Local Linear Fitting under Near Epoch Dependence:
Uniform Consistency with Convergence Rate[*]

Degui Li[†], Zudi Lu[‡] and Oliver Linton[δ]

The Suntory Centre
Suntory and Toyota International Centres for
Economics and Related Disciplines
London School of Economics and Political Science
Discussion paper no.:   Houghton Street
EM/2010/549   London WC2A 2AE
August 2010   Tel: 020 7955 6674

[†] School of Economics, The University of Adelaide, SA 5005, Australia.
[‡] Section of Statistics, School of Mathematical Sciences.
[δ] Department of Economics, London School of Economics, London WC2A 2AE.

Abstract

Local linear fitting is a popular nonparametric method in nonlinear statistical and econometric modelling. Lu and Linton (2007) established the point wise asymptotic distribution (central limit theorem) for the local linear estimator of nonparametric regression function under the condition of near epoch dependence. We further investigate the uniform consistency of this estimator. The uniformly strong and weak consistencies with convergence rates for the local linear fitting are established under mild conditions. Furthermore, general results of uniform convergence rates for nonparametric kernel-based estimators are provided. Applications of our results to conditional variance function estimation and some economic time series models are also discussed. The results of this paper will be of widely potential interest in time series semiparametric modelling.

JEL subject classifications: C13, C14, C22.

Keywords: $\alpha-$mixing, local linear fitting, near epoch dependence, convergence rates, uniform consistency.

1. Introduction

Local linear fitting is a popular nonparametric method in nonlinear statistical and econometric modelling. See, for example, Fan and Gijbels (1996), Fan and Yao (2003) and Li and Racine (2007). Lu and Linton (2007) recently established the pointwise asymptotic distribution (central limit theorem) for the local linear estimator of a nonparametric regression function under the weak assumption of near epoch dependence, which cover a wide range of popular time series econometric models. In this paper, we further investigate the uniform consistency of this nonparametric estimator for near epoch dependent processes. The results of this paper will be of widely potential interest in time series semiparametric modelling (c.f., Andrews, 1995).

Uniform consistency results of nonparametric kernel–based estimators have been studied by many authors, as they are useful in many applications such as the specification testing issue based on the nonparametric kernel method and semiparametric estimator with first–stage nonparametric estimator. For recent development, the reader is referred to Liebscher (1996), Masry (1996), Bosq (1998), Fan and Yao (2003), Hansen (2008), Kristensen (2009) and the references therein. A rather obvious feature of the above literature is that the observed time series are assumed to be α–mixing (i.e., strongly mixing). It is well–known that α–mixing dependence has been one of the most popular dependence conditions in statistic and econometric literature. Indeed, the stationary solutions of many linear and nonlinear time series models are α–mixing under some suitable conditions; see Tjøstheim (1990), Tong (1990), Masry and Tjøstheim (1995), Lu (1998), Cline and Pu (1999) for example.

However, from a practical point of view, the α–mixing dependence suffers from many undesirable features. As pointed out by Davidson (1994) and Lu (2001), α–mixing condition is difficult to verify in practice, especially in the cases of compound processes. For example, the ARCH models and its generalized version GARCH models have been proved to be α–mixing under some mild conditions (Bollerslev 1986, Lu 1996a, b, Carrasco and Chen 2002). But for the compound processes such as ARMA process with ARCH or GARCH errors, it is still difficult to establish whether they are α–mixing or not except in some very special cases. In fact, even very simple autoregressive processes may not be α–mixing for some

1

cases. Andrews (1984) showed that the stationary solution to a simple linear AR(1) model of the form

$$X_t = 1/2 X_{t-1} + e_t, \tag{1.1}$$

with e_t's being independent symmetric Bernoulli random variables taking values -1 and 1, is not α–mixing. Hence, it is natural to consider a more generalized version of stochastic processes beyond α–mixing process in both linear and nonlinear time series analysis.

In this paper, we consider the stationary near epoch dependent (NED) or stable process, which includes the α–mixing process as a particularly special case. One can allow some types of nonstationarity, but this complicates the notation considerably, so we don't formally consider this but we discuss below some special cases. Let both $\{Y_t\}$ and $\{\mathbf{X}_t\}$ be stationary processes of $\mathbb{R}^1$– and $\mathbb{R}^d$–valued, respectively. Based on a stationary process $\{\varepsilon_t\}$, $\{Y_t\}$ and $\{\mathbf{X}_t\}$ are defined by

$$
\begin{aligned}
Y_t &= \Psi_Y(\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots), \\
\mathbf{X}_t &= (X_{t1}, \cdots, X_{td})^\tau = \Psi_{\mathbf{X}}(\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots),
\end{aligned}
\tag{1.2}
$$

where $\mathbf{X}^\tau$ denotes the transpose of $\mathbf{X}$, $\Psi_Y : \mathbb{R}^\infty \longrightarrow \mathbb{R}^1$ and $\Psi_{\mathbf{X}} : \mathbb{R}^\infty \longrightarrow \mathbb{R}^d$ are two Borel measurable functions and $\{\varepsilon_t\}$ may be vector–valued. The definition of NED process is provided as follows.

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) with respect to a stationary α–mixing process $\{\varepsilon_t\}$,* if

$$v_\nu(m) = \mathsf{E}|Y_t - Y_t^{(m)}|^\nu + \mathsf{E}\|\mathbf{X}_t - \mathbf{X}_t^{(m)}\|^\nu \to 0, \quad \nu > 0, \tag{1.3}$$

as $m \to \infty$, where $|\cdot|$ and $\|\cdot\|$ are the absolute value and the Euclidean norm of $\mathbb{R}^d$, respectively, $Y_t^{(m)} = \Psi_{Y,m}(\varepsilon_t, \ldots, \varepsilon_{t-m+1})$, $\mathbf{X}_t^{(m)} = (X_{t1}^{(m)}, \cdots, X_{td}^{(m)})^\tau = \Psi_{\mathbf{X},m}(\varepsilon_t, \ldots, \varepsilon_{t-m+1})$, and $\Psi_{Y,m}$ and $\Psi_{\mathbf{X},m}$ 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 $\{(Y_t, \mathbf{X}_t)\}$.

The above concept of NED process dated back to Ibragimov (1962) and was further developed by Billingsley (1968), McLeish (1975a,1975b,1977) and Lin (2004). Basically, most of them assumed that $\{\varepsilon_t\}$ is martingale difference or φ–mixing. It has been used in

econometrics following Bierens (1981); see, for example, Gallant (1987), Gallant and White (1988), and Andrews (1995). In this paper, we are concerned with NED process with respect to the stationary α–mixing process $\{\varepsilon_t\}$. The NED process can easily cover some important compounded econometric processes and many nonlinear non–α–mixing processes; see Section 4 for example.

