Econometric Theory, **23**, 2007, 37–70. Printed in the United States of America.
DOI: 10.1017/S0266466607070028

LOCAL LINEAR FITTING UNDER NEAR EPOCH DEPENDENCE

ZUDI LU
London School of Economics
Chinese Academy of Sciences
Curtin University of Technology

OLIVER LINTON
London School of Economics

Local linear fitting of nonlinear processes under strong (i.e., α-) 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 autoregressive moving average (ARMA) and generalized autoregressive conditionally heteroskedastic (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.

1. INTRODUCTION

We consider local linear modeling 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 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 an $\mathbb{R}^{1+d}$-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 ϕ-, ρ-, β-, and α-mixing, α-mixing is no doubt the weakest and most popular in the econometric literature. Under some suitable conditions, the stationary solutions of many time series econo-

We thank Yuichi Kitamura and two referees for helpful comments. This research was partially supported by a Leverhulme Trust research grant, the National Natural Science Foundation of China, and the Economic and Social Science Research Council of the UK. Address correspondence to Oliver Linton, Department of Economics, London School of Economics, London, WC2A 2AE, UK; e-mail: o.linton@lse.ac.uk.

© 2007 Cambridge University Press 0266-4666/07 $12.00

https://doi.org/10.1017/S0266466607070028 Published online by Cambridge University Press

metric models (linear or nonlinear) are α-mixing; see, for example, Goródetskii (1977), Pham (1986), Pham and Tran (1985), Tjøstheim (1990), and Tong (1990) on nonlinear autoregressive (AR) models and see Masry and Tjøstheim (1995), Lu (1998), Cline and Pu (1999), Lu and Jiang (2001), Carrasco and Chen (2002), and Saikkonen (2001) on nonlinear autoregressive conditionally heteroskedastic/generalized autoregressive conditionally heteroskedastic (ARCH/GARCH) models. This mixing condition is used extensively in the time series literature (cf. 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* $\{X_t,\ t = 0,\ \pm 1, \ldots\}$ *is said to be* α*-mixing if*

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

as $k \to \infty$*, where* $\mathcal{F}_{-\infty}^{n}$ *and* $\mathcal{F}_{n+k}^{\infty}$ *are two* σ*-fields generated by* $\{X_t,\ t \le n\}$ *and* $\{X_t,\ t \ge n + k\}$*, respectively. We call* $\alpha(\cdot)$ *the mixing coefficient.*

However, from a practical point of view, the α-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 α-mixing and (ii) α-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 $X_t = \frac{1}{2}X_{t-1} + e_t$, with e_t's being independent symmetric Bernoulli random variables taking values -1 and 1, is not α-mixing (but it is near epoch dependent [NED] defined in Definition 1, which follows). For the latter case, the autoregressive moving average (ARMA) process with ARCH/GARCH errors, discussed in Engle (1982) and Weiss (1984) and also Ling and Li (1997), is well applied in financial econometrics, where the model is composed of two time series (ARMA and ARCH/GARCH) models:

$$X_t = a_1 X_{t-1} + \cdots + a_p X_{t-p} + \varepsilon_t - b_1 \varepsilon_{t-1} - \cdots - b_q \varepsilon_{t-q}, \tag{1.2}$$

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

where a_i and b_i are the coefficients in the ARMA(p,q) model, and α_i and β_i are the coefficients in the GARCH(P,Q) model, with e_t being independent and identically distributed (i.i.d.) innovation with mean 0 and variance 1. Although the ARCH model and its generalized version, GARCH (see Bollerslev, 1986), have been proved to be α-mixing under some mild conditions (cf. Lu, 1996a, 1996b; Carrasco and Chen, 2002), no results exist to establish whether com-

https://doi.org/10.1017/S0266466607070028 Published online by Cambridge University Press

pound processes are α-mixing. For this reason we use a generalized version of mixing processes, called stable or NED processes, which can easily cover the compounded processes and many nonlinear/non-α-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) and Nze and Doukhan (2004) have investigated an alternative class of dependent processes they call "weak dependent."[1] 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, $\mathbb{R}^1$- and $\mathbb{R}^d$-valued, respectively, defined based on a stationary process $\{\varepsilon_t\}$ by

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

$$\mathbf{X}_t = (X_{t1}, \ldots, X_{td})^\tau = \Psi_X(\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots), \tag{1.5}$$

where $\mathbf{X}^\tau$ denotes the transpose of $\mathbf{X}$ (a vector or matrix), $\Psi_Y : \mathbb{R}^\infty \to \mathbb{R}^1$ and $\Psi_X : \mathbb{R}^\infty \to \mathbb{R}^d$ 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_ν* norm *(NED in L_ν for simplicity) with respect to a stationary α-mixing process $\{\varepsilon_t\}$ if*

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

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

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

$$\alpha_m(k) \le \begin{cases} \alpha(k-m) & k \ge m+1, \\ 1 & k \le m. \end{cases} \tag{1.7}$$

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, and near epoch dependence is sometimes easier to verify in the case of these models.

Our limiting results will resemble conventional limiting results for local linear estimators under more standard conditions. That is, under some restrictions

https://doi.org/10.1017/S0266466607070028 Published online by Cambridge University Press

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 (for near-integrated processes, see also Bandi, 2004) for which the rates of convergence are slower and limiting distributions are nonnormal. Moloche (2000) also discusses local polynomial estimation for recurrent diffusions.

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 (cf. (1.6)), with order $\nu = 2 + \delta/2$, with respect to the α-mixing process $\{\varepsilon_t\}$, where the constant $\delta > 0$ is specified in Assumption (A2), which follows. 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) (i) The regression function $g(\mathbf{x}) = E(Y_t|\mathbf{X}_t = x)$ 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 $\mathbf{x}$. (ii) The density function $f(\mathbf{x})$ is continuous at $\mathbf{x}$. (iii) The conditional variance function $\mathrm{Var}(Y_t|\mathbf{X}_t = \mathbf{x})$ is continuous at $\mathbf{x}$.

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

(A4) For the mixing coefficient of ε_t, the function α is such that

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

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

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

https://doi.org/10.1017/S0266466607070028 Published online by Cambridge University Press

(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 $C > 0$,

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

This assumption allows an unbounded support for the kernel function; compare this with Condition 2(i) of Masry and Fan (1997, p. 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 $(Y_t, \mathbf{X}_t)$, $t = 1,2,\ldots,T$. For convenient reference, we list here some conditions on the asymptotic behavior, as $T \to \infty$, of the bandwidth b_T that will be used in the discussion that follows (cf. Lemma 3.4). Assumption (B1), which follows, is standard, whereas Assumptions (B2)–(B4) look complex: some simple and verifiable conditions on the stability and mixing coefficients to ensure that 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 $Tb_T^d \to \infty$ as $T \to \infty$.

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

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

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 $m = m_T \to \infty$ such that $p = p_T = o((Tb_T^d)^{1/2})$, $q/p \to 0$ and $T/p \to \infty$, and $Tp^{-1}\alpha(m) \to 0$.

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

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

Remark. Assumption (B1) is standard on the bandwidth, the same 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 that are associated with the bandwidth, among which

https://doi.org/10.1017/S0266466607070028 Published online by Cambridge University Press

(B3) together with (B1) is similar to the conditions specified for the strongly mixing processes in Condition 3 of Masry and Fan (1997, p. 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)–(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 Squares Local Linear Fitting

Although the Nadaraya–Watson method is central in most nonparametric regression methods in the traditional i.i.d. series case, it has been well documented (see, e.g., 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—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 α-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 of approximating, in a neighborhood of $\mathbf{x}$, the unknown function g by a linear function. Under Assumption (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 $(a_0, \mathbf{a}_1^{\tau}) = (g(\mathbf{x}), g'(\mathbf{x}))$, hence constructing an estimator of g from

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

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 $\mathbb{R}^+$.

https://doi.org/10.1017/S0266466607070028 Published online by Cambridge University Press

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 ith row $(1, b_T^{-1}(\mathbf{X}_i - \mathbf{x})^\tau)$, $\mathcal{W}(\mathbf{x}) = b_T^{-1}\operatorname{diag}(K((\mathbf{X}_1 - \mathbf{x})/b_T), \ldots, K((\mathbf{X}_T - \mathbf{x})/b_T))$, and $\mathcal{Y} = (Y_1, \ldots, Y_T)^\tau$ (see, e.g., Fan and Gijbels, 1996). In the generalized NED series case, though such an expression still holds, we instead write the solution to (2.3) in the following form, which is very convenient for the purpose of characterizing its limiting distribution:

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

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

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

and

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

It follows that

$$\begin{aligned} \mathbf{H}_T &:= \begin{pmatrix} \hat{a}_0 - a_0 \\ \hat{\mathbf{a}}_1 b_T - \mathbf{a}_1 b_T \end{pmatrix} = \begin{pmatrix} g_T(\mathbf{x}) - g(\mathbf{x}) \\ (g_T'(\mathbf{x}) - g'(\mathbf{x}))b_T \end{pmatrix} \\ &= \mathbf{U}_T^{-1}\left\{\mathbf{V}_\mathbf{T} - \mathbf{U}_\mathbf{T}\begin{pmatrix} a_0 \\ a_1 b_T \end{pmatrix}\right\} := \mathbf{U}_T^{-1}\mathbf{W}_T, \end{aligned} \tag{2.4}$$

where

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

$$i = 0, \ldots, d, \tag{2.5}$$

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 conditions (C1)–(C3) (to be presented subsequently) apply, then (2.4) and Slutsky's classical argument imply that, for all $\mathbf{x}$ (all quantities involved indeed depend on $\mathbf{x}$),

https://doi.org/10.1017/S0266466607070028 Published online by Cambridge University Press

$$(Tb_T^d)^{1/2}\left[\begin{pmatrix} g_T(\mathbf{x}) - g(\mathbf{x}) \\ (g_T'(\mathbf{x}) - g'(\mathbf{x}))b_T \end{pmatrix} - \mathbf{U}^{-1}\boldsymbol{\mu}_T\right]$$

$$= (Tb_T^d)^{1/2}[\mathbf{H}_T - \mathbf{U}^{-1}\boldsymbol{\mu}_T] \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, \mathbf{U}^{-1}\boldsymbol{\Sigma}\mathbf{U}^{-1\tau}).$$

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.

