X(x)\| f_X(y) \|f'_X(z)\| + |k - s|^{2d_X - 1} f_X(x) \|f'_X(y)\| \|f'_X(z)\| \Big)$$

for any $x, y, z \in \mathbf{R^p}$ as $|k - t| \to \infty$, $|k - s| \to \infty$ and $|s - t| \to \infty$, where $D_{t,s,k}(x, y, z) = |f_{X,t,s,k}(x, y, z) - f_{X,t,s}(x, y) f_X(z)|$, and G'_X is a positive constant.

(A5) The regression functions $m_j(\cdot)$, $1 \le j \le p$, are bounded, integrable and three times differentiable with bounded derivatives.

(A6) The kernel density K is bounded with compact support, and $u^l K(u)$ has Fourier transform $\Phi_l(r) = 2\pi \int e^{iru} u^l K(u) du$ that satisfies $\int |\Phi_l(r)| dr < \infty$ for $l = 0, 1, 2, 3$.

(A7) The functions $\sigma(\cdot)$ and $\sigma_j(\cdot)$, $1 \le j \le p$, are bounded and integrable.

(A8) $n^{d_X - 1/2} h^{-1} \to 0$ and $n^{2d_\eta - 1} h^{-1} \to 0$.

(A9) $n^{1/2 - \max(d_\varepsilon, d_\eta)} h \to 0$, $n^{d_X - \max(d_\varepsilon, d_\eta)} h^{-1} \to 0$ and $n^{2\max(d_\varepsilon, d_\eta) - 1} h^{-1} \log^2 n \to 0$.

(A10) The interval $[a_n, q_n]$ is such that $-a_n$ and q_n tend to infinity slowly enough so that $\log n \inf_{x_j\in[a_n,q_n]} f_j(x_j) \geq C_j$ for $0 < C_j < \infty$, $j = 1, \cdots, p$.

The assumptions (A3), (A5)-(A7) for the kernel, the density functions and the regression functions are the usual standard conditions for local linear estimation.

In addition, Assumption (A8) is used for proving the weak consistency of the local linear estimator $\hat{m}_j(\cdot)$ (Theorem 2.1). Assumptions (A9) and (A10) are needed to obtain the asymptotic distributions of the least squares estimator $\hat{w}$ and the proposed nonparametric estimator $\hat{m}(x)$ (Theorems 2.2 and 2.3). Assumption (A10) is analogous to [29] who proves uniform convergence rates for local linear estimators based on independent data.

Similar long memory conditions in Assumptions (A1), (A2) and (A4) are also imposed in [28] and [31]. Many commonly used densities including Gaussian can be shown to satisfy Assumption (A3). The long memory property of a stationary random process can be characterized in various ways (see, for example [19]). In this paper the error processes $\{\varepsilon_t\}$ and $\{\eta_t\}$ have long memory in the covariance sense, whereas the long memory property of the explanatory process $\{X_t\}$ is characterized by its joint distributions (Assumption (A4)), which is similar to Assumption B10 in [31].

When $p = 1$, from the proof of Lemma 2 of [16], we can show that, if $\{X_t\}$ is a one-dimensional linear long memory process, then under some mild conditions,

$$f_{X,t,s}(x, y) - f_X(x)f_X(y) = \gamma_X(s - t)f'_X(x)f'_X(y) + o(|s - t|^{2d_X-1}),$$

and

$$\begin{aligned} &f_{X,t,s,k}(x, y, z) - f_{X,t,s}(x, y)f_X(z)\\ &= \gamma_X(k - t)\frac{\partial f_{X,t,s}(x, y)}{\partial x} f'_X(z) + \gamma_X(k - s)\frac{\partial f_{X,t,s}(x, y)}{\partial y} f'_X(z)\\ &\quad + o(|k - t|^{2d_X-1}) + o(|k - s|^{2d_X-1}), \end{aligned}$$

for any $x, y, z \in \mathbf{R}$ as $|s - t| \to \infty$, $|k - t| \to \infty$ and $|k - s| \to \infty$, where $\partial f_{X,t,s}(x, y)/\partial x$ and $\partial f_{X,t,s}(x, y)/\partial y$ are the partial derivatives of $f_{X,t,s}(x, y)$, and $\gamma_X(k) = \mathrm{Cov}(X_0, X_k) \sim C_X|k|^{2d_X-1}$ as $|k| \to \infty$. These equations obviously imply Assumption (A4).

Let

$$\mu_l = \int u^l K(u)du, \quad l = 1, 2, 3, \quad \mu = (\mu_2, \mu_3)^T,$$

$$U = \begin{pmatrix} 1 & \mu_1 \\ \mu_1 & \mu_2 \end{pmatrix},$$

and $V^*_n := V^*_n(X_{tj}) = (V^*_0, V^*_1)^T$, where

$$V^*_l := V^*_l(X_{tj}) = \frac{1}{n}\sum_{s=1}^{n}\left(\frac{X_{sj} - X_{tj}}{h}\right)^l K_h(X_{sj} - X_{tj})\sigma_j(X_s)\eta_s, \quad l = 0, 1.$$

We first show the weak consistency of the local linear estimator $\hat{m}_j(\cdot)$. Let $m''_j(\cdot)$ be the second derivative of the function $m_j(\cdot)$.

Theorem 2.1. *Under Assumptions (A2)-(A8), we have, for any fixed t,*

$$\begin{aligned}\hat{m}_j(X_{tj})-m_j(X_{tj}) &= (1,0)\Big\{(f_j(X_{tj})U)^{-1}V_n^*+\frac{h^2}{2}U^{-1}\mu m_j''(X_{tj})\Big\}\\ &\quad +O_P(n^{d_X-1/2}h^{-1})+O_P(h) \end{aligned}\tag{4}$$

and

$$\hat{m}_j(X_{tj})-m_j(X_{tj})\xrightarrow{\mathcal{P}}0.$$

Next we establish the asymptotic distribution of the least squares estimator $\hat{w}$. As we will see, under long memory of the processes $\{\varepsilon_t\}$, $\{\eta_t\}$ and $\{X_t\}$, the limiting distribution and the convergence rate of $\hat{w}$ depend on the relative strength of dependence in ε_t and η_t but do not depend on the dimension p.

Define

$$M=\begin{pmatrix} m_1(X_{11}) & \cdots & m_p(X_{1p})\\ \vdots & \vdots & \vdots\\ m_1(X_{n1}) & \cdots & m_p(X_{np})\end{pmatrix},$$

$$\Lambda=\begin{pmatrix} E(m_1(X_{t1})m_1(X_{t1})) & \cdots & E(m_1(X_{t1})m_p(X_{tp}))\\ \vdots & \vdots & \vdots\\ E(m_p(X_{tp})m_1(X_{t1})) & \cdots & E(m_p(X_{tp})m_p(X_{tp}))\end{pmatrix},$$

and $\varepsilon^*=(\sigma(X_1)\varepsilon_1,\cdots,\sigma(X_n)\varepsilon_n)^T$. To simplify the formulas we assume that the kernel K is symmetric. This implies that $\mu_1=\mu_3=0$.

Theorem 2.2. *Assume that Assumption (A) holds, the kernel K is symmetric and the matrix Λ is positive definite.*

(a). If $d_\eta<d_\varepsilon$ and $n^{-1/2-d_\varepsilon}\sum_{t=1}^n\varepsilon_t$ converges in distribution to a random variable Z_1, then

$$n^{1/2-d_\varepsilon}(\hat{w}-w)\xrightarrow{\mathcal{D}}\Lambda^{-1}\Theta_1Z_1,$$

where $\Theta_1=(\Theta_{11},\cdots,\Theta_{1p})^T$, and $\Theta_{1j}=E\Big(m_j(X_{tj})\sigma(X_t)\Big)$.