There has been some literature on estimation and testing issues under NED processes. Andrews (1995) established uniform convergence with rates for nonparametric density and regression estimators based on the local constant paradigm under NED conditions. Lu (2001) established asymptotic normality for kernel density estimators for NED processes. Ling (2007) developed a strong law of large numbers and a strong invariance principle for NED sequences when $\{\varepsilon_t\}$ is independent and used the results to test for change points. Lu and Linton (2007) established the pointwise asymptotic distribution of local linear estimators for NED process. In this paper, we further establish the uniform strong and weak convergence rates of the local linear estimators. Furthermore, we develop some general results on uniform convergence rates of nonparametric kernel–based estimators, which could be seen as an extension of Hansen (2008) from α-mixing to NED processes. In particular, we obtain the uniform rate over expanding subsets of the covariate support. The rate we obtain is constrained by the amount of dependence but does not explicitly depend on it, as it does in Andrews (1995). This means that in some special cases our rate is optimal, Stone (1980). We then discuss the application of the obtained results to establish the uniform convergence rate of local linear estimator of conditional variance function. We also apply the uniform consistency results with convergence rates to some interesting time series models as well as some compounded processes.

We also provide new results on estimation of a countable number of regression functions, for example $m_j(\mathbf{x}) = E(Y_t|\mathbf{X}_{t-j} = \mathbf{x})$, $j = 1, 2, \ldots$. This application occurs naturally in a number of time series settings (Hong (2000) and Linton and Mammen (2005)) but does not appear to have been formally treated before at this level of generality. We establish the uniform rate of convergence of the local linear estimators uniformly over j as well.

Before the end of this section, we remark that some alternative extension of dependence beyond mixing can also be found in Nze, Bühlmann, and Doukhan (2002) and Nze and

3

Doukhan (2004) who investigated a class of dependent processes they call "weak depen-dent", the definition of which is quite involved. They established the asymptotic normality and uniform convergence of local constant nonparametric regression estimators under their conditions. Differently from their work of uniform consistency over a finite compact set for local constant estimators, we are concerned with more desirable local linear fitting with the uniform convergence under NED, over a compact set depending on sample size T, which approaches to an infinite set as $T \to \infty$.

The rest of the paper is organized as follows. The local linear fitting and the uniform convergence rates of the local linear estimators are presented in Section 2. The general results of uniform convergence rates for nonparametric kernel-based estimators are provided in Section 3. Some applications are given in Section 4. The technical lemmas and the proofs of the main results are collected in an Appendix.

2. Uniform convergence rate of local linear fitting

In this section, we study the local linear estimator of the regression function

$$g(\mathbf{x}) := \mathsf{E}\left(Y_t | \mathbf{X}_t = \mathbf{x}\right). \tag{2.1}$$

Local linear fitting is a widely–used nonparametric estimation method and it has advantages over the popular Nadaraya–Watson kernel method, in terms of the ability of design adaption and high asymptotic efficiency (e.g., the Best Linear Minimax property). The local linear estimation method can adapt to almost all regression settings and cope well with the edge effects. See Fan and Gijbels (1996) for detailed account on this subject.

The main idea of local linear fitting consists in approximating, in a neighborhood of $\mathbf{x}$, the unknown regression function $g(\cdot)$ by a linear function. Under the condition that $g(\cdot)$ has continuous derivatives up to the second order, 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 $(a_0, \mathbf{a}_1^\tau) = (g(\mathbf{x}), (g'(\mathbf{x}))^\tau)$ by

$$\begin{pmatrix} \widehat{a}_0 \\ \widehat{\mathbf{a}}_1 \end{pmatrix} := \arg \min_{(a_0, \mathbf{a}_1) \in \mathbb{R}^{d+1}} \sum_{t=1}^{T} (Y_t - a_0 - \mathbf{a}_1^\tau (\mathbf{X}_t - \mathbf{x}))^2 K\left(\frac{\mathbf{X}_t - \mathbf{x}}{h}\right), \tag{2.2}$$

where $h := h_T$ is a sequence of bandwidths tending to zero at appropriate rate as T tends to infinity, and $K(\cdot)$ is a kernel function with values in $\mathbb{R}^+$. Denote the local linear estimators of $(g(\mathbf{x}), (g'(\mathbf{x}))^\tau)$ by $(\widehat{g}(\mathbf{x}), (\widehat{g}'(\mathbf{x}))^\tau)$, where $\widehat{a}_0 = \widehat{g}(\mathbf{x})$ and $\widehat{\mathbf{a}}_1 = \widehat{g}'(\mathbf{x})$.

There has been rich literature on the uniform convergence rate for the local linear estimators under mixing dependent condition, see Masry (1995), Fan and Yao (2003) and Hansen (2008) for example. Lu and Linton (2007) established the pointwise asymptotic distribution for the local linear estimators under the NED condition. In this section, we will provide the uniform convergence rate for $\widehat{g}(\mathbf{x})$ over the set

$$\{\mathbf{x} : \|\mathbf{x}\| \leq C_T\}, \tag{2.3}$$

where $C_T = (\log T)^{1/d} T^{1/\tau_0}$, for some $\tau_0 > 0$.

We first introduce some regularity conditions to establish the uniform convergence rate for the proposed estimators.

A1 The kernel function $K(\cdot)$ is positive, bounded and Lipschitz continuous such that

$$|K(\mathbf{x}_1) - K(\mathbf{x}_2)| \leq C_K \|\mathbf{x}_1 - \mathbf{x}_2\|,$$

where C_K is some positive constant. Furthermore, $\int_{\mathbb{R}^d} \|u\|^j K(u) du < \infty$ for $j = 0, 1, 2$.

A2 (i) $f(\cdot)$ is continuous and $a_T(f) := \inf_{\|\mathbf{x}\| \leq C_T} f(\mathbf{x}) > 0$, where $f(\cdot)$ is the density function of the stationary process $\{\mathbf{X}_t\}$. Furthermore, the joint density function $f_{ij}(\cdot, \cdot)$ of $(\mathbf{X}_i, \mathbf{X}_j)$ satisfies $f_{ij}(\mathbf{x}_1, \mathbf{x}_2) < C_f$ for all $i \neq j$ and $(\mathbf{x}_1, \mathbf{x}_2) \in \mathbb{R}^{2d}$, $C_f < \infty$.

(ii) The regression function $g(\cdot)$ has continuous derivatives up to the second order.

A3 (i) $\{Y_t, \mathbf{X}_t\}$ is stationary NED in L_{p_0} norm with respect to a stationary α–mixing $\{\varepsilon_t\}$, $\mathsf{E}|Y_t|^{p_0} < \infty$ and $\mathsf{E}\|\mathbf{X}_t\|^{p_0} < \infty$, $p_0 = 2 + \varepsilon^*$, $\varepsilon^* > 0$.

(ii) The mixing coefficient α_t of the stationary α–mixing $\{\varepsilon_t\}$ satisfies $\alpha_t \leq C_\alpha t^{-\theta_0}$, $C_\alpha < \infty$, $\theta_0 > \beta_1$, $\beta_1 = \left(\frac{3p_0+6}{4p_0} + (\frac{1}{2} + \frac{1}{\tau_0})d \right) / \left(\frac{1}{2} - \frac{1}{p_0} \right)$, where τ_0 is defined in C_T of (2.3).