Conditions (C1)–(C2) are now presented.

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

(C2) $(Tb_T^d)^{1/2}(E\mathbf{W}_T - \boldsymbol{\mu}_T) \to \mathbf{0}$ and $\mathrm{Var}((Tb_T^d)^{1/2}\mathbf{W}_T) \to \boldsymbol{\Sigma}$ for some positive definite matrix Σ, and

(C3) $\mathbf{U}_T \xrightarrow{\mathrm{P}} \mathbf{U}$ for some nonsingular matrix $\mathbf{U}$.

2.3. Approximations

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

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

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

where

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

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

$$\alpha_m(k) \le 1 \quad \text{for } k = 0,1,\ldots,m, \quad \text{and } \alpha_m(k) = \alpha(k-m) \quad \text{for } k \ge m+1, \tag{2.9}$$

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

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

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

https://doi.org/10.1017/S0266466607070028 Published online by Cambridge University Press

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

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

and

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

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 ∞ and $L = L_T$ a positive real number tending to ∞ as $T \to \infty$. Then under Assumptions (A2) and (A5), as $T \to \infty$,*

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

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

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

Proof. See Section A.2 in the Appendix.

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, which follows, and is therefore omitted.

LEMMA 3.1. *Assume that Assumptions (A1), (A4), and (A5) hold and that b_T satisfies Assumptions (B1) and (B2). Then*

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

as $T \to \infty$.

https://doi.org/10.1017/S0266466607070028 Published online by Cambridge University Press

The remainder of this Section 3.1 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 $\mathbf{c}$ and hence that of $(\mathbf{T}b_T^d)^{1/2}\mathbf{W}_T$.

LEMMA 3.2. *Assume that Assumptions (A1), (A2), (A3)(ii) and (iii), (A4), and (A5) hold and that b_T satisfies Assumptions (B1) and (B2). Then*

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

where

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

Hence, $\lim_{T\to\infty} \mathrm{Var}((\mathbf{T}b_T^d)^{1/2}\mathbf{W}_T) = \mathbf{\Sigma}$.

Proof. See Section A.2 in the Appendix.

Next, we consider the asymptotic behavior of $E[A_T]$.

LEMMA 3.3. *Assume that Assumptions (A3)(i) and (ii) and (A5) hold. Then*

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

where

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

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

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

Proof. The proof is routine and is omitted.

https://doi.org/10.1017/S0266466607070028 Published online by Cambridge University Press

3.2. Asymptotic Normality

The asymptotic normality of our estimators relies on the following lemma (see (2.5) for the definition of $\mathbf{W}_T(\mathbf{x})$).

LEMMA 3.4. *Suppose that Assumptions (A1), (A2), (A3)(ii) and (iii), (A4), and (A5) hold with* $f(\mathbf{x}) > 0$ *and that the bandwidth* b_T *satisfies conditions (B1)–(B4). Denote by* σ^2 *the asymptotic variance (3.1). Then* $(Tb_T^d)^{1/2}(\mathbf{c}^\tau[\mathbf{W}_T(\mathbf{x}) - E\mathbf{W}_T(\mathbf{x})]/\sigma)$ *is asymptotically standard normal as* $T \to \infty$.

Proof. See Section A.3 in the Appendix.

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

THEOREM 3.1. *Let Assumptions (A1), (A2), (A3), and (A5) hold, with* $f(\mathbf{x}) > 0, v_{2+\delta/2}(x) = O(x^{-\mu})$, *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* $T^{2+4/\delta}b_T^{\mu/\kappa_2-\kappa_1} \to 0$, $Tb_T^{[1+2/\{a(1+4/\delta))]d} / \log T \to \infty$, *and* $Tb_T^{[2\lambda/\{a(1+4/\delta)\}-1]d} \log T \to 0$ *as* $T \to \infty$, *where* $\kappa_1 = 2 + d + 2d/\delta$, $\kappa_2 = a(1 + 4/\delta)(1 + \delta/4)/d$, *and* $\kappa_3 = 2 + 2d + \delta/2 + 4d/\delta$. *Then,*

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

as $T \to \infty$, *where* $\mathbf{U}$, $\boldsymbol{\Sigma}$, *and* $B_0(\mathbf{x})$ *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 strengthened to*

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

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

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

and

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

https://doi.org/10.1017/S0266466607070028 Published online by Cambridge University Press

Proof. See Section A.3 in the Appendix.

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

COROLLARY 3.1. *Let Assumptions (A1)–(A3) and (A5) hold with* $f(\mathbf{x}) > 0, v_{2+\delta/2}(x) = O(x^{-\mu})$, *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* $b_T = O(T^{-1/(4+d)})$, *where* κ_1, κ_2, *and* κ_3 *are specified in Theorem 3.1 and* $\kappa_4 = (4 + d)(2 + 4/\delta)$. *Then the conclusion of Theorem 3.1 holds.*

Proof. See Section A.3 in the Appendix.

Remark.

(i) In Theorem 3.1 and Corollary 3.1, the positive constants μ and λ specify the decay rates for the stability and mixing coefficients tending to zero (the larger μ and λ, the faster the decay rates), which are related to the positive constant δ 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 μ and λ are sufficiently large, δ can be equal to a small number close to zero. This condition is automatically satisfied when both the stability and mixing coefficients decay at geometric rates (cf. the examples in Section 4).

(ii) When the model (1.2) with (1.3) being an integrated GARCH model is considered, then the second-order moments 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 $E|Y_t| < \infty$ holds.

The asymptotic distribution is as if the sequence (Y_j, X_j) were i.i.d. with the same marginal distributions. That these results are expected for such weakly dependent stationary processes as ours has already been shown by Masry and Fan (1997) for α-mixing processes. By contrast, for nonstationary or strongly dependent time series slower convergence rates and even nonnormal 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 $\mathrm{Var}(Y_j|\mathbf{X}_j = \mathbf{x}) = E(Y_j^2|\mathbf{X}_j = \mathbf{x}) - E^2(Y_j|\mathbf{X}_j = \mathbf{x})$ 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 earlier. Consistency of the standard errors follows under our conditions.

https://doi.org/10.1017/S0266466607070028 Published online by Cambridge University Press

The optimal bandwidth can be found using the same formula given in Masry and Fan (1997, Sect. 3.1). The practical choice of the bandwidth can be made by the conventional cross-validation (CV) rule, which is often computationally intensive, especially for large sample size, or the generalized cross-validation (GCV) rule of Wahba (1977) and Craven and Wahba (1979) for less intensive computation. In Section 4, we use the bandwidth selected by GCV in the simulations and by CV in the empirical application.

One of the important advantages of local polynomial (and linear) fitting over the 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

$$B_g := \frac{1}{2}\left(\frac{\partial^2 g}{\partial x^2}\right)_{0^+} \int_{-c}^{\infty} u^2 K(u)\, du,$$

$$\sigma_0^2 := \frac{\mathrm{Var}(Y_j | X_j = 0^+) \int_{-c}^{\infty} K^2(u)\, du}{f(0^+)}, \tag{3.4}$$

and

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

4. NUMERICAL RESULTS

4.1. Simulation

In this section we report the results of a small 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 α-mixing processes can work reasonably well for the processes for which α-mixing is not guaranteed but near epoch dependence is satisfied.

Model 1 (AR(1)-GARCH(1) model). In financial econometrics, a commonly used model is the following compound model:

$$r_t = a_0 + a r_{t-1} + \varepsilon_t, \tag{4.1}$$

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

https://doi.org/10.1017/S0266466607070028 Published online by Cambridge University Press

where r_t is the daily return of some equity on day t, modeled by an AR(1) model, (4.1), and h_t is the conditional variance of r_t, given the past information up to day $t-1$, modeled by a GARCH(1,1) model, (4.2), with $\alpha_0 > 0$, $\alpha_1 \geq 0$, $\beta_1 \geq 0$, and $\{e_t\}$ being an i.i.d. random sequence with $Ee_t = 0$ and $Ee_t^2 = 1$ (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 (cf. Carrasco and Chen, 2002), the ε_t in the GARCH(1,1) model (4.2) is strongly (α-) 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 (α-) mixing, but it can be shown that r_t is NED of order $2 + \delta$ with respect to a strongly (α-) mixing process, if $E|\varepsilon_t|^{2+\delta} < \infty$, with stable coefficients (because of the convex property of $|\cdot|^{2+\delta}$)

$$v_{2+\delta}(k) = E|r_t - r_t^{(k)}|^{2+\delta} = w_k^{2+\delta} E\left| \sum_{j=k+1}^{\infty} \frac{a^j}{w_k} \varepsilon_{t-j} \right|^{2+\delta}$$

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

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 whereas the theory based on the strong (α-) mixing in the literature would not do if r_t is non-strongly mixing. The box plots 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 Figures 1a–d, respectively, where e_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 the FT100 Index given in Section 4.2 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).

https://doi.org/10.1017/S0266466607070028 Published online by Cambridge University Press

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 Figures 1a and c for the estimates of the conditional mean, which are compatible with the actual line in Model 1, and Figures 1b and d for the estimates of the conditional variance. The sample size equal to 500 seems to work very well in all cases.

