

# LOCAL LINEAR FITTING UNDER NEAR EPOCH DEPENDENCE \*

Zudi Lu<sup>1,2</sup>     Oliver Linton<sup>3</sup>

<sup>1</sup>Department of Statistics, London School of Economics, London, WC2A 2AE, UK

<sup>2</sup>Institute of Systems Science, Academy of Mathematics and Systems Sciences  
Chinese Academy of Sciences, Beijing 100080, China

<sup>3</sup>Department of Economics, London School of Economics, London, WC2A 2AE, UK

## Abstract

Local linear fitting in modelling of nonlinear processes under strong (i.e.,  $\alpha$ -) mixing conditions has been investigated extensively. However, it is often a difficult step to establish the strong mixing of a nonlinear process composed of several parts such as the popular combination of ARMA and GARCH models. In this paper we develop an asymptotic theory of local linear fitting for near-epoch dependent (NED) processes. We establish the pointwise asymptotic normality of the local linear kernel estimators under some restrictions on the amount of dependence. Simulations and application examples illustrate that the proposed approach can work quite well for the medium size of economic time series.

*Key words:*  $\alpha$ -mixing, kernel estimation, near-epoch dependence stable process, central limit theorem, local linear fitting, asymptotic normality, simulation.

---

\*Partially supported by a Leverhulme Trust research grant and the National Natural Science Foundation of China and the Economic and Social Science Research Council of UK.

# 1 Introduction

In this paper we consider local linear modelling in a time series context under near epoch dependence. Andrews (1995) established uniform convergence with rates for nonparametric density and regression estimators based on the local constant estimation paradigm also under near epoch dependence conditions. The purpose of this study is to provide a Central Limit Theorem for the more desirable class of local linear estimators (Fan and Gijbels (1996)) under similar weak dependence conditions.

Assume that  $\{(Y_t, \mathbf{X}_t)\}$  is a  $\mathbb{R}^{1+d}$ -valued stationary sequence (with  $Y_t$  being  $\mathbb{R}^1$ -valued and  $\mathbf{X}_t$  being  $\mathbb{R}^d$ -valued), defined on some probability space  $(\Omega, \mathcal{F}, P)$  (throughout the paper all the random variables are defined on this space). Among the widely used mixing conditions, such as  $\phi$ -,  $\rho$ -,  $\beta$ - and  $\alpha$ -mixings,  $\alpha$ -mixing is no doubt the weakest and most popular in the econometric literature. Under some suitable conditions, the stationary solutions of many time series econometric models (linear or nonlinear) are  $\alpha$ -mixing; see, for example, Goródeskii (1977), Pham (1986), Pham and Tran (1985), Tøstheim (1990) and Tong (1990) on nonlinear AR models, and Masry and Tøstheim (1995), Lu (1998), Cline and Pu (1999), Lu and Jiang (2001), Carrasco and Chen (2002), and Saikkonen (2001) on nonlinear ARCH/GARCH models. This mixing condition is used extensively in the time series literature (c.f., Fan and Yao, 2003). Masry and Fan (1997) establish the asymptotic normality of local polynomial regression estimators under this condition. For reference, its definition is stated as follows.

**Definition 0.** A stationary sequence  $\{X_t, t = 0, \pm 1, \dots\}$  is said to be  $\alpha$ -mixing if

$$\alpha(k) = \sup_{A \in \mathcal{F}_{-\infty}^n, B \in \mathcal{F}_{n+k}^\infty} |P(AB) - P(A)P(B)| \rightarrow 0 \quad (1.1)$$

as  $k \rightarrow \infty$ , where  $\mathcal{F}_{-\infty}^n$  and  $\mathcal{F}_{n+k}^\infty$  are two  $\sigma$ -fields generated by  $\{X_t, t \leq n\}$  and  $\{X_t, t \geq n+k\}$ , respectively. We call  $\alpha(\cdot)$  the mixing coefficient.

However, from a practical point of view, the  $\alpha$ -mixing concept suffers from many undesirable features. As Davidson (1994) points out, the following shortcomings are often serious (see also Lu (2001)): (i) even simple autoregressive processes might not be  $\alpha$ -mixing, (ii)  $\alpha$ -mixing is hard to verify in practice, especially in the case of compound processes. For the former case, Andrews (1984) showed that the stationary solution to a simple linear AR(1) model of the form

$X_t = 1/2X_{t-1} + e_t$ , with  $e_t$ 's being independent symmetric Bernoulli random variables taking values  $-1$  and  $1$ , is not  $\alpha$ -mixing (but it is near epoch dependent, defined in Definition 1 below). For the latter case, the ARMA process with ARCH/GARCH errors, discussed in Engle (1982) and Weiss (1984) as well as Lin and Li (1997), is well applied in financial econometrics, where the model is composed of two time series (ARMA and ARCH/GARCH) models:

$$X_t = a_1X_{t-1} + \cdots + a_pX_{t-p} + \varepsilon_t - b_1\varepsilon_{t-1} - \cdots - b_q\varepsilon_{t-q}, \quad (1.2)$$

$$\varepsilon_t = e_t h_t^{1/2}, \quad h_t = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \cdots + \alpha_P\varepsilon_{t-P}^2 + \beta_1 h_{t-1} + \cdots + \beta_Q h_{t-Q} \quad (1.3)$$

where  $a_i$  and  $b_i$  are the coefficients in the  $ARMA(p, q)$  model, and  $\alpha_i$  and  $\beta_i$  are the coefficients in the  $GARCH(P, Q)$  model, with  $e_t$  being *i.i.d* innovation with mean 0 and variance 1. Although the ARCH model and its generalized version, GARCH, by Bollerslev (1986) have been proved to be  $\alpha$ -mixing under some mild conditions (c.f., Lu (1996a, b), Carrasco and Chen (2002)), no results exist to examine whether the compound processes are  $\alpha$ -mixing. For this reason we use a generalized version of mixing processes, called stable or near epoch dependent processes, which can easily cover the compounded processes and many nonlinear/non- $\alpha$ -mixing processes. This concept was introduced in Ibragimov (1962) and was developed further by Billingsley (1968) and McLeish (1975a, 1975b, 1977). It has been used extensively in econometrics following Bierens (1981), see for example Gallant (1987), Gallant and White (1988), and Andrews (1995). Lu (2001) established asymptotic normality for kernel density estimators under this condition. Nze, Bühlmann and Doukhan (2002) have investigated an alternative class of dependent processes they call 'weak dependent'. They establish the asymptotic normality of local constant nonparametric regression estimators under their conditions.

Let  $Y_t$  and  $\mathbf{X}_t$  be both stationary processes, of  $\mathbb{R}^1$ - and  $\mathbb{R}^d$ - valued, respectively, defined based on a stationary process  $\{\varepsilon_t\}$  by

$$Y_t = \Psi_Y(\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots), \quad (1.4)$$

$$\mathbf{X}_t = (X_{t1}, \dots, X_{td})^T = \Psi_X(\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots), \quad (1.5)$$

where  $\mathbf{X}^T$  denotes the transpose of  $\mathbf{X}$  (a vector or matrix),  $\Psi_Y : \mathbb{R}^\infty \longrightarrow \mathbb{R}^1$  and  $\Psi_X : \mathbb{R}^\infty \longrightarrow \mathbb{R}^d$  are two Borel measurable functions, respectively, and  $\{\varepsilon_t\}$  may be vector-valued. Let  $\nu > 0$  be a positive real number.

**Definition 1.** The stationary process  $\{(Y_t, \mathbf{X}_t)\}$  is said to be *near epoch dependent in  $L_\nu$  norm* (NED in  $L_\nu$  for simplicity) with respect to a stationary  $\alpha$ -mixing process  $\{\varepsilon_t\}$ , if

$$v_\nu(m) = E|Y_t - Y_t^{(m)}|^\nu + E\|\mathbf{X}_t - \mathbf{X}_t^{(m)}\|^\nu \rightarrow 0 \quad (1.6)$$

as  $m \rightarrow \infty$ , where  $|\cdot|$  and  $\|\cdot\|$  are the absolute value and the Euclidean norm of  $\mathbb{R}^d$ , respectively,  $Y_t^{(m)} = \Psi_{Y,m}(\varepsilon_t, \dots, \varepsilon_{t-m+1})$ ,  $\mathbf{X}_t^{(m)} = (X_{t1}^{(m)}, \dots, X_{td}^{(m)})^\tau = \Psi_{X,m}(\varepsilon_t, \dots, \varepsilon_{t-m+1})$ , and  $\Psi_{Y,m}$  and  $\Psi_{X,m}$  are  $\mathbb{R}^1$ - and  $\mathbb{R}^d$ -valued Borel measurable functions with  $m$  arguments involved, respectively. We will call  $v_\nu(m)$  the *stability coefficients of order  $\nu$*  of the process  $\{(Y_t, \mathbf{X}_t)\}$ .

Clearly,  $\{(Y_t^{(m)}, \mathbf{X}_t^{(m)})\}$  is an  $\alpha$ -mixing process with mixing coefficient

$$\alpha_m(k) \leq \begin{cases} \alpha(k-m) & k \geq m+1, \\ 1 & k \leq m. \end{cases} \quad (1.7)$$

The type of setting where our results are useful is for models with complicated dynamics in both mean and variance for which the usual mixing conditions do not necessarily apply. These sorts of models are common in finance and economics.

Our limiting results will resemble conventional limiting results for local linear estimators under more standard conditions. That is, under some restrictions on  $\alpha(k)$  and  $v_\nu(m)$  we obtain the optimal pointwise rate of convergence for i.i.d. data (Stone (1980)) along with the same asymptotic distribution that would obtain were the data i.i.d. with the same marginal distribution. This type of result is to be contrasted with those obtained by Phillips and Park (1998) and Karlsen and Tjøstheim (2001) for unit root or null recurrent processes (see also Bandi (2004) for near-integrated processes) for which the rates of convergence are slower and limiting distributions are non-normal.

In the next section we define the setting and estimator we shall examine. The main asymptotic results are given in section 3. In section 4 we provide some numerical results based on some common econometric models.

## 2 Methodology

### 2.1 Notation and main assumptions

We summarize here the main assumptions we are making on the data generating process (DGP) (1.6) and the kernel  $K$  to be used in the estimation method. Assumptions (A1)-(A4) are related

to the nonlinear process itself.

(A1) The DGP is a strictly stationary NED process (c.f., (1.6)), with order  $\nu = 2 + \delta/2$ , with respect to the  $\alpha$ -mixing process  $\{\varepsilon_t\}$ , where the constant  $\delta > 0$  is specified in Assumption (A2) below. For all  $i$  and  $j$  in  $\mathbb{Z}$ , the vectors  $\mathbf{X}_i$  and  $\mathbf{X}_j$  admit a joint density  $f_{ij}$ ; moreover,  $f_{ij}(\mathbf{x}', \mathbf{x}'') \leq C$  for all  $i, j \in \mathbb{Z}$ , all  $\mathbf{x}', \mathbf{x}'' \in \mathbb{R}^d$ , where  $C > 0$  is some constant, and  $f$  denotes the marginal density of  $\mathbf{X}_i$ .

(A2) The random variable  $Y_i$  has finite absolute moment of order  $(2 + \delta)$ , that is,  $E[|Y_i|^{2+\delta}] < \infty$  for some  $\delta > 0$ .

(A3) The regression function  $g(\mathbf{x}) = E(Y_t | \mathbf{X}_t = \mathbf{x})$  is twice differentiable. Denoting by  $g'(\mathbf{x})$  and  $g''(\mathbf{x})$  its gradient and the matrix of its second derivatives (at  $\mathbf{x}$ ), respectively,  $\mathbf{x} \mapsto g''(\mathbf{x})$  is continuous at all  $\mathbf{x}$ .

Assumption (A4) is an assumption of the mixing coefficients.

(A4) For the mixing coefficient of  $\varepsilon_t$ , the function  $\alpha$  is such that

$$\lim_{k \rightarrow \infty} k^a \sum_{j=k}^{\infty} \{\alpha(j)\}^{\delta/(4+\delta)} = 0 \quad \text{for some constant } a > \delta/(4 + \delta).$$

Assumption (A5) deals with the kernel function  $K : \mathbb{R}^d \rightarrow \mathbb{R}$ , to be used in the estimation method. For any  $\mathbf{c} := (c_0, \mathbf{c}_1^\tau)^\tau \in \mathbb{R}^{d+1}$ , define

$$K_{\mathbf{c}}(\mathbf{u}) := (c_0 + \mathbf{c}_1^\tau \mathbf{u})K(\mathbf{u}). \quad (2.1)$$

(A5)(i) For any  $\mathbf{c} \in \mathbb{R}^{d+1}$ ,  $|K_{\mathbf{c}}(\mathbf{u})|$  is uniformly bounded by some constant  $K_{\mathbf{c}}^+$ , and is integrable:

$$\int_{\mathbb{R}^d} |K_{\mathbf{c}}(\mathbf{x})| d\mathbf{x} < \infty.$$

(ii) For any  $\mathbf{c} \in \mathbb{R}^{d+1}$ ,  $|K_{\mathbf{c}}|$  has an integrable second order radial majorant, that is,  $Q_{\mathbf{c}}^K(\mathbf{x}) := \sup_{\|\mathbf{y}\| \geq \|\mathbf{x}\|} [\|\mathbf{y}\|^2 K_{\mathbf{c}}(\mathbf{y})]$  is integrable.

(iii) For any  $\mathbf{c} \in \mathbb{R}^{d+1}$ ,  $K_{\mathbf{c}}$  is Lipschitz continuous of order 1, that is, for some constant  $C > 0$ ,

$$|K_{\mathbf{c}}(\mathbf{u}) - K_{\mathbf{c}}(\mathbf{v})| \leq C \|\mathbf{u} - \mathbf{v}\| \text{ for any } \mathbf{u}, \mathbf{v} \in \mathbb{R}^d.$$

This assumption allows an unbounded support for the kernel function; compare this with Condition 2(i) of Masry and Fan (1997, page 170) who require the kernel function to have a bounded support.

Throughout, we assume that the observations of the NED process  $\{(Y_t, \mathbf{X}_t)\}$  are  $(Y_t, \mathbf{X}_t)$ ,  $t = 1, 2, \dots, T$ . For convenient reference, we are listing here some conditions on the asymptotic behavior, as  $T \rightarrow \infty$ , of the bandwidth  $b_T$  that will be used for generality in the sequel (c.f., Lemma 3.4 below), where Assumption (B1) below is standard, while Assumptions (B2) through (B4) below look complex: some simple and verifiable conditions on the stability and mixing coefficients to ensure they hold will be given in the main theorem (Theorem 3.1) and its corollary (Corollary 3.1).

(B1) The bandwidth  $b_T$  tends to zero in such a way that  $Tb_T^d \rightarrow \infty$  as  $T \rightarrow \infty$ .

(B2) There is a positive integer  $m = m_T \rightarrow \infty$  such that the stability coefficients, defined in (1.6) with  $\nu = 2$  and  $\nu = 2 + \delta/2$ , satisfy

$$T^{2+4/\delta} b_T^{-(2+d+2d/\delta)} v_2(m) \rightarrow 0, \quad \text{and} \quad b_T^{-4(1+d+2d/\delta)} v_2(m) = O(1),$$

and

$$b_T^{-(2+2d+\delta/2+4d/\delta)} v_{2+\delta/2}(m) = O(1).$$

(B3) There exist two sequences of positive integer vectors,  $p = p_T \in \mathbb{Z}$  and  $q = q_T = 2m_T \in \mathbb{Z}$ , with  $m = m_T \rightarrow \infty$  such that  $p = p_T = o((Tb_T^d)^{1/2})$ ,  $q/p \rightarrow 0$  and  $T/p \rightarrow \infty$ , and  $Tp^{-1}\alpha(m) \rightarrow 0$ .

(B4)  $b_T$  tends to zero in such a manner that  $qb_T^d = O(1)$  such that

$$b_T^{-\delta d/(4+\delta)} \sum_{t=q}^{\infty} \{\alpha_m(t)\}^{\delta/(4+\delta)} \rightarrow 0 \quad \text{as } T \rightarrow \infty. \quad (2.2)$$

*Remark.* Assumption (B1) is standard on the bandwidth as in the *i.i.d.* case, Assumption (B2) is concerned with the conditions on the stability coefficients related to the bandwidth, and Assumptions (B3) and (B4) are on the mixing coefficients which are associated with the bandwidth, among which (B3) together with (B1) is similar to the conditions specified for the strongly mixing processes in Condition 3 of Masry and Fan (1997, page 172). Assumptions B2-B4

are phrased as restrictions on the decay rates of the stability and mixing coefficients for a given bandwidth, although one could rewrite these conditions as restrictions on the bandwidth (and hence the implied rate of convergence of the estimator) for a given decay rate thereby allowing greater dependence at the cost of slower convergence. Although Assumptions (B2) through (B4) look somewhat complex, some milder and more specific conditions can be derived from them with the bandwidth set as a power function of the number of observations, as is generally the case in practice. For the details, see Theorem 3.1 together with Corollary 3.1 and the remark there in Section 3.