A4 (i) Let $m_T = \left[\sqrt{\frac{T^{1-2/p_0} h^d}{\log T}} \right]$. The bandwidth h and the stability coefficient defined in (1.3) satisfy:

$$h^{-(d+1)} v_1(m_T) \rho_T^{-1} = O(1), \quad h^{-4(1+d+d/(p_0-2))} v_2(m_T) = O(1) \tag{2.4}$$

5

$$h^{-(2d+p_0+2d/(p_0-2))}v_{p_0}(m_T) = O(1),$$

where

$$\rho_T = \sqrt{\frac{\log T}{Th^d}}.$$

(ii) The bandwidth h satisfies $h \to 0$,

$$\frac{T^{1-2/p_0}h^d}{\log T} \to \infty, \quad (\log T)^{\frac{\theta_0}{2}+\beta_3}h^{-\frac{\theta_0 d}{2}-\beta_2}T^{(\beta_1-\theta_0)(\frac{1}{2}-\frac{1}{p_0})} = o(1), \qquad (2.5)$$

where $\beta_2 = \frac{7(2+d)d}{4}$ and $\beta_3 = \frac{5-2d}{4}$.

REMARK 2.1. A1 is mild and some commonly–used kernel functions such as the standard normal kernel function can be shown to satisfy A1. By contrast, Masry (1996) requires kernels that have compact support. A2 (i) and (ii) are some conditions on the density functions and the regression function and they are similar to the corresponding assumptions in Lu and Linton (2007); if the regression function g is less smooth than assumed here, one obtains a different magnitude of the bias terms but otherwise the argument goes through. A3 provides the moment conditions on $\{Y_t, \mathbf{X}_t\}$ and the mixing coefficient condition for $\{\varepsilon_t\}$. A4 (i) is on the stability coefficient defined by (1.3) in Section 1. As A4 (i) looks complex, we will give some specific examples in Section 4 to show that it holds under mild conditions on the bandwidth. Note that

$$
\begin{aligned}
h^{-(d+1)}v_1(m_T)\rho_T^{-1} &= (h^{-2(d+1)}v_2(m_T)\rho_T^{-2})^{1/2} \\
&= \left(h^{-4(1+d+d/(p_0-2))}v_2(m_T)h^{2(d+1)+4d/(p_0-2)}\rho_T^{-2}\right)^{1/2}.
\end{aligned}
$$

If $h^{2(d+1)+4d/(p_0-2)}\rho_T^{-2} = O(1)$, we can show that the first term in (2.4) is a particular case of the second term of (2.4). A4 (ii) is the regularity condition to establish the uniform convergence rate for nonparametric kernel–based estimators under α–mixing dependent assumption. Meanwhile, A4 (ii) can be simplified for some particular cases. For example, if $\theta_0 \to \infty$ (α–mixing process decays with the exponential rate), the second term in (2.5) can be re–written as

$$\frac{T^{\frac{1}{2}-\frac{1}{p_0}}h^{\frac{d}{2}}}{(\log T)^{\frac{1}{2}}}\left((\log T)^{-\beta_3/\theta_0}h^{\beta_2/\theta_0}T^{-\beta_1(\frac{1}{2}-\frac{1}{p_0})/\theta_0}\right) \to \infty.$$

As β_1, β_2 and β_3 are some constant, it means that

$$\beta_1(\frac{1}{2} - \frac{1}{p_0})/\theta_0 \to 0, \quad \beta_2/\theta_0 \to 0, \quad \beta_3/\theta_0 \to 0, \quad \text{as} \quad \theta_0 \to \infty.$$

Hence, for the case of $\theta_0 \to \infty$, the second term in (2.5) is just slightly stronger than the first term in (2.5). Similarly, as $p_0 \to \infty$, the first term in (2.5) is slightly stronger than the condition $Th^d/\log T \to \infty$.

Finally, we discuss the case where one of the covariates is rescaled time, i.e., $X_t = t/T$ (or some stochastic perturbation thereof). In this case, the covariate is nonstationary and compactly supported, and although the above conditions are not satisfied, the subsequent results go through as stated.

We first give the uniform convergence rate of local linear estimator $\widehat{g}(\mathbf{x})$ in probability.

THEOREM 2.1. *Suppose that the conditions A1–A4 are satisfied. Then, we have*

$$\sup_{\|\mathbf{x}\| \leq C_T} |\widehat{g}(\mathbf{x}) - g(\mathbf{x})| = O_P\left(\frac{\rho_T}{a_T(f)} + \frac{b_T(g)h^2}{a_T(f)}\right), \tag{2.6}$$

where $b_T(g) = \sup\limits_{\|\mathbf{x}\| \leq C_T} \|g''(\mathbf{x})\|$, $a_T(f)$ *and* ρ_T *are defined in A2 (i) and A4 (i), respectively.*

REMARK 2.2. The above theorem can be regarded as an extension of Theorem 10 in Hansen (2008) from α–mixing process to NED process. If the second order derivatives of $g(\mathbf{x})$ are uniformly bounded as in Masry (1996), $b_T(g) < C_g$ for some $0 < C_g < \infty$. Then (2.6) becomes

$$\sup_{\|\mathbf{x}\| \leq C_T} |\widehat{g}(\mathbf{x}) - g(\mathbf{x})| = O_P\left(\frac{\rho_T + h^2}{a_T(f)}\right). \tag{2.7}$$

Furthermore, if we let $C_T = C^*$ and $a_T(f) > c_0 > 0$, (2.7) becomes

$$\sup_{\|\mathbf{x}\| \leq C^*} |\widehat{g}(\mathbf{x}) - g(\mathbf{x})| = O_P\left(\rho_T + h^2\right). \tag{2.8}$$

Taking $h \propto (\log T/T)^{1/(4+d)}$, the right hand side becomes $(\log T/T)^{2/(4+d)}$, which is the optimal rate in the compactly supported i.i.d. case, Stone (1980). This bandwidth is consistent with A4 under restrictions on p_0, d, θ_0, and the stability coefficients v_j, $j = 1, 2, \ldots, p_0$. For example, we require that $1 - 2/p_0 - d/(4+d) > 0$. (2.8) can be regarded

7

as the extension of some existing results under the mixing dependence assumption such as Theorem 6.5 in Fan and Yao (2003).

We next establish the uniform strong convergence rate of local linear estimator $\widehat{g}(\mathbf{x})$.