(b). If $d_\eta>d_\varepsilon$ and $n^{-1/2-d_\eta}\sum_{t=1}^n\eta_t$ converges in distribution to a random variable Z_2, then

$$n^{1/2-d_\eta}(\hat{w}-w)\xrightarrow{\mathcal{D}}-\Lambda^{-1}\Theta_2Z_2,$$

where $\Theta_2=(\Theta_{21},\cdots,\Theta_{2p})^T$, $\Theta_{2j}=\sum_{k=1}^p w_kE(\sigma_k(X_t)\zeta_j(X_{tk}))$ and $\zeta_j(X_{tk})=E(m_j(X_{tj})|X_{tk})$.

(c). If $d_\eta=d_\varepsilon=d$, $n^{-1/2-d}\sum_{t=1}^n\varepsilon_t$ converges in distribution to a random variable Z_1, and $n^{-1/2-d}\sum_{t=1}^n\eta_t$ converges in distribution to a random variable Z_2, then

$$n^{1/2-d}(\hat{w}-w)\xrightarrow{\mathcal{D}}\Lambda^{-1}(\Theta_1Z_1-\Theta_2Z_2).$$

Theorem 2.2 examines the asymptotic behavior of the parametric estimator of the weight and shows that the estimator can achieve the $n^{1/2-\max\{d_\varepsilon,d_\eta\}}$ convergence rate when we replace $m_j(X_{tj})$ by its nonparametric local linear estimator. We can also see that the estimator is asymptotically unbiased and its asymptotic distribution is a scaled distribution of a fixed random variable. Actually, from the proof we see that the behavior of the estimator is asymptotically governed by the sample mean of the errors $\{\varepsilon_t\}$ and $\{\eta_t\}$. If $\{\varepsilon_t\}$ and $\{\eta_t\}$ are linear or Gaussian random processes with long memory, the random variables Z_1 and Z_2 will have normal distributions (cf. [17] and [30]), hence the limiting distribution of the estimator is also normal.

Next we establish the limiting distribution of the proposed nonparametric estimator $\hat{m}(x)$.

Theorem 2.3. *(a). Under the conditions of Theorem 2.2 (a), for any* $x \in [a_n, q_n]^p$,

$$n^{1/2-d_\varepsilon}(\hat{m}(x) - m_w(x)) \xrightarrow{\mathcal{D}} (\Lambda^{-1}\Theta_1)^T \Upsilon(x) Z_1,$$

where $\Upsilon(x) = (m_1(x_1), \cdots, m_p(x_p))^T$.

(b). Under the conditions of Theorem 2.2 (b), for any $x \in [a_n, q_n]^p$,

$$n^{1/2-d_\eta}(\hat{m}(x) - m_w(x)) \xrightarrow{\mathcal{D}} (w^T\Psi(x) - (\Lambda^{-1}\Theta_2)^T \Upsilon(x)) Z_2,$$

where $\Psi(x) = (g_1(x_1), \cdots, g_p(x_p))^T$, *and* $g_j(x_j) = (f_j(x_j))^{-1} \int \sigma_j(x_j, x') f_X(x_j, x') dx'$.

(c). Under the conditions of Theorem 2.2 (c), for any $x \in [a_n, q_n]^p$,

$$n^{1/2-d}(\hat{m}(x) - m_w(x)) \xrightarrow{\mathcal{D}} (\Lambda^{-1}\Theta_1)^T \Upsilon(x) Z_1 + (w^T\Psi(x) - (\Lambda^{-1}\Theta_2)^T \Upsilon(x)) Z_2.$$

Typically if one uses $h \propto n^{-\delta}$, then the assumptions about h in Theorem 2.3 will be satisfied as long $1/2 - \max\{d_\varepsilon, d_\eta\} < \delta < \min\{1 - 2\max(d_\varepsilon, d_\eta), \max\{d_\varepsilon, d_\eta\} - d_X\}$. For example, if $d_\varepsilon = d_\eta = 0.4$ and $d_X = 0.2$, then Theorems 2.1-2.3 hold for $0.1 < \delta < 0.2$.

From Theorem 2.3 we see that the estimator $\hat{m}(x)$ is asymptotically unbiased and have the same convergence rate as in the univariate nonparametric regression setting up (cf. [28]). Moreover, unlike in the independent or weakly dependent cases, the limiting distribution of the estimator is a scaled distribution of a fixed random variable. When $\{\varepsilon_t\}$ and $\{\eta_t\}$ are linear or Gaussian random processes with long memory, the limiting distribution of the estimator will be normal.

As evidenced by Theorem 2.3, although the asymptotic variance of the estimator for the long memory multivariate semiparametric additive model is more complex than its univariate counterpart, the central limit theorems similar to those studied for the univariate case still hold and the convergence rate does not depend on the dimension p. Therefore the curse of dimensionality is effectively avoided.

§3 A real data example

We now report the results of a small real data example to demonstrate the finite sample performance of the proposed local linear estimation for semiparametric additive long memory time series models. We will see that the main advantage of the proposed method is computational and its performance in the case of high-dimensional problems.

We consider a real financial data set from the S&P 500 Index series starting from April 11, 2005 to March 23, 2018. The data consists of daily trading volume v_t and the absolute daily return $R_t = |100\log(c_t/c_{t-1})|$, where c_t is the daily close price. It is well known that most daily financial time series exhibit quite persistent autocorrelation in their absolute returns, e.g. [18]

Table 1: The MSE of the residuals under six circumstances

model	(1)	(2)	(3)
ARFIMA	1.2547	1.2528	1.2414
WAL	0.5527	0.4876	0.4540
model	(4)	(5)	(6)
ARFIMA	1.2547	1.2528	1.2414
WAL	0.0034	0.0016	0.0014

analyzed the long memory properties of absolute stock returns of the S&P 500 Index.

We are interested in fitting the data using the proposed weighted nonparametric additive model with a long lag. We will also check if the volume lags would be helpful in improving the precision of the estimation and model fitting. We explore the following six circumstances with $Y_t = R_t$. (1). p=30 and $X_t = (R_{t-1}, \cdots, R_{t-p})^T$; (2). p=45 and $X_t = (R_{t-1}, \cdots, R_{t-p})^T$; (3). p=60 and $X_t = (R_{t-1}, \cdots, R_{t-p})^T$; (4). p=30 and $X_t = (R_{t-1}, \cdots, R_{t-p}, v_{t-1}, \cdots, v_{t-p})^T$; (5). p=45 and $X_t = (R_{t-1}, \cdots, R_{t-p}, v_{t-1}, \cdots, v_{t-p})^T$; (6). p=60 and $X_t = (R_{t-1}, \cdots, R_{t-p}, v_{t-1}, \cdots, v_{t-p})^T$. We compare the mean squared error (MSE) of the fitted residuals by using two methods, the linear ARFIMA $(p, d, 0)$ model and the weighted nonparametric additive model of order p by local linear regression estimation (WAL) method. The MSE is defined by

$$MSE = \frac{1}{n-p} \sum_{t=p+1}^{n} (Y_t - \hat{Y}_t)^2,$$

where $\hat{Y}_t = \hat{m}(X_t) = \sum_{j=1}^{p} \hat{w}_j \hat{m}_j(X_{t-j})$.