## 2.2 Least-square local linear fitting

Although the Nadaraya-Watson method is central in most nonparametric regression method in the traditional *i.i.d.* series case, it has been well documented (see, for instance, Fan and Gijbels 1996) that this approach suffers from several severe drawbacks, such as poor boundary performances, excessive bias and low efficiency, and that the local polynomial fitting methods developed by Stone (1977) and Cleveland (1979) are generally preferable. Local polynomial fitting, and particularly its special case—local linear fitting—recently have become increasingly popular in the light of recent work by Cleveland and Loader (1996), Fan (1992), Fan and Gijbels (1992, 1995), Hastie and Loader (1993), Ruppert and Wand (1994), and several others. Masry and Fan (1997) have studied the asymptotics of local polynomial fitting for regression under general  $\alpha$ -mixing conditions; see also Fan and Yao (2003). In this paper, we extend this approach to the context of our generalized mixing dependence—NED processes by defining an estimator of  $g$  based on local linear fitting and establishing its asymptotic properties.

The idea of local linear fitting consists in approximating, in a neighborhood of  $\mathbf{x}$ , the unknown function  $g$  by a linear function. Under (A3), we have

$$g(\mathbf{z}) \approx g(\mathbf{x}) + (\mathbf{g}'(\mathbf{x}))^\tau(\mathbf{z} - \mathbf{x}) := a_0 + \mathbf{a}_1^\tau(\mathbf{z} - \mathbf{x}).$$

Locally, this suggests estimating  $(a_0, \mathbf{a}_1^\tau) = (g(\mathbf{x}), \mathbf{g}'(\mathbf{x}))$ , hence constructing an estimator of  $g$  from

$$\begin{pmatrix} g_T(\mathbf{x}) \\ g'_T(\mathbf{x}) \end{pmatrix} = \begin{pmatrix} \hat{a}_0 \\ \hat{\mathbf{a}}_1 \end{pmatrix} := \arg \min_{(a_0, \mathbf{a}_1) \in \mathbb{R}^{d+1}} \sum_{j=1}^T (Y_j - a_0 - \mathbf{a}_1^\tau(\mathbf{X}_j - \mathbf{x}))^2 K\left(\frac{\mathbf{X}_j - \mathbf{x}}{b_T}\right), \quad (2.3)$$

where  $b_T$  is a sequence of bandwidths tending to zero at appropriate rate as  $T$  tends to infinity, and  $K(\cdot)$  is a (bounded) kernel with values in  $\mathbb{R}^+$ .

In the classical *i.i.d.* and time series case, the solution of the minimization problem (2.3) is easily shown to be  $\{\mathcal{X}(\mathbf{x})^\tau \mathcal{W}(\mathbf{x}) \mathcal{X}(\mathbf{x})\}^{-1} \{\mathcal{X}(\mathbf{x})^\tau \mathcal{W}(\mathbf{x}) \mathcal{Y}\}$ , where  $\mathcal{X}(\mathbf{x})$  is an  $n \times (d+1)$  matrix with  $i$ -th row  $(1, b_T^{-1}(\mathbf{X}_i - \mathbf{x})^\tau)$ ,  $\mathcal{W}(\mathbf{x}) = b_T^{-1} \text{diag} \left( K \left( \frac{\mathbf{X}_1 - \mathbf{x}}{b_T} \right), \dots, K \left( \frac{\mathbf{X}_T - \mathbf{x}}{b_T} \right) \right)$ , and  $\mathcal{Y} = (Y_1, \dots, Y_T)^\tau$  (see, e.g., Fan and Gijbels 1996). In the generalized NED series case, though such an expression still holds, we rather write the solution to (2.3) in the following form, which is very convenient for the purpose of characterizing its limiting distribution:

$$\begin{pmatrix} \hat{a}_0 \\ \hat{\mathbf{a}}_1 b_T \end{pmatrix} = \mathbf{U}_T^{-1} \mathbf{V}_T, \quad \text{where} \quad \mathbf{V}_T := \begin{pmatrix} v_{T0} \\ \mathbf{v}_{T1} \end{pmatrix} \quad \text{and} \quad \mathbf{U}_T := \begin{pmatrix} u_{T00} & \mathbf{u}_{T01} \\ \mathbf{u}_{T10} & \mathbf{u}_{T11} \end{pmatrix},$$

with (letting  $\left(\frac{\mathbf{X}_j - \mathbf{x}}{b_T}\right)_0 := 1$ )

$$(\mathbf{V}_T)_i := (Tb_T^d)^{-1} \sum_{t=1}^T Y_j \left( \frac{\mathbf{X}_j - \mathbf{x}}{b_T} \right)_i K \left( \frac{\mathbf{X}_j - \mathbf{x}}{b_T} \right), \quad i = 0, \dots, d,$$

and

$$(\mathbf{U}_T)_{i\ell} := (Tb_T^d)^{-1} \sum_{t=1}^T \left( \frac{\mathbf{X}_j - \mathbf{x}}{b_T} \right)_i \left( \frac{\mathbf{X}_j - \mathbf{x}}{b_T} \right)_\ell K \left( \frac{\mathbf{X}_j - \mathbf{x}}{b_T} \right), \quad i, \ell = 0, \dots, d.$$

It follows that

$$\mathbf{H}_T := \begin{pmatrix} \hat{a}_0 - a_0 \\ \hat{\mathbf{a}}_1 b_T - \mathbf{a}_1 b_T \end{pmatrix} = \begin{pmatrix} g_T(\mathbf{x}) - g(\mathbf{x}) \\ (g'_T(\mathbf{x}) - g'(\mathbf{x})) b_T \end{pmatrix} = \mathbf{U}_T^{-1} \left\{ \mathbf{V}_T - \mathbf{U}_T \begin{pmatrix} a_0 \\ \mathbf{a}_1 b_T \end{pmatrix} \right\} := \mathbf{U}_T^{-1} \mathbf{W}_T, \quad (2.4)$$

where

$$\mathbf{W}_T := \begin{pmatrix} w_{T0} \\ \mathbf{w}_{T1} \end{pmatrix}, \quad (\mathbf{W}_T)_i := (Tb_T^d)^{-1} \sum_{j=1}^T Z_j \left( \frac{\mathbf{X}_j - \mathbf{x}}{b_T} \right)_i K \left( \frac{\mathbf{X}_j - \mathbf{x}}{b_T} \right), \quad i = 0, \dots, d, \quad (2.5)$$

and  $Z_j := Y_j - a_0 - \mathbf{a}_1^\tau (\mathbf{X}_j - \mathbf{x})$ .

The organization of the paper is as follows. If, under adequate conditions, we are able to show that:

(C1)  $(Tb_T^d)^{1/2}(\mathbf{W}_T - \mathbf{E}\mathbf{W}_T)$  is asymptotically normal,

(C2)  $(Tb_T^d)^{1/2}(\mathbf{E}\mathbf{W}_T - \boldsymbol{\mu}_T) \longrightarrow \mathbf{0}$  and  $\text{Var} \left( (Tb_T^d)^{1/2} \mathbf{W}_T \right) \longrightarrow \boldsymbol{\Sigma}$ , and

(C3)  $\mathbf{U}_T \xrightarrow{\text{P}} \mathbf{U}$ ,

then (2.4) and Slutsky's classical argument imply that, for all  $\mathbf{x}$  (all quantities involved indeed depend on  $\mathbf{x}$ )

$$(Tb_T^d)^{1/2} \left[ \begin{pmatrix} g_T(\mathbf{x}) - g(\mathbf{x}) \\ (g'_T(\mathbf{x}) - g'(\mathbf{x}))b_T \end{pmatrix} - \mathbf{U}^{-1}\boldsymbol{\mu}_T \right] = (Tb_T^d)^{1/2} [\mathbf{H}_T - \mathbf{U}^{-1}\boldsymbol{\mu}_T] \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, \mathbf{U}^{-1}\boldsymbol{\Sigma}\mathbf{U}^{-1\tau}).$$

This asymptotic normality result (with explicit values of  $\boldsymbol{\mu}_T$ ,  $\boldsymbol{\Sigma}$  and  $\mathbf{U}$ ), under various forms (depending on the stability assumption, the mixing assumption, the choice of the bandwidth  $b_T$ , etc.), is the main contribution of this paper; see Theorem 3.1 and its Corollary 3.1.

### 2.3 Approximations

A fundamental technique which will be widely used to study (C1)–(C3) is the following approximation to the NED process  $\{(Y_t, \mathbf{X}_t)\}$  by the  $\alpha$ -mixing process  $\{(Y_t^{(m)}, \mathbf{X}_t^{(m)})\}$  defined in Definition 1, that is

$$Y_t = Y_t^{(m)} + (Y_t - Y_t^{(m)}) := Y_t^{(m)} + \delta_{Y,t}^{(m)}, \quad (2.6)$$

$$\mathbf{X}_t = \mathbf{X}_t^{(m)} + (\mathbf{X}_t - \mathbf{X}_t^{(m)}) := \mathbf{X}_t^{(m)} + \delta_{X,t}^{(m)}, \quad (2.7)$$

where

$$E[\delta_{Y,t}^{(m)}]^2 = O(v_2(m)), \text{ and } E[\delta_{X,t}^{(m)}]^2 = O(v_2(m)), \text{ as } m \rightarrow \infty, \quad (2.8)$$

and the mixing coefficients of  $\{(Y_t^{(m)}, \mathbf{X}_t^{(m)})\}$  satisfy

$$\alpha_m(k) \leq 1 \text{ for } k = 0, 1, \dots, m, \quad \text{and } \alpha_m(k) = \alpha(k - m) \text{ for } k \geq m + 1, \quad (2.9)$$

with  $\alpha(\cdot)$  defined in Definition 0.

On the bases of (2.6) and (2.7), we can construct the approximations to  $\mathbf{U}_T$  and  $\mathbf{W}_T$ , respectively, by:

$$\mathbf{U}_T^{(m)} := \begin{pmatrix} u_{T00}^{(m)} & \mathbf{u}_{T01}^{(m)} \\ \mathbf{u}_{T10}^{(m)} & \mathbf{u}_{T11}^{(m)} \end{pmatrix}, \quad \text{and} \quad \mathbf{W}_T^{(m)} := \begin{pmatrix} w_{T0}^{(m)} \\ \mathbf{w}_{T1}^{(m)} \end{pmatrix},$$

with (letting  $\left(\frac{\mathbf{X}_j^{(m)} - \mathbf{x}}{b_T}\right)_0 := 1$ )

$$\left(\mathbf{U}_T^{(m)}\right)_{i\ell} := (Tb_T^d)^{-1} \sum_{j=1}^T \left(\frac{\mathbf{X}_j^{(m)} - \mathbf{x}}{b_T}\right)_i \left(\frac{\mathbf{X}_j^{(m)} - \mathbf{x}}{b_T}\right)_\ell K\left(\frac{\mathbf{X}_j^{(m)} - \mathbf{x}}{b_T}\right), \quad i, \ell = 0, \dots, d;$$

and

$$\left(\mathbf{W}_T^{(m)}\right)_i := (Tb_T^d)^{-1} \sum_{j=1}^T Z_j^{(m)} \left(\frac{\mathbf{X}_j^{(m)} - \mathbf{x}}{b_T}\right)_i K\left(\frac{\mathbf{X}_j^{(m)} - \mathbf{x}}{b_T}\right), \quad i = 0, \dots, d, \quad (2.10)$$

and  $Z_j^{(m)} := Y_j^{(m)} - a_0 - \mathbf{a}_1^\tau (\mathbf{X}_j^{(m)} - \mathbf{x})$ .

We have the following lemma on the approximations.

**Lemma 2.1** *Let  $m = m_T$  be a positive integer tending to  $\infty$  and  $L = L_T$  a positive real number tending to  $\infty$  as  $T \rightarrow \infty$ . Then under Assumptions (A2) and (A5), as  $T \rightarrow \infty$ ,*

$$E \left| \left(\mathbf{W}_T^{(m)}\right)_i - (\mathbf{W}_T)_i \right| = O \left( b_T^{-d-1} L \sqrt{v_2(m)} \right) + o \left( b_T^{-d} L^{-(1+\delta)} \right), \quad (2.11)$$

$$E \left| \left(\mathbf{U}_T^{(m)}\right)_{i\ell} - (\mathbf{U}_T)_{i\ell} \right| = O \left( b_T^{-d-1} L \sqrt{v_2(m)} \right) + o \left( b_T^{-d} L^{-(1+\delta)} \right), \quad (2.12)$$

for  $i, \ell = 0, 1, \dots, d$ .

The proof of this lemma is relegated to Section 5.2. □

### 3 Asymptotic results

We begin with some preliminaries on the asymptotic bias and variance.

#### 3.1 Preliminaries

Claim (C3) is easily established from the following lemma, the proof of which is similar to that of Lemma 3.2 below, and is therefore omitted.

**Lemma 3.1** *Assume that Assumptions (A1), (A4), and (A5) hold, that  $b_T$  satisfies Assumptions (B1) and (B2). Then, for all  $\mathbf{x}$ ,*

$$\mathbf{U}_T \xrightarrow{\mathbb{P}} \mathbf{U} := \begin{pmatrix} f(\mathbf{x}) \int K(\mathbf{u}) d\mathbf{u} & f(\mathbf{x}) \int \mathbf{u}^\tau K(\mathbf{u}) d\mathbf{u} \\ f(\mathbf{x}) \int \mathbf{u} K(\mathbf{u}) d\mathbf{u} & f(\mathbf{x}) \int \mathbf{u} \mathbf{u}^\tau K(\mathbf{u}) d\mathbf{u} \end{pmatrix}$$

as  $T \rightarrow \infty$ .

The remainder of this subsection is devoted to claim (C2). The usual Cramér-Wold device will be adopted. For all  $\mathbf{c} := (c_0, \mathbf{c}_1^\tau)^\tau \in \mathbb{R}^{1+d}$ , let

$$A_T := (Tb_T^d)^{1/2} \mathbf{c}^\tau \mathbf{W}_T = (Tb_T^d)^{-1/2} \sum_{j=1}^T Z_j K_{\mathbf{c}} \left( \frac{\mathbf{X}_j - \mathbf{x}}{b_T} \right),$$

with  $K_{\mathbf{c}}(\mathbf{u})$  defined in (2.1). The following lemma provides the asymptotic variance of  $A_T$  for all  $\mathbf{c}$ , hence that of  $(\mathbf{T}b_T^d)^{1/2} \mathbf{W}_T$ .

**Lemma 3.2** *Assume that Assumptions (A1), (A2), (A4), and (A5) hold, and that  $b_T$  satisfies Assumptions (B1) and (B2). Then,*

$$\lim_{T \rightarrow \infty} \text{Var}[A_T] = \text{Var}(Y_j | \mathbf{X}_j = \mathbf{x}) f(\mathbf{x}) \int_{\mathbb{R}^d} K_{\mathbf{c}}^2(\mathbf{u}) d\mathbf{u} = \mathbf{c}^\tau \boldsymbol{\Sigma} \mathbf{c}, \quad (3.1)$$

where

$$\boldsymbol{\Sigma} := \text{Var}(Y_j | \mathbf{X}_j = \mathbf{x}) f(\mathbf{x}) \begin{pmatrix} \int K^2(\mathbf{u}) d\mathbf{u} & \int \mathbf{u}^\tau K^2(\mathbf{u}) d\mathbf{u} \\ \int \mathbf{u} K^2(\mathbf{u}) d\mathbf{u} & \int \mathbf{u} \mathbf{u}^\tau K^2(\mathbf{u}) d\mathbf{u} \end{pmatrix}.$$

Hence,  $\lim_{T \rightarrow \infty} \text{Var} \left( (\mathbf{T}b_T^d)^{1/2} \mathbf{W}_T \right) = \boldsymbol{\Sigma}$ .

**Proof.** See Section 5.2 □

Next, we consider the asymptotic behavior of  $\mathbb{E}[A_T]$ .