THEOREM 2.2. *Suppose that the conditions in Theorem 2.1 are satisfied,* $\mathsf{E}|Y_t|^s < \infty$ *and* $\mathsf{E}\|\mathbf{X}_t\|^s < \infty$, $s > 2p_0$,

$$Th^{-(d+1)}v_1(m_T)\rho_T^{-1} = O\left((\log T)^{-(1+\varsigma)}\right), \quad \varsigma > 0, \tag{2.9}$$

and

$$(\log T)^{\frac{\theta_0}{2}+\beta_3}h^{-\frac{\theta_0 d}{2}-\beta_2}T^{1+(\beta_1-\theta_0)(\frac{1}{2}-\frac{1}{p_0})} = O\left((\log T)^{-(1+\varsigma)}\right). \tag{2.10}$$

Then, we have

$$\sup_{\|\mathbf{x}\| \leq C_T} |\widehat{g}(\mathbf{x}) - g(\mathbf{x})| = O\left(\frac{\rho_T}{a_T(f)} + \frac{b_T(g)h^2}{a_T(f)}\right) \quad a.s. \tag{2.11}$$

3. General uniform convergence

Before we prove the above theorems, in this section we provide some general uniform convergence for the following W_T under our NED assumption, from which we can derive the above theorems conveniently. Let $\{Y_t, \mathbf{X}_t\}$ be a stationary NED sequence defined in Section 1. We next consider the weighted average form

$$W_T(\mathbf{x}) = \frac{1}{Th^d} \sum_{t=1}^{T} \Psi(Y_t) K_T\left(\frac{\mathbf{X}_t - \mathbf{x}}{h}\right), \tag{3.1}$$

where h is the bandwidth and $K_T(\cdot) : \mathbb{R}^d \to \mathbb{R}$ is a kernel–based weight function. By suitable choice of $K_T(\cdot)$ and $\Psi(\cdot)$, many kernel–based nonparametric estimators such as the kernel density estimator, Nadaraya–Watson estimator and local polynomial estimator can be written as the form of (3.1). When $\Psi(y) \equiv y$, Hansen (2008) established the weak and strong uniform convergence rate of $W_T(\cdot)$ for stationary α–mixing process. We will provide the uniform convergence rate for $W_T(\cdot)$ defined by (3.1) when α–mixing dependence is replaced by NED condition, which is a generalized version of α–mixing dependence.

To establish the uniform convergence rate of $W_T(\mathbf{x})$, we provide the following regularity conditions.

8

A5 (i) The function $\Psi(\cdot)$ has continuous derivatives up to the q_0–th order and

$$\mathsf{E}\left|\Psi^{(k)}(Y_t)\right|^s < \infty, \quad 0 \le k < q_0, \quad s > p_0 = 2 + \varepsilon^*, \quad \varepsilon^* > 0, \quad \sup_{y \in \mathbb{R}}\left|\Psi^{(q_0)}(y)\right| < \infty.$$

(ii) $K_T(\cdot)$ is integrable, bounded and Lipschitz continuous satisfying

$$\sup_{T \ge 1}|K_T(\mathbf{x}_1) - K_T(\mathbf{x}_2)| \le C_K^* \|\mathbf{x}_1 - \mathbf{x}_2\|,$$

where C_K^* is some positive constant.

A6 A4 (i) and (ii) are satisfied, and

$$h^{-d}\rho_T^{-1}\left(\sum_{k=1}^{q_0-1} v_{k\delta_p}^{1/\delta_p}(m_T) + v_{q_0}(m_T)\right) = O(1),$$

where $\delta_p = p_0/(p_0 - 1)$.

The uniform convergence rate results for $W_T(\mathbf{x})$ are provided in the following two theorems.

THEOREM 3.1. *Suppose that the conditions A5–A6 are satisfied. Then, we have*

$$\sup_{\|\mathbf{x}\| \le C_T}|W_T(\mathbf{x}) - \mathsf{E}[W_T(\mathbf{x})]| = O_P(\rho_T). \tag{3.2}$$

THEOREM 3.2. *Suppose that the conditions in Theorem 3.1 and (2.10) are satisfied and $s > 2p_0$ in A5 (i),*

$$Th^{-d}\rho_T^{-1}\left(\sum_{k=1}^{q_0-1} v_{k\delta_p}^{1/\delta_p}(m_T) + v_{q_0}(m_T)\right) = O\left((\log T)^{-(1+\varsigma)}\right), \quad \varsigma > 0, \tag{3.3}$$

Then, we have

$$\sup_{\|\mathbf{x}\| \le C_T}|W_T(\mathbf{x}) - \mathsf{E}[W_T(\mathbf{x})]| = O(\rho_T) \quad a.s. \tag{3.4}$$

REMARK 3.1. The above theorems established the weak and strong convergences for $W_T(\mathbf{x})$. We remark that under some suitable conditions, an L_{Q_0} convergence of $W_T(\mathbf{x})$, for some

$Q_0 > 1$, can also be established. For example, let $Q_1 > Q_0 > 1$. Suppose that the function $\Psi(\cdot)$ has continuous derivatives up to the q_0–th order and

$$\mathsf{E}\left|\Psi^{(k)}(Y_t)\right|^{Q_1} < \infty, \quad 0 \leq k < q_0, \quad \sup_{y \in \mathbb{R}}\left|\Psi^{(q_0)}(y)\right| < \infty,$$

and the mixing coefficient

$$\alpha_t \leq C_\alpha^* t^{-\theta_0^*}, \quad \theta_0^* > (Q_1 Q_0)/2(Q_1 - Q_0).$$

Then, applying Theorem 4.1 in Shao and Yu (1996) and following the proof of Lemma A.2, we can show that if $\{(\mathbf{X}_t, Y_t)\}$ is NED in $L_{\tau_{q_0-1} \vee \tau^*}$,

$$\sup_{\|\mathbf{x}\| \leq C_T} \left(\mathsf{E}\left|W_T(\mathbf{x}) - \mathsf{E}\left[W_T(\mathbf{x})\right]\right|^{Q_0}\right)^{1/Q_0} = O\left(T^{-\frac{1}{2}} h^{\frac{(1-Q_1)d}{Q_1}} + \sum_{k=1}^{q_0-1} v_{\tau_k}^{\tau_0}(m_T) + v_{\tau^*}^{\frac{1}{Q_0}}(m_T)\right), \tag{3.5}$$

where $\tau_0 = \frac{Q_1 - Q_0}{Q_1 Q_0}$, $\tau_k = \frac{kQ_1}{Q_1 - Q_0}$ and $\tau^* = q_0 Q_0$. When the stable coefficient decays at a geometric rate as in Sections 4.3 and 4.4, $O(T^{-\frac{1}{2}} h^{\frac{(1-Q_1)d}{Q_1}})$ is the leading term for convergence rate on the right hand side of (3.5).