4.2. An Empirical Application

We investigate the UK FT100 Index, with sample size 602, from January 1, 1990, to April 21, 1992, for an illustration. 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 Figures 2a 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 Figures 2c and d, where the bandwidths used for the conditional mean and conditional variance are 0.34 and 0.5, respectively, chosen by the cross-validation rule. In addition the estimates of the following parametric linear model:

$$r_t = \tilde{a}_0 + \tilde{a} r_{t-1} + e_t \sqrt{\tilde{\alpha}_0 + \tilde{\alpha}_1 r_{t-1}^2} \tag{4.3}$$

with e_t the same as before, leading to $E(r_t|r_{t-1} = x) = \tilde{a}_0 + \tilde{a}x$ and $\mathrm{Var}(r_t|r_{t-1} = x) = \tilde{\alpha}_0 + \tilde{\alpha}_1 x^2$, are also plotted in Figures 2c 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 Figures 2c and d, we can observe that both the conditional mean and conditional variance functions appear to be nonlinear. Roughly, the conditional mean first increases before x around -0.4, then decreases between -0.4 and around 0.7, and increases again when $x > 0.7$. Correspondingly, the conditional variance mostly follows a similar pattern to that of the conditional mean, increasing with x when $x < -0.75$ and $x > 0.8$ and decreasing with x when $0.25 < x < 0.9$, which appears to be consistent with the "high return, high risk" rule, but the pattern is 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 1d in the simulation example. However, the pointwise standard errors reveal that the mean effect is never significantly different from the linear, whereas the variance effect is only rarely different from a quadratic function.

https://doi.org/10.1017/S0266466607070028 Published online by Cambridge University Press

https://doi.org/10.1017/S0266466607070028 Published online by Cambridge University Press

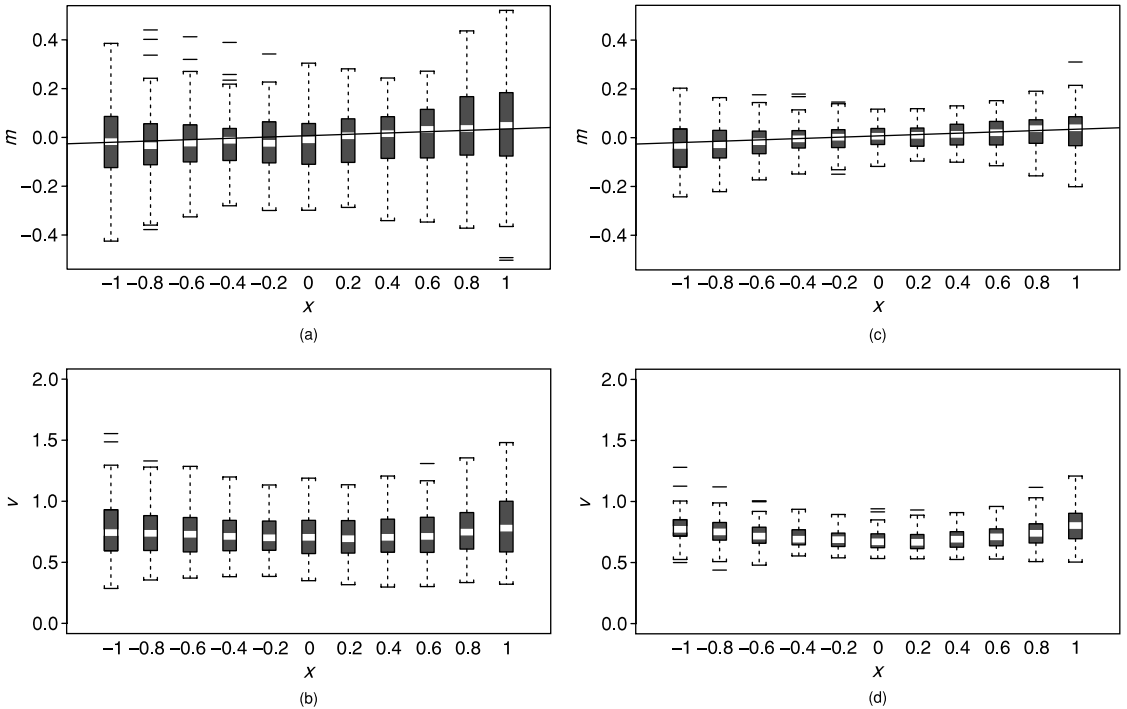


FIGURE 1. Simulation for Model 1—Box plots of the local linear fitting for the conditional mean and conditional variance of 100 replications of the linear AR(1)-GARCH(1,1) model (Model 1) with $a_0 = 0.001682$, $a = 0.020602$, $\alpha_0 = 0.137526$, $\alpha_1 = 0.094518$, and $\beta_1 = 0.726777$, and $e_t \sim N(0,1)$, 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 $m(x) = a_0 + ax$.

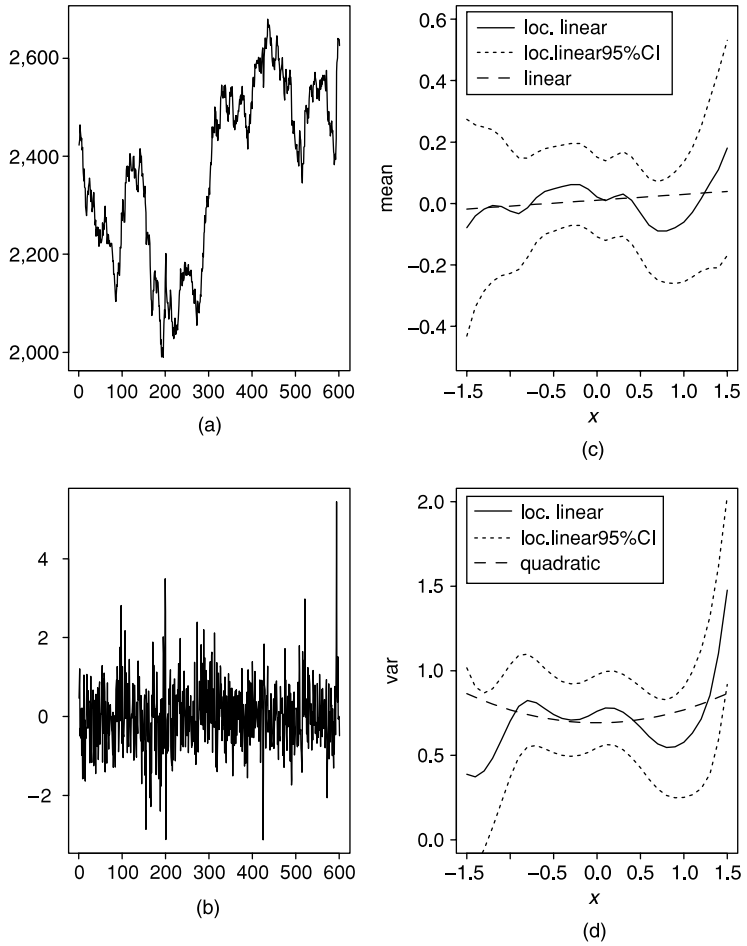


FIGURE 2. Real example—The local linear fitting for the conditional mean and conditional variance of the return series of the UK FT100 Index from January 1, 1990, to April 21, 1992, with 602 observations: (a) FT100 Index series, (b) return series of FT100 Index, (c) conditional mean, and (d) conditional 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).

NOTE

1. Let E be some normed measurable space with norm $\|\cdot\|$, for example, the Euclidean space $E = \mathbb{R}^D$ with the Euclidean norm and $D = 1 + d$ in the setting of this paper. An E-valued stochastic sequence $\{Z_t\}_{t\in\mathbb{Z}}$ is called $(\theta, \mathcal{L}, \psi)$-weak dependent (cf. Nze et al., 2002, p. 399) if for some monotone sequence $\theta = (\theta_r)_{r\in\mathbb{N}}$ decreasing to zero at infinity and some real-valued function ψ with arguments $(h, k, u, v) \in \mathcal{L}^2 \times \mathbb{N}^2$,

https://doi.org/10.1017/S0266466607070028 Published online by Cambridge University Press

$$|\mathrm{Cov}(h(Z_{i_1},\ldots,Z_{i_u}),k(Z_{j_1},\ldots,Z_{j_v}))| \le \psi(h,k,u,v)\theta_r$$

for any u-tuple $(i_1,\ldots,i_u)$, any v-tuple $(j_1,\ldots,j_v)$ with $i_1 \le \cdots \le i_u < i_u + r \le j_1 \le \cdots \le j_v$ and all $u,v \in \mathbb{N}$, where $\mathbb{N} = \{1,2,3,\ldots\}$, $\mathbb{Z} = \{0,\pm1,\pm2,\ldots\}$, and $\mathcal{L} = \bigcup_{u=1}^{\infty}\{h \in \mathbb{L}^{\infty}(E^u) : \mathrm{Lip}(h) < \infty, \|h\|_{\infty} \le 1\}$. Here $\mathbb{L}^{\infty}(E^u)$ $(u \in \mathbb{N})$ is the set of measurable bounded functions on E^u and set $\mathbb{L}^{\infty} = \bigcup_{u=1}^{\infty} \mathbb{L}^{\infty}(E^u)$, $\|h\|_{\infty} = \sup_{x \in E^u}|h(x)|$ for $h \in \mathbb{L}^{\infty}(E^u)$, and $\mathrm{Lip}(h) = \sup_{x \ne y}|h(x) - h(y)|/\|x - y\|_1$ is the Lipschitz modulus of a function $h : E^u \to \mathbb{R}$ with respect to the ℓ^1-norm in E^u, defined by $\|x\|_1 = \|(x_1,\ldots,x_u)\|_1 = \|x_1\| + \cdots + \|x_u\|$ for $x \in E^u$.

REFERENCES

Andrews, D.W.K. (1984) Non-strong mixing autoregressive processes. *Journal of Applied Probability* 21, 930–934.

Andrews, D.W.K. (1995) Nonparametric kernel estimation for semiparametric models. *Econometric Theory* 11, 560–596.