The results are displayed in Table 1. From the table we see that the MSEs of WAL method are much smaller than those of the ARFIMA method. Moreover, it seems that the volume lags contribute much to the estimation and model fitting. It is clear that the weighted nonlinear additive model performs better than the linear ARFIMA method. The simulation evidence indicates that the weighted nonparametric additive model can effectively avoid the curse of dimensionality and solve the model fitting precision and computation burden problems for the high-dimensional nonlinear models.

§4 Proofs

Proof of Theorem 2.1. We first show that, for $l = 0, 1, 2$,

$$\sup_{x_j \in \mathbf{R}} |U_l - f_j(x_j)\mu_l| = O_P(n^{d_X - 1/2} h^{-1}) + O_P(h), \tag{5}$$

where $\mu_0 = 1$.

By Assumptions (A3) and (A4), we have

$$\begin{aligned} D_{j,t,s}(x_j, y_j) &:= |f_{j,t,s}(x_j, y_j) - f_j(x_j) f_j(y_j)| \\ &= \left| \int (f_{X,t,s}(x_j, x', y_j, y') - f_X(x_j, x') f_X(y_j, y')) dx' dy' \right| \\ &\leq G_X |s-t|^{2d_X - 1} \int \|f'_X(x_j, x')\| \|f'_X(y_j, y')\| dx' dy' \end{aligned} \tag{6}$$

for any $x_j, y_j \in \mathbf{R}$ as $|s-t| \to \infty$, where $f_{j,t,s}$ is the joint density function of the random variables X_{tj} and X_{sj}, and $x' = (x_1 \cdots, x_{j-1}, x_{j+1}, \cdots, x_p)^T$, and $y' = (y_1 \cdots, y_{j-1}, y_{j+1}, \cdots, y_p)^T$. Similar arguments also yield

$$\begin{aligned} & D_{j,t,s,k}(x_j, y_j, z_j) := |f_{j,t,s,k}(x_j, y_j, z_j) - f_{j,t,s}(x_j, y_j) f_j(z_j)| \\ \leq\ & G'_X\Big(|k-t|^{2d_X-1} f_j(y_j) \int \|f'_X(x_j, x')\| \|f'_X(z_j, z')\| dx' dz' \\ & + |k-s|^{2d_X-1} f_j(x_j) \int \|f'_X(y_j, y')\| \|f'_X(z_j, z')\| dy' dz'\Big) \qquad (7) \end{aligned}$$

for any $x_j, y_j, z_j \in \mathbf{R}$ as $|s-t| \to \infty$, $|k-t| \to \infty$ and $|k-s| \to \infty$, where $f_{j,t,s,k}$ is the joint density function of the random variables X_{tj}, X_{sj} and X_{kj}.

First

$$E(U_l) = \int u^l K(u) f_j(x_j + hu) du.$$

Hence

$$\begin{aligned} \sup_{x_j \in \mathbf{R}} |E(U_l) - f_j(x_j)\mu_l| &= \sup_{x_j \in \mathbf{R}} \left| \int u^l K(u)(f_j(x_j + hu) - f_j(x_j)) du \right| \\ &\leq \int |u^l K(u)| \sup_{x_j \in \mathbf{R}} |f_j(x_j + hu) - f_j(x_j)| du \\ &= O(h). \qquad (8) \end{aligned}$$

Moreover, since $U_l = \frac{1}{nh} \sum_{t=1}^n \int e^{-ir(X_{tj} - x_j)/h} \Phi_l(r) dr$, then

$$\begin{aligned} & \sup_{x_j \in \mathbf{R}} |U_l - E(U_l)| \\ =\ & \sup_{x_j \in \mathbf{R}} \left| \frac{1}{nh} \sum_{t=1}^n \left\{ \int e^{-ir(X_{tj} - x_j)/h} \Phi_l(r) dr - E\left[\int e^{-ir(X_{tj} - x_j)/h} \Phi_l(r) dr \right] \right\} \right| \\ \leq\ & \int \left| \frac{1}{nh} \sum_{t=1}^n \left\{ e^{-irX_{tj}/h} - E[e^{-irX_{tj}/h}] \right\} \right| \sup_{x_j \in \mathbf{R}} |e^{irx_j/h}| |\Phi_l(r)| dr \\ \leq\ & \int \left| \frac{1}{n} \sum_{t=1}^n \left\{ e^{-irX_{tj}} - E[e^{-irX_{tj}}] \right\} \right| |\Phi_l(rh)| dr. \end{aligned}$$

By (6), for some large enough N,

$$\begin{aligned} & Var\left(\frac{1}{n} \sum_{t=1}^n \cos(rX_{tj}) \right) \\ =\ & \frac{1}{n^2} \sum_{t=1}^n Var(\cos(rX_{tj})) + \frac{1}{n^2} \sum_{t \neq s} \mathrm{Cov}(\cos(rX_{tj}), \cos(rX_{sj})) \\ \leq\ & Cn^{-1} + \frac{1}{n^2} \left(\sum_{0 < |s-t| \leq N} + \sum_{|s-t| > N} \right) \int \cos(rx) \cos(ry) \{f_{j,t,s}(x,y) - f_j(x) f_j(y)\} dx dy \\ \leq\ & Cn^{-1} + Cn^{-2} \sum_{|s-t| > N} \left(|s-t|^{2d_X-1} + o(|s-t|^{2d_X-1}) \right) = O(n^{2d_X-1}). \end{aligned}$$

The same inequality holds with cos(·) replaced by sin(·). Hence,

$$\sup_{r\in\mathbf{R}} E\Big|\frac{1}{n}\sum_{t=1}^{n}\{e^{-irX_{tj}}-E[e^{-irX_{tj}}]\}\Big| = O(n^{d_X-1/2}).$$

This, together with the fact that $\int|\Phi_l(rh)|dr = O(h^{-1})$ by Assumption (A6), implies

$$\sup_{x_j\in\mathbf{R}}|U_l - E(U_l)| = O_P(n^{d_X-1/2}h^{-1}). \tag{9}$$

Thus (5) follows from (8) and (9). Now (5) implies that

$$U_n(X_{tj}) - f_j(X_{tj})U = O_P(n^{d_X-1/2}h^{-1}) + O_P(h) \tag{10}$$

in the sense that each element converges in probability. From (2) and (3), we get

$$\hat{m}_j(X_{tj}) - m_j(X_{tj}) = (1,0)\{U_n^{-1}(X_{tj})V_n^* + U_n^{-1}(X_{tj})R_n\},$$

where $R_n = (R_0, R_1)^T$, $R_l = \frac{1}{n}\sum_{s=1}^{n}\left(\frac{X_{sj}-X_{tj}}{h}\right)^l K_h(X_{sj}-X_{tj})\big(m_j(X_{sj})-m_j(X_{tj})-m_j'(X_{tj})$ $(X_{sj}-X_{tj})\big)$, $l=0,1$.

Using the Taylor expansion for $m_j(\cdot)$ and Assumptions (A5)-(A6), and along similar lines of the proof of (10), we have that,

$$R_n - \frac{1}{2}\mu h^2 f_j(X_{tj})m_j''(X_{tj}) = O_P(n^{d_X-1/2}h^{-1}) + O_P(h). \tag{11}$$

This, together with (10), implies the first result of Theorem 2.1.