REMARK 3.2. In (3.1), Ψ is only a function of Y_t. Sometimes, it might be useful to allow Ψ to depend on some other parameter like y, say $\Psi(Y_t, y) = I_{\{Y_t \leq y\}}$, which may not be differentiable as a function of Y_t, where I_A is an indicator function of set A. In this way we can handle conditional cumulative distribution function estimation and so things like conditional quantiles and expected shortfall. The outcomes of the above theorems can still apply to this case under some suitably modified conditions, with the proof going through as done in Appendices A and B by replacing (A.9) in Appendix A below with

$$\mathsf{E}|I_{\{Y_t \leq y\}} - I_{\{Y_t^{(m)} \leq y\}}| = \mathsf{E}|I_{\{y < Y_t\}} - I_{\{y < Y_t^{(m)}\}}| \leq \mathsf{E}I_{\{|y - Y_t| \leq |Y_t - Y_t^{(m)}|\}}$$
$$\leq \mathsf{P}\{|y - Y_t| \leq |Y_t - Y_t^{(m)}|\} \leq O(v_2^{1/3}(m)),$$

where the final inequality follows by noticing

$$\mathsf{P}\{|y - Y_t| \leq |Y_t - Y_t^{(m)}|\} \leq \mathsf{P}\{|y - Y_t| \leq v_2^{1/3}(m)\} + \mathsf{P}\{v_2^{1/3}(m) \leq |Y_t - Y_t^{(m)}|\},$$

which is controlled by $2f_Y(y)O(v_2^{1/3}(m)) + v_2^{-2/3}(m)\mathsf{E}|Y_t - Y_t^{(m)}|^2 = O(v_2^{1/3}(m))$ if we assume the probability density function of Y_t, $f_Y(y)$, is bounded away from the infinity.

4. Applications

In this section, we provide some applications of the main results established in Sections 2 and 3. We not only study the uniform convergence rate for local linear estimator of the conditional variance function under NED assumption, but also establish the uniform convergence rate results for some interesting time series models.

4.1. Estimation of conditional variance function

Define $\sigma^2(\mathbf{x}) := \mathsf{Var}(Y_t|\mathbf{X}_t = \mathbf{x})$. There exists an extensive literature on estimating $\sigma^2(\mathbf{x})$ in stationary time series, where $\mathbf{X}_t$ contains lagged values of Y_t; see Pagan and Hong (1991), Fan and Yao (1998), Yu and Jones (2004), and Chen *et. al.* (2009) for example. We next use the local linear method to estimate $\sigma^2(\mathbf{x})$ under NED assumption, and establish the uniform convergence rate for the resulting estimator.

Let $e_t = Y_t - g(\mathbf{X}_t)$ and $\widehat{e}_t = Y_t - \widehat{g}(\mathbf{X}_t)$, where $\widehat{g}(\mathbf{X}_t)$ is the local linear estimator of $g(\mathbf{X}_t)$ defined by (2.2). It is easy to check that $\sigma^2(\mathbf{x}) = \mathsf{Var}(Y_t|\mathbf{X}_t = \mathbf{x}) = \mathsf{E}(e_t^2|\mathbf{X}_t = \mathbf{x})$. We then estimate $\sigma^2(\mathbf{x})$ by $\widehat{\sigma}^2(\mathbf{x}) = \widehat{a}_0^*$, where

$$
\begin{pmatrix} \widehat{a}_0^* \\ \widehat{\mathbf{a}}_1^* \end{pmatrix} = \arg \min_{(a_0, \mathbf{a}_1) \in \mathbb{R}^{d+1}} \sum_{t=1}^{T} \left(\widehat{e}_t^2 - a_0 - \mathbf{a}_1^\tau (\mathbf{X}_t - \mathbf{x}) \right)^2 W \left(\frac{\mathbf{X}_t - \mathbf{x}}{h_*} \right), \qquad (4.1)
$$

where $W(\cdot)$ is a kernel function and h_* is a bandwidth. The following condition is necessary to obtain the uniform convergence rate for $\sigma^2(\mathbf{x})$.

A7 (i) The kernel function $W(\cdot)$ is positive, bounded and Lipschitz continuous. Furthermore, $\int_{\mathbb{R}^d} \|u\|^j W(u) du < \infty$ for $j = 0, 1, 2$.

(ii) $\mathsf{E}|Y_t|^{2p_0} < \infty$ and $\mathsf{E}\|\mathbf{X}_t\|^{2p_0} < \infty$, $2p_0 > d$.

(iii) The conditional variance function $\sigma^2(\cdot)$ has continuous derivatives up to the second order and $b_T(\sigma^2) \sim b_T(g)$, where $b_T(\sigma^2) = \sup_{\|\mathbf{x}\| \leq C_T} \left| (\sigma^2(\mathbf{x}))'' \right|$ and $a_T \sim b_T$ means that $a_T/b_T \to 1$ as $T \to \infty$.

(iv) $h \sim h_*$ and A4 holds if h is replaced by h_*.

We next give the uniform weak and strong convergence rate of local linear estimator $\widehat{\sigma}^2(\mathbf{x})$ in the following two propositions.

PROPOSITION 4.1. *Suppose that the conditions A1–A4 and A7 are satisfied. Then, we have*

$$\sup_{\|\mathbf{x}\|\leq C_T} \left|\widehat{\sigma}^2(\mathbf{x}) - \sigma^2(\mathbf{x})\right| = O_P\left(\frac{\rho_T}{a_T(f)} + \frac{b_T(\sigma^2)h^2}{a_T(f)}\right). \tag{4.2}$$

PROPOSITION 4.2. *Suppose that the conditions in Proposition 4.1 and (2.10) are satisfied and*

$$Th^{-d}\rho_T^{-1}\left(v_{\delta_p}^{1/\delta_p}(m_T) + v_2(m_T)\right) = O\left((\log T)^{-(1+\varsigma)}\right), \quad \varsigma > 0.$$

Then, we have

$$\sup_{\|\mathbf{x}\|\leq C_T} \left|\widehat{\sigma}^2(\mathbf{x}) - \sigma^2(\mathbf{x})\right| = O\left(\frac{\rho_T}{a_T(f)} + \frac{b_T(\sigma^2)h^2}{a_T(f)}\right) \quad a.s. \tag{4.3}$$

4.2. Estimation of a countable number of conditional expectations

Define the quantities $m_j(\mathbf{x}) = \mathsf{E}(Y_t|\mathbf{X}_{t-j} = \mathbf{x})$, $j = 1, 2, \ldots$. There are many cases of interest that require estimation of this whole family of regression functions. For example, consider the quantity

$$m(\mathbf{x}) = \sum_{j=1}^{\infty} w_j m_j(\mathbf{x}), \tag{4.4}$$

where we assume that this sum is well defined. This quantity is of interest in a number of applications. For example, Hong (2000) proposed a test of serial independence of an observed scalar series X_t. In practice checking the independence of X_t from $X_{t-1}, X_{t-2}, \ldots$ is very difficult due to the curse of dimensionality. He proposed to check all pairwise joint relationships (X_t, X_{t-j}) for departures from the null. An alternative approach is to check all pairwise conditional relationships $X_t|X_{t-j}$. For example, to check whether all functions $m_j(x) = \mathsf{E}(\Psi(X_t)|X_{t-j} = x)$ are constant. This can be done by evaluating an empirical version of the weighted sum $\sum_{j=1}^{\infty} w_j|m_j(x) - m|$, where w_j are summable weights and m_j are average values. Linton and Mammen (2005) considered the semiparametric volatility model for observed returns $X_t = \sigma_t\varepsilon_t$ with ε_t and $\varepsilon_t^2 - 1$ martingale difference sequences and