Bandi, F. (2004) On Persistence and Nonparametric Estimation (with an application to stock return predictability). Working paper, Graduate School of Business, University of Chicago.

Bierens, H. (1981) *Robust Methods and Asymptotic Theory*. Lecture Notes in Economics and Mathematical Systems 192. Springer.

Billingsley, P. (1968) *Convergence of Probability Measures*. Wiley.

Bollerslev, T. (1986) Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics* 31, 307–327.

Carrasco, M. & X. Chen (2002) Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory* 18, 17–39.

Cleveland, W.S. (1979) Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association* 74, 829–836.

Cleveland, W.S. & C. Loader (1996) Smoothing by local regression: Principles and methods. In W. Härdle & M.G. Schimek (eds.), *Statistical Theory and Computational Aspects of Smoothing*, 10–49. Physica-Verlag.

Cline, D.B.H. & H.M.H. Pu (1999) Geometric ergodicity of nonlinear time series. *Statistica Sinica* 9, 1103–1118.

Craven, P. & G. Wahba (1979) Smoothing noise data with spline functions: Estimating the correct degree of smoothing by the method of generalized cross-validation. *Numerical Mathematics* 31, 377–403.

Davidson, J. (1994) *Stochastic Limit Theory*. Oxford University Press.

Deo, C.M. (1973) A note on empirical processes of strong mixing sequences. *Annals of Probability* 1, 870–875.

Devroye, L. & L. Györfi (1985) *Nonparametric Density Estimation: The L1 View*. Wiley.

Engle, R.F. (1982) Autoregressive conditional heteroscedasticity with estimates of United Kingdom inflation. *Econometrica* 50, 987–1007.

Fan, J. (1992) Design-adaptive nonparametric regression. *Journal of the American Statistical Association* 87, 998–1004.

Fan, J. & I. Gijbels (1992) Variable bandwidth and local linear regression smoothers. *Annals of Statistics* 20, 2008–2036.

Fan, J. & I. Gijbels (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.

Fan, J. & I. Gijbels (1996) *Local Polynomial Modeling and Its Applications*. Chapman and Hall.

Fan, J. & Q. Yao (2003) *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer-Verlag.

Gallant, A.R. (1987) *Nonlinear Statistical Models*. Wiley.

Gallant, A.R. & H. White (1988) *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. Blackwell.

https://doi.org/10.1017/S0266466607070028 Published online by Cambridge University Press

Goródetskii, V.V. (1977) On the strong mixing properties for linear sequences. *Theory of Probability and Its Applications* 22, 411–413.
Hastie, T. & C. Loader (1993) Local regression: Automatic kernel carpentry. *Statistical Science* 8, 120–143.
Ibragimov, I.A. (1962) Some limit theorems for stationary processes. *Theory of Probability and Its Applications* 7, 349–382.
Ibragimov, I.A. & Y.V. Linnik (1971) *Independent and Stationary Sequences of Random Variables.* Wolters-Noordhoff.
Karlsen, H.A. & D. Tjøstheim (2001) Nonparametric estimation in null recurrent time series. *Annals of Statistics* 29, 372–416.
Ling, S. & W.K. Li (1997) On fractionally integrated autoregressive moving-average time series models with conditional heteroscedasticity. *Journal of the American Statistical Association* 92, 1184–1194.
Lu, Z. (1996a) A note on geometric ergodicity of autoregressive conditional heteroscedasticity (ARCH) model. *Statistics and Probability Letters* 30, 305–311.
Lu, Z. (1996b) Geometric ergodicity of a general ARCH type model with applications to some typical models. *Chinese Science Bulletin* 41, 1630.
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.
Lu, Z. (2001) Asymptotic normality of kernel density estimators under dependence. *Annals of the Institute of Statistical Mathematics* 53, 447–468.
Lu, Z. & Z. Jiang (2001) L_1 geometric ergodicity of a multivariate nonlinear AR model with an ARCH term. *Statistics and Probability Letters* 51, 121–130.
Masry, E. (1986) Recursive probability density estimation from weakly dependent processes. *IEEE Transactions on Information Theory* IT-32, 254–267.
Masry, E. & J. Fan (1997) Local polynomial estimation of regression functions for mixing processes. *Scandinavian Journal of Statistics* 24, 165–179.
Masry, E. & D. Tjøstheim (1995) Nonparametric estimation and identification of nonlinear ARCH time series. *Econometric Theory* 11, 258–289.
McLeish, D.L. (1975a) A maximal inequality and dependent strong laws. *Annals of Probability* 3, 826–836.
McLeish, D.L. (1975b) Invariance principles for dependent variables. *Zeitschrift fur Wahrscheinlichskeittheorie und verwandte Gebiete* 32, 165–178.
McLeish, D.L. (1977) On the invariance principle for nonstationary mixingales. *Annals of Probability* 5, 616–621.
Moloche, G. (2000) Local Nonparametric Estimation of Scalar Diffusions. Working paper, Department of Economics, MIT.
Nakhapetyan, B.S. (1987) An approach to proving limit theorems for dependent random variables. *Theory of Probability and Its Applications* 32, 535–539.
Nze, P.A., P. Bühlmann, & P. Doukhan (2002) Weak dependence beyond mixing and asymptotics for nonparametric regression. *Annals of Statistics* 30, 397–430.
Nze, P.A. & P. Doukhan (2004) Weak dependence: Models and applications to econometrics. *Econometric Theory* 20, 995–1045.
Pham, T.D. (1986) The mixing properties of bilinear and generalized random coefficient autoregressive models. *Stochastic Processes and Their Applications* 23, 291–300.
Pham, D.T. & L.T. Tran (1985) Some strong mixing properties of time series models. *Stochastic Processes and Their Applications* 19, 297–303.
Phillips, P.C.B. & J.Y. Park (1998) Nonstationary Density Estimation and Kernel Autoregression. Cowles Foundation Discussion paper 1181.
Robinson, P.M. (1983) Nonparametric estimators for time series. *Journal of Time Series Analysis* 4, 185–207.
Ruppert, D. & M.P. Wand (1994) Multivariate locally weighted least squares regression. *Annals of Statistics* 22, 1346–1370.

https://doi.org/10.1017/S0266466607070028 Published online by Cambridge University Press

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/.
Stone, C.J. (1977) Consistent nonparametric regression. *Annals of Statistics* 5, 595–620.
Stone, C.J. (1980) Optimal rates of convergence for nonparametric estimators. *Annals of Statistics* 8, 1348–1360.
Tjøstheim, D. (1990) Nonlinear time series and Markov chains. *Advances in Applied Probability* 22, 587–611.
Tong, H. (1990) *Nonlinear Time Series: A Dynamical System Approach*. Oxford University Press.
Wahba, G. (1977) A survey of some smoothing problems and the method of generalized cross-validation for solving them. In P.R. Krisnaiah (ed.), *Applications of Statistics*, pp. 507–523. North-Holland.
Weiss, A.A. (1984) ARMA model with ARCH errors. *Journal of Time Series Analysis* 5, 129–143.

APPENDIX: Proofs

A.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 for a proof.

LEMMA A.1.

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

$$|E[XY] - E[X]E[Y]| \le C\|X\|_r \|Y\|_s [\alpha(\mathcal{F}_1, \mathcal{F}_2)]^{1/h}, \tag{A.1}$$

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 (A.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 α-mixing samples $(\mathbf{X}_i^{(m)}, Y_i^{(m)})$, which will play important roles throughout the proofs that follow.

LEMMA A.2. *Let $m = m_T$ be a positive integer tending to ∞ as $T \to \infty$. Then under Assumptions (A2) and (A5), it holds that for any positive real number $L = L_T$ tending to ∞, as $T \to \infty$,*

$$EY_i^{(m)} K((\mathbf{x} - \mathbf{X}_i^{(m)})/b_T) = EY_i K((\mathbf{x} - \mathbf{X}_i)/b_T) + O(b_T^{-1} L v_2^{1/2}(m)) + o(L^{-(1+\delta)}), \tag{A.2}$$

$$EY_i^{(m)2} K^2((\mathbf{x} - \mathbf{X}_i^{(m)})/b_T) = EY_i^2 K^2((\mathbf{x} - \mathbf{X}_i)/b_T) + O(b_T^{-2} L^2 v_2(m)) + O(b_T^{-1} L^2 v_2^{1/2}(m)) + o(L^{-\delta}), \tag{A.3}$$

https://doi.org/10.1017/S0266466607070028 Published online by Cambridge University Press

for $i \neq j$,

$$EY_i^{(m)} Y_j^{(m)} K((\mathbf{x} - \mathbf{X}_i^{(m)})/b_T) K((\mathbf{x} - X_j^{(m)})/b_T)$$
$$= EY_i Y_j K((\mathbf{x} - \mathbf{X}_i)/b_T) K((\mathbf{x} - \mathbf{X}_j)/b_T) + O(b_T^{-1} L^2 v_2^{1/2}(m)) + o(L^{-\delta}), \quad \textbf{(A.4)}$$

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

$$E|Y_i^{(m)}|^{2+\delta'} K^{2+\delta'}((\mathbf{x} - \mathbf{X}_i^{(m)})/b_T)$$
$$\leq CE|Y_i|^{2+\delta'} K^{2+\delta'}((\mathbf{x} - \mathbf{X}_i)/b_T) + O(b_T^{-2-\delta'} L^{2+\delta'} v_{2+\delta'}(m)) + o(L^{-\delta+\delta'}), \quad \textbf{(A.5)}$$

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

Proof. We prove (A.2)–(A.5) one by one. To prove (A.2), note first that

$$EY_i^{(m)} K((\mathbf{x} - \mathbf{X}_i^{(m)})/b_T) = EY_i K((\mathbf{x} - \mathbf{X}_i^{(m)})/b_T) + E(Y_i^{(m)} - Y_i) K((\mathbf{x} - \mathbf{X}_i^{(m)})/b_T)$$
$$=: EY_i K((\mathbf{x} - \mathbf{X}_i^{(m)})/b_T) + \delta_{1T}. \quad \textbf{(A.6)}$$