For any $c=(c_0,c_1)^T\in\mathbf{R^2}$, let $K_c(u)=(c_0+c_1u)K(u)$. Then, by (6), (7) and Assumptions (A2), (A3), and (A6)-(A7), for any fixed t, we have,

$$\begin{aligned}
& Var(c^TV_n^*)\\
\le\ & \frac{2c_0^2K^2(0)}{n^2h^2}E\eta_t^2E(\sigma_j(X_t))^2 + \frac{2}{n^2h^2}\sum_{\substack{1\le s\le n\\ s\ne t}}E(K_c((X_{sj}-X_{tj})/h)\sigma_j(X_s))^2E\eta_s^2\\
& +\frac{2}{n^2h^2}\sum_{\substack{1\le s\le n\\ s\ne t}}\sum_{l=1}^{n}E(\eta_s\eta_{s+l})E\left(K_c\left(\frac{X_{sj}-X_{tj}}{h}\right)K_c\left(\frac{X_{s+l,j}-X_{tj}}{h}\right)\sigma_j(X_s)\sigma_j(X_{s+l})\right)\\
\le\ & C(nh)^{-2}\\
& +\frac{C}{n^2h}\sum_{l=1}^{n}\left\{E(\sigma_j(X_s))^4\int K_c^4(u)(D_{j,s,s+l}(x_j,x_j+hu)+f_j(x_j)f_j(x_j+hu))dx_jdu\right\}^{1/2}\\
& +\frac{C}{n^2h^2}\sum_{\substack{1\le s\le n\\ s\ne t}}\sum_{l=1}^{n}|\gamma_\eta(l)|\left\{E(\sigma_j(X_s))^4E\left(K_c\left(\frac{X_{sj}-X_{tj}}{h}\right)K_c\left(\frac{X_{s+l,j}-X_{tj}}{h}\right)\right)^2\right\}^{1/2}\\
\le\ & C(nh)^{-2}+C(nh)^{-1}+C(nh)^{-1}n^{d_X-1/2}\\
& +\frac{C}{n^2h}\sum_{\substack{1\le s\le n\\ s\ne t}}\sum_{l=1}^{n}|\gamma_\eta(l)|\left\{\int K_c^2(u)K_c^2(v)f_{j,t,s,s+l}(x_j,x_j+hu,x_j+hv)dx_jdudv\right\}^{1/2}\\
\le\ & O((nh)^{-1}) + \frac{C}{n^2h}\sum_{\substack{1\le s\le n\\ s\ne t}}\sum_{l=1}^{n}l^{2d_\eta-1}\Big\{\int K_c^2(u)K_c^2(v)\big(D_{j,t,s,s+l}(x_j,x_j+hu,x_j+hv)\\
& \quad +D_{j,t,s}(x_j,x_j+hu)f_j(x_j+hv)+f_j(x_j)f_j(x_j+hu)f_j(x_j+hv)\big)dx_jdudv\Big\}^{1/2}.
\end{aligned}$$

Since Assumption (A6) implies that $\int K_c^2(u)du < \infty$, by (6), (7) and Assumption (A3), for any fixed t,

$$\frac{C}{n^2h}\sum_{\substack{1\le s\le n\\ s\ne t}}\sum_{l=1}^{n} l^{2d_\eta-1}\Big\{\int K_c^2(u)K_c^2(v)D_{j,t,s,s+l}(x_j,x_j+hu,x_j+hv)dx_jdudv\Big\}^{1/2}$$
$$\le \frac{C}{n^2h}\sum_{\substack{1\le s\le n\\ s\ne t}}\sum_{l=1}^{n} l^{2d_\eta-1}\big(l^{d_X-1/2}+|s+l-t|^{d_X-1/2}\big)=O(h^{-1}n^{2d_\eta+d_X-3/2}),$$
$$\frac{C}{n^2h}\sum_{\substack{1\le s\le n\\ s\ne t}}\sum_{l=1}^{n} l^{2d_\eta-1}\Big\{\int K_c^2(u)K_c^2(v)D_{j,t,s}(x_j,x_j+hu)f_j(x_j+hv)dx_jdudv\Big\}^{1/2}$$
$$\le \frac{C}{n^2h}\sum_{\substack{1\le s\le n\\ s\ne t}}\sum_{l=1}^{n} l^{2d_\eta-1}|s-t|^{d_X-1/2}=O(h^{-1}n^{2d_\eta+d_X-3/2}),$$

and

$$\frac{C}{n^2h}\sum_{\substack{1\le s\le n\\ s\ne t}}\sum_{l=1}^{n} l^{2d_\eta-1}\Big\{\int K_c^2(u)K_c^2(v)f_j(x_j)f_j(x_j+hu)f_j(x_j+hv)dx_jdudv\Big\}^{1/2}$$
$$\le \frac{C}{n^2h}\sum_{\substack{1\le s\le n\\ s\ne t}}\sum_{l=1}^{n} l^{2d_\eta-1}=O(h^{-1}n^{2d_\eta-1}).$$

Therefore Assumption (A8) implies that

$$Var(c^TV_n^*)=O((nh)^{-1})+O(h^{-1}n^{2d_\eta+d_X-3/2})+O(h^{-1}n^{2d_\eta-1})=o(1). \tag{12}$$

Combining this with (4) and the fact that $E(V_n^*)=0$, we have $\hat m_j(X_{tj})-m_j(X_{tj})$ converges to 0 in probability. □

Proof of Theorem 2.2. The proof will follow some of the arguments used in [22]. Hence we shall mainly indicate the extra steps that are needed for us to achieve our goal. Note that

$$\hat w=(\hat M^T\hat M)^{-1}\hat M^Ty=(\hat M^T\hat M)^{-1}\hat M^T(Mw+\varepsilon^*)=w+I_1+I_2, \tag{13}$$

where $I_1=(\hat M^T\hat M)^{-1}\hat M^T(M-\hat M)w$, and $I_2=(\hat M^T\hat M)^{-1}\hat M^T\varepsilon^*$. From Theorem 2.1, we obtain that $\hat M\sim M$ and $\hat M^T\hat M\sim M^TM$ in the sense that each element converges in probability. Moreover, we will show

$$\frac{1}{n}M^TM\xrightarrow{\mathcal P}\Lambda. \tag{14}$$

In fact,

$$\frac{1}{n}M^TM=\frac{1}{n}\sum_{t=1}^{n}M_tM_t^T,$$

and

$$E(M_tM_t^T)=\Lambda,$$

where $M_t=(m_1(X_{t1}),\cdots,m_p(X_{tp}))^T$, $t=1,\cdots,n$. Therefore, to prove (14), it suffices to show that, for any $1\le i,j\le p$,

$$\frac{1}{n}\sum_{t=1}^{n}m_i(X_{ti})m_j(X_{tj})-E(m_i(X_{ti})m_j(X_{tj}))\xrightarrow{\mathcal P}0. \tag{15}$$

By Assumptions (A3) and (A4),

$$
\begin{aligned}
&Var\left(\frac{1}{n}\sum_{t=1}^{n} m_i(X_{ti})m_j(X_{tj})\right)\\
&=\frac{1}{n}Var(m_i(X_{ti})m_j(X_{tj}))+\frac{1}{n^2}\sum_{s\neq t}\mathrm{Cov}(m_i(X_{ti})m_j(X_{tj}),m_i(X_{si})m_j(X_{sj}))\\
&\leq O(n^{-1})+\frac{C}{n}\sum_{l=1}^{n}\int m_i(x_i)m_i(y_i)m_j(x_j)m_j(y_j)\\
&\qquad\cdot(f_{X,t,t+l}(x_i,x_j,x',y_i,y_j,y')-f_X(x_i,x_j,x')f_X(y_i,y_j,y'))dx'dy'dx_idx_jdy_idy_j\\
&\leq O(n^{-1})\\
&\quad+\frac{C}{n}\sum_{l=1}^{n} l^{2d_X-1}\int |m_i(x_i)m_i(y_i)m_j(x_j)m_j(y_j)|\\
&\qquad\qquad\qquad\qquad\cdot\|f'_X(x_i,x_j,x')\|\|f'_X(y_i,y_j,y')\|dx'dy'dx_idx_jdy_idy_j\\
&\leq O(n^{-1})+\frac{C}{n}\sum_{l=1}^{n} l^{2d_X-1}=O(n^{-1})+O(n^{2d_X-1}),
\end{aligned}
$$

which implies (15).