$$\sigma_t^2 = \sum_{j=1}^{\infty} \psi_j(\theta)g(X_{t-j}),$$

where g is an unknown function and the parametric family $\{\psi_j(\theta), \theta \in \Theta\}_{j=1}^{\infty}$ satisfy some regularity conditions. This model includes the GARCH(1,1) as a special case. They assumed

that $\{X_t\}$ is stationary and geometrically mixing. They obtained a characterization of the function g that involves a weighted sum of the form (4.4). They proposed an estimation strategy for the unknown quantities, which requires as input the estimation of $m_j(x) = \mathsf{E}(X_t^2 | X_{t-j} = x)$ for $j = 1, 2, \ldots, J(T)$, where $J(T) = c \log T$ for some $c > 0$. They required to bound the estimation error of $m_j(x)$ uniformly over x and over $j = 1, 2, \ldots, J(T)$. They provided only a sketch proof of this result in the case where the process is assumed to have compact support and to be strong mixing with geometric decay. Here we give more definitive results under weaker conditions. As a final motivation, consider the nonparametric prediction of a future value X_{T+1} given a sample $\{X_1, \ldots, X_T\}$. Linton and Sancetta (2009) establish consistency of estimators of $E(X_{T+1} | X_T, \ldots, X_1)$ under weak conditions but rates of convergence are not available and practical performance is likely to be poor. Instead, it makes sense to use lower dimensional predictors, but which one? Consider the following model averaging approach, which makes use of a large number of low dimensional predictors. Let

$$\widehat{\Psi}_{T+1|T} = \sum_{j=1}^{J(T)} w_{T,j} \widehat{m}_j(X_{T-j}),$$

where $w_{T,j}$ are weights such that $\sum_{j=1}^{J(T)} w_{T,j} = 1$, while $\widehat{m}_j(.)$ are the nonparametric regression fits described above, and $J(T)$ is an increasing sequence. Let $m(x_1, x_2, \ldots)$ denote $\mathsf{E}(\Psi(X_t) | X_{t-1} = x_1, X_{t-2} = x_2, \ldots)$ the best prediction function. Then $m_w(x_1, x_2, \ldots) = \sum_{j=1}^{J(T)} w_{T,j} m_j(x_j)$ can be considered an approximation to $m(x_1, x_2, \ldots)$. One can choose the weights according to several criteria, which we do not go into here. In this case, to show the rate of uniform convergence of $\widehat{m}_w(x_1, x_2, \ldots)$ to $m_w(x_1, x_2, \ldots)$, where $\widehat{m}_w(x_1, x_2, \ldots) = \sum_{j=1}^{J(T)} w_{T,j} \widehat{m}_j(x_j)$, it suffices to control the rate for each $\widehat{m}_j(x_j)$ uniformly over $j = 1, \ldots, J(T)$.

We next give a result that establishes the same rate of convergence as in Theorem 2.1 but uniformly over j as well. We just need some restriction on the rate at which $J(T)$ can increase to infinity. Our result allows $J(T)$ to grow at a polynomial rate in some cases.

PROPOSITION 4.3. *Suppose that the conditions in Theorem 2.1 are satisfied and*

$$J(T) h^{-(d+1)} v_1(m_T) \rho_T^{-1} = O(1),$$
$$J(T)(\log T)^{\frac{\theta_0}{2} + \beta_3} h^{-\frac{\theta_0 d}{2} - \beta_2} T^{(\beta_1 - \theta_0)(\frac{1}{2} - \frac{1}{p_0})} = o(1).$$

13

Then, we have

$$\max_{1 \leq j \leq J(T)} \sup_{\|\mathbf{x}\| \leq C_T} |\widehat{m}_j(\mathbf{x}) - m_j(\mathbf{x})| = O_P \left(\frac{\rho_T}{a_T(f)} + \frac{b_T h^2}{a_T(f)} \right),$$

where f is the density function of X_t and $b_T = \max_{1 \leq j \leq J(T)} b_T(m_j)$.

4.3. AR(1)–NARCH(1,1) model

We next consider the compound model which is commonly–used in financial econometrics

$$\begin{aligned} X_t &= b_0 + b_1 X_{t-1} + \varepsilon_t, \\ \varepsilon_t &= \mu(\varepsilon_{t-1}) + \sigma(\varepsilon_{t-1}) e_t, \end{aligned} \tag{4.5}$$

where X_t is the daily return of some equity on day t, modelled by an AR(1) model and the error is modelled by a nonlinear ARCH (NARCH) model with $\{e_t\}$ being independent and identically distributed (i.i.d.) random sequence with $E[e_t] = 0$ and $E[e_t^2] = 1$.

If $|b_1| < 1$, it is well–known that the AR (1) model in (4.5) can be expressed as

$$X_t = b_0/(1 - b_1) + \varepsilon_t + \sum_{j=1}^{\infty} b_1^j \varepsilon_{t-j}. \tag{4.6}$$

On the other hand, under some suitably regular conditions on $\mu(\cdot)$ and $\sigma(\cdot)$, the ε_t in the NARCH(1,1) model is α–mixing with a geometrically decaying mixing coefficient, see, for example, Masry and Tjøstheim (1995) and Lu (1998). To the best of our knowledge, it is difficult to show that $\{X_t\}$ is α–mixing under some mild conditions. However, it can be shown that $\{X_t\}$ is NED of order p_0 with respect to a α–mixing process, if $\mathsf{E}|\varepsilon_t|^{p_0} < \infty$ ($p_0 > 2$), with stable coefficients (owing to the convex property of $|\cdot|^{p_0}$)

$$\begin{aligned} v_{p_0}(m) = \mathsf{E} \left| X_t - X_t^{(m)} \right|^{p_0} &= w_m^{p_0} \mathsf{E} \left| \sum_{j=m+1}^{\infty} \frac{b_1^j}{w_m} \varepsilon_{t-j} \right|^{p_0} \\ &\leq w_m^{p_0} \mathsf{E} \left[\sum_{j=m+1}^{\infty} \frac{b_1^{jp_0}}{w_m^{p_0}} |\varepsilon_{t-j}|^{p_0} \right] = O\left(|b_1|^{mp_0} \right), \end{aligned}$$

decaying at a geometric rate, where:

$$X_t^{(m)} = \varepsilon_t + \sum_{j=1}^{m} b_1^j \varepsilon_{t-j}$$

14

$$w_m = \sum_{j=m+1}^{\infty} b_1^j = O(b_1^m).$$

We are concerned with estimation of the autoregression function $g(\mathbf{x}) = \mathsf{E}(X_t|X_{t-1} = x)$ and the conditional variance function $\sigma^2(\mathbf{x}) = \mathsf{Var}(X_t|X_{t-1} = x)$. As the stable coefficient decays at a geometric rate, it is easy to check that the condition A4 (i) is satisfied if A4 (ii) holds. Hence, by Theorems 2.1, 2.2 and Propositions 4.1, 4.2, we can obtain the uniform strong and weak convergence rate of local linear fitting in AR(1)–NARCH(1,1) model (4.5).