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

$$|\delta_{1T}| \leq E|Y_i^{(m)} - Y_i| K((\mathbf{x} - \mathbf{X}_i^{(m)})/b_T) \leq CE|Y_i^{(m)} - Y_i|$$
$$\leq C[E|Y_i^{(m)} - Y_i|^2]^{1/2} = O(\sqrt{v_2(m)}). \quad \textbf{(A.7)}$$

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

$$EY_i K((\mathbf{x} - \mathbf{X}_i^{(m)})/b_T) = EY_{i,L} K((\mathbf{x} - \mathbf{X}_i^{(m)})/b_T) + EY_{i,U} K((\mathbf{x} - \mathbf{X}_i^{(m)})/b_T)$$
$$= EY_{i,L} K((\mathbf{x} - \mathbf{X}_i)/b_T) + EY_{i,L} K((\mathbf{x} - \mathbf{X}_i^{(m)})/b_T)$$
$$- K((\mathbf{x} - \mathbf{X}_i)/b_T) + EY_{i,U} K((\mathbf{x} - \mathbf{X}_i^{(m)})/b_T)$$
$$=: \delta_{2T} + \delta_{3T} + \delta_{4T}, \quad \textbf{(A.8)}$$

where, as a result of $|EY_{i,U} K((\mathbf{x} - \mathbf{X}_i)/b_T)| = o(L^{-(1+\delta)})$ as $T \to \infty$ (in an argument similar to (A.11), which follows),

$$\delta_{2T} = EY_i K((\mathbf{x} - \mathbf{X}_i)/b_T) - EY_{i,U} K((\mathbf{x} - \mathbf{X}_i)/b_T)$$
$$= EY_i K((\mathbf{x} - \mathbf{X}_i)/b_T) + o(L^{-(1+\delta)}); \quad \textbf{(A.9)}$$

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

$$|\delta_{3T}| \leq E|Y_{i,L}| |K((\mathbf{x} - \mathbf{X}_i^{(m)})/b_T) - K((\mathbf{x} - \mathbf{X}_i)/b_T)|$$
$$\leq CLE \left\| \frac{\mathbf{X}_i^{(m)} - \mathbf{X}_i}{b_T} \right\| \leq Cb_T^{-1} L[E\|\mathbf{X}_i^{(m)} - \mathbf{X}_i\|^2]^{1/2} = O(b_T^{-1} L\sqrt{v_2(m)}); \quad \textbf{(A.10)}$$

https://doi.org/10.1017/S0266466607070028 Published online by Cambridge University Press

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

$$|\delta_{4T}| \le E|Y_i|I_{\{|Y_i|>L\}}K((\mathbf{x}-\mathbf{X}_i^{(m)})/b_T) \le L^{-(1+\delta)}E|Y_i|^{2+\delta}I_{\{|Y_i|>L\}}K((\mathbf{x}-\mathbf{X}_i^{(m)})/b_T)$$
$$\le CL^{-(1+\delta)}E|Y_i|^{2+\delta}I_{\{|Y_i|>L\}} = o(L^{-(1+\delta)}). \quad \textbf{(A.11)}$$

Then it follows from (A.8) together with (A.9)–(A.11) that

$$EY_iK((\mathbf{x}-\mathbf{X}_i^{(m)})/b_T) = EY_iK((\mathbf{x}-\mathbf{X}_i)/b_T) + O(b_T^{-1}L\sqrt{v_2(m)}) + o(L^{-(1+\delta)}). \quad \textbf{(A.12)}$$

Finally, (A.2) follows from (A.6) together with (A.7) and (A.12).

The proof of (A.3) can be proved in an argument similar to that in the preceding discussion. First,

$$E[Y_i^{(m)}]^2K^2((\mathbf{x}-\mathbf{X}_i^{(m)})/b_T) = EY_i^2K^2((\mathbf{x}-\mathbf{X}_i^{(m)})/b_T)$$
$$+ 2EY_i(Y_i^{(m)}-Y_i)K^2((\mathbf{x}-\mathbf{X}_i^{(m)})/b_T)$$
$$+ E(Y_i^{(m)}-Y_i)^2K^2((\mathbf{x}-\mathbf{X}_i^{(m)})/b_T)$$
$$=: \delta_{5T}+\delta_{6T}+\delta_{7T}. \quad \textbf{(A.13)}$$

Here, similarly to (A.7) by the boundedness of $K(\cdot)$,

$$|\delta_{7T}| \le E|Y_i^{(m)}-Y_i|^2K^2((\mathbf{x}-\mathbf{X}_i^{(m)})/b_T) \le CE|Y_i^{(m)}-Y_i|^2 = O(v_2(m)); \quad \textbf{(A.14)}$$

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

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

Next,

$$\delta_{5T} = EY_{i,L}^2K^2((\mathbf{x}-\mathbf{X}_i^{(m)})/b_T) + EY_{i,U}^2K^2((\mathbf{x}-\mathbf{X}_i^{(m)})/b_T)$$
$$= EY_{i,L}^2K^2((\mathbf{x}-\mathbf{X}_i)/b_T) + 2EY_{i,L}^2K((\mathbf{x}-\mathbf{X}_i)/b_T)$$
$$\times (K((\mathbf{x}-\mathbf{X}_i^{(m)})/b_T) - K((\mathbf{x}-\mathbf{X}_i)/b_T))$$
$$+ EY_{i,L}^2(K((\mathbf{x}-\mathbf{X}_i^{(m)})/b_T) - K((\mathbf{x}-\mathbf{X}_i)/b_T))^2 + EY_{i,U}^2K^2((\mathbf{x}-\mathbf{X}_i^{(m)})/B_T)$$
$$=: \delta_{8T}+\delta_{9T}+\delta_{10T}+\delta_{11T}, \quad \textbf{(A.15)}$$

where, similarly to (A.11),

$$|\delta_{11T}| \le E|Y_i|^2I_{\{|Y_i|>L\}}K^2((\mathbf{x}-\mathbf{X}_i^{(m)})/B_T) \le CL^{-\delta}E|Y_i|^{2+\delta}I_{\{|Y_i|>L\}} = o(L^{-\delta}); \quad \textbf{(A.16)}$$

as a result of $|EY_{i,U}^2K^2((\mathbf{x}-\mathbf{X}_i)/b_T)| = o(L^{-\delta})$ (the same argument as (A.16)),

$$\delta_{8T} = EY_i^2K^2((\mathbf{x}-\mathbf{X}_i)/b_T) - EY_{i,U}^2K^2((\mathbf{x}-\mathbf{X}_i)/b_T)$$
$$= EY_i^2K^2((\mathbf{x}-\mathbf{X}_i)/b_T) + o(L^{-\delta}); \quad \textbf{(A.17)}$$

https://doi.org/10.1017/S0266466607070028 Published online by Cambridge University Press

the same argument as (A.10) leads to

$$|\delta_{10T}| \le E|Y_{i,L}|^2 |K((\mathbf{x}-\mathbf{X}_i^{(m)})/b_T) - K((\mathbf{x}-\mathbf{X}_i)/b_T)|^2$$
$$\le CL^2E\left\| \frac{\mathbf{X}_i^{(m)}-\mathbf{X}_i}{b_T} \right\|^2 \le Cb_T^{-2}L^2E\|\mathbf{X}_i^{(m)}-\mathbf{X}_i\|^2 = O(b_T^{-2}L^2v_2(m)) \quad \textbf{(A.18)}$$

and

$$|\delta_{9T}| \le CL^2E\left\| \frac{\mathbf{X}_i^{(m)}-\mathbf{X}_i}{b_T} \right\| \le Cb_T^{-1}L^2[E\|\mathbf{X}_i^{(m)}-\mathbf{X}_i\|^2]^{1/2} = O(b_T^{-1}L^2\sqrt{v_2(m)}). \quad \textbf{(A.19)}$$

Finally, (A.3) follows from (A.13)–(A.19).

For (A.4), first note that

$$E[Y_i^{(m)}Y_j^{(m)}K((\mathbf{x}-\mathbf{X}_i^{(m)})/b_T)K((\mathbf{x}-X_j^{(m)})/b_T)]$$
$$= EY_iY_jK((\mathbf{x}-\mathbf{X}_i^{(m)})/b_T)K((\mathbf{x}-X_j^{(m)})/b_T)$$
$$+ E[Y_i(Y_j^{(m)}-Y_j)+(Y_i^{(m)}-Y_i)Y_j]K((\mathbf{x}-\mathbf{X}_i^{(m)})/b_T)K((\mathbf{x}-X_j^{(m)})/b_T)$$
$$+ E(Y_i^{(m)}-Y_i)(Y_j^{(m)}-Y_j)K((\mathbf{x}-\mathbf{X}_i^{(m)})_T)K((\mathbf{x}-X_j^{(m)})/b_T)$$
$$=: \delta_{12T}+\delta_{13T}+\delta_{14T}. \quad \textbf{(A.20)}$$

Here, similarly to (A.7), by the boundedness of $K(\cdot)$, and then by Cauchy inequality,

$$|\delta_{14T}| \le CE|Y_i^{(m)}-Y_i||Y_j^{(m)}-Y_j| = O(v_2(m)); \quad \textbf{(A.21)}$$

again,

$$|\delta_{13T}| \le \{[EY_i^2]^{1/2}[E|Y_j^{(m)}-Y_j|^2]^{1/2} + [E|Y_i^{(m)}-Y_i|^2]^{1/2}[EY_j^2]^{1/2}\}$$
$$= O(\sqrt{v_2(m)}). \quad \textbf{(A.22)}$$