On the other hand, Theorem 2.1 yields that, for $1\leq j\leq p$,

$$
\begin{aligned}
&(M^T(M-\hat{M})w)_j\\
&=\sum_{t=1}^{n}\sum_{k=1}^{p} w_k m_j(X_{tj})(m_k(X_{tk})-\hat{m}_k(X_{tk}))\\
&=\sum_{t=1}^{n}\sum_{k=1}^{p} w_k m_j(X_{tj})\Big(-\frac{1}{n}\sum_{s=1}^{n}(f_k(X_{tk}))^{-1}K_h(X_{sk}-X_{tk})\sigma_k(X_s)\eta_s\\
&\qquad\qquad\qquad\qquad+O_P(n^{d_X-1/2}h^{-1})+O_P(h)\Big)\\
&=-\sum_{s=1}^{n}\sum_{k=1}^{p} w_k\sigma_k(X_s)\eta_s\frac{1}{n}\sum_{t=1}^{n} m_j(X_{tj})(f_k(X_{tk}))^{-1}K_h(X_{sk}-X_{tk})\\
&\quad+O_P(n^{d_X+1/2}h^{-1})+O_P(nh).
\end{aligned}
\tag{16}
$$

When $k=j$,

$$
\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n} m_j(X_{tj})(f_k(X_{tk}))^{-1}K_h(X_{sk}-X_{tk})\\
&=\frac{K(0)}{nh}m_j(X_{sj})(f_j(X_{sj}))^{-1}+\frac{1}{n}\sum_{t\neq s} m_j(X_{tj})(f_j(X_{tj}))^{-1}K_h(X_{sj}-X_{tj}).
\end{aligned}
$$

By Assumption (A10),

$$
\Big|\frac{K(0)}{nh}m_j(X_{sj})(f_j(X_{sj}))^{-1}\Big|\leq\frac{C}{nh\inf_{1\leq s\leq n} f_j(X_{sj})}=O_P((nh)^{-1}\log n). \tag{17}
$$

Next, similarly to the proof of (12), we obtain

$$\begin{aligned}
&E\left(\frac{1}{n}\sum_{t\neq s}m_j(X_{tj})(f_j(X_{tj}))^{-1}K_h(X_{sj}-X_{tj})-m_j(X_{sj})\right)^2\\
&=\int m_j^2(x)f_j(x)dx-\frac{2}{n}\sum_{t\neq s}E(m_j(X_{tj})(f_j(X_{tj}))^{-1}K_h(X_{sj}-X_{tj})m_j(X_{sj}))\\
&\quad+\frac{1}{n^2}\sum_{t,l\neq s}E\left(m_j(X_{tj})m_j(X_{lj})(f_j(X_{tj})f_j(X_{lj}))^{-1}K_h(X_{sj}-X_{tj})K_h(X_{sj}-X_{lj})\right)\\
&=\int m_j^2(x)f_j(x)dx+R_1+R_2,\quad\text{say.}
\end{aligned}\tag{18}$$

Note that

$$\begin{aligned}
R_1 &= -\frac{2(n-1)}{n}\int m_j(x)m_j(x+hu)f_j(x+hu)K(u)dxdu\\
&\quad-\frac{2}{n}\sum_{t\neq s}\int\frac{m_j(x)m_j(x+hu)}{f_j(x)}K(u)(f_{j,t,s}(x,x+hu)-f_j(x)f_j(x+hu))dxdu\\
&\leq -2\int m_j^2(x)f_j(x)dx+O(h)+C\frac{1}{n}\sum_{l=1}^{n}l^{2d_X-1}\\
&= -2\int m_j^2(x)f_j(x)dx+O(h)+O(n^{2d_X-1}),
\end{aligned}\tag{19}$$

and

$$\begin{aligned}
R_2 &= \frac{1}{n^2}\sum_{t\neq s}\sum_{l=1}^{n}\int\frac{m_j(x-hu)m_j(x-hv)}{f_j(x-hu)f_j(x-hv)}K(u)K(v)f_{j,s,t,t+l}(x,x-hu,x-hv)dxdudv\\
&\leq \frac{1}{n^2}\sum_{t\neq s}\sum_{l=1}^{n}\int\frac{m_j(x-hu)m_j(x-hv)}{f_j(x-hu)f_j(x-hv)}K(u)K(v)f_j(x)f_j(x-hu)f_j(x-hv)dxdudv\\
&\quad+\frac{C}{n^2}\sum_{t\neq s}\sum_{l=1}^{n}\Big\{|t+l-s|^{2d_X-1}+l^{2d_X-1}\\
&\quad+\int\frac{m_j(x-hu)m_j(x-hv)}{f_j(x-hu)}K(u)K(v)\Big(f_{j,s,t}(x,x-hu)-f_j(x)f_j(x-hu)\Big)dxdudv\Big\}\\
&= \int m_j^2(x)f_j(x)dx+O(h)+O(n^{2d_X-1}).
\end{aligned}\tag{20}$$

Combining (17)–(20), we arrive at

$$\frac{1}{n}\sum_{t=1}^{n}m_j(X_{tj})(f_j(X_{tj}))^{-1}K_h(X_{sj}-X_{tj})=m_j(X_{sj})+o_P(1).\tag{21}$$

If $k\neq j$, along the same but more tedious lines of the proof of (15) and (21), we get

$$\frac{1}{n}\sum_{t=1}^{n}m_j(X_{tj})(f_k(X_{tk}))^{-1}K_h(X_{sk}-X_{tk})=\zeta_j(X_{sk})+o_P(1).\tag{22}$$

Obviously, $\zeta_j(X_{sj})=m_j(X_{sj})$. Hence, (14), together with (16), (21) and (22), yields

$$I_1=-\Lambda^{-1}\frac{1}{n}\sum_{t=1}^{n}\beta_t(1+o_P(1))+O_P(n^{d_X-1/2}h^{-1})+O_P(h),\tag{23}$$

where $\beta_t=(\beta_{t1},\cdots,\beta_{tp})^T$, and $\beta_{tj}=\eta_t\sum_{k=1}^p w_k\sigma_k(X_t)\zeta_j(X_{tk})$.

For I_2, we have

$$I_2=\Big\{\Lambda^{-1}\frac{1}{n}\sum_{t=1}^n\beta_{tj}^*+\Lambda^{-1}\frac{1}{n}(\hat M-M)^T\varepsilon^*\Big\}(1+o_P(1)),\tag{24}$$

where $\beta_t^*=(\beta_{t1}^*,\cdots,\beta_{tp}^*)^T$, and $\beta_{tj}^*=m_j(X_{tj})\sigma(X_t)\varepsilon_t$.