**Lemma 3.3** *Under Assumptions (A3) and (A5),*

$$\begin{aligned} \mathbb{E}[A_T] &= \sqrt{Tb_T^d b_T^2} \frac{1}{2} f(\mathbf{x}) \text{tr} \left[ g''(\mathbf{x}) \int \mathbf{u} \mathbf{u}^\tau K_{\mathbf{c}}(\mathbf{u}) d\mathbf{u} \right] + o \left( \sqrt{Tb_T^d b_T^2} \right) \\ &= \sqrt{Tb_T^d b_T^2} [c_0 B_0(\mathbf{x}) + \mathbf{c}_1^\tau \mathbf{B}_1(\mathbf{x})] + o \left( \sqrt{Tb_T^d b_T^2} \right), \end{aligned} \quad (3.2)$$

where

$$B_0(\mathbf{x}) := \frac{1}{2} f(\mathbf{x}) \sum_{i=1}^d \sum_{j=1}^d g_{ij}(\mathbf{x}) \int u_j u_i K(\mathbf{u}) d\mathbf{u}, \quad \mathbf{B}_1(\mathbf{x}) := \frac{1}{2} f(\mathbf{x}) \sum_{i=1}^d \sum_{j=1}^d g_{ij}(\mathbf{x}) \int u_j u_i \mathbf{u} K(\mathbf{u}) d\mathbf{u},$$

$g_{ij}(\mathbf{x}) = \partial^2 g(\mathbf{x}) / \partial x_i \partial x_j$ ,  $i, j = 1, \dots, d$ , and  $\mathbf{u} := (u_1, \dots, u_d)^\tau \in \mathbb{R}^d$ .

**Proof.** The proof of Lemma 3.3 will be given in Section 5.2. □

### 3.2 Asymptotic normality

The asymptotic normality of our estimators relies on the following lemma (see (2.10) for the definition of  $\mathbf{W}_T(\mathbf{x})$ ).

**Lemma 3.4** *Suppose that Assumptions (A1), (A2), (A4) and (A5) hold, and that the bandwidth  $b_T$  satisfies conditions (B1)–(B4). Denote by  $\sigma^2$  the asymptotic variance (3.1). Then  $(Tb_T^d)^{1/2}(\mathbf{c}^\tau[\mathbf{W}_T(\mathbf{x}) - \mathbf{E}\mathbf{W}_T(\mathbf{x})]/\sigma)$  is asymptotically standard normal as  $T \rightarrow \infty$ .*

The proof of this lemma will be given in Section 5.3. □

We now turn to the main consistency and asymptotic normality result.

**Theorem 3.1** *Let Assumptions (A1), (A2), (A3), and (A5) hold, with  $v_{2+\delta/2}(x) = O(x^{-\mu})$  and  $\alpha(x) = O(x^{-\lambda})$  for some  $\mu \geq \max\{4(\kappa_1 - 1), \kappa_3/(1 + \delta/4)\}\kappa_2$  and some  $\lambda > (a + 1)(1 + 4/\delta)$  with  $a > \delta/(4 + \delta)$ , such that  $T^{2+4/\delta}b_T^{\mu/\kappa_2 - \kappa_1} \rightarrow 0$ ,  $Tb_T^{[1+2/\{a(1+4/\delta)\}]d}/\log T \rightarrow \infty$ , and  $Tb_T^{[2\lambda/\{a(1+4/\delta)\}-1]d}\log T \rightarrow 0$  as  $T \rightarrow \infty$ , where  $\kappa_1 = 2 + d + 2d/\delta$  and  $\kappa_2 = a(1 + 4/\delta)(1 + \delta/4)/d$ ,  $\kappa_3 = 2 + 2d + \delta/2 + 4d/\delta$ . Then,*

$$(Tb_T^d)^{1/2} \left[ \begin{pmatrix} g_T(\mathbf{x}) - g(\mathbf{x}) \\ b_T(g'_T(\mathbf{x}) - g'(\mathbf{x})) \end{pmatrix} - \mathbf{U}^{-1} \begin{pmatrix} B_0(\mathbf{x}) \\ \mathbf{B}_1(\mathbf{x}) \end{pmatrix} b_T^2 \right] \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, \mathbf{U}^{-1}\boldsymbol{\Sigma}(\mathbf{U}^{-1})^\tau), \quad (3.3)$$

as  $T \rightarrow \infty$ , where  $\mathbf{U}$ ,  $\boldsymbol{\Sigma}$ ,  $B_0(\mathbf{x})$  and  $\mathbf{B}_1(\mathbf{x})$  are defined in Lemmas 3.1, 3.2, and 3.3, respectively.

If furthermore the kernel  $K(\cdot)$  is a symmetric density function, then (3.3) can be reinforced into

$$\begin{pmatrix} (Tb_T^d)^{1/2} [g_T(\mathbf{x}) - g(\mathbf{x}) - B_g(\mathbf{x})b_T^2] \\ (Tb_T^{d+2})^{1/2} [g'_T(\mathbf{x}) - g'(\mathbf{x})] \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N} \left( \mathbf{0}, \begin{pmatrix} \sigma_0^2(\mathbf{x}) & 0 \\ 0 & \boldsymbol{\sigma}_1^2(\mathbf{x}) \end{pmatrix} \right)$$

(so that  $g_T(\mathbf{x})$  and  $g'_T(\mathbf{x})$  are asymptotically independent), where

$$B_g(\mathbf{x}) := \frac{1}{2} \sum_{i=1}^d g_{ii}(\mathbf{x}) \int (\mathbf{u})_i^2 K(\mathbf{u}) d\mathbf{u}, \quad \sigma_0^2(\mathbf{x}) := \frac{\text{Var}(Y_j | \mathbf{X}_j = \mathbf{x}) \int K^2(\mathbf{u}) d\mathbf{u}}{f(\mathbf{x})},$$

and

$$\boldsymbol{\sigma}_1^2(\mathbf{x}) := \frac{\text{Var}(Y_j | \mathbf{X}_j = \mathbf{x})}{f(\mathbf{x})} \left[ \int \mathbf{u}\mathbf{u}^\tau K(\mathbf{u}) d\mathbf{u} \right]^{-1} \left[ \int \mathbf{u}\mathbf{u}^\tau K^2(\mathbf{u}) d\mathbf{u} \right] \left[ \int \mathbf{u}\mathbf{u}^\tau K(\mathbf{u}) d\mathbf{u} \right]^{-1}.$$

From this theorem, we can derive the following corollary which gives the conditions under which the usually used optimal bandwidth,  $b_T = O\left(T^{-1/(4+d)}\right)$ , is achievable.

**Corollary 3.1** *Let Assumptions (A1), (A2), (A3), and (A5) hold, with  $v_{2+\delta/2}(x) = O(x^{-\mu})$  and  $\alpha(x) = O(x^{-\lambda})$  for some  $\mu \geq \max\{4(\kappa_1 - 1), \kappa_3/(1 + \delta/4), (\kappa_4 + \kappa_1)\}\kappa_2$  and some  $\lambda > \max\{(a + 1), a(1 + 2/d)\}(1 + 4/\delta)$  with  $a > \max\{1, d/2\}\delta/(4 + \delta)$ , and  $b_T = O\left(T^{-1/(4+d)}\right)$ , where  $\kappa_1, \kappa_2$  and  $\kappa_3$  are specified in Theorem 3.1, and  $\kappa_4 = (4 + d)(2 + 4/\delta)$ . Then the conclusion of Theorem 3.1 holds.*

The proof of Theorem 3.1 and its corollary is relegated to the Appendix.

*Remark.* (i) In Theorem 3.1 and its Corollary 3.1, the positive constants  $\mu$  and  $\lambda$  specify the decay rates for the stability and mixing coefficients tending to 0 (the larger  $\mu$  and  $\lambda$ , the faster the decay rates), which are related to the positive constant  $\delta$  determining the moment order of  $Y_t$ . For example, if we let  $a = 2d\delta/(4 + \delta)$  and  $d = 1$  in Corollary 3.1, then simple calculation leads to the requirement that  $\mu > 37 + 6.5\delta + 44/\delta$  and  $\lambda > \max\{3 + 4/\delta, 6\}$ . Therefore, when  $\mu$  and  $\lambda$  are sufficiently large,  $\delta$  can be equal to a small number close to 0. This condition is automatically satisfied when both the stability and mixing coefficients decay at geometric rates (c.f., examples in Section 4). (ii) When the model (1.2) with (1.3) being an integrated GARCH model, then the second order moment of  $Y_t$  are unavailable, for which the asymptotic normality for the estimates stated in Theorem 3.1 and Corollary 3.1 cannot be ensured but the consistency of the estimates is still obtainable if  $E|Y_t| < \infty$  is available.

The asymptotic distribution is as if the sequence  $(Y_j, X_j)$  were i.i.d. with the same marginal distributions. These results are expected for such weakly dependent stationary processes as ours has already been shown by Masry and Fan (1997) for  $\alpha$ -mixing processes. By contrast for nonstationary or strongly dependent time series slower convergence rates and even non-normal limiting distributions can hold, see Phillips and Park (1998), Karlsen and Tjøstheim (2001), and Bandi (2004). Consistent standard errors can be computed by estimating the conditional variance and marginal covariate density. For the conditional variance we can use that  $\text{Var}(Y_j|\mathbf{X}_j = \mathbf{x}) = E(Y_j^2|\mathbf{X}_j = \mathbf{x}) - E^2(Y_j|\mathbf{X}_j = \mathbf{x})$  and then compute the additional regression estimator of  $E(Y_j^2|\mathbf{X}_j = \mathbf{x})$ . We can estimate the marginal density by the kernel estimator  $(\mathbf{U}_T)_{00}$  defined above. Consistency of the standard errors follows under our conditions. The optimal

bandwidth can be found using the same formula given in Masry and Fan (1997, section 3.1).

One of the important advantages of local polynomial (and linear) fitting over more traditional Nadaraya-Watson approach is that it has much better boundary behavior. This advantage often has been emphasized in the usual regression and time-series settings when the regressors take values on a compact subset of  $\mathbb{R}^d$ . For example, as Fan and Gijbels (1996) and Fan and Yao (2003) illustrate, for a univariate regressor  $X$  with bounded support ( $[0, 1]$ , say; here,  $d = 1$ ), it can be proved, using an argument similar to the one we are developing in the proof of Theorem 3.1, that asymptotic normality still holds at the boundary point  $x = cb_T$  (here  $c$  is a positive constant), but with asymptotic bias and variance

$$B_g := \frac{1}{2} \left( \frac{\partial^2 g}{\partial x^2} \right)_{0^+} \int_{-c}^{\infty} u^2 K(u) du, \quad \sigma_0^2 := \frac{\text{Var}(Y_j | X_j = 0^+) \int_{-c}^{\infty} K^2(u) du}{f(0^+)}, \quad (3.4)$$

and

$$\sigma_1^2 := \frac{\text{Var}(Y_j | X_j = 0^+)}{f(0^+)} \left[ \int_{-c}^{\infty} u^2 K(u) du \right]^{-2} \left[ \int_{-c}^{\infty} u^2 K^2(u) du \right]. \quad (3.5)$$

## 4 Numerical results

### 4.1 Simulation

In this section we report the results of a brief Monte Carlo study of the method described in this paper, the purpose of which is to illustrate that local linear estimates with a conventional choice of bandwidth as pointed out by Masry and Fan (1997) for  $\alpha$ -mixing processes can work reasonably well for the processes for which alpha-mixing is not guaranteed but near-epoch dependence is satisfied.

*Model 1 (AR(1)-GARCH(1) model):* In financial econometrics, a well applied model is the following compound model:

$$r_t = a_0 + ar_{t-1} + \varepsilon_t, \quad (4.1)$$

$$\varepsilon_t = e_t h_t^{1/2}, \quad h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}, \quad (4.2)$$

where  $r_t$  is the daily return of some equity on day  $t$ , modelled by an AR(1) model (4.1), and  $h_t$  is the conditional variance of  $r_t$ , given the past information up to day  $t - 1$ , modelled by a GARCH(1,1) model (4.2), with  $\alpha_0 > 0$ ,  $\alpha_1 \geq 0$ ,  $\beta_1 \geq 0$ , and  $\{e_t\}$  being *i.i.d.* random sequence

with  $E\varepsilon_t = 0$  and  $E\varepsilon_t^2 = 1$  (taken to be standard normally distributed in this example). This is a special case of the general ARMA( $p,q$ )-GARCH( $P,Q$ ) model given in (1.2) and (1.3).

If  $|a| < 1$ , it is well-known that model (4.1) can be expressed as

$$r_t = a_0/(1-a) + \varepsilon_t + \sum_{j=1}^{\infty} a^j \varepsilon_{t-j},$$

and under  $\alpha_1 + \beta_1 < 1$  with some suitably regular conditions (c.f. Carrasco and Chen (2002)), the  $\varepsilon_t$  in the GARCH(1,1) model (4.2) is strongly ( $\alpha$ -) mixing with a geometrically decaying mixing coefficient. Here it is difficult to show under such natural and mild conditions  $|a| < 1$  and  $\alpha_1 + \beta_1 < 1$  (to the best of our knowledge) that  $r_t$  is strongly ( $\alpha$ -) mixing, but it can be shown that  $r_t$  is NED of order  $2 + \delta$  w.r.t. a strongly ( $\alpha$ -) mixing process, if  $E|\varepsilon_t|^{2+\delta} < \infty$ , with stable coefficients (owing to the convex property of  $|\cdot|^{2+\delta}$ )

$$\begin{aligned} v_{2+\delta}(k) &= E|r_t - r_t^{(k)}|^{2+\delta} = w_k^{2+\delta} E \left| \sum_{j=k+1}^{\infty} \frac{a^j}{w_k} \varepsilon_{t-j} \right|^{2+\delta} \\ &\leq w_k^{2+\delta} E \left[ \sum_{j=k+1}^{\infty} \frac{a^j}{w_k} |\varepsilon_{t-j}|^{2+\delta} \right] = O(|a|^{(2+\delta)k}), \end{aligned}$$

decaying at a geometric rate, where  $r_t^{(k)} = a_0/(1-a) + \varepsilon_t + \sum_{j=1}^k a^j \varepsilon_{t-j}$ , and  $w_k = \sum_{j=k+1}^{\infty} a^j = O(a^k)$ . Here the conditions to ensure  $E|\varepsilon_t|^{2+\delta} < \infty$  can be found in Carrasco and Chen (2002), and therefore  $E|r_t|^{2+\delta} < \infty$  can be guaranteed.

We are concerned with estimation of the autoregression function  $m(\mathbf{x}) = E(r_t | r_{t-1} = x) = a_0 + ax$  and the conditional variance function  $v(\mathbf{x}) = E((r_t - m(\mathbf{x}))^2 | r_{t-1} = x)$ , where our theory developed in this paper applies obviously while the theory based on the strong ( $\alpha$ -) mixing in the literature would not do if  $r_t$  is non-strongly mixing. The boxplots of the local linear estimators of  $m$  and  $v$ , based on 100 replications with each sample size equal to 100 and 500, are depicted in Figure 1(a)-(d), respectively, where  $\varepsilon_t$  is with standard normal distribution, and as suggested by a referee, we took the parameters  $a_0 = 0.001682$ ,  $a = 0.020602$ ,  $\alpha_0 = 0.137526$ ,  $\alpha_1 = 0.094518$ , and  $\beta_1 = 0.726777$ , which are the parameter estimates of the model (4.1) with (4.2) obtained from the real data of FT100 index given in Subsection 4.2 below by the maximum likelihood method procedure in the GARCH module of S-plus. In the simulations, the bandwidth  $b_T$  was chosen by the conventional generalized cross-validation rule of Wahba (1977) and Craven and Wahba (1979).