Next,

$$\delta_{12T} = EY_iY_jK((\mathbf{x}-\mathbf{X}_i)/b_T)K((\mathbf{x}-X_j)/b_T)$$
$$+ EY_iY_j[K((\mathbf{x}-\mathbf{X}_i^{(m)})/b_T)-K((\mathbf{x}-\mathbf{X}_i)/b_T)]K((\mathbf{x}-X_j)/b_T)$$
$$+ EY_iY_jK((\mathbf{x}-\mathbf{X}_i^{(m)})/b_T)[K((\mathbf{x}-X_j^{(m)})/b_T)-K((\mathbf{x}-X_j)/b_T)]$$
$$=: EY_iY_jK((\mathbf{x}-\mathbf{X}_i)/b_T)K((\mathbf{x}-X_j)/b_T)+\delta_{15T}+\delta_{16T}, \quad \textbf{(A.23)}$$

https://doi.org/10.1017/S0266466607070028 Published online by Cambridge University Press

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

$$\begin{aligned}\delta_{16T} &= EY_i(I_{\{|Y_i|\le L\}} + I_{\{|Y_i|>L\}})Y_j(I_{\{|Y_j|\le L\}} + I_{\{|Y_j|>L\}})K((\mathbf{x}-\mathbf{X}_i^{(m)})/b_T) \\ &\quad \times [K((\mathbf{x}-X_j^{(m)})/b_T) - K((\mathbf{x}-X_j)/b_T)] \\ &\le CL^2 \left\| \frac{X_j^{(m)} - X_j}{b_T} \right\| + 2CLL^{-(1+\delta)}E|Y_i|^{2+\delta}I_{\{|Y_i|>L\}} + L^{-\delta}E|Y_i|^{2+\delta}I_{\{|Y_i|>L\}} \\ &= O(b_T^{-1}L^2 v_2^{1/2}(m)) + o(L^{-\delta}), \end{aligned} \tag{A.24}$$

and similarly,

$$\delta_{15T} = O(b_T^{-1}L^2 v_2^{1/2}(m)) + o(L^{-\delta}). \tag{A.25}$$

Finally, (A.4) follows from (A.20)–(A.25).

For the proof of (A.5), by noting that

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

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

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

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

$$\begin{aligned} E|YK_2|^{2+\delta'} &\le CE|Y_tI_{\{|Y_t|\le L\}}\{K((x-\mathbf{X}_t^{(m)})/b_T) - K((x-\mathbf{X}_t)/b_T)\}|^{2+\delta'} \\ &\quad + CE|Y_tI_{\{|Y_t|\ge L\}}|^{2+\delta'} \\ &\le C\{Lb_T^{-1}\}^{2+\delta'}\|\mathbf{X}_t - \mathbf{X}_t^{(m)}\|^{2+\delta'}(m) + CL^{\delta'-\delta}E|Y_t|^{2+\delta}I_{\{|Y_t|\ge L\}} \\ &\le C\{Lb_T^{-1}\}^{2+\delta'}v_{2+\delta'}(m) + o(L^{\delta'-\delta}), \end{aligned} \tag{A.26}$$

and hence (A.5) holds. ■

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

LEMMA A.3. *(Cross term lemma). Let* $\{(Z_j^{(m)}, \mathbf{X}_j^{(m)});\ 1 \le j \le q\}$ *be a stationary sequence with mixing coefficient*

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

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* $\Delta_j^{(m)}(x) = \eta_j^{(m)}(x) - E\eta_j^{(m)}(x)$, $\tilde{V}_{12}(\mathbf{x}) = \sum_{1\le i<j\le q}\Delta_i^{(m)}(x)\Delta_j^{(m)}(x)$. *Then, for any positive integer* c_T,

$$|\tilde{V}_{12}(\mathbf{x})| \le Cqb_T^d[\tilde{J}_1(\mathbf{x}) + \tilde{J}_2(\mathbf{x})], \tag{A.27}$$

https://doi.org/10.1017/S0266466607070028 Published online by Cambridge University Press

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. First, we note that

$$E\Delta_j^{(m)}(\mathbf{x})\Delta_i^{(m)}(\mathbf{x}) = \{E(\eta_j^{(m)}(\mathbf{x})\eta_i^{(m)}(\mathbf{x})) - E(\eta_j^{(m)}(\mathbf{x}))E(\eta_i^{(m)}(\mathbf{x}))\}. \tag{A.28}$$

Then applying Lemma A.2,

$$\begin{aligned} E(\eta_j^{(m)}(\mathbf{x})\eta_i^{(m)}(\mathbf{x})) &= E(Z_i^{(m)} K((\mathbf{x} - \mathbf{X}_i^{(m)})/b_T))(Z_j^{(m)} K((\mathbf{x} - \mathbf{X}_j^{(m)})/b_T)) \\ &= E(Z_i K((\mathbf{x} - \mathbf{X}_i)/b_T))(Z_j K((\mathbf{x} - \mathbf{X}_j)/b_T)) \\ &\quad + O(b_T^{-1} L^2 \sqrt{v_2(m)}) + o(L^{-\delta}), \end{aligned} \tag{A.29}$$

$$\begin{aligned} E\eta_j^{(m)}(\mathbf{x}) &= E(Z_j^{(m)} K((\mathbf{x} - \mathbf{X}_j^{(m)})/b_T)) \\ &= E(Z_j K((\mathbf{x} - \mathbf{X}_j)/b_T)) + O(b_T^{-1} L \sqrt{v_2(m)}) + o(L^{-(1+\delta)}). \end{aligned} \tag{A.30}$$

Therefore it follows from (A.28)–(A.30) that (taking $L = b_T^{-2d/\delta}$)

$$\begin{aligned} &E[\Delta_j^{(m)}(\mathbf{x})\Delta_i^{(m)}(\mathbf{x})] \\ &\quad = [EZ_i K((\mathbf{x} - \mathbf{X}_i)/b_T)(Z_j K((\mathbf{x} - \mathbf{X}_j)/b_T) \\ &\qquad\qquad - EZ_i K((\mathbf{x} - \mathbf{X}_i)/b_T))EZ_j K((\mathbf{x} - \mathbf{X}_j)/b_T)] \\ &\qquad + [O(b_T^{-2} L^2 v_2(m)) + O(b_T^{-1} L^2 \sqrt{v_2(m)}) + o(L^{-\delta}) \\ &\qquad\quad + b_T^d \{O(b_T^{-1} L \sqrt{v_2(m)}) + o(L^{-(1+\delta)})\}] \\ &\quad = b_T^{2d} + O(b_T^{-2} L^2 v_2(m)) + O(b_T^{-1} L^2 \sqrt{v_2(m)}) + o(L^{-\delta}) \\ &\qquad + b_T^d \{O(b_T^{-1} L \sqrt{v_2(m)}) + o(L^{-(1+\delta)})\} \\ &\quad \le C b_T^{2d} [1 + O(b_T^{-2-2d} L^2 \sqrt{v_2(m)} + 1)]; \end{aligned} \tag{A.31}$$

therefore,

$$\begin{aligned} \sum_{i=1}^{q} \sum_{j-i=1}^{c_T} E[\Delta_j^{(m)}(\mathbf{x})\Delta_i^{(m)}(\mathbf{x})] &\le C b_T^{2d} (q c_T) \max\{1, b_T^{-2(1+d+2d/\delta)} v_2^{1/2}(m)\} \\ &= q b_T^d \tilde{J}_1(\mathbf{x}). \end{aligned} \tag{A.32}$$

https://doi.org/10.1017/S0266466607070028 Published online by Cambridge University Press

On the other hand, for $i < j$, by Lemma A.1(i) with $r = s = 1/(2 + \delta')$ and (A.5) of Lemma 5.2 and then taking $L = b_T^{-2d/\delta}$ and $\delta' = \delta/2$,

$$
\begin{aligned}
&E[\Delta_j^{(m)}(\mathbf{x})\Delta_i^{(m)}(\mathbf{x})] \\
&\quad \le C\|\eta_i^{(m)}(\mathbf{x})\|_{2+\delta'}\|\eta_j^{(m)}(\mathbf{x})\|_{2+\delta'}\alpha_m^{1-2/(2+\delta')}(j-i) \\
&\quad \le C\{b_T^d + (b_T^{-1}L)^{2+\delta'}v_{2+\delta'}(m) + L^{-\delta+\delta'}\}^{2/(2+\delta')}\alpha_m^{\delta'/(2+\delta')}(j-i) \\
&\quad \le Cb_T^{2d/(2+\delta')}\{1 + b_T^{-d}(b_T^{-1}L)^{2+\delta'}v_{2+\delta'}(m) + L^{-\delta+\delta'}b_T^{-d}\}^{2/(2+\delta')}\alpha_m^{\delta'/(2+\delta')}(j-i) \\
&\quad \le Cb_T^{4d/(4+\delta)}\{1 + b_T^{-2d-2-\delta/2-4d/\delta}v_{2+\delta/2}(m) + 1\}^{4/(4+\delta)}\alpha_m^{\delta/(4+\delta)}(j-i); \qquad \textbf{(A.33)}
\end{aligned}
$$

therefore,

$$
\begin{aligned}
&\sum_{i=1}^{q}\sum_{j-i=c_T}^{q} E[\Delta_j^{(m)}(\mathbf{x})\Delta_i^{(m)}(\mathbf{x})] \\
&\quad \le Cb_T^{4d/(4+\delta)}q\left(\sum_{j=c_T}^{q}\{\alpha_m(j)\}^{\delta/(4+\delta)}\right)\max\{1, b_T^{-2(4+4d+\delta+8d/\delta)/(4+\delta)}v_{2+\delta/2}^{4/(4+\delta)}(m)\} \\
&\quad = Cqb_T^d\tilde{J}_2(\mathbf{x}). \qquad \textbf{(A.34)}
\end{aligned}
$$