Again by Theorem 2.1, we obtain, for $1\le j\le p$,

$$\begin{aligned}&\Big(\frac{1}{n}(\hat M-M)^T\varepsilon^*\Big)_j\\ &=\frac{1}{n^2}\sum_{t,s=1}^n(f_j(X_{tj}))^{-1}K_h(X_{sj}-X_{tj})\sigma_j(X_s)\sigma(X_t)\varepsilon_t\eta_s+O_P(n^{d_X-1/2}h^{-1})+O_P(h)\\ &=\frac{K(0)}{n^2h}\sum_{t=1}^n(f_j(X_{tj}))^{-1}\sigma_j(X_t)\sigma(X_t)\varepsilon_t\eta_t\\ &\quad+\frac{1}{n^2}\sum_{t\ne s}(f_j(X_{tj}))^{-1}K_h(X_{sj}-X_{tj})\sigma_j(X_s)\sigma(X_t)\varepsilon_t\eta_s\\ &\quad+O_P(n^{d_X-1/2}h^{-1})+O_P(h)\\ &=Q_1+Q_2+O_P(n^{d_X-1/2}h^{-1})+O_P(h).\end{aligned}\tag{25}$$

First, Assumption (A10) implies

$$|Q_1|\le\frac{C}{n^2h\inf_{1\le j\le n}f_j(X_{tj})}\sum_{t=1}^n\sigma_j(X_t)\sigma(X_t)|\varepsilon_t\eta_t|=O_P((nh)^{-1}\log n).\tag{26}$$

In a same way, we have

$$\begin{aligned}n^4Q_2^2\quad&=O_P(n^2h^{-2}\log^2 n)+\sum_{t_1\ne s_1}\sum_{t_2\ne s_2}\varepsilon_{t_1}\varepsilon_{t_2}\eta_{s_1}\eta_{s_2}\sigma_j(X_{s_1})\sigma(X_{t_1})\sigma_j(X_{s_2})\sigma(X_{t_2})\\ &\qquad\cdot(f_j(X_{t_1,j}))^{-1}(f_j(X_{t_2,j}))^{-1}K_h(X_{s_1,j}-X_{t_1,j})K_h(X_{s_2,j}-X_{t_2,j})\\ &=O_P(n^2h^{-2}\log^2 n)+Q_{22}.\end{aligned}$$

Note that

$$\begin{aligned}EQ_{22}\quad&=\sum_{t_1\ne s_1}\sum_{t_2\ne s_2}\gamma_\varepsilon(t_1-t_2)\gamma_\eta(s_1-s_2)E\big(\sigma_j(X_{s_1})\sigma(X_{t_1})\sigma_j(X_{s_2})\sigma(X_{t_2})\\ &\qquad\cdot(f_j(X_{t_1,j}))^{-1}(f_j(X_{t_2,j}))^{-1}K_h(X_{s_1,j}-X_{t_1,j})K_h(X_{s_2,j}-X_{t_2,j})\big)\\ &\le Cn^2h^{-1}\log^2 n\sum_{l_1,l_2=1}^n l_1^{2d_\varepsilon-1}l_2^{2d_\eta-1}=O(n^{2d_\varepsilon+2d_\eta+2}h^{-1}\log^2 n).\end{aligned}$$

This implies that

$$Q_2=O_P((nh)^{-1}\log n)+O_P(n^{d_\varepsilon+d_\eta-1}h^{-1/2}\log n).\tag{27}$$

(24) to (27) lead to

$$I_2=\Lambda^{-1}\frac{1}{n}\sum_{t=1}^n\beta_t^*(1+o_P(1))+O_P(n^{d_\varepsilon+d_\eta-1}h^{-1/2}\log n)+O_P(n^{d_X-1/2}h^{-1})+O_P(h).\tag{28}$$

By (13), (23) and (28), we obtain

$$\begin{aligned}\hat{w}-w &= \Lambda^{-1}\frac{1}{n}\sum_{t=1}^{n}(\beta_t^*-\beta_t)(1+o_P(1))\\ &\quad +O_P(n^{d_\varepsilon+d_\eta-1}h^{-1/2}\log n)+O_P(n^{d_X-1/2}h^{-1})+O_P(h). \qquad (29)\end{aligned}$$

We are now ready to prove (a). For any $c=(c_1,\cdots,c_p)^T$,

$$n^{1/2-d_\varepsilon}\frac{1}{n}\sum_{t=1}^{n}c^T(\beta_t^*-\beta_t)=n^{-1/2-d_\varepsilon}\sum_{t=1}^{n}\varepsilon_t c^T M_t\sigma(X_t)-n^{-1/2-d_\varepsilon}\sum_{t=1}^{n}\eta_t W_t,$$

where $W_t=\sum_{j=1}^{p}\sum_{k=1}^{p}w_k\sigma_k(X_t)c_j\zeta_j(X_{tk})$. Observe that

$$\begin{aligned}Var\left(\sum_{t=1}^{n}\eta_t W_t\right) &= nE\eta_t^2EW_t^2+\sum_{t\neq s}\gamma_\eta(t-s)E(W_tW_s)\\ &\le O(n)+Cn\sum_{l=1}^{n}l^{2d_\eta-1}=O(n)+O(n^{2d_\eta+1}).\end{aligned}$$

Since $d_\eta<d_\varepsilon$, we have

$$n^{-1/2-d_\varepsilon}\sum_{t=1}^{n}\eta_tW_t=O_P(n^{-d_\varepsilon})+O_P(n^{d_\eta-d_\varepsilon})=o_P(1). \qquad (30)$$

By (29), (30), the Slutsky theorem and the assumption that $n^{1/2-d_\varepsilon}h=o(1)$, $n^{d_X-d_\varepsilon}h^{-1}=o(1)$ and $n^{d_\eta-1/2}h^{-1/2}\log n=o(1)$, it suffices to show that

$$n^{-1/2-d_\varepsilon}\sum_{t=1}^{n}\varepsilon_tc^TM_t\sigma(X_t)\xrightarrow{\mathcal{D}}c^T\Theta_1Z_1. \qquad (31)$$

Let $S_t=\sum_{j=1}^{p}c_jm_j(X_{tj})\sigma(X_t)$. We obtain

$$\sum_{t=1}^{n}\varepsilon_tc^TM_t\sigma(X_t)=\sum_{t=1}^{n}\varepsilon_tES_t+\sum_{t=1}^{n}\varepsilon_t(S_t-ES_t). \qquad (32)$$

First we show that

$$\sum_{t=1}^{n}\varepsilon_t(S_t-ES_t)=o_P(n^{1/2+d_\varepsilon}). \qquad (33)$$

Notice that $ES_t^2=E(c^TM_t\sigma(X_t))^2<\infty$ and

$$\begin{aligned}&E\left(\sum_{t=1}^{n}\varepsilon_t(S_t-ES_t)\right)^2\\ &\le\sum_{t=1}^{n}E\varepsilon_t^2ES_t^2+\sum_{t\neq s}\gamma_\varepsilon(t-s)\mathrm{Cov}(S_t,S_s)\\ &\le O(n)+\sum_{t\neq s}\gamma_\varepsilon(t-s)\sum_{j,k=1}^{p}c_jc_k\int m_j(x_j)m_k(y_k)\sigma(x_j,x')\sigma(y_k,y')\\ &\qquad\cdot(f_{X,t,s}(x_j,x',y_k,y')-f_X(x_j,x')f_X(y_k,y'))dx'dy'dx_jdy_k\\ &\le O(n)+Cn\sum_{l=1}^{n}l^{2d_\varepsilon-1}l^{2d_X-1}=O(n)+O(n^{2d_\varepsilon+2d_X}),\end{aligned}$$

which implies (33).

Moreover,

$$E(S_t) = c^T\Theta_1. \tag{34}$$

Combining (32)–(34), we show (31) and this completes the proof of the first result of Theorem 2.2.