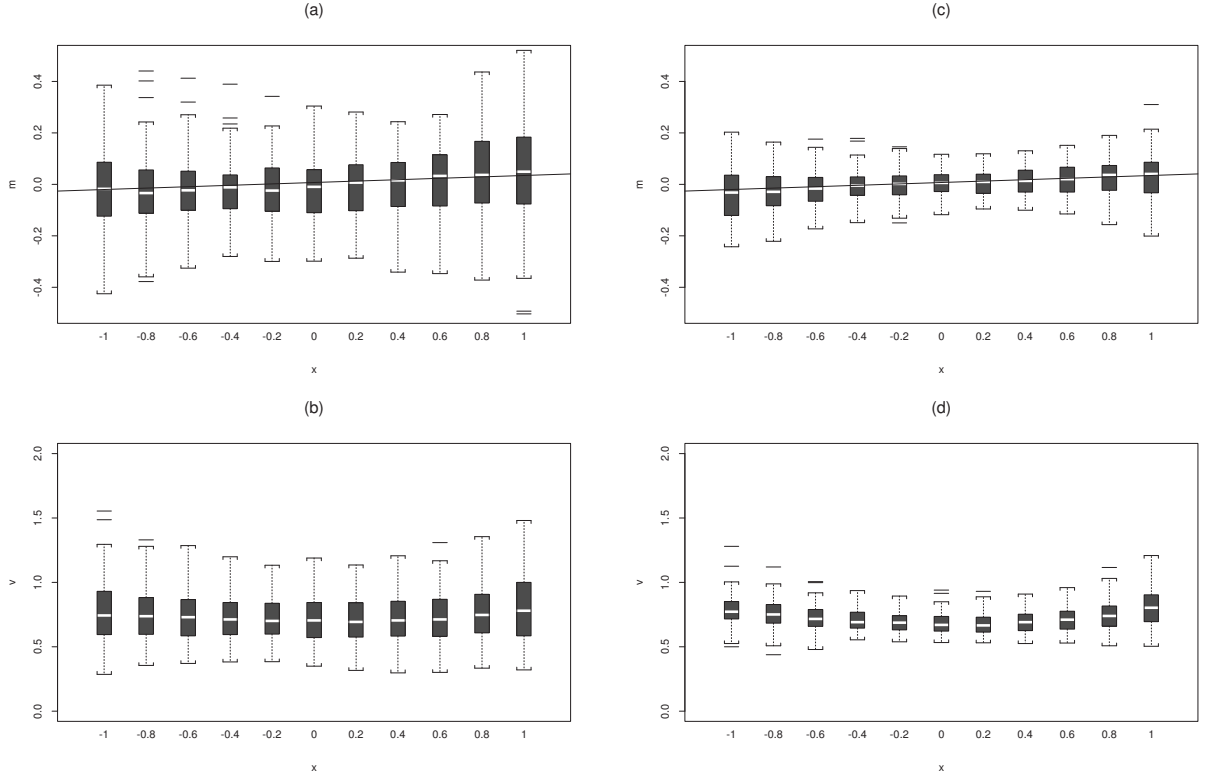


Figure 1: *Simulation for Model 1 – Boxplots of the local linear fitting for the conditional mean and conditional variance of 100 replications of the linear AR(1)-GARCH(1,1) model (Model 1) with  $a_0 = 0.001682$ ,  $a = 0.020602$ ,  $\alpha_0 = 0.137526$ ,  $\alpha_1 = 0.094518$ , and  $\beta_1 = 0.726777$ , and  $e_t \sim N(0, 1)$ , for different sample sizes: (a) conditional mean  $m(\mathbf{x}) = E(r_t|r_{t-1} = x) = a_0 + ax$ , sample size= 100, (b) conditional variance  $v(\mathbf{x})$ , sample size= 100, (c) conditional mean  $m(\mathbf{x}) = E(r_t|r_{t-1} = x) = a_0 + ax$ , sample size= 500, and (d) conditional variance  $v(\mathbf{x})$ , sample size= 500. In (a) and (c), the solid line stands for the true conditional mean  $m(x) = a_0 + ax$ .*

Overall, the simulation results in the example of Model 1 adapt very well to our asymptotic theory: with the sample size increasing, the local linearly estimated curves with a conventional choice of bandwidth become more stable and fit better to the actual curve lines both for the conditional mean and conditional variance functions; see Figure 1(a)(c) for the estimates of the conditional mean, which are compatible to the actual line in Model 1, and Figure 1(b)(d) for the estimates of the conditional variance. Impressively, the sample size equal to 500 seems to work very well in all cases.

## 4.2 A real example

We consider the UK FT100 Index, with sample size 602 from 1st January 1990 to 21st April 1992, as our real example for an illustration of application. In Figure 2, the index series  $P_t$  and the return series  $r_t$ , defined by

$$r_t = \log(P_t/P_{t-1}) \times 100,$$

are plotted in (a) and (b), respectively. The local linear estimates of the conditional mean  $m(x) = E(r_t|r_{t-1} = x)$  and the conditional variance function  $v(x) = E((r_t - m(x))^2|r_{t-1} = x)$  together with the 95% confidence intervals based on the asymptotic normality in Section 3 are plotted in Figure 2(c) and (d), where the bandwidths used for the conditional mean and conditional variance are 0.34 and 0.5, respectively, chosen by cross-validation rule. In addition the estimates of the following parametric linear model

$$r_t = \tilde{a}_0 + \tilde{a}r_{t-1} + e_t\sqrt{\tilde{\alpha}_0 + \tilde{\alpha}_1r_{t-1}^2} \quad (4.3)$$

with  $e_t$  the same as in the above, leading to  $E(r_t|r_{t-1} = x) = \tilde{a}_0 + \tilde{a}x$  and  $Var(r_t|r_{t-1} = x) = \tilde{\alpha}_0 + \tilde{\alpha}_1x^2$ , are also plotted in Figure 2(c) and (d), respectively, where the estimated parameters are  $\tilde{a}_0 = 0.01069$ ,  $\tilde{a} = 0.01906$ ,  $\tilde{\alpha}_0 = 0.69223$  and  $\tilde{\alpha}_1 = 0.07707$ .

From Figure 2(c) and (d), we can observe that both the conditional mean and conditional variance functions look to be nonlinear. Roughly, the conditional mean first increases before  $x$  taking value around  $-0.4$ , then decreases between  $-0.4$  and  $0.7$  around, and increases again when  $x > 0.7$ . Correspondingly, the conditional variance mostly follows the similar pattern to that of the conditional mean, increasing with  $x$  when  $x < -0.75$  and  $x > 0.8$  and decreasing with

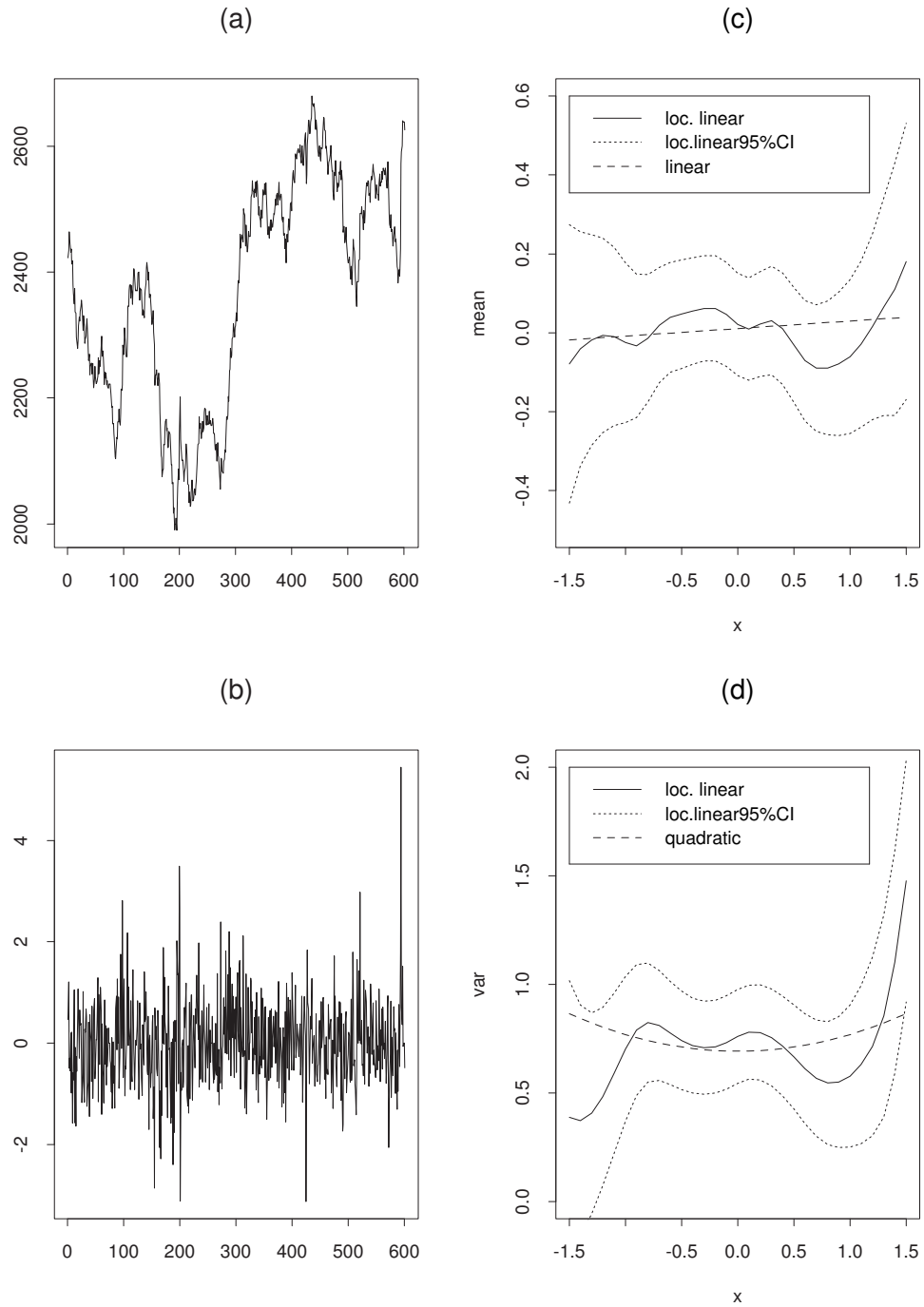


Figure 2: *Real example – The local linear fitting for the conditional mean and conditional variance of the return series of the UK FT100 Index from 1st January 1990 to 21st April 1992, with 602 observations: (a) FT100 Index series, (b) return series of FT100 Index, (c) conditional mean, and (d) conditional variance. The dashed lines in (c) and (d) are the estimated linear mean and quadratic variance functions, respectively, from the parametric linear model (4.3).*

$x$  when  $0.25 < x < 0.9$ , which appears to be consistent with the “high return, high risk” rule, but the pattern completely different from that of the conditional mean when  $-0.75 < x < 0.25$ , which looks to be of “U” shape as observed in Figure 1(d) in simulation example. However, the pointwise standard errors reveal that the mean effect is never significantly different from linear, while the variance effect is only rarely different from a quadratic function.

## 5 Appendix: proofs

### 5.1 Some basic lemmas

The proof of the main results relies on some intermediate results. The first one is a lemma borrowed from Ibragimov and Linnik (1971) or Deo (1973), where we refer to for a proof.

**Lemma 5.1** (i) Let  $\mathcal{L}_r(\mathcal{F})$  denote the class of  $\mathcal{F}$ -measurable random variables  $\xi$  satisfying  $\|\xi\|_r := (\mathbb{E}|\xi|^r)^{1/r} < \infty$ . Let  $X \in \mathcal{L}_r(\mathcal{F}_1)$  and  $Y \in \mathcal{L}_s(\mathcal{F}_2)$ . Then, for any  $1 \leq r, s, h < \infty$  such that  $r^{-1} + s^{-1} + h^{-1} = 1$ ,

$$\left| \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \right| \leq C \|X\|_r \|Y\|_s [\alpha(\mathcal{F}_1, \mathcal{F}_2)]^{1/h}, \quad (5.1)$$

where  $\alpha(\mathcal{F}_1, \mathcal{F}_2) = \sup_{A \in \mathcal{F}_1, B \in \mathcal{F}_2} |P(AB) - P(A)P(B)|$ .

(ii) If moreover  $|X|$  and  $|Y|$  are P-a.s. bounded, the right-hand side of (5.1) can be replaced with  $C\alpha(\mathcal{F}_1, \mathcal{F}_2)$ .

□

The following lemma establishes the relationship between the related quantities based on the original samples  $(\mathbf{X}_i, Y_i)$  and on the approximated  $\alpha$ -mixing samples  $(\mathbf{X}_i^{(m)}, Y_i^{(m)})$ , which will play important roles throughout the proofs in the sequel.

**Lemma 5.2** Let  $m = m_T$  be a positive integer tending to  $\infty$  as  $T \rightarrow \infty$ . Then under Assumptions (A2) and (A5), it holds that for any positive real number  $L = L_T$  tending to  $\infty$ , as  $T \rightarrow \infty$ ,

$$EY_i^{(m)} K((\mathbf{x} - \mathbf{X}_i^{(m)})/b) = EY_i K((\mathbf{x} - \mathbf{X}_i)/b) + O\left(b_T^{-1} L v_2^{1/2}(m)\right) + o\left(L^{-(1+\delta)}\right), \quad (5.2)$$

$$\begin{aligned} & EY_i^{(m)2} K^2((\mathbf{x} - \mathbf{X}_i^{(m)})/b) \\ &= EY_i^2 K^2((\mathbf{x} - \mathbf{X}_i)/b) + O\left(b_T^{-2} L^2 v_2(m)\right) + O\left(b_T^{-1} L^2 v_2^{1/2}(m)\right) + o\left(L^{-\delta}\right), \end{aligned} \quad (5.3)$$

for  $i \neq j$ ,

$$\begin{aligned} & EY_i^{(m)}Y_j^{(m)}K((\mathbf{x} - \mathbf{X}_i^{(m)})/b)K((\mathbf{x} - \mathbf{X}_j^{(m)})/b) \\ &= EY_iY_jK((\mathbf{x} - \mathbf{X}_i)/b)K((\mathbf{x} - \mathbf{X}_j)/b) + O\left(b_T^{-1}L^2v_2^{1/2}(m)\right) + o\left(L^{-\delta}\right), \end{aligned} \quad (5.4)$$

and for any  $0 < \delta' < \delta$ ,

$$\begin{aligned} & E|Y_i^{(m)}|^{2+\delta'}K^{2+\delta'}((\mathbf{x} - \mathbf{X}_i^{(m)})/b) \\ & \leq CE|Y_i|^{2+\delta'}K^{2+\delta'}((\mathbf{x} - \mathbf{X}_i)/b) + O\left(b_T^{-2-\delta'}L^{2+\delta'}v_{2+\delta'}(m)\right) + o\left(L^{-\delta+\delta'}\right), \end{aligned} \quad (5.5)$$

where  $O(\cdot)$  and  $o(\cdot)$  hold uniformly with respect to  $x$ .

*Proof.* We prove (5.2)–(5.5) one by one. To prove (5.2), note first that

$$\begin{aligned} EY_i^{(m)}K((\mathbf{x} - \mathbf{X}_i^{(m)})/b_T) &= EY_iK((\mathbf{x} - \mathbf{X}_i^{(m)})/b_T) + E(Y_i^{(m)} - Y_i)K((\mathbf{x} - \mathbf{X}_i^{(m)})/b_T) \\ &=: EY_iK((\mathbf{x} - \mathbf{X}_i^{(m)})/b_T) + \delta_{1T}. \end{aligned} \quad (5.6)$$

Here, using the bounded property of the kernel function  $K(\cdot)$ ,

$$|\delta_{1T}| \leq E|Y_i^{(m)} - Y_i|K((\mathbf{x} - \mathbf{X}_i^{(m)})/b_T) \leq CE|Y_i^{(m)} - Y_i| \leq C[E|Y_i^{(m)} - Y_i|^2]^{1/2} = O\left(\sqrt{v_2(m)}\right). \quad (5.7)$$

Next, let  $Y_{i,L} = Y_iI_{\{|Y_i| \leq L\}}$  and  $Y_{i,U} = Y_iI_{\{|Y_i| > L\}}$ , where  $I_A$  is an indicator function of set  $A$ .

Then

$$\begin{aligned} EY_iK((\mathbf{x} - \mathbf{X}_i^{(m)})/b_T) &= EY_{i,L}K((\mathbf{x} - \mathbf{X}_i^{(m)})/b_T) + EY_{i,U}K((\mathbf{x} - \mathbf{X}_i^{(m)})/b_T) \\ &= EY_{i,L}K((\mathbf{x} - \mathbf{X}_i)/b_T) + EY_{i,L}(K((\mathbf{x} - \mathbf{X}_i^{(m)})/b_T) - K((\mathbf{x} - \mathbf{X}_i)/b_T)) + EY_{i,U}K((\mathbf{x} - \mathbf{X}_i^{(m)})/b_T) \\ &=: \delta_{2T} + \delta_{3T} + \delta_{4T}, \end{aligned} \quad (5.8)$$

where, owing to  $|EY_{i,U}K((\mathbf{x} - \mathbf{X}_i)/b_T)| = o\left(L^{-(1+\delta)}\right)$  as  $T \rightarrow \infty$  (in an argument similar to (5.11) below),

$$\delta_{2T} = EY_iK((\mathbf{x} - \mathbf{X}_i)/b_T) - EY_{i,U}K((\mathbf{x} - \mathbf{X}_i)/b_T) = EY_iK((\mathbf{x} - \mathbf{X}_i)/b_T) + o\left(L^{-(1+\delta)}\right); \quad (5.9)$$

owing to the Lipschitz continuity of the kernel function  $K(\cdot)$ ,

$$\begin{aligned} |\delta_{3T}| &\leq E|Y_{i,L}| |K((\mathbf{x} - \mathbf{X}_i^{(m)})/b_T) - K((\mathbf{x} - \mathbf{X}_i)/b_T)| \\ &\leq CLE \left\| \frac{\mathbf{X}_i^{(m)} - \mathbf{X}_i}{b} \right\| \leq Cb_T^{-1}L[E\|\mathbf{X}_i^{(m)} - \mathbf{X}_i\|^2]^{1/2} = O\left(b_T^{-1}L\sqrt{v_2(m)}\right); \end{aligned} \quad (5.10)$$

using the boundedness of  $K(\cdot)$ ,

$$\begin{aligned} |\delta_{4T}| &\leq E|Y_i|I_{\{|Y_i|>L\}}K((\mathbf{x} - \mathbf{X}_i^{(m)})/b_T) \leq L^{-(1+\delta)}E|Y_i|^{2+\delta}I_{\{|Y_i|>L\}}K((\mathbf{x} - \mathbf{X}_i^{(m)})/b_T) \\ &\leq CL^{-(1+\delta)}E|Y_i|^{2+\delta}I_{\{|Y_i|>L\}} = o\left(L^{-(1+\delta)}\right). \end{aligned} \quad (5.11)$$

Then it follows from (5.8) together with (5.9), (5.10) and (5.11) that

$$EY_iK((\mathbf{x} - \mathbf{X}_i^{(m)})/b_T) = EY_iK((\mathbf{x} - \mathbf{X}_i)/b_T) + O\left(b_T^{-1}L\sqrt{v_2(m)}\right) + o\left(L^{-(1+\delta)}\right). \quad (5.12)$$

Finally, (5.2) follows from (5.6) together with (5.7) and (5.12).