The result of this lemma therefore follows from (A.32) and (A.34). ■

A.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

$$
\begin{aligned}
&E|(\mathbf{W}_T^{(m)})_i - (\mathbf{W}_T)_i| \\
&\quad = E\left|(Tb_T^d)^{-1}\sum_{j=1}^{T}\left[Z_j^{(m)}K_i\left(\frac{\mathbf{X}_j^{(m)} - \mathbf{x}}{b_T}\right) - Z_jK_i\left(\frac{\mathbf{X}_j - \mathbf{x}}{b_T}\right)\right]\right| \\
&\quad \le (Tb_T^d)^{-1}\sum_{j=1}^{T}E\left|Z_j^{(m)}K_i\left(\frac{\mathbf{X}_j^{(m)} - \mathbf{x}}{b_T}\right) - Z_jK_i\left(\frac{\mathbf{X}_j - \mathbf{x}}{b_T}\right)\right| \\
&\quad = b_T^{-d}E|Z_j^{(m)}K_i((\mathbf{X}_j^{(m)} - \mathbf{x})/b_T) - Z_jK_i((\mathbf{X}_j - \mathbf{x})/b_T)| \\
&\quad \le b_T^{-d}E|Z_j^{(m)} - Z_j|K_i((\mathbf{X}_j^{(m)} - \mathbf{x})/b_T) \\
&\qquad + b_T^{-d}E|Z_j|(I_{\{|Z_j|\le L\}} + I_{\{|Z_j|>L\}})|K_i((\mathbf{X}_j^{(m)} - \mathbf{x})/b_T) - K_i((\mathbf{X}_j - \mathbf{x})/b_T)| \\
&\quad = O(b_T^{-d}\sqrt{v_2(m)}) + O(b_T^{-d-1}L\sqrt{v_2(m)}) + o(b_T^{-d}L^{-(1+\delta)}) \\
&\quad = O(b_T^{-d-1}L\sqrt{v_2(m)}) + o(b_T^{-d}L^{-(1+\delta)}), \qquad \textbf{(A.35)}
\end{aligned}
$$

which is the desired result. ■

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

https://doi.org/10.1017/S0266466607070028 Published online by Cambridge University Press

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

$$\begin{aligned}\mathrm{Var}[A_T] &= (Tb_T^d)^{-1} \left\{ \sum_{j=1}^T E\Delta_j^2 + 2 \sum_{1\le i<j\le T} E\Delta_i \Delta_j \right\} \\ &= b_T^{-d} E\Delta_j^2 + 2(Tb_T^d)^{-1} \sum_{1\le i<j\le T} E\Delta_i \Delta_j := A_{T1} + 2A_{T2}. \end{aligned} \tag{A.36}$$

It easily follows from the Lebesgue density theorem (see Devroye and Györfi, 1985, Ch. 5) that the first term of (A.36) is convergent to the right-hand side of (3.1). Therefore, to complete the proof of this lemma, it suffices to show that $A_{T2} \to 0$ as $T \to \infty$. By noting that $E\Delta_i \Delta_j = E\Delta_i^{(m)} \Delta_j^{(m)} + E\Delta_i^{(m)}(\Delta_j - \Delta_j^{(m)}) + E(\Delta_i - \Delta_i^{(m)})\Delta_j$, we can further separate A_{T2} into three parts: $A_{T2} = A_{T21} + A_{T22} + A_{T23}$, where, by Lemma A.3 (taking $q = T$ and $c_T^a = b_T^{-\delta d/(4+\delta)}$) together with Assumption (B2),

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

by Assumption (A4);

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

and as a result of $E(\Delta_i^{(m)})^2 \le E(\eta_j^{(m)})^2 = O(b_T^d) + O(b_T^{-2-2d/\delta} v_2(m)) + O(b_T^{-1-2d/\delta} v_2^{1/2}(m)) + o(b_T^d) = O(b_T^d)$, following from (A.3) of Lemma A.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 $\tilde{L}^\delta = T^2/b_T^d$,

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

https://doi.org/10.1017/S0266466607070028 Published online by Cambridge University Press

we have

$$A_{T22} \le CTb_T^{-d/2}[\tilde{L}b_T^{-1}v_2^{1/2}(m) + o(\tilde{L}^{-\delta/2})]$$

$$\le C[T^{1+2/\delta}b_T^{-1-d/2-d/\delta}v_2^{1/2}(m) + o(1)] \to 0;$$

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

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

The proof of the lemma is completed. ■

A.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 $m = m_T \to \infty$ (to be specified later),

$$\mathbf{W}_T(\mathbf{x}) = \mathbf{W}_T^{(m)}(\mathbf{x}) + [\mathbf{W}_T(\mathbf{x}) - \mathbf{W}_T^{(m)}(\mathbf{x})], \tag{A.37}$$

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

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

$$= O_P((Tb_T^d)^{1/2}b_T^{-d-1}L\sqrt{v_2(m)}) + o_P((Tb_T^d)^{1/2}b_T^{-d}L^{-(1+\delta)})$$

$$= O_P(\{T^{1+1/(1+\delta)}b_T^{-(1+1/(1+\delta))\,d-2}v_2(m)\}^{1/2}) + o_P(1)$$

$$\le O_P(\{T^{2+4/\delta}b_T^{-(2+d+2d/\delta)}v_2(m)\}^{1/2}) + o_P(1) \to_P 0, \tag{A.38}$$

following from Assumption (B2); and similarly

$$(Tb_T^d)^{1/2}E[\mathbf{W}_T(\mathbf{x}) - \mathbf{W}_T^{(m)}(\mathbf{x})] \to 0. \tag{A.39}$$

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

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

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

Recalling

$$\eta_j^{(m)}(\mathbf{x}) := Z_j^{(m)}K_{\mathbf{c}}(\mathbf{x} - \mathbf{X}_j^{(m)}) \quad \text{and} \quad \Delta_j^{(m)}(\mathbf{x}) := \eta_j^{(m)}(\mathbf{x}) - \mathrm{E}\eta_j^{(m)}(\mathbf{x}), \tag{A.40}$$

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

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

https://doi.org/10.1017/S0266466607070028 Published online by Cambridge University Press

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 $r \to \infty$. For each integer $1 \le i \le 2$, 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 large or small blocks, can be considered. This term will not change the proof much. The general approach consists of showing that, as $T \to \infty$,

$$Q_1^{(m)} := \left| E\exp[iu\Upsilon^{(m)}(T,\mathbf{x},1)] - \prod_{j=0}^{r-1} E\exp[iuU^{(m)}(1,T,\mathbf{x},j)] \right| \to 0, \tag{A.41}$$

$$Q_2^{(m)} := T^{-1}E(\Upsilon^{(m)}(T,\mathbf{x},2))^2 \to 0, \tag{A.42}$$

$$Q_3^{(m)} := T^{-1}\sum_{j=0}^{r-1} E[U^{(m)}(1,T,\mathbf{x},j)]^2 \to \sigma^2, \tag{A.43}$$

$$Q_4^{(m)} := T^{-1}\sum_{j=0}^{r-1} E[(U^{(m)}(1,T,\mathbf{x},j))^2 I\{|U^{(m)}(1,T,\mathbf{x},j)| > \varepsilon\sigma T^{1/2}\}] \to 0 \tag{A.44}$$

for every $\varepsilon > 0$. Note that

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

The term $\Upsilon^{(m)}(T,\mathbf{x},2)/(\sigma T^{1/2})$ is asymptotically negligible by (A.42). The random variables $U^{(m)}(1,T,\mathbf{x},j)$ are asymptotically mutually independent by (A.41). The asymptotic normality of $\Upsilon^{(m)}(T,\mathbf{x},1)/(\sigma T^{1/2})$ follows from (A.43) and the Lindeberg–Feller condition (A.44). The lemma thus follows if we can prove (A.41)–(A.44). This proof is

https://doi.org/10.1017/S0266466607070028 Published online by Cambridge University Press

given subsequently. The arguments are reminiscent of those used by Robinson (1983), Masry (1986), and Nakhapetyan (1987), but, differently from those references, because of $m = m_T$ depending on T in the α-mixing process $\{(Y_t^{(m)}, \mathbf{X}_t^{(m)})\}$, the details become much more complex and involved (cf. Lu, 2001) and heavily depend on Lemmas 2.1 and A.3 established previously.

Proof of (A.41). Let

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

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

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

Because $|E[a_j]| \le 1$,

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

Note that $d(I_\ell, I_j) \ge q$ for any $\ell \neq j$, and set $q = 2m$. It follows by applying Lemma A.1(ii) to each term on the right-hand side that

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

which tends to zero by condition (B3). ■

Proof of (A.42). For notational simplicity, refer to the random variables $U^{(m)}(2,T,\mathbf{x},j)$, $j = 0,1,\dots,r-1$, as $\hat{U}_1, \dots, \hat{U}_r$. We have

$$E[\Upsilon^{(m)}(T,\mathbf{x},2)]^2 = \sum_{i=1}^{r} \mathrm{Var}(\hat{U}_i) + 2 \sum_{1 \le i < j \le r} \mathrm{Cov}(\hat{U}_i, \hat{U}_j) := \hat{V}_1 + \hat{V}_2, \quad \text{say.} \tag{A.45}$$

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

$$\mathrm{Var}(\hat{U}_i) = \sum_{i=1}^{q} E[(\zeta_{Ti}^{(m)}(\mathbf{x}))^2] + \sum_{1 \le i < j \le q} E[\zeta_{Tj}^{(m)}(\mathbf{x})\zeta_{Ti}^{(m)}(\mathbf{x})] := \hat{V}_{11} + \hat{V}_{12}.$$

https://doi.org/10.1017/S0266466607070028 Published online by Cambridge University Press