4.4. GARCH(1,1) model with α–mixing innovations

Consider the GARCH(1,1) process defined by

$$X_t = \sigma_t e_t, \quad \sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2. \tag{4.7}$$

When $\{e_t\}$ is i.i.d. random sequence with $E[e_t] = 0$ and $E[e_t^2] = 1$, under $\alpha_1 + \beta_1 < 1$ with some suitably regular conditions, $\{X_t\}$ in the GARCH (1,1) model (4.7) is α–mixing with a geometrically decaying mixing coefficient, see Carrasco and Chen (2002) for example.

We next relax the i.i.d. condition on $\{e_t\}$ by allowing it to be stationary and α–mixing with $E[e_t|\mathcal{F}_{t-1}] = 0$ and $E[e_t^2|\mathcal{F}_{t-1}] = 1$, where $\mathcal{F}_t = \sigma(e_s, \ s \leq t)$ is the σ-field generated as usual. Then the GARCH(1,1) model (4.7) is a semi-strong GARCH model, which is of central importance in the theory of estimation, Drost and Nijman (1993) and Lee and Hansen (1994). We show that $\{X_t\}$ is NED as in Definition 1. By elementary calculation, we have

$$X_t = e_t \sqrt{\alpha_0 + \alpha_0 \sum_{j=1}^{\infty} Z_t(j)}, \tag{4.8}$$

where $Z_t(j) = \prod_{k=1}^{j}(\alpha_1 e_{t-k}^2 + \beta_1)$. Let

$$X_t^{(m)} = e_t \sqrt{\alpha_0 + \alpha_0 \sum_{j=1}^{m} Z_t(j)}. \tag{4.9}$$

Assuming that $\mathsf{E}[|e_t|^{2p_0}] < \infty$ and

$$\mathsf{E}\left[|\alpha_1 e_t^2 + \beta_1|^{2p_0}|\mathcal{F}_{t-1}\right] \leq \rho_* < 1, \tag{4.10}$$

then, by Minkowski inequality and Cauchy–Schwarz inequality, we have

$$\left(\mathsf{E}\left|X_t - X_t^{(m)}\right|^{p_0}\right)^{1/p_0} = \left(\mathsf{E}\left[\left|e_t\left(\sqrt{\alpha_0 + \alpha_0 \sum_{j=1}^{\infty} Z_t(j)} - \sqrt{\alpha_0 + \alpha_0 \sum_{j=1}^{m} Z_t(j)}\right)\right|^{p_0}\right]\right)^{1/p_0}$$

$$\leq \left(\mathsf{E}[|e_t|^{2p_0}]\right)^{1/(2p_0)} O\left(\sum_{j=m+1}^{\infty} \left(\mathsf{E}|Z_t(j)|^{2p_0}\right)^{1/(2p_0)}\right)$$

$$= O\left(|\rho_*|^{m/(2p_0)}\right).$$

Hence, $\{X_t\}$ is NED with respect to the α–mixing process $\{e_t\}$ and the stable coefficient decays at a geometric rate. Analogously, $\{\sigma_t^2\}$ is also NED with stable coefficient decaying at a geometric rate. This result is similar to that obtained by Hansen (1991).

It is interesting in practice to estimate the functions $m_j(x) = E(X_t^2|X_{t-j} = x)$ for $j = 1, 2, \ldots$ as this gives some diagnostic about the shape of the conditional variance. It is not known what is the functional form of $m_j(x)$ and it may not be quadratic, see Tong (1990, p13). However, the functions m_j exist and are likely to be smooth under some conditions. As the stable coefficient decays at a geometric rate, it is easy to check that the condition A4 (i) is satisfied if A4 (ii) holds. Hence, by Propositions 4.1 and 4.2, we can obtain the uniform strong and weak convergence rate of local linear fitting for conditional variance functions $m_j(x)$'s in the GARCH(1,1) model (4.7).

4.5. Linear process

We next study the case that both $\{\mathbf{X}_t\}$ and $\{Y_t\}$ are generated by the linear processes

$$\mathbf{X}_t = \sum_{s=0}^{\infty} A_s \varepsilon_{t-s}, \quad Y_t = \sum_{s=0}^{\infty} B_s \varepsilon_{t-s}, \tag{4.11}$$

where $A_s = (a_{ij}(s))_{d \times q}$, $B_s = (b_1(s), \cdots, b_q(s))^\tau$, $\{\varepsilon_t := (\varepsilon_{t1}, \cdots, \varepsilon_{tq})^\tau\}$ is q–dimensional sequence of stationary α–mixing random vectors. If

$$\max\{\|A_t\|, \|B_t\|\} = O(t^{-\lambda}), \quad \lambda > 0,$$

where $\|A_t\| = \left(\sum_{i=1}^{d}\sum_{j=1}^{q} a_{ij}^2(t)\right)^{1/2}$, we can show that both $\{\mathbf{X}_t\}$ and $\{Y_t\}$ are NED with respect to $\{\varepsilon_t\}$ and the stable coefficient decays with a polynomial rate.

To derive the uniform convergence rate of the local linear fitting for the conditional regression function $g(\mathbf{x})$, we need to check whether A4 (i) holds under some mild conditions

on the bandwidth h. For the simple case of $d = 1$, we take the pointwise optimal bandwidth with $h = O(T^{-1/5})$. Then, if

$$\lambda > \frac{2p_0(2p_0 - 3)}{(p_0 - 2)(2p_0 - 5)} + 1, \quad p_0 > 5/2,$$

we can show that A4 (i) is satisfied.

4.6. Semiparametric ARCH(∞) model

As mentioned in Section 4.2, Linton and Mammen (2005) considered the model $X_t = \sigma_t \varepsilon_t$ with ε_t and $\varepsilon_t^2 - 1$ being martingale difference sequences and

$$\sigma_t^2 = \sum_{j=1}^{\infty} \psi_j(\theta) g(X_{t-j}),$$

where $g(\cdot)$ is an unknown function and $\{\psi_j(\theta), \ \theta \in \Theta\}_{j=1}^{\infty}$ satisfies some regularity conditions. The above model includes GARCH(1,1) as a special case. Linton and Mammen (2005) assumed that $\{X_t\}$ is stationary and geometrically mixing but they did not give primitive conditions. We next will show that $\{\sigma_t^2\}$ is NED with respect to $\{\varepsilon_t\}$ under some mild conditions on $g(\cdot)$ and the weights.