(b). In a similar way, we obtain

$$n^{-1/2-d_\eta}\sum_{t=1}^n \varepsilon_t c^T M_t\sigma(X_t) = O_P(n^{-d_\eta}) + O_P(n^{d_\varepsilon-d_\eta}) = o_P(1) \tag{35}$$

due to the assumption that $d_\varepsilon < d_\eta$, and

$$n^{-1/2-d_\eta}\sum_{t=1}^n \eta_t W_t = n^{-1/2-d_\eta}\sum_{t=1}^n \eta_t EW_t + o_P(1). \tag{36}$$

Then (b) follows from (35), (36) and the fact that $\mathrm{E}W_t = c^T\Theta_2$.

(c). Again, for any $c = (c_1, \cdots, c_p)^T$,

$$n^{1/2-d}\frac{1}{n}\sum_{t=1}^n c^T(\beta_t^* - \beta_t) = n^{-1/2-d}\sum_{t=1}^n(\varepsilon_t ES_t - \eta_t EW_t) + o_P(1).$$

Therefore (c) is proved due to the independence of $\{\varepsilon_t\}$ and $\{\eta_t\}$. □

Proof of Theorem 2.3. We give the detailed proof of (a). (b) and (c) follow by the same arguments as in part (a), hence we only give a sketch of the proof for (b) and omit the proof of (c).

In light of Theorems 2.1 and 2.2, by (29) and (30), we have

$$\begin{aligned}
&n^{1/2-d_\varepsilon}(\hat m(x) - m_w(x))\\
&= n^{1/2-d_\varepsilon}\sum_{j=1}^p(\hat w_j - w_j)\hat m_j(x_j) + n^{1/2-d_\varepsilon}\sum_{j=1}^p w_j(\hat m_j(x_j) - m_j(x_j))\\
&= n^{-1/2-d_\varepsilon}\sum_{t=1}^n \varepsilon_t\sum_{j=1}^p \sigma(X_t)(\Lambda^{-1}M_t)_j\hat m_j(x_j)\\
&\quad + n^{-1/2-d_\varepsilon}\sum_{t=1}^n\eta_t\sum_{j=1}^p w_j(f_j(x_j))^{-1}K_h(X_{tj}-x_j)\sigma_j(X_t) + o_P(1)\\
&= n^{-1/2-d_\varepsilon}\sum_{t=1}^n \varepsilon_t\sum_{j=1}^p \sigma(X_t)(\Lambda^{-1}M_t)_j\\
&\qquad\cdot\left(m_j(x_j) + (f_j(x_j))^{-1}\frac{1}{n}\sum_{s=1}^n K_h(X_{sj}-x_j)\sigma_j(X_s)\eta_s + O_P(n^{d_X-1/2}h^{-1}) + O_P(h)\right)\\
&\quad + n^{-1/2-d_\varepsilon}\sum_{t=1}^n\eta_t\sum_{j=1}^p w_j(f_j(x_j))^{-1}K_h(X_{tj}-x_j)\sigma_j(X_t) + o_P(1).
\end{aligned}$$

Similarly to the proof of (30), we can prove that

$$n^{-1/2-d_\varepsilon}\sum_{t=1}^n\eta_t\sum_{j=1}^p w_j(f_j(x_j))^{-1}K_h(X_{tj}-x_j)\sigma_j(X_t) = O_P(n^{-d_\varepsilon}h^{-1/2}\log n) + O_P(n^{d_\eta-d_\varepsilon}\log n).$$

Then (a) will follow if we show that

$$n^{-1/2-d_\varepsilon}\sum_{t=1}^n \varepsilon_t \sum_{j=1}^p \sigma(X_t)(\Lambda^{-1}M_t)_j \left(m_j(x_j)+(f_j(x_j))^{-1}\frac{1}{n}\sum_{s=1}^n K_h(X_{sj}-x_j)\sigma_j(X_s)\eta_s\right) \xrightarrow{\mathcal{D}} (\Lambda^{-1}\Theta_1)^T\Upsilon(x)Z_1.$$

Following the same arguments as in proving (31), we need to show that

$$n^{-1/2-d_\varepsilon}\sum_{t=1}^n \varepsilon_t(P_t-EP_t)=o_P(1), \tag{37}$$

where $P_t=\sum_{j=1}^p \sigma(X_t)(\Lambda^{-1}M_t)_j\left(m_j(x_j)+(f_j(x_j))^{-1}\frac{1}{n}\sum_{s=1}^n K_h(X_{sj}-x_j)\sigma_j(X_s)\eta_s\right)$. Observe that

$$\begin{aligned}
E(P_t^2) &\le 2p^2\sum_{j,k=1}^p \Lambda_{jk}^2 m_j^2(x_j)E(\sigma(X_t)m_k^2(X_{tk}))\\
&\quad +2p^2\sum_{j,k=1}^p \Lambda_{jk}^2(f_j(x_j))^{-2}E\left(\frac{1}{n}\sum_{s=1}^n m_k(X_{tk})K_h(X_{sj}-x_j)\sigma(X_t)\sigma_j(X_s)\eta_s\right)^2\\
&\le O(1)+C(nh)^{-1}\log^2 n + C\log^2 n\sum_{j,k=1}^p\frac{1}{n^2}\sum_{s_1,s_2=1}^n \gamma_\eta(s_1-s_2)\\
&\qquad \cdot E(m_k^2(X_{tk})\sigma^2(X_t)\sigma_j(X_{s_1})\sigma_j(X_{s_2})K_h(X_{s_1,j}-x_j)K_h(X_{s_2,j}-x_j))\\
&\le O(1)+O((nh)^{-1}\log^2 n)+Cn^{-1}\log^2 n\sum_{l=1}^n l^{2d_\eta-1}=O(1),
\end{aligned}$$

where Λ_{jk} is the element of the matrix Λ^{-1} in row j and column k. Thus we obtain

$$\begin{aligned}
&E\left(\sum_{t=1}^n \varepsilon_t(P_t-EP_t)\right)^2\\
&=O(n)+\sum_{t\neq s}\gamma_\varepsilon(t-s)(1+o(1))\\
&\qquad \cdot\sum_{j_1,j_2,k_1,k_2=1}^p \Lambda_{j_1,k_1}\Lambda_{j_2,k_2}m_{j_1}(x_{j_1})m_{j_2}(x_{j_2})\mathrm{Cov}(\sigma(X_t)m_{k_1}(X_{tk_1}),\sigma(X_s)m_{k_2}(X_{sk_2}))\\
&\le O(n)+\sum_{t\neq s}\gamma_\varepsilon(t-s)\sum_{j_1,j_2,k_1,k_2=1}^p |\Lambda_{j_1k_1}\Lambda_{j_2k_2}m_{j_1}(x_{j_1})m_{j_2}(x_{j_2})|\\
&\quad \cdot\int \sigma(x_{k_1},x')\sigma(y_{k_2},y')|f_{X,t,s}(x_{k_1},y_{k_2},x',y')-f_X(x_{k_1},x')f_X(y_{k_2},y')|dx'dy'dx_{k_1}dy_{k_2}\\
&\le O(n)+Cn\sum_{l=1}^n l^{2d_\varepsilon-1}l^{2d_X-1}=O(n)+O(n^{2d_\varepsilon+2d_X})
\end{aligned}$$

which implies (37). On the other hand,

$$EP_t=(\Lambda^{-1}\Theta_1)^T\Upsilon(x).$$

This completes the proof of (a).