From (A.3) and the Lebesgue density theorem (see Devroye and Györfi, 1985, Ch. 2), and then taking $L = b_T^{-d/\delta}$,

$$\begin{aligned}
\hat{V}_{11} &= q \operatorname{Var}\{\zeta_{Ti}^{(m)}(\mathbf{x})\} = q\{b_T^{-d} E(\Delta_i^{(m)}(\mathbf{x}))^2\} \\
&\le q\{b_T^{-d} E(\eta_i^{(m)}(\mathbf{x}))^2\} = q\{b_T^{-d} E(Z_i^{(m)} K((\mathbf{x} - \mathbf{X}_i^{(m)})/b_T))^2\} \\
&= q\{b_T^{-d} E(Z_i K((\mathbf{x} - \mathbf{X}_i)/b_T))^2 + O(b_T^{-2-d} L^2 v_2(m)) \\
&\qquad + O(b_T^{-1-d} L^2 \sqrt{v_2(m)}) + o(b_T^{-d} L^{-\delta})\} \\
&= Cq\{1 + b_T^{-2-d-2d/\delta} v_2(m) + (b_T^{-2-2d-4d/\delta} v_2(m))^{1/2} + o(1)\} \\
&= O(q), \qquad \textbf{(A.46)}
\end{aligned}$$

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

We need the cross term lemma, Lemma A.3, for $\hat{V}_{12}$. Thus, applying Lemma A.3 together with (B2) and then taking $c_T = q$ yields

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

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

$$T^{-1}\hat{V}_1 = T^{-1} r(\hat{V}_{11} + \hat{V}_{12}) \le T^{-1} rCq[1 + \pi_T] \le C(q/p)[1 + \pi_T]. \qquad \textbf{(A.47)}$$

Set

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

Then $U^{(m)}(2, T, \mathbf{x}, j) = \sum_{i \in \mathcal{I}(2,T,\mathbf{x},j)} \zeta_{Ti}^{(m)}$. Because $p > q$, 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 (A.45) and (A.33) and then Assumption (B2), we obtain

$$\begin{aligned}
|\hat{V}_2| &\le C \sum_{\{i,j : |i-j| \ge q,\ 1 \le i,j \le T\}} \sum |E[\zeta_{Ti}^{(m)}(\mathbf{x}) \zeta_{Tj}^{(m)}(\mathbf{x})]| \\
&\le Cb_T^{-d} \sum_{\{i,j : |i-j| \ge q,\ 1 \le i,j \le T\}} \sum |E[\Delta_{Ti}^{(m)}(\mathbf{x}) \Delta_{Tj}^{(m)}(\mathbf{x})]| \\
&\le Cb_T^{-d} \sum_{\{i,j : |i-j| \ge q,\ 1 \le i,j \le T\}} \sum b_T^{4d/(4+\delta)} \{\alpha_m(j-i)\}^{\delta/(4+\delta)} \\
&\le Cb_T^{-\delta d/(4+\delta)} T \left(\sum_{t=q}^{\infty} \{\alpha_m(t)\}^{\delta/(4+\delta)} \right). \qquad \textbf{(A.48)}
\end{aligned}$$

https://doi.org/10.1017/S0266466607070028 Published online by Cambridge University Press

Condition (B4) implies that $qb_T^d = O(1)$ and $\pi_T = O(1)$. Thus, from (A.45), (A.47), and (A.48),

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

which tends to zero by $q/p \to 0$ and condition (B4); (A.42) follows. ■

Proof of (A.43). Let $S_T^{(m)\prime} := \Upsilon^{(m)}(T,\mathbf{x},1)$ and $S_T^{(m)\prime\prime} := \Upsilon^{(m)}(T,\mathbf{x},2)$. Then $S_T^{(m)\prime}$ is a sum of $\zeta_j^{(m)}$'s over the "large" blocks, $S_T^{(m)\prime\prime}$ over the "small" ones. Lemma 3.2 and its argument together with Lemma A.2 imply $T^{-1}E[|S_T^{(m)}|^2] \to \sigma^2$. This, combined with (A.42), entails $T^{-1}E[|S_T^{(m)\prime}|^2] \to \sigma^2$. Now,

$$T^{-1}E[|S_T^{(m)\prime}|^2] = T^{-1}\sum_{j=0}^{r-1}E[U^{(m)}(1,T,\mathbf{x},j)]^2 + T^{-1}\sum_{i\ne j\in\mathcal{J}^*}\mathrm{Cov}(U^{(m)}(1,T,\mathbf{x},j),U^{(m)}(1,T,\mathbf{x},i)), \tag{A.49}$$

where $\mathcal{J}^* = \mathcal{J}^*(p,q) := \{i,j : 1 \le i,j \le r-1\}$. Observe that (A.43) follows from (A.49) if the last sum in the right-hand side of (A.49) tends to zero as $T \to \infty$. Using the same argument as in the derivation of the bound (A.45) for $\hat{V}_2$, this sum can be bounded by

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

which tends to zero by condition (B4). ■

Proof of (A.44). We need a truncation argument because $Z_i^{(m)}$ is not necessarily bounded. Set $Z_i^{(m)M} := Z_i^{(m)}I_{\{|Z_i^{(m)}|\le M\}}$, $\eta_i^{(m)M} := Z_i^{(m)M}K_{\mathbf{c}}((\mathbf{X}_i^{(m)}-\mathbf{x})/b_T)$, $\Delta_i^{(m)M} := \eta_i^{(m)M} - E\eta_i^{(m)M}$, $\zeta_{Ti}^{(m)M} := b_T^{-d/2}\Delta_i^{(m)M}$, 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

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

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

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

Now, $U^{(m)M}(1,T,\mathbf{x},j)/(\sigma T^{1/2}) \le Cp(Tb_T^d)^{-1/2} \to 0$, because of Assumption (B3). Thus $P[U^{(m)M}(1,T,\mathbf{x},j) > \varepsilon\sigma T^{1/2}] = 0$ at all j for sufficiently large T. Thus $Q_4^{(m)M} = 0$ for large T, and (A.44) holds for the truncated variables. Hence,

$$T^{-1/2}S_T^{(m)M} := T^{-1/2}\sum_{j=1}^{T}\zeta_{Tj}^{(m)M} \xrightarrow{\mathcal{L}} N(0,\sigma_M^2), \tag{A.50}$$

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

https://doi.org/10.1017/S0266466607070028 Published online by Cambridge University Press

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

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

Letting $T \to \infty$, E_1 tends to zero by (A.50) and the dominated convergence theorem. Letting M go to infinity, the dominated convergence theorem also implies that $\sigma_M^2 := \mathrm{Var}(Z_i^M|\mathbf{X}_i = \mathbf{x})f(\mathbf{x})\int K_{\mathbf{c}}^2(\mathbf{u})\,d\mathbf{u}$ converges to

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

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

$$
E[(S_T^{(m)M*}/T^{1/2})^2] \to \mathrm{Var}(|Z_i|I_{\{|Z_i|>M\}}|\mathbf{X}_i = \mathbf{x})f(\mathbf{x})\int K_{\mathbf{c}}^2(\mathbf{u})\,d\mathbf{u} \quad \text{as } T \to \infty.
$$

This follows along the routine argument of Lemma 3.3 together with Lemma A.2. The proof of Lemma 3.4 is thus complete. ■

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

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 $v_2(m) = O(m^{-\mu/(1+\delta/4)})$.

$$
\begin{aligned}
T^{2+4/\delta}b_T^{-(2+d+2d/\delta)}v_2(m) &\le CT^{2+4/\delta}b_T^{-(2+d+2d/\delta)}m^{-\mu/(1+\delta/4)} \\
&\le CT^{2+4/\delta}b_T^{-(2+d+2d/\delta)}b_T^{\mu\delta d/\{a(4+\delta)(1+\delta/4)\}} \\
&\le T^{2+4/\delta}b_T^{\mu\delta d/\{a(4+\delta)(1+\delta/4)\}-(2+d+2d/\delta)} \\
&= T^{2+4/\delta}b_T^{\mu/\kappa_2-\kappa_1}, \qquad \textbf{(A.51)}
\end{aligned}
$$

which tends to zero by the condition of this theorem;

$$
\begin{aligned}
b_T^{-4(1+d+2d/\delta)}v_2(m) &\le Cb_T^{-4(1+d+2d/\delta)}m^{-\mu/(1+\delta/4)} \\
&\le Cb_T^{-4(1+d+2d/\delta)}b_T^{\mu\delta d/\{a(4+\delta)(1+\delta/4)\}} \\
&\le b_T^{\mu\delta d/\{a(4+\delta)(1+\delta/4)\}-4(1+d+2d/\delta)} = b_T^{\mu/\kappa_2-4(\kappa_1-1)} = O(1), \qquad \textbf{(A.52)}
\end{aligned}
$$

https://doi.org/10.1017/S0266466607070028 Published online by Cambridge University Press

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

$$\begin{aligned} b_T^{-(2+2d+\delta/2+4d/\delta)} v_{2+\delta/2}(m) &\leq C b_T^{-(2+2d+\delta/2+4d/\delta)} m^{-\mu} \\ &\leq C b_T^{-(2+2d+\delta/2+4d/\delta)} b_T^{\mu\delta d/\{a(4+\delta)\}} \\ &\leq b_T^{\mu\delta d/\{a(4+\delta)\}-(2+2d+\delta/2+4d/\delta)} \\ &= b_T^{\mu(1+\delta/4)/\kappa_2-\kappa_3} = O(1), \end{aligned} \tag{A.53}$$

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 $T/p = \{T \log T/b_T^d\}^{1/2} \to \infty$ clearly; and

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

and

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

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 $a > \delta/(4 + \delta)$,

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

as $k \to \infty$. Therefore Assumption (A4) holds. The theorem thus follows from Lemmas 3.1–3.4. ■

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

https://doi.org/10.1017/S0266466607070028 Published online by Cambridge University Press