For (5.3), it can be proved in an argument similar to that in the above. First,

$$\begin{aligned} E[Y_i^{(m)}]^2K^2((\mathbf{x} - \mathbf{X}_i^{(m)})/b_T) &= EY_i^2K^2((\mathbf{x} - \mathbf{X}_i^{(m)})/b_T) + 2EY_i(Y_i^{(m)} - Y_i)K^2((\mathbf{x} - \mathbf{X}_i^{(m)})/b_T) \\ &\quad + E(Y_i^{(m)} - Y_i)^2K^2((\mathbf{x} - \mathbf{X}_i^{(m)})/b_T) \\ &=: \delta_{5T} + \delta_{6T} + \delta_{7T}. \end{aligned} \quad (5.13)$$

Here, similarly to (5.7) by the boundedness of  $K(\cdot)$ ,

$$|\delta_{7T}| \leq E|Y_i^{(m)} - Y_i|^2 K^2((\mathbf{x} - \mathbf{X}_i^{(m)})/b_T) \leq CE|Y_i^{(m)} - Y_i|^2 = O(v_2(m)); \quad (5.14)$$

again by the boundedness of  $K(\cdot)$  together with Cauchy inequality,

$$|\delta_{6T}| \leq 2 \left[ EY_i^2 \right]^{1/2} \left[ E|Y_i^{(m)} - Y_i|^2 \right]^{1/2} = O\left(\sqrt{v_2(m)}\right). \quad (5.15)$$

Next,

$$\begin{aligned} \delta_{5T} &= EY_{i,L}^2K^2((\mathbf{x} - \mathbf{X}_i^{(m)})/b_T) + EY_{i,U}^2K^2((\mathbf{x} - \mathbf{X}_i^{(m)})/b_T) \\ &= EY_{i,L}^2K^2((\mathbf{x} - \mathbf{X}_i)/b_T) + 2EY_{i,L}^2K((\mathbf{x} - \mathbf{X}_i)/b_T)(K((\mathbf{x} - \mathbf{X}_i^{(m)})/b_T) - K((\mathbf{x} - \mathbf{X}_i)/b_T)) \\ &\quad + EY_{i,L}^2(K((\mathbf{x} - \mathbf{X}_i^{(m)})/b_T) - K((\mathbf{x} - \mathbf{X}_i)/b_T))^2 + EY_{i,U}^2K^2((\mathbf{x} - \mathbf{X}_i^{(m)})/B_T) \\ &=: \delta_{8T} + \delta_{9T} + \delta_{10T} + \delta_{11T}, \end{aligned} \quad (5.16)$$

where, similarly to (5.11),

$$|\delta_{11T}| \leq E|Y_i|^2I_{\{|Y_i|>L\}}K^2((\mathbf{x} - \mathbf{X}_i^{(m)})/B_T) \leq CL^{-\delta}E|Y_i|^{2+\delta}I_{\{|Y_i|>L\}} = o\left(L^{-\delta}\right); \quad (5.17)$$

owing to  $|EY_{i,U}^2K^2((\mathbf{x} - \mathbf{X}_i)/b_T)| = o\left(L^{-\delta}\right)$  (the same argument to (5.17)),

$$\delta_{8T} = EY_i^2K^2((\mathbf{x} - \mathbf{X}_i)/b_T) - EY_{i,U}^2K^2((\mathbf{x} - \mathbf{X}_i)/b_T) = EY_i^2K^2((\mathbf{x} - \mathbf{X}_i)/b_T) + o\left(L^{-\delta}\right); \quad (5.18)$$

the same argument to (5.10)) leads to

$$\begin{aligned} |\delta_{10T}| &\leq E|Y_{i,L}|^2 |K((\mathbf{x} - \mathbf{X}_i^{(m)})/b_T) - K((\mathbf{x} - \mathbf{X}_i)/b_T)|^2 \\ &\leq CL^2 E \left\| \frac{\mathbf{X}_i^{(m)} - \mathbf{X}_i}{b} \right\|^2 \leq Cb_T^{-2} L^2 E \|\mathbf{X}_i^{(m)} - \mathbf{X}_i\|^2 = O\left(b_T^{-2} L^2 v_2(m)\right), \end{aligned} \quad (5.19)$$

and

$$|\delta_{9T}| \leq CL^2 E \left\| \frac{\mathbf{X}_i^{(m)} - \mathbf{X}_i}{b} \right\| \leq Cb_T^{-1} L^2 [E\|\mathbf{X}_i^{(m)} - \mathbf{X}_i\|^2]^{1/2} = O\left(b_T^{-1} L^2 \sqrt{v_2(m)}\right). \quad (5.20)$$

Finally, (5.3) follows from (5.13)–(5.20).

For (5.4), first note that

$$\begin{aligned} &E[Y_i^{(m)} Y_j^{(m)} K((\mathbf{x} - \mathbf{X}_i^{(m)})/b_T) K((\mathbf{x} - X_j^{(m)})/b_T)] \\ &= EY_i Y_j K((\mathbf{x} - \mathbf{X}_i^{(m)})/b_T) K((\mathbf{x} - X_j^{(m)})/b_T) \\ &\quad + E[Y_i(Y_j^{(m)} - Y_j) + (Y_i^{(m)} - Y_i)Y_j] K((\mathbf{x} - \mathbf{X}_i^{(m)})/b_T) K((\mathbf{x} - X_j^{(m)})/b_T) \\ &\quad + E(Y_i^{(m)} - Y_i)(Y_j^{(m)} - Y_j) K((\mathbf{x} - \mathbf{X}_i^{(m)})/b_T) K((\mathbf{x} - X_j^{(m)})/b_T) \\ &=: \delta_{12T} + \delta_{13T} + \delta_{14T}. \end{aligned} \quad (5.21)$$

Here, similarly to (5.7), by the boundedness of  $K(\cdot)$ , and then by Cauchy inequality,

$$|\delta_{14T}| \leq CE|Y_i^{(m)} - Y_i| |Y_j^{(m)} - Y_j| = O(v_2(m)); \quad (5.22)$$

again,

$$|\delta_{13T}| \leq \left\{ [EY_i^2]^{1/2} [E|Y_j^{(m)} - Y_j|^2]^{1/2} + [E|Y_i^{(m)} - Y_i|^2]^{1/2} [EY_j^2]^{1/2} \right\} = O\left(\sqrt{v_2(m)}\right). \quad (5.23)$$

Next,

$$\begin{aligned} \delta_{12T} &= EY_i Y_j K((\mathbf{x} - \mathbf{X}_i)/b_T) K((\mathbf{x} - X_j)/b_T) \\ &\quad + EY_i Y_j \left[ K((\mathbf{x} - \mathbf{X}_i^{(m)})/b_T) - K((\mathbf{x} - \mathbf{X}_i)/b_T) \right] K((\mathbf{x} - X_j)/b_T) \\ &\quad + EY_i Y_j K((\mathbf{x} - \mathbf{X}_i^{(m)})/b_T) \left[ K((\mathbf{x} - X_j^{(m)})/b_T) - K((\mathbf{x} - X_j)/b_T) \right] \\ &=: EY_i Y_j K((\mathbf{x} - \mathbf{X}_i)/b_T) K((\mathbf{x} - X_j)/b_T) + \delta_{15T} + \delta_{16T}, \end{aligned} \quad (5.24)$$

where, by (A2) and the assumed properties of  $K(\cdot)$  in (A5),

$$\begin{aligned}
\delta_{16T} &= EY_i(I_{\{|Y_i|\leq L\}} + I_{\{|Y_i|>L\}})Y_j(I_{\{|Y_j|\leq L\}} + I_{\{|Y_j|>L\}})K((\mathbf{x} - \mathbf{X}_i^{(m)})/b_T) \\
&\quad \times \left[ K((\mathbf{x} - X_j^{(m)})/b_T) - K((\mathbf{x} - X_j)/b_T) \right] \\
&\leq CL^2 \left\| \frac{X_j^{(m)} - X_j}{b_T} \right\| + 2CLL^{-(1+\delta)}E|Y_i|^{2+\delta}I_{\{|Y_i|>L\}} + L^{-\delta}E|Y_i|^{2+\delta}I_{\{|Y_i|>L\}} \\
&= O(b_T^{-1}L^2v_2^{1/2}(m)) + o(L^{-\delta}), \tag{5.25}
\end{aligned}$$

and similarly,

$$\delta_{15T} = O(b_T^{-1}L^2v_2^{1/2}(m)) + o(L^{-\delta}). \tag{5.26}$$

Finally, (5.4) follows from (5.21)–(5.26).

For the proof of (5.5), by noticing

$$Y_t^{(m)}K((x - \mathbf{X}_t^{(m)})/b) = YK_1 + YK_2 + YK_3,$$

with  $YK_1 = (Y_t^{(m)} - Y_t)K((x - \mathbf{X}_t^{(m)})/b)$ ,  $YK_2 = Y_t\{K((x - \mathbf{X}_t^{(m)})/b) - K((x - \mathbf{X}_t)/b)\}$  and  $YK_3 = Y_tK((x - \mathbf{X}_t)/b)$ , it follows that

$$|Y_t^{(m)}K((x - \mathbf{X}_t^{(m)})/b)|^{2+\delta'} \leq 3^{1+\delta'}\{|YK_1|^{2+\delta'} + |YK_2|^{2+\delta'} + |YK_3|^{2+\delta'}\},$$

where  $E|YK_1|^{2+\delta'} \leq C|Y_t^{(m)} - Y_t|^{2+\delta'} \leq Cv_{2+\delta'}(m)$ , and

$$\begin{aligned}
E|YK_2|^{2+\delta'} &\leq CE|Y_tI_{\{|Y_t|\leq L\}}\{K((x - \mathbf{X}_t^{(m)})/b) - K((x - \mathbf{X}_t)/b)\}|^{2+\delta'} + CE|Y_tI_{\{|Y_t|\geq L\}}|^{2+\delta'} \\
&\leq C\{Lb_T^{-1}\}^{2+\delta'}\|\mathbf{X}_t - \mathbf{X}_t^{(m)}\|^{2+\delta'}(m) + CL^{\delta'-\delta}E|Y_t|^{2+\delta}I_{\{|Y_t|\geq L\}} \\
&\leq C\{Lb_T^{-1}\}^{2+\delta'}v_{2+\delta'}(m) + o(L^{\delta'-\delta}), \tag{5.27}
\end{aligned}$$

and hence (5.5) holds.  $\square$

To cope with the approximation terms of  $\mathbf{U}_T^{(m)}$  and  $\mathbf{W}_T^{(m)}$  defined in Subsection 2.3, we will need the following lemma, termed *cross term lemma*, which is of independent interest and will play important roles in the proof of the theorems below.

**Lemma 5.3** (*Cross term lemma*) *Let  $\{(Z_j^{(m)}, \mathbf{X}_j^{(m)}); 1 \leq j \leq q\}$  be a stationary sequence with mixing coefficient*

$$\alpha_m(j) := \sup \left\{ |\mathbb{P}(AB) - \mathbb{P}(A)\mathbb{P}(B)| : A \in \mathcal{B}(\{Z_i^{(m)}, \mathbf{X}_i^{(m)}\}), B \in \mathcal{B}(\{Z_{i+j}^{(m)}, \mathbf{X}_{i+j}^{(m)}\}) \right\},$$

and set  $\eta_j^{(m)}(x) = Z_j^{(m)} \tilde{K}((\mathbf{x} - \mathbf{X}_j^{(m)})/b_T)$ , where  $\tilde{K}$  is a kernel function satisfying Assumption (A5). Denote by  $\Delta_j^{(m)}(x) = \eta_j^{(m)}(x) - E\eta_j^{(m)}(x)$ ,  $\tilde{V}_{12}(\mathbf{x}) = \sum_{1 \leq i < j \leq q} \Delta_i^{(m)}(x) \Delta_j^{(m)}(x)$ . Then, for any positive integer  $c_T$ ,

$$|\tilde{V}_{12}(\mathbf{x})| \leq Cqb_T^d [\tilde{J}_1(\mathbf{x}) + \tilde{J}_2(\mathbf{x})]. \quad (5.28)$$

where  $\tilde{J}_1(\mathbf{x}) := c_T b_T^d \max\{1, b_T^{-2(1+d+2d/\delta)} v_2^{1/2}(m)\}$  and

$$\tilde{J}_2(\mathbf{x}) := b_T^{-\delta d/(4+\delta)} \left( \sum_{j=c_T}^{\infty} \{\alpha_m(j)\}^{\delta/(4+\delta)} \right) \max\{1, b_T^{-2(4+4d+\delta+8d/\delta)/(4+\delta)} v_{2+\delta/2}^{4/(4+\delta)}(m)\}.$$

**Proof.** Firstly, we note that

$$E\Delta_j^{(m)}(\mathbf{x})\Delta_i^{(m)}(\mathbf{x}) = \left\{ E\left(\eta_j^{(m)}(\mathbf{x})\eta_i^{(m)}(\mathbf{x})\right) - E\left(\eta_j^{(m)}(\mathbf{x})\right)E\left(\eta_i^{(m)}(\mathbf{x})\right) \right\} \quad (5.29)$$

Then applying Lemma 5.2,

$$\begin{aligned} E\left(\eta_j^{(m)}(\mathbf{x})\eta_i^{(m)}(\mathbf{x})\right) &= E\left(Z_i^{(m)}K((\mathbf{x} - \mathbf{X}_i^{(m)})/b_T)\right)\left(Z_j^{(m)}K((\mathbf{x} - \mathbf{X}_j^{(m)})/b_T)\right) \\ &= E\left(Z_iK((\mathbf{x} - \mathbf{X}_i)/b_T)\right)\left(Z_jK((\mathbf{x} - \mathbf{X}_j)/b_T)\right) \\ &\quad + O\left(b_T^{-1}L^2\sqrt{v_2(m)}\right) + o\left(L^{-\delta}\right), \end{aligned} \quad (5.30)$$

$$\begin{aligned} E\eta_j^{(m)}(\mathbf{x}) &= E\left(Z_j^{(m)}K((\mathbf{x} - \mathbf{X}_j^{(m)})/b_T)\right) \\ &= E\left(Z_jK((\mathbf{x} - \mathbf{X}_j)/b_T)\right) + O\left(b_T^{-1}L\sqrt{v_2(m)}\right) + o\left(L^{-(1+\delta)}\right). \end{aligned} \quad (5.31)$$