(b). Again

$$n^{1/2-d_\eta}(\hat{m}(x)-m_w(x)) = n^{1/2-d_\eta}\sum_{j=1}^{p}(\hat{w}_j-w_j)\hat{m}_j(x_j) + n^{1/2-d_\eta}\sum_{j=1}^{p}w_j(\hat{m}_j(x_j)-m_j(x_j))$$

$$= -n^{-1/2-d_\eta}\sum_{t=1}^{n}\eta_t\sum_{j,l=1}^{p}\Lambda_{jl}m_j(x_j)\sum_{k=1}^{p}w_k\sigma_k(X_t)\zeta_l(X_{tk})$$

$$-n^{-1/2-d_\eta}\sum_{t=1}^{n}\eta_t\sum_{j,l=1}^{p}\Lambda_{jl}\sum_{k=1}^{p}w_k\sigma_k(X_t)\zeta_l(X_{tk})(f_j(x_j))^{-1}\frac{1}{n}\sum_{s=1}^{n}K_h(X_{sj}-x_j)\sigma_j(X_s)\eta_s$$

$$+n^{-1/2-d_\eta}\sum_{t=1}^{n}\eta_t\sum_{j=1}^{p}w_j(f_j(x_j))^{-1}K_h(X_{tj}-x_j)\sigma_j(X_t)+o_P(1).$$

Using arguments similar to those in the proof of (a), we show that

$$n^{-1/2-d_\eta}\sum_{t=1}^{n}\eta_t(P_t^*-EP_t^*)=o_P(1),$$

and

$$EP_t^* = w^T\Psi(x)-(\Lambda^{-1}\Theta_2)^T\Upsilon(x),$$

where $P_t^* = -\sum_{j,l,k=1}^{p}w_k\Lambda_{jl}\sigma_k(X_t)\zeta_l(X_{tk})\Big(m_j(x_j)+(f_j(x_j))^{-1}\frac{1}{n}\sum_{s=1}^{n}K_h(X_{sj}-x_j)\,\sigma_j(X_s)\eta_s\Big)$.
This completes the proof of (b). □

References

[1] D Aydin, S E Ahmed, E Yilmaz. *Estimation of semiparametric regression model with right-censored high-dimensional data*, J Stat Comput Simul, 2019, 89: 985-1004.

[2] K Benhenni, A H Hassan, Y Su. *Local polynomial estimation of regression operators from functional data with correlated errors*, J Multivariate Anal, 2019, 170: 80-94.

[3] J Beran. *Statistics for Long-Memory Processes*, Chapman and Hall, New York, 1994.

[4] J Beran, Y Feng, S Ghosh, R Kulik. *Long-memory processes, Probabilistic properties and statistical methods*, Springer, Hiedelberg, 2013.

[5] J Beran, B Steffens, S Ghosh. *On Local Trigonometric Regression Under Dependence*, J Time Ser Anal, 2018, 39: 592-617.

[6] A Berlinet, A Elamine, A Mas. *Local linear regression for functional data,* Ann Inst Statist Math, 2011, 63: 1047-1075.

[7] P Bühlmann, S van de Geer. *Statistics for High-Dimensional Data: Methods, Theory and Applications,* In: Springer Series in Statistics, Springer, 2011.

[8] X Chen, Y He. *Inference of high-dimensional linear models with time-varying coefficients,* Statist Sinica, 2018, 28: 255-276.

[9] P Doukhan, G Oppenheim, M S Taqqu. *Theory and Applications of Long-Range Dependence,* Birkhäuser, Boston, 2003.

[10] J Fan. *Design-adaptive Nonparametric Regression*, J Amer Statist Assoc, 1992, 87: 998-1004.

[11] J Fan, T Gijbels. *Local Polynomial Modelling and its Applications*, Chapman and Hall, London, 1996.

[12] J Fan, J Lv. *A selective overview of variable selection in high dimensional feature space,* Stat Sinica, 2010, 20: 101-148.

[13] J Fan, Q Yao. *Nonlinear Time Series: Nonparametric and Parametric Methods*, Springer, New York, 2003.

[14] M Feng, Z Chen, X Cheng. *Two-step variable selection in partially linear additive models with time series data,* Comm Statist Simulation Comput, 2018, 47: 661-671.

[15] S Gaiffas, G Lecué. *Optimal rates and adaptation in the single-index model using aggregation*, Electron J Stat, 2007, 1: 538-573.

[16] L Giraitis, H L Koul, D Surgailis. *Asymptotic normality of regression estimators with long memory errors*, Statist Probab Lett, 1996, 29: 317-335.

[17] L Giraitis, H L Koul, D Surgailis. *Large sample inference for long memory processes*, Imperial College Press, London, 2012.

[18] C W J Granger, N Hyung. *Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns*, J Empirical Finance, 2004, 11: 399-421.

[19] D Guégan. *How can we define the concept of long memory? An econometric survey*, Econometric Rev, 2005, 24: 113-149.

[20] W Härdle. *Applied Nonparametric Regression*, In: *Econometric Society Monograph Series*, vol 19, Cambridge University Press, Boston, 1990.

[21] T Honda. *Nonparametric estimation of conditional medians for linear and related processes.*, Ann Inst Statist Math, 2010, 62: 995-1021.

[22] D Li, O Linton, Z Lu. *A flexible semiparametric forecasting model for time series*, J Econometrics, 2015, 187: 345-357.

[23] Q Li, J Racine. *Cross-validated local linear nonparametric regression,* Stat Sinica, 2004, 14: 485-512.

[24] X Li, L Wang, D Nettleton. *Sparse model identification and learning for ultra-high-dimensional additive partially linear models,* J Multivariate Anal, 2019, 173: 204-228.

[25] H Liang, G H Zou, A T K Wan, X Y Zhang. *Optimal weight choice for frequentist model average estimators*, J Amer Statist Assoc, 2011, 106: 1053-1066.

[26] S Ma, L Yang. *Spline-backfitted kernel smoothing of partially linear additive model*, J Statist Plann Inference, 2011, 141: 204-219.

[27] E Masry, J Fan. *Local polynomial estimation of regression functions for mixing processes*, Scand J Statist, 1997, 24: 165-179.

[28] E Masry, J Mielniczuk. *Local linear regression estimation for time series with long-range dependence*, Stochastic Process Appl, 1999, 82: 173-193.

[29] U U Müller, A Schick, W Wefelmeyer. *Estimating the innovation distribution in nonparametric autoregression*, Probab Theory Related Fields, 2009, 144: 53-77.

[30] W Palma. *Long-Memory Time Series, Theory and Methods*, In: *Wiley Series in Probability and Statistics*, John Wiley and Sons, Hoboken, 2007.

[31] P M Robinson. *Asymptotic theory for nonparametric regression with spatial data*, J Econometrics, 2011, 165: 5-19.

[32] Y K Truong. *Nonparametric curve estimation with time series errors*, J Statist Plann Inference, 1991, 28: 167-183.

[33] H Wang, J Wang. *Estimation of trend function for spatio-temporal models*, J Nonparametr Stat, 2009, 21: 567-588.

[34] L Wang. *The nonparametric estimation of long memory spatio-temporal random field models*, Sci China Math, 2015, 58: 1115-1128.

[35] T Teräsvirta, D Tjϕstheim, C Granger. *Modelling nonlinear Economic Time Series*, Oxford University Press, London, 2010.

[36] L Yang, Y Fang, J Wang, Y Shao. *Variable selection for partially linear models via learning gradients,* Electron J Stat, 2017, 11: 2907-2930.

Department of Mathematics, Nanjing University, Nanjing 210093, China.
Email: lhwang@nju.edu.cn