Suppose that there exists a Borel measurable function $\mathcal{F}$ such that X_t can be rewritten as $X_t = \mathcal{F}(\varepsilon_t, \varepsilon_{t-1}, \cdots)$ and $g(\cdot)$ is Lipschitz continuous. Let

$$\mathcal{M}_{t,s} = \sigma(\varepsilon_t, \varepsilon_{t-1}, \cdots, \varepsilon_s), \ t \geq s.$$

Furthermore, assume that $E|\epsilon_t|^{p_0} < \infty$, $p_0 \geq 1$, $0 < \inf_x g(x) < \sup_x g(x) < \infty$, $\sup_{\theta \in \Theta} \sum_{j=1}^{\infty} \left| \psi_j(\theta) \right| < \infty$, and

$$\mathsf{E} \left| X_t - \mathsf{E}(X_t | \mathcal{M}_{t,t-m}) \right|^{p_0} \to 0, \quad m \to \infty.$$

Define

$$\sigma_t^2(m) = \sum_{j=1}^{m} \psi_j(\theta) g\left(X_{t-j}^{(m)}\right), \quad X_t^{(m)} = \mathsf{E}(X_t | \mathcal{M}_{t,t-m}).$$

As $g(\cdot)$ is bounded and the weighs are summable, it is easy to check that

$$\left(\mathsf{E} \left| \sum_{j=\frac{m}{2}+1}^{\infty} \psi_j(\theta) g(X_{t-j}) - \sum_{j=\frac{m}{2}+1}^{\infty} \psi_j(\theta) g\left(X_{t-j}^{(m)}\right) \right|^{p_0} \right)^{1/p_0}$$

$$= O\left(\sum_{j=\frac{m}{2}+1}^{\infty} |\psi_j(\theta)| \right) = o(1) \quad m \to \infty.$$

17

On the other hand, by the Lipschitz continuity of $g(\cdot)$, we have

$$
\begin{aligned}
& \mathsf{E}\left|g(X_{t-j}) - g\left(X_{t-j}^{(m)}\right)\right|^{p_0} \\
= \quad & O\left(\mathsf{E}\left|X_{t-j} - X_{t-j}^{(m)}\right|^{p_0}\right) \\
= \quad & O\left(\mathsf{E}\left|X_{t-j} - \mathsf{E}\left(X_{t-j}|\mathcal{M}_{t-j,\ t-j-(m-j)}\right)\right|^{p_0}\right) \\
= \quad & o(1), \qquad m - j \to 0.
\end{aligned}
$$

Hence, we can show that $\{\sigma_t^2\}$ is NED with respect to $\{\varepsilon_t\}$.

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.

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

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

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

Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction*, 2nd Edition. Lecture Notes in Statistics 110, Springer–Verlag.

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

Chen, L., Cheng, M. and Peng, L. (2009). Conditional variance estimation in heteroscedastic regression models. *Journal of Statistical Planning and Inference* **139**, 236-245.

Chen, X., Linton, O. and Van Keilegom, I. (2003). Estimation of semiparametric models when the criterion is not smooth. *Econometrica* **71**, 1591-1608.

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

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

Drost, F. C. and Nijman, T. E. (1993). Temporal aggregation of GARCH processes. *Econometrica* **61**, 909-927.

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

Fan, J. and Yao, Q. (1998). Efficient estimation of conditional variance functions in stochastic regression. *Biometrika* **85**, 645–660.

Fan, J. and Yao, Q. (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer, New York.

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

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

Hansen, B. E. (1991). GARCH (1,1) processes are near epoch dependent. *Economics Letters* **36**, 181–186.

Hansen, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory* **24**, 726–748.

Hong, Y. (2000). Generalized spectral tests for serial dependence. Journal of the Royal Statistical Society: Series B (Methodological) 62, 3, pp 557-574

Ibragimov, I. A. (1962). Some limit theorems for stationary processes. *Theory of Probability and its Applications* **7**, 349–382.

Kristensen, D. (2009). Uniform convergence rates of kernel estimators with heterogenous dependent data. *Econometric Theory* **25**, 1433–1445.

Lee, S. W. and Hansen, B. E. (1994). Asymptotic theory for the GARCH(1,1) quasi–maximum likelihood estimator. *Econometric Theory* **10**, 29-52.

Li, Q. and Racine, J.S. (2007). *Nonparametric econometrics: Theory and Practice.* Princeton University Press, Princeton.

Liebscher, E. (1996). Strong convergence of sums of α–mixing random variables with applications to density estimation. *Stochastic Processes and Their Applications* **65**, 69–80.

Lin, Z. (2004). Strong near epoch dependence. *Science in China* (Series A) **47**, 497–507.

Ling, S. (2007). Testing for change points in time series models and limiting theorems for NED sequences. *Annals of Statistics* **35**, 1213–1237.

Linton, O. B. and Mammen, E. (2005). Estimating semiparametric ARCH(∞) models by kernel smoothing methods. *Econometrica* **73**, 771–836.

Linton, O. B. and A. Sancetta (2009). Consistent Estimation of A General Nonparametric Regression Function in Time Series. Journal of Econometrics 152, 70-78.

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. and Linton, O. (2007). Local linear fitting under near epoch dependence. *Econometric Theory* **23**, 37–70.

Masry, E. (1996). Multivariate local polynomial regression for time series: uniform strong consistency and rates. *Journal of Time Series Analysis* **17**, 571–599.

Masry, E. and Tjøstheim, D. (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 Wahrscheinlichskeit-theorie und verwandte Gebiete* **32**, 165–178.

McLeish, D. L. (1977). On the invariance principle for nonstationary mixingales. *Annals of Probability* **5**, 616–621.

Nze, P.A., Bühlamann, P. and Doukhan, P. (2002). Weak dependence beyond mixing and asymptotics for nonparametric regression. *Annals of Statistics*, **30**, 397–430.

Nze, P.A., and Doukhan, P. (2004). Weak dependence: Models and applications to econometrics. *Econometric Theory*, **20**, 995–1045.

Pagan, A. R. and Schwert, G. W. (1990). Alternative models for conditional stock volatility. *Journal of Econometrics* **45**, 267–290.

Pagan, A. R. and Hong, Y. S. (1991). Nonparametric Estimation and the Risk Premium. in W. Barnett, J. Powell, and G.E. Tauchen (eds.) *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, Cambridge University Press.

Shao, Q. and Yu, H. (1996). Weak convergence for weighted empirical processes of dependent sequences. *Annals of Probability* **24**, 2098–2127.

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, Oxford.

Yu, K. and Jones, M. C. (2004). Likelihood–based local linear estimation of the conditional variance function. *Journal of the American Statistical Association* **99**, 139-144.

Appendix A: Some useful lemmas

We next provide some critical lemmas, which are necessary for the proofs of the main results. The first one is the Bernstein inequality for α–mixing process, which can be found in several books such as Fan and Yao (2003).

LEMMA A.1 *Let $\{Z_t\}$ be a zero–mean real–valued α–mixing process satisfying*

$$\mathsf{P}(|Z_t| \leq B) = 1$$

for all $t \geq 1$. Then for each integer $q \in [1, \frac{T}{2}]$ and each $\epsilon > 0$, we have

$$\mathsf{P}\left(\left|\sum_{t=1}^{T} Z_t\right| > T\epsilon\right) \leq 4\exp\left(-\frac{\epsilon^2 q}{8v^2(q)}\right) + 22\left(1 + \frac{4B}{\epsilon