Therefore it follows from (5.29), (5.30) and (5.31) that (taking  $L = b_T^{-2d/\delta}$ )

$$\begin{aligned} E\left[\Delta_j^{(m)}(\mathbf{x})\Delta_i^{(m)}(\mathbf{x})\right] &= [EZ_iK((\mathbf{x} - \mathbf{X}_i)/b_T)]\left(Z_jK((\mathbf{x} - \mathbf{X}_j)/b_T) - EZ_iK((\mathbf{x} - \mathbf{X}_i)/b_T)\right)E\left(Z_jK((\mathbf{x} - \mathbf{X}_j)/b_T)\right) \\ &\quad + \left[O\left(b_T^{-2}L^2v_2(m)\right) + O\left(b_T^{-1}L^2\sqrt{v_2(m)}\right) + o\left(L^{-\delta}\right) + b_T^d\left\{O\left(b_T^{-1}L\sqrt{v_2(m)}\right) + o\left(L^{-(1+\delta)}\right)\right\}\right] \\ &= b_T^{2d} + O\left(b_T^{-2}L^2v_2(m)\right) + O\left(b_T^{-1}L^2\sqrt{v_2(m)}\right) + o\left(L^{-\delta}\right) + b_T^d\left\{O\left(b_T^{-1}L\sqrt{v_2(m)}\right) + o\left(L^{-(1+\delta)}\right)\right\} \\ &\leq Cb_T^{2d}\left[1 + O\left(b_T^{-2-2d}L^2\sqrt{v_2(m)} + 1\right)\right], \end{aligned} \quad (5.32)$$

therefore,

$$\begin{aligned} \sum_{i=1}^q \sum_{j=i+1}^{c_T} E\left[\Delta_j^{(m)}(\mathbf{x})\Delta_i^{(m)}(\mathbf{x})\right] &\leq Cb_T^{2d}(qc_T) \max\{1, b_T^{-2(1+d+2d/\delta)} v_2^{1/2}(m)\} \\ &= qb_T^d \tilde{J}_1(\mathbf{x}). \end{aligned} \quad (5.33)$$

On the other hand, for  $i < j$ , by the Lemma 5.1(i) with  $r = s = 1/(2 + \delta')$  and (5.5) of Lemma 5.2 and then taking  $L = b_T^{-2d/\delta}$  and  $\delta' = \delta/2$ ,

$$\begin{aligned}
\mathbb{E} \left[ \Delta_j^{(m)}(\mathbf{x}) \Delta_i^{(m)}(\mathbf{x}) \right] &\leq C \left\| \eta_i^{(m)}(\mathbf{x}) \right\|_{2+\delta'} \left\| \eta_j^{(m)}(\mathbf{x}) \right\|_{2+\delta'} \alpha_m^{1-2/(2+\delta')}(j-i) \\
&\leq C \left\{ b_T^d + (b_T^{-1}L)^{2+\delta'} v_{2+\delta'}(m) + L^{-\delta+\delta'} \right\}^{2/(2+\delta')} \alpha_m^{\delta'/(2+\delta')}(j-i) \\
&\leq C b_T^{2d/(2+\delta')} \left\{ 1 + b_T^{-d} (b_T^{-1}L)^{2+\delta'} v_{2+\delta'}(m) + L^{-\delta+\delta'} b_T^{-d} \right\}^{2/(2+\delta')} \alpha_m^{\delta'/(2+\delta')}(j-i) \\
&\leq C b_T^{4d/(4+\delta)} \left\{ 1 + b_T^{-2d-2-\delta/2-4d/\delta} v_{2+\delta/2}(m) + 1 \right\}^{4/(4+\delta)} \alpha_m^{\delta/(4+\delta)}(j-i), \tag{5.34}
\end{aligned}$$

therefore,

$$\begin{aligned}
&\sum_{i=1}^q \sum_{j=i+c_T}^q \mathbb{E} \left[ \Delta_j^{(m)}(\mathbf{x}) \Delta_i^{(m)}(\mathbf{x}) \right] \\
&\leq C b_T^{4d/(4+\delta)} q \left( \sum_{j=c_T}^q \{ \alpha_m(j) \}^{\delta/(4+\delta)} \right) \max \left\{ 1, b_T^{-2(4+4d+\delta+8d/\delta)/(4+\delta)} v_{2+\delta/2}^{4/(4+\delta)}(m) \right\} \\
&= C q b_T^d \tilde{J}_2(\mathbf{x}). \tag{5.35}
\end{aligned}$$

The result of this lemma therefore follows from (5.33) and (5.35).  $\square$

## 5.2 Proof for Sections 2 and 3.1

**Proof of Lemma 2.1.** Denote by  $K_i(\mathbf{x}) = (\mathbf{x})_i K(\mathbf{x})$ . Then it can be noted that

$$\begin{aligned}
\mathbb{E} \left| \left( \mathbf{W}_T^{(m)} \right)_i - (\mathbf{W}_T)_i \right| &= \mathbb{E} \left| (T b_T^d)^{-1} \sum_{j=1}^T \left[ Z_j^{(m)} K_i \left( \frac{\mathbf{X}_j^{(m)} - \mathbf{x}}{b_T} \right) - Z_j K_i \left( \frac{\mathbf{X}_j - \mathbf{x}}{b_T} \right) \right] \right| \\
&\leq (T b_T^d)^{-1} \sum_{j=1}^T \mathbb{E} \left| Z_j^{(m)} K_i \left( \frac{\mathbf{X}_j^{(m)} - \mathbf{x}}{b_T} \right) - Z_j K_i \left( \frac{\mathbf{X}_j - \mathbf{x}}{b_T} \right) \right| \\
&= b_T^{-d} \mathbb{E} \left| Z_j^{(m)} K_i \left( (\mathbf{X}_j^{(m)} - \mathbf{x})/b_T \right) - Z_j K_i \left( (\mathbf{X}_j - \mathbf{x})/b_T \right) \right| \\
&\leq b_T^{-d} \mathbb{E} \left| Z_j^{(m)} - Z_j \right| K_i \left( (\mathbf{X}_j^{(m)} - \mathbf{x})/b_T \right) \\
&\quad + b_T^{-d} \mathbb{E} |Z_j| (I_{\{|Z_j| \leq L\}} + I_{\{|Z_j| > L\}}) \left| K_i \left( (\mathbf{X}_j^{(m)} - \mathbf{x})/b_T \right) - K_i \left( (\mathbf{X}_j - \mathbf{x})/b_T \right) \right| \\
&= O \left( b_T^{-d} \sqrt{v_2(m)} \right) + O \left( b_T^{-d-1} L \sqrt{v_2(m)} \right) + o \left( b_T^{-d} L^{-(1+\delta)} \right) \\
&= O \left( b_T^{-d-1} L \sqrt{v_2(m)} \right) + o \left( b_T^{-d} L^{-(1+\delta)} \right), \tag{5.36}
\end{aligned}$$

which is the desired.  $\square$

**Proof of Lemma 3.1.** This lemma easily follows from Lu (2001); see also the proof of Lemma 3.2. The detail is omitted.  $\square$

**Proof of Lemma 3.2.** Set  $\eta_j = Z_j K_c ((\mathbf{X}_j - \mathbf{x})/b_T)$  and  $\Delta_j = \eta_j - E\eta_j$ . Then we note that  $A_T$  defined in Subsection 3.1 can be expressed as  $A_T = (Tb_T^d)^{-1/2} \sum_{j=1}^T \eta_j$ , and therefore

$$\begin{aligned} \text{Var}[A_T] &= (Tb_T^d)^{-1} \left\{ \sum_{j=1}^T E\Delta_j^2 + 2 \sum_{1 \leq i < j \leq T} E\Delta_i \Delta_j \right\} \\ &= b_T^{-d} E\Delta_j^2 + 2(Tb_T^d)^{-1} \sum_{1 \leq i < j \leq T} E\Delta_i \Delta_j := A_{T1} + 2A_{T2}. \end{aligned} \quad (5.37)$$

It easily follows from the Lebesgue density theorem [see Chapter 2 of Devroye and Györfi (1985)] that the first term of (5.37) is convergent to the right-hand side of (3.1). Therefore, to complete the proof of this lemma, it suffices to show that  $A_{T2} \rightarrow 0$  as  $T \rightarrow \infty$ . By noticing  $E\Delta_i \Delta_j = E\Delta_i^{(m)} \Delta_j^{(m)} + E\Delta_i^{(m)} (\Delta_j - \Delta_j^{(m)}) + E(\Delta_i - \Delta_i^{(m)}) \Delta_j$ , we can further separate  $A_{T2}$  into three parts:  $A_{T2} = A_{T21} + A_{T22} + A_{T23}$ , where, by Lemma 5.3 (taking  $q = T$  and  $c_T^a = b_T^{-\delta d/(4+\delta)}$ ) together with assumption (B2),

$$\begin{aligned} A_{T21} &:= (Tb_T^d)^{-1} \sum_{1 \leq i < j \leq T} E\Delta_i^{(m)} \Delta_j^{(m)} = (Tb_T^d)^{-1} (Tb_T^d) [\tilde{J}_1(\mathbf{x}) + \tilde{J}_2(\mathbf{x})] \\ &= O(1) \left[ b_T^d b_T^{-\delta d/\{a(4+\delta)\}} + c_T^a \sum_{j=c_T}^{\infty} \{\alpha(j)\}^{\delta/(4+\delta)} \right] \rightarrow 0 \end{aligned}$$

by assumption (A4);

$$\begin{aligned} A_{T22} &:= (Tb_T^d)^{-1} \sum_{1 \leq i < j \leq T} E\Delta_i^{(m)} (\Delta_j - \Delta_j^{(m)}) \\ &= (Tb_T^d)^{-1} \sum_{1 \leq i < j \leq T} \left\{ E(\Delta_i^{(m)})^2 \right\}^{1/2} \left\{ E(\Delta_j - \Delta_j^{(m)})^2 \right\}^{1/2} \\ &= (Tb_T^d)^{-1} \frac{T(T-1)}{2} \left\{ E(\Delta_i^{(m)})^2 \right\}^{1/2} \left\{ E(\Delta_j - \Delta_j^{(m)})^2 \right\}^{1/2}, \end{aligned}$$

and owing to  $E(\Delta_i^{(m)})^2 \leq E(\eta_j^{(m)})^2 = O(b_T^d) + O(b_T^{-2-2d/\delta} v_2(m)) + O(b_T^{-1-2d/\delta} v_2^{1/2}(m)) + o(b_T^d) = O(b_T^d)$ , following from (5.3) of Lemma 5.2 with  $L = b_T^{-d/\delta}$  and the condition (B2), and,

by using the Lipschitz continuity and boundedness of the kernel  $K(\cdot)$  and taking  $\tilde{L}^\delta = T^2/b_T^d$ ,

$$\begin{aligned}
E \left( \Delta_j - \Delta_j^{(m)} \right)^2 &\leq E \left( \eta_j - \eta_j^{(m)} \right)^2 \\
&= E \left[ (Z_j - Z_j^{(m)}) K_{\mathbf{c}} \left( (\mathbf{X}_j^{(m)} - \mathbf{x})/b_T \right) + Z_j \left\{ K_{\mathbf{c}} \left( (\mathbf{X}_j - \mathbf{x})/b_T \right) - K_{\mathbf{c}} \left( (\mathbf{X}_j^{(m)} - \mathbf{x})/b_T \right) \right\} \right]^2 \\
&\leq 2 \left[ E(Z_j - Z_j^{(m)})^2 K_{\mathbf{c}}^2 \left( (\mathbf{X}_j^{(m)} - \mathbf{x})/b_T \right) \right. \\
&\quad \left. + E Z_j^2 (I_{\{|Z_j| \leq \tilde{L}\}} + I_{\{|Z_j| > \tilde{L}\}}) \left\{ K_{\mathbf{c}} \left( (\mathbf{X}_j - \mathbf{x})/b_T \right) - K_{\mathbf{c}} \left( (\mathbf{X}_j^{(m)} - \mathbf{x})/b_T \right) \right\}^2 \right] \\
&\leq C \left[ E(Z_j - Z_j^{(m)})^2 + \tilde{L}^2 b_T^{-2} E \left\| \mathbf{X}_j - \mathbf{X}_j^{(m)} \right\|^2 + E Z_j^2 I_{\{|Z_j| > \tilde{L}\}} \right] \\
&\leq C \left[ v_2(m) + \tilde{L}^2 b_T^{-2} v_2(m) + o(\tilde{L}^{-\delta}) \right],
\end{aligned}$$

we have

$$A_{T22} \leq C T b_T^{-d/2} \left[ \tilde{L} b_T^{-1} v_2^{1/2}(m) + o(\tilde{L}^{-\delta/2}) \right] \leq C \left[ T^{1+2/\delta} b_T^{-1-d/2-d/\delta} v_2^{1/2}(m) + o(1) \right] \rightarrow 0;$$

and similarly to  $A_{T22}$ , it can be proved that

$$A_{T23} := (T b_T^d)^{-1} \sum_{1 \leq i < j \leq T} E \Delta_i (\Delta_j - \Delta_j^{(m)}) \rightarrow 0.$$

The proof of the lemma is completed.  $\square$

**Proof of Lemma 3.3.** The proof of this lemma is routine and omitted.  $\square$

### 5.3 Proofs for Section 3.2

**Proof of Lemma 3.4.** The fundamental idea to prove the asymptotic normality of  $\mathbf{W}_T(\mathbf{x})$  is to divide  $\mathbf{W}_T(\mathbf{x})$  into two parts: with  $m = m_T \rightarrow \infty$  (to be specified later),

$$\mathbf{W}_T(\mathbf{x}) = \mathbf{W}_T^{(m)}(\mathbf{x}) + \left[ \mathbf{W}_T(\mathbf{x}) - \mathbf{W}_T^{(m)}(\mathbf{x}) \right], \quad (5.38)$$

where  $\mathbf{W}_T^{(m)}(\mathbf{x})$  is defined in Subsection 2.3. Then applying the approximation lemma (Lemma 2.1) with  $L = L_T = (T b_T^{-d})^{1/[2(1+\delta)]}$ ,

$$\begin{aligned}
(T b_T^d)^{1/2} \left[ \mathbf{W}_T(\mathbf{x}) - \mathbf{W}_T^{(m)}(\mathbf{x}) \right] &= O_P \left( (T b_T^d)^{1/2} b_T^{-d-1} L \sqrt{v_2(m)} \right) + o_P \left( (T b_T^d)^{1/2} b_T^{-d} L^{-(1+\delta)} \right) \\
&= O_P \left( \left\{ T^{1+1/(1+\delta)} b_T^{-(1+1/(1+\delta))d-2} v_2(m) \right\}^{1/2} \right) + o_P(1) \\
&\leq O_P \left( \left\{ T^{2+4/\delta} b_T^{-(2+d+2d/\delta)} v_2(m) \right\}^{1/2} \right) + o_P(1) \rightarrow_P 0,
\end{aligned} \quad (5.39)$$

following from (B2); and similarly

$$(Tb_T^d)^{1/2} E \left[ \mathbf{W}_T(\mathbf{x}) - \mathbf{W}_T^{(m)}(\mathbf{x}) \right] \rightarrow 0. \quad (5.40)$$

Therefore, to have the conclusion of this lemma, it suffices to prove that

$$(Tb_T^d)^{1/2} (\mathbf{c}^\tau [\mathbf{W}_T^{(m)}(\mathbf{x}) - E\mathbf{W}_T^{(m)}(\mathbf{x})] / \sigma)$$

is asymptotically standard normal as  $T \rightarrow \infty$ , which is the main effort we will made in this paper.

Recalling

$$\eta_j^{(m)}(\mathbf{x}) := Z_j^{(m)} K_{\mathbf{c}}(\mathbf{x} - \mathbf{X}_j^{(m)}) \quad \text{and} \quad \Delta_j^{(m)}(\mathbf{x}) := \eta_j^{(m)}(\mathbf{x}) - E\eta_j^{(m)}(\mathbf{x}), \quad (5.41)$$

define  $\zeta_{Tj}^{(m)} := b_T^{-d/2} \Delta_j^{(m)}$ , and let  $S_T^{(m)} := \sum_{j=1}^T \zeta_{Tj}^{(m)}$ . Then,

$$T^{-1/2} S_T^{(m)} = (Tb_T^d)^{1/2} \mathbf{c}^\tau (W_T^{(m)}(\mathbf{x}) - E W_T^{(m)}(\mathbf{x})) = A_T^{(m)} - EA_T^{(m)},$$

where

$$A_T^{(m)} := (Tb_T^d)^{1/2} \mathbf{c}^\tau \mathbf{W}_T^{(m)} = (Tb_T^d)^{-1/2} \sum_{j=1}^T Z_j^{(m)} K_{\mathbf{c}} \left( \frac{\mathbf{X}_j^{(m)} - \mathbf{x}}{b_T} \right).$$

Now, let us decompose  $T^{-1/2} S_T^{(m)}$  into smaller pieces involving “large” and “small” blocks.

More specifically, consider

$$U^{(m)}(1, T, \mathbf{x}, j) := \sum_{i=j(p+q)+1}^{j(p+q)+p} \zeta_{Ti}^{(m)}(\mathbf{x}),$$

$$U^{(m)}(2, T, \mathbf{x}, j) := \sum_{i=j(p+q)+p+1}^{(j+1)(p+q)} \zeta_{Ti}^{(m)}(\mathbf{x}),$$

where  $p = p_T$  and  $q = q_T$  are specified in Assumption (B3). Without loss of generality, assume that, for some integer  $r = r_T$ ,  $T$  is such that  $T = r(p+q)$ , with  $r \rightarrow \infty$ . For each integer  $1 \leq i \leq 2$ , define

$$\Upsilon^{(m)}(T, \mathbf{x}, i) := \sum_{j=0}^{r-1} U^{(m)}(i, T, \mathbf{x}, j).$$

Clearly  $S_T^{(m)} = \Upsilon^{(m)}(T, \mathbf{x}, 1) + \Upsilon^{(m)}(T, \mathbf{x}, 2)$ . Note that  $\Upsilon^{(m)}(T, \mathbf{x}, 1)$  is the sum of the random variables  $\zeta_{Ti}^{(m)}$  over “large” blocks, whereas  $\Upsilon^{(m)}(T, \mathbf{x}, 2)$  are sums over “small” blocks. If it is not the case that  $T = r(p+q)$  for some integer  $r$ , then an additional term  $\Upsilon^{(m)}(T, \mathbf{x}, 3)$ , say,

containing all the  $\zeta_{Tj}^{(m)}$ 's that are not included in the big or small blocks, can be considered. This term will not change the proof much. The general approach consists in showing that, as  $T \rightarrow \infty$ ,

$$Q_1^{(m)} := \left| \mathbb{E} \exp[iu \Upsilon^{(m)}(T, \mathbf{x}, 1)] - \prod_{j=0}^{r-1} \mathbb{E} \exp[iu U^{(m)}(1, T, \mathbf{x}, j)] \right| \longrightarrow 0, \quad (5.42)$$

$$Q_2^{(m)} := T^{-1} \mathbb{E} \left( \Upsilon^{(m)}(T, \mathbf{x}, 2) \right)^2 \longrightarrow 0, \quad (5.43)$$

$$Q_3^{(m)} := T^{-1} \sum_{j=0}^{r-1} \mathbb{E} [U^{(m)}(1, T, \mathbf{x}, j)]^2 \longrightarrow \sigma^2, \quad (5.44)$$

$$Q_4^{(m)} := T^{-1} \sum_{j=0}^{r-1} \mathbb{E} [(U^{(m)}(1, T, \mathbf{x}, j))^2 I\{|U^{(m)}(1, T, \mathbf{x}, j)| > \varepsilon \sigma T^{1/2}\}] \longrightarrow 0 \quad (5.45)$$

for every  $\varepsilon > 0$ . Note that

$$\begin{aligned} [A_T^{(m)} - \mathbb{E} A_T^{(m)}] / \sigma &= (T b_T^d)^{1/2} \mathbf{c}^\tau [W_T^{(m)}(\mathbf{x}) - \mathbb{E} W_T^{(m)}(\mathbf{x})] / \sigma = S_T^{(m)} / (\sigma T^{1/2}) \\ &= \Upsilon^{(m)}(T, \mathbf{x}, 1) / (\sigma T^{1/2}) + \Upsilon^{(m)}(T, \mathbf{x}, 2) / (\sigma T^{1/2}). \end{aligned}$$

The term  $\Upsilon^{(m)}(T, \mathbf{x}, 2) / (\sigma T^{1/2})$  is asymptotically negligible by (5.43). The random variables  $U^{(m)}(1, T, \mathbf{x}, j)$  are asymptotically mutually independent by (5.42). The asymptotic normality of  $\Upsilon^{(m)}(T, \mathbf{x}, 1) / (\sigma T^{1/2})$  follows from (5.44) and the Lindeberg-Feller condition (5.45). The lemma thus follows if we can prove (5.42)-(5.45). This proof is given below. The arguments are reminiscent of those used by Robinson (1983), Masry (1986) and Nakhapetyan (1987), but, differently from those references, owing to  $m = m_T$  depending on  $T$  in the  $\alpha$ -mixing process  $\{(Y_t^{(m)}, \mathbf{X}_t^{(m)})\}$ , the details become much more complex and involved (c.f., Lu 2001), and heavily depend on Lemmas 2.1 and 5.3 established in the above.

*Proof of (5.42).* Let

$$\mathcal{I}_j = \mathcal{I}(1, T, \mathbf{x}, j) := \{i : j(p+q) + 1 \leq i \leq j(p+q) + p\}.$$

The distance,  $d(\mathcal{I}_j, \mathcal{I}_{j'})$ , between two distinct sets  $\mathcal{I}(1, T, \mathbf{x}, j)$  and  $\mathcal{I}(1, T, \mathbf{x}, j')$  is at least  $q$  if  $j \neq j'$ . Clearly,  $\mathcal{I}(1, T, \mathbf{x}, j)$  is the set of indices involved in  $U^{(m)}(1, T, \mathbf{x}, j)$ , which contains  $p$  indices.

Let  $a_j := \exp\{iuU^{(m)}(1, T, \mathbf{x}, j)\}$ , where  $i^2 = -1$ . Note that

$$\begin{aligned} \mathbb{E}[a_1 \dots a_r] - \mathbb{E}[a_1] \dots \mathbb{E}[a_r] &= \mathbb{E}[a_1 \dots a_r] - \mathbb{E}[a_1] \mathbb{E}[a_2 \dots a_r] \\ &\quad + \mathbb{E}[a_1] \{\mathbb{E}[a_2 \dots a_r] - \mathbb{E}[a_2] \mathbb{E}[a_3 \dots a_r]\} + \dots \\ &\quad + \mathbb{E}[a_1] \mathbb{E}[a_2] \dots \mathbb{E}[a_{r-2}] \{\mathbb{E}[a_{r-1} a_r] - \mathbb{E}[a_{r-1}] \mathbb{E}[a_r]\}. \end{aligned}$$

Since  $|\mathbb{E}[a_j]| \leq 1$ ,

$$\begin{aligned} Q_1^{(m)} = |\mathbb{E}[a_1 \dots a_r] - \mathbb{E}[a_1] \dots \mathbb{E}[a_r]| &\leq |\mathbb{E}[a_1 \dots a_r] - \mathbb{E}[a_1] \mathbb{E}[a_2 \dots a_r]| \\ &\quad + |\mathbb{E}[a_2 \dots a_r] - \mathbb{E}[a_2] \mathbb{E}[a_3 \dots a_r]| + \dots \\ &\quad + |\mathbb{E}[a_{r-1} a_r] - \mathbb{E}[a_{r-1}] \mathbb{E}[a_r]|. \end{aligned}$$

Note that  $d(I_\ell, I_j) \geq q$  for any  $\ell \neq j$ , and set  $q = 2m$ . It follows by applying Lemma 5.1(ii) to each term on the right hand side that

$$Q_1^{(m)} \leq C \sum_{k=1}^{r-1} \alpha_m(q) \leq C r \alpha(q - m) = C r \alpha(m),$$

which tends to zero by condition (B3).

*Proof of (5.43).* For notational simplicity, refer to the random variables  $U^{(m)}(2, T, \mathbf{x}, j)$ ,  $j = 0, 1, \dots, r-1$ , as  $\widehat{U}_1, \dots, \widehat{U}_r$ . We have

$$\mathbb{E}[\Upsilon^{(m)}(T, \mathbf{x}, 2)]^2 = \sum_{i=1}^r \text{Var}(\widehat{U}_i) + 2 \sum_{1 \leq i < j \leq r} \text{Cov}(\widehat{U}_i, \widehat{U}_j) := \widehat{V}_1 + \widehat{V}_2, \quad \text{say.} \quad (5.46)$$

Since  $\mathbf{X}_n$  is stationary (recall that  $\zeta_{Tj}^{(m)}(\mathbf{x}) := b_T^{-d/2} \Delta_j^{(m)}(\mathbf{x})$ ),

$$\text{Var}(\widehat{U}_i) = \sum_{i=1}^q \mathbb{E} \left[ \left( \zeta_{Ti}^{(m)}(\mathbf{x}) \right)^2 \right] + \sum_{1 \leq i < j \leq q} \mathbb{E} \left[ \zeta_{Tj}^{(m)}(\mathbf{x}) \zeta_{Ti}^{(m)}(\mathbf{x}) \right] := \widehat{V}_{11} + \widehat{V}_{12}.$$

From (5.3) and the Lebesgue density theorem (see Chapter 2 of Devroye and Györfi 1985), and then taking  $L = b_T^{-d/\delta}$ ,

$$\begin{aligned} \widehat{V}_{11} &= q \text{Var}\{\zeta_{Ti}^{(m)}(\mathbf{x})\} = q \{b_T^{-d} \mathbb{E} \left( \Delta_i^{(m)}(\mathbf{x}) \right)^2\} \\ &\leq q \{b_T^{-d} \mathbb{E} \left( \eta_i^{(m)}(\mathbf{x}) \right)^2\} = q \{b_T^{-d} \mathbb{E} \left( Z_i^{(m)} K((\mathbf{x} - \mathbf{X}_i^{(m)})/b_T) \right)^2\} \\ &= q \left\{ b_T^{-d} \mathbb{E} \left( Z_i K((\mathbf{x} - \mathbf{X}_i)/b_T) \right)^2 + O(b_T^{-2-d} L^2 v_2(m)) + O(b_T^{1-d} L^2 \sqrt{v_2(m)}) + o(b_T^{-d} L^{-\delta}) \right\} \\ &= C q \left\{ 1 + b_T^{-2-d-2d/\delta} v_2(m) + \left( b_T^{-2-2d-4d/\delta} v_2(m) \right)^{1/2} + o(1) \right\} \\ &= O(q), \end{aligned} \quad (5.47)$$

where the final equality follows from  $b_T^{-2-d-2d/\delta} v_2(m) \leq b_T^{-2-2d-4d/\delta} v_2(m) \leq b_T^{-4(1+d+2d/\delta)} v_2(m) = O(1)$  by Assumption (B2).

We need the cross lemma, Lemma 5.3, for  $\widehat{V}_{12}$ . Thus, applying Lemma 5.3 together with (B2) and then taking  $c_T = q$  yields

$$\begin{aligned} \widehat{V}_{12} &= b_T^{-d} \sum_{1 \leq i < j \leq q} \mathbb{E} \left[ \Delta_j^{(m)}(\mathbf{x}) \Delta_i^{(m)}(\mathbf{x}) \right] \\ &\leq C b_T^{-d} q b_T^d \left[ b_T^d c_T + b_T^{-\delta d/(4+\delta)} \left( \sum_{t=c_T}^T \{\alpha_m(t)\}^{\delta/(4+\delta)} \right) \right] \\ &= C q \left[ b_T^d q T + b_T^{-\delta d/(4+\delta)} \left( \sum_{t=q_T}^{\infty} \{\alpha_m(t)\}^{\delta/(4+\delta)} \right) \right] \\ &:= C q \pi_T. \end{aligned}$$

It follows from assumption (B4) that  $\pi_T = O(1)$  and

$$T^{-1} \widehat{V}_1 = T^{-1} r \left( \widehat{V}_{11} + \widehat{V}_{12} \right) \leq T^{-1} r C q [1 + \pi_T] \leq C(q/p) [1 + \pi_T]. \quad (5.48)$$

Set

$$\mathcal{I}(2, T, \mathbf{x}, j) := \{i : j(p+q) + p + 1 \leq i \leq (j+1)(p+q)\}.$$

Then  $U^{(m)}(2, T, \mathbf{x}, j) = \sum_{i \in \mathcal{I}(2, T, \mathbf{x}, j)} \zeta_{Ti}^{(m)}$ . Since  $p > q$ , if  $i$  and  $i'$  belong to two distinct sets  $\mathcal{I}(2, T, \mathbf{x}, j)$  and  $\mathcal{I}(2, T, \mathbf{x}, j')$ , then  $|i - i'| > q$ . In view of (5.46) and (5.34) and then assumption (B2), we obtain

$$\begin{aligned} |\widehat{V}_2| &\leq C \sum_{\{i, j : |i-j| \geq q, 1 \leq i, j \leq T\}} \sum_{1 \leq i, j \leq T} \left| \mathbb{E} \left[ \zeta_{Ti}^{(m)}(\mathbf{x}) \zeta_{Tj}^{(m)}(\mathbf{x}) \right] \right| \\ &\leq C b_T^{-d} \sum_{\{i, j : |i-j| \geq q, 1 \leq i, j \leq T\}} \sum_{1 \leq i, j \leq T} \left| \mathbb{E} \left[ \Delta_{Ti}^{(m)}(\mathbf{x}) \Delta_{Tj}^{(m)}(\mathbf{x}) \right] \right| \\ &\leq C b_T^{-d} \sum_{\{i, j : |i-j| \geq q, 1 \leq i, j \leq T\}} \sum_{1 \leq i, j \leq T} b_T^{4d/(4+\delta)} \{\alpha_m(j-i)\}^{\delta/(4+\delta)} \\ &\leq C b_T^{-\delta d/(4+\delta)} T \left( \sum_{t=q}^{\infty} \{\alpha_m(t)\}^{\delta/(4+\delta)} \right). \end{aligned} \quad (5.49)$$

Condition (B4) implies that  $q b_T^d = O(1)$  and  $\pi_T = O(1)$ . Thus, from (5.46), (5.48), and (5.49),

$$T^{-1} \mathbb{E}[\Upsilon^{(m)}(T, \mathbf{x}, 2)]^2 \leq C(q/p) [1 + \pi_T] + C b_T^{-\delta d/(4+\delta)} \left( \sum_{t=q}^{\infty} \{\alpha_m(t)\}^{\delta/(4+\delta)} \right),$$

which tends to zero by  $q/p \rightarrow 0$  and condition (B4); (5.43) follows.

*Proof of (5.44).* Let  $S_T^{(m)'} := \Upsilon^{(m)}(T, \mathbf{x}, 1)$  and  $S_T^{(m)''} := \Upsilon^{(m)}(T, \mathbf{x}, 2)$ . Then  $S_T^{(m)'}$  is a sum of  $\zeta_j^{(m)}$ 's over the “large” blocks,  $S_T^{(m)''}$  over the “small” ones. Lemma 3.2 and its argument together with Lemma 5.2 imply  $T^{-1}\mathbb{E} \left[ |S_T^{(m)}|^2 \right] \rightarrow \sigma^2$ . This, combined with (5.43), entails  $T^{-1}\mathbb{E} \left[ |S_T^{(m)'}|^2 \right] \rightarrow \sigma^2$ . Now,

$$T^{-1}\mathbb{E} \left[ |S_T^{(m)'}|^2 \right] = T^{-1} \sum_{j=0}^{r-1} \mathbb{E}[U^{(m)}(1, T, \mathbf{x}, j)]^2 + T^{-1} \sum_{i \neq j \in \mathcal{J}^*} \text{Cov} \left( U^{(m)}(1, T, \mathbf{x}, j), U^{(m)}(1, T, \mathbf{x}, i) \right). \quad (5.50)$$

where  $\mathcal{J}^* = \mathcal{J}^*(p, q) := \{i, j : 1 \leq i, j \leq r-1\}$ . Observe that (5.44) follows from (5.50) if the last sum in the right-hand side of (5.50) tends to zero as  $T \rightarrow \infty$ . Using the same argument as in the derivation of the bound (5.46) for  $\widehat{V}_2$ , this sum can be bounded by

$$T^{-1} C b_T^{-\delta d/(4+\delta)} \sum_{i>p} \sum_{j=1}^{r-1} \{\alpha_m(i)\}^{\delta/(4+\delta)} \leq C b_T^{-\delta d/(4+\delta)} \left( \sum_{t=p}^{\infty} \{\alpha_m(t)\}^{\delta/(4+\delta)} \right),$$

which tends to zero by condition (B4).

*Proof of (5.45).* We need a truncation argument because  $Z_i^{(m)}$  is not necessarily bounded. Set  $Z_i^{(m)M} := Z_i^{(m)} I_{\{|Z_i^{(m)}| \leq M\}}$ ,  $\eta_i^{(m)M} := Z_i^{(m)M} K_{\mathbf{c}}((\mathbf{X}_i^{(m)} - \mathbf{x})/b_T)$ ,  $\Delta_i^{(m)M} := \eta_i^{(m)M} - \mathbb{E}\eta_i^{(m)M}$ ,  $\zeta_{Ti}^{(m)M} := b_T^{-d/2} \Delta_i^{(m)M}$ , where  $M$  is a fixed positive constant, and define  $U^{(m)M}(1, T, \mathbf{x}, j) := \sum_{i \in \mathcal{I}(1, T, \mathbf{x}, j)} \zeta_{Ti}^{(m)M}$ . Put

$$Q_4^{(m)M} := T^{-1} \sum_{j=0}^{r-1} \mathbb{E} \left[ (U^{(m)M}(1, T, \mathbf{x}, j))^2 I_{\{|U^{(m)M}(1, T, \mathbf{x}, j)| > \varepsilon \sigma T^{1/2}\}} \right].$$

Clearly,  $|\zeta_{Ti}^{(m)M}| \leq C M b_T^{-d/2}$ . Therefore  $|U^{(m)M}(1, T, \mathbf{x}, j)| < C M p b_T^{-d/2}$ . Hence

$$Q_4^{(m)M} \leq C p^2 b_T^{-d} T^{-1} \sum_{j=0}^{r-1} \mathbb{P}[U^{(m)M}(1, T, \mathbf{x}, j) > \varepsilon \sigma T^{1/2}].$$

Now,  $U^{(m)M}(1, T, \mathbf{x}, j)/(\sigma T^{1/2}) \leq C p (T b_T^d)^{-1/2} \rightarrow 0$ , owing to assumption (B3). Thus  $\mathbb{P}[U^{(m)M}(1, T, \mathbf{x}, j) > \varepsilon \sigma T^{1/2}] = 0$  at all  $j$  for sufficiently large  $T$ . Thus  $Q_4^{(m)M} = 0$  for large  $T$ , and (5.45) holds for the truncated variables. Hence,

$$T^{-1/2} S_T^{(m)M} := T^{-1/2} \sum_{j=1}^T \zeta_{Tj}^{(m)M} \xrightarrow{\mathcal{L}} N(0, \sigma_M^2), \quad (5.51)$$

where  $\sigma_M^2 := \text{Var}(Z_i^M | \mathbf{X}_i = \mathbf{x}) f(\mathbf{x}) \int K_{\mathbf{c}}^2(\mathbf{u}) d\mathbf{u}$ , and  $Z_i^M = Z_i I_{\{|Z_i| \leq M\}}$ .

Defining  $S_T^{(m)M*} := \sum_{j=1}^T (\zeta_{Tj} - \zeta_{Tj}^{(m)M})$ , we have  $S_T^{(m)} = S_T^{(m)M} + S_T^{(m)M*}$ . Note that

$$\begin{aligned}
\left| \mathbb{E} \left[ \exp(iuS_T^{(m)})/T^{1/2} \right] - \exp(-u^2\sigma^2/2) \right| &\leq \left| \mathbb{E}[\exp(iuS_T^{(m)M}/T^{1/2}) - \exp(-u^2\sigma_M^2/2)] \exp(iuS_T^{(m)M^*}/T^{1/2}) \right| \\
&+ \left| \mathbb{E}[\exp(iuS_T^{(m)M^*}/T^{1/2}) - 1] \exp(-u^2\sigma_M^2/2) \right| + \left| \exp(-u^2\sigma_M^2/2) - \exp(-u^2\sigma^2/2) \right| \\
&= E_1 + E_2 + E_3, \quad \text{say.}
\end{aligned}$$

Letting  $T \rightarrow \infty$ ,  $E_1$  tends to 0 by (5.51) and the dominated convergence theorem. Letting  $M$  go to infinity, the dominated convergence theorem also implies that  $\sigma_M^2 := \text{Var}(Z_i^M | \mathbf{X}_i = \mathbf{x}) f(\mathbf{x}) \int K_c^2(\mathbf{u}) d\mathbf{u}$  converges to

$$\text{Var}(Z_i | \mathbf{X}_i = \mathbf{x}) f(\mathbf{x}) \int K_c^2(\mathbf{u}) d\mathbf{u} = \text{Var}(Y_i | \mathbf{X}_i = \mathbf{x}) f(\mathbf{x}) \int K_c^2(\mathbf{u}) d\mathbf{u} := \sigma^2,$$

and hence that  $E_3$  tends to 0. Finally, in order to prove that  $E_2$  also tends to 0, it suffices to show that  $S_T^{(m)M^*}/T^{1/2} \rightarrow 0$  in probability as first  $T \rightarrow \infty$  and then  $M \rightarrow \infty$ , which in turn would follow if we could show that

$$\mathbb{E} \left[ (S_T^{(m)M^*}/T^{1/2})^2 \right] \rightarrow \text{Var}(|Z_i| I_{\{|Z_i| > M\}} | \mathbf{X}_i = \mathbf{x}) f(\mathbf{x}) \int K_c^2(\mathbf{u}) d\mathbf{u} \quad \text{as } T \rightarrow \infty.$$

This follows along the same lines as the argument of Lemma 3.3 together with Lemma 5.2. The proof of Lemma 3.4 is thus complete.  $\square$

**Proof of Theorem 3.1.** We establish that the bandwidth conditions (B1)-(B4) hold. First, Assumption (B1) holds clearly by the condition  $Tb_T^{[1+2/\{(a(1+4/\delta)\}]^d}/\log T \rightarrow \infty$ .

Next, take  $m = m_T = \lfloor b_T^{-\delta d/\{a(4+\delta)\}} \rfloor$  and  $q = 2m$ , where  $\lfloor a \rfloor$  stands for the integer part of  $a$ . Then it easily follows from assumption (A4) that (B4) holds.

On the other hand, note that  $v_{2+\delta/2}(m) = O(m^{-\mu})$  implies  $v_2(m) = O(m^{-\mu/(1+\delta/4)})$ .

$$\begin{aligned}
T^{2+4/\delta} b_T^{-(2+d+2d/\delta)} v_2(m) &\leq CT^{2+4/\delta} b_T^{-(2+d+2d/\delta)} m^{-\mu/(1+\delta/4)} \\
&\leq CT^{2+4/\delta} b_T^{-(2+d+2d/\delta)} b_T^{\mu\delta d/\{a(4+\delta)(1+\delta/4)\}} \\
&\leq T^{2+4/\delta} b_T^{\mu\delta d/\{a(4+\delta)(1+\delta/4)\} - (2+d+2d/\delta)} = T^{2+4/\delta} b_T^{\mu/\kappa_2 - \kappa_1}, \tag{5.52}
\end{aligned}$$

which tends to 0 by the condition of this theorem ;

$$\begin{aligned}
b_T^{-4(1+d+2d/\delta)} v_2(m) &\leq C b_T^{-4(1+d+2d/\delta)} m^{-\mu/(1+\delta/4)} \\
&\leq C b_T^{-4(1+d+2d/\delta)} b_T^{\mu\delta d/\{a(4+\delta)(1+\delta/4)\}} \\
&\leq b_T^{\mu\delta d/\{a(4+\delta)(1+\delta/4)\} - 4(1+d+2d/\delta)} = b_T^{\mu/\kappa_2 - 4(\kappa_1 - 1)} = O(1), \tag{5.53}
\end{aligned}$$

which is deduced by the condition  $\mu \geq 4(\kappa_1 - 1)\kappa_2$ ; and

$$\begin{aligned} b_T^{-(2+2d+\delta/2+4d/\delta)} v_{2+\delta/2}(m) &\leq C b_T^{-(2+2d+\delta/2+4d/\delta)} m^{-\mu} \\ &\leq C b_T^{-(2+2d+\delta/2+4d/\delta)} b_T^{\mu\delta/\{a(4+\delta)\}} \\ &\leq b_T^{\mu\delta/\{a(4+\delta)\}-(2+2d+\delta/2+4d/\delta)} = b_T^{\mu(1+\delta/4)/\kappa_2-\kappa_3} = O(1), \end{aligned} \quad (5.54)$$

which is deduced by the condition  $\mu \geq \kappa_2\kappa_3/(1 + \delta/4)$ . Therefore Assumption (B2) holds.

Set  $p = (Tb_T^d/\log T)^{1/2}$ . Then  $p = o((Tb_T^d)^{1/2})$  and  $T/p = \{T \log T/b_T^d\}^{1/2} \rightarrow \infty$  clearly; and

$$q/p = \left\{ \frac{\log T}{T b_T^{(1+2/\{a(1+4/\delta)\})d}} \right\}^{1/2} \rightarrow 0$$

and

$$\frac{T}{p} \alpha(m) = \left\{ T \log T/b_T^d \right\}^{1/2} m^{-\lambda} = \left[ T \log T b_T^{(2\lambda/\{a(1+4/\delta)\}-1)d} \right]^{1/2} \rightarrow 0$$

by the conditions of this theorem. Therefore Assumption (B3) holds.

Finally, as  $\alpha(x) = O(x^{-\lambda})$  for some  $\lambda > (a+1)(1+4/\delta)$  with  $a > \delta/(4+\delta)$ ,

$$k^a \sum_{j=k}^{\infty} \{\alpha(j)\}^{\delta/(4+\delta)} = k^a \sum_{j=k}^{\infty} j^{-\lambda\delta/(4+\delta)} = O(1) k^{a-\lambda\delta/(4+\delta)+1} \rightarrow 0$$

as  $k \rightarrow \infty$ . Therefore Assumption (A4) holds. The theorem thus follows from Lemmas 3.1–3.3, and 3.4.  $\square$

**Proof of Corollary 3.1.** This corollary easily follows by checking the conditions on the bandwidth in Theorem 3.1. The detail is omitted.  $\square$

## References

- [1] Andrews, D. W. K. (1984). Non-strong mixing autoregressive processes. *Journal of Applied Probability* **21**, 930–934.
- [2] Andrews, D.W.K. (1995). Nonparametric Kernel Estimation for Semiparametric Models. *Econometric Theory* **11**, 560-596.
- [3] Bandi, F. (2004). *On persistence and nonparametric estimation (with an application to stock return predictability)*. Working paper.

- [4] Bierens, H. (1981). *Robust Methods and Asymptotic Theory*. Lecture Notes in Economics and Mathematical Systems no. 192. Berlin, Springer.
- [5] Billingsley, P. (1968). *Convergence of Probability Measures*. New York, Wiley
- [6] Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics* **31**, 307–327.
- [7] Carrasco, M., Chen, X., (2002). Mixing and moment properties of various GARCH and stochastic volatility models, *Econometric Theory* **18**, 17-39.
- [8] Cleveland, W.S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association* **74**, 829–836.
- [9] Cleveland, W.S. and Loader, C. (1996). Smoothing by local regression: principles and methods. In: *Statistical Theory and Computational Aspects of Smoothing*, W. Härdle and M.G. Schimek, Eds. Heidelberg: Physica-Verlag.
- [10] Cline, D.B.H. and Pu, H.M.H. (1999). Geometric ergodicity of nonlinear time series, *Statistica Sinica* **9**, 1103-1118.
- [11] Craven, P. and Wahba, G. (1979) Smoothing noise data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation. *Numeri. Math.* **31**, 377-403.
- [12] Davidson (1994). *Stochastic limit theory*. Oxford University Press.
- [13] Deo, C. M. (1973). A note on empirical processes of strong mixing sequences. *Annals of Probability* **1**, 870–875.
- [14] Devroye, L., and Györfi, L. (1985). *Nonparametric Density Estimation: The L1 View*. New York: John Wiley.
- [15] Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of United Kingdom inflation. *Econometrica* **50**, 987-1007.
- [16] Fan, J. (1992). Design-adaptive nonparametric regression. *Journal of the American Statistical Association* **87**, 998–1004.

- [17] Fan, J. and Gijbels, I. (1992). Variable bandwidth and local linear regression smoothers. *Annals of Statistics* **20**, 2008–2036.
- [18] Fan, J. and Gijbels, I. (1995). Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaption. *Journal of the Royal Statistical Society, Series B* **57**, 371–394.
- [19] Fan, J. and Gijbels, I. (1996). *Local polynomial modeling and its applications*. London: Chapman and Hall.
- [20] Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: nonparametric and parametric methods*. New York: Springer-Verlag.
- [21] Gallant, A.R. (1987). *Nonlinear Statistical Models*. New York, Wiley
- [22] Gallant, A.R., and H. White (1988). *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. New York, Blackwell.
- [23] Gorodetskii, V. V. (1977). On the strong mixing properties for linear sequences. *Theory of Probability and its Applications* **22**, 411–413.
- [24] Hastie, T. and Loader, C. (1993). Local regression: automatic kernel carpentry. *Statistical Science* **8**, 120–143.
- [25] Ibragimov, I.A. (1962). Some limit theorems for stationary processes. *Theory of Probability and its Applications* **7**, 349–382.
- [26] Ibragimov, I. A., and Linnik, Yu. V. (1971). *Independent and Stationary Sequences of Random Variables*. Groningen: Wolters-Noordhoff.
- [27] Karlsen, H.A. and D. Tjøstheim (2001). Nonparametric estimation in null recurrent time series. *The Annals of Statistics* **29**, 372–416.
- [28] Ling, S. and Li, W. K. (1997). On fractionally integrated autoregressive moving-average time series models with conditional heteroscedasticity, *Journal of the American Statistical Association* **92**, 1184–1194.
- [29] Lu, Z. (1996a). A note on geometric ergodicity of autoregressive conditional heteroscedasticity (ARCH) model. *Statistics and Probability Letters*, **30**, 305–311.

- [30] Lu, Z. (1996b). Geometric ergodicity of a general ARCH type model with applications to some typical models. *Chinese Science Bulletin*, **41**, 1630.
- [31] Lu, Z. (1998). On geometric ergodicity of a non-linear autoregressive (AR) model with an autoregressive conditional heteroscedastic (ARCH) term. *Statistica Sinica* **8**, 1205–1217.
- [32] Lu, Z. (2001). Asymptotic normality of kernel density estimators under dependence. *Annals of the Institute of Statistical Mathematics* **53**, 447-468.
- [33] Lu, Z. and Jiang, Z. (2001).  $L_1$  geometric ergodicity of a multivariate nonlinear AR model with an ARCH term, *Statistics & Probability Letters*, **51**, 121-130.
- [34] Masry, E. (1986). Recursive probability density estimation from weakly dependent processes. *IEEE Transactions on Information Theory* **IT-32**, 254-267.
- [35] Masry, E. and Fan, J. (1997). Local polynomial estimation of regression functions for mixing processes. *Scandinavian Journal of Statistics* **24**, 165-179.
- [36] Masry, E. and Györfi, L. (1987). Strong consistency and rates for recursive density estimators for stationary mixing processes. *Journal of Multivariate Analysis* **22**, 79–93.
- [37] McLeish, D.L. (1975a). A maximal inequality and dependent strong laws. *Annals of Probability* **3**, 826-836.
- [38] McLeish, D.L. (1975b). Invariance principles for dependent variables. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **32**, 165-178.
- [39] McLeish, D.L. (1977). On the invariance principle for nonstationary mixingales. *Annals of Probability* **5**, 616-621
- [40] Moloche, G. (2000). *Local nonparametric estimation of scalar diffusions*. Working paper.
- [41] Nakhapetyan, B. S. (1987). An approach to proving limit theorems for dependent random variables. *Theory of Probability and its Applications* **32**, 535–539.
- [42] Nze, P.A., P. Bühlmann, and P. Doukhan (2002). Weak dependence beyond mixing and asymptotics for nonparametric regression. *Annals of Statistics* **30**, 397-430.

- [43] Nze, P.A., and P. Doukhan (2003). Weak dependence: models and applications to econometrics. Manuscript, CNRS.
- [44] Pham, T. D. (1986). The mixing properties of bilinear and generalized random coefficient autoregressive models. *Stochastic Processes and their Applications* **23**, 291–300.
- [45] Pham, D. T., and Tran, L. T. (1985). Some strong mixing properties of time series models. *Stochastic Processes and their Applications* **19** 297–303.
- [46] Phillips, P.C.B. and J.Y. Park (1998). Nonstationary density estimation and kernel autoregression. Cowles Foundation Discussion Paper no. 1181
- [47] Robinson, P.M. (1983). Nonparametric estimators for time series. *Journal of Time Series Analysis* **4**, 185–207.
- [48] Ruppert D, and Wand, M P. (1994). Multivariate locally weighted least squares regression. *Annals of Statistics* **22**, 1346–1370.
- [49] Saikkonen, P. (2001). Stability Results for Nonlinear Vector Autoregressions with an Application to a Nonlinear Error Correction Model. DISCUSSION PAPERS OF INTERDISCIPLINARY RESEARCH PROJECT 373, <http://sfb.wiwi.hu-berlin.de/>
- [50] Stone, C.J. (1977). Consistent nonparametric regression. *Annals of Statistics* **5**, 595–620.
- [51] Stone, C.J. (1980), “Optimal rates of convergence for nonparametric estimators,” *Annals of Statistics*, **8**, 1348-1360.
- [52] Tjøstheim, D. (1990) Nonlinear time series and Markov chains. *Advances in Applied Probability* **22**, 587–611.
- [53] Tong, H. (1990). *Nonlinear time series: a dynamical system approach*. Oxford University Press, Oxford.
- [54] Wahba, G. (1977) A survey of some smoothing problems and the method of generalized cross-validation for solving them. In: *Applications of Statistics* (ed. P.R. Krishnaiah), pp. 507-523. Amsterdam: North-Holland.
- [55] Weiss, A. A. (1984). ARMA model with ARCH errors, *Journal of Time Series Analysis* **5**, 129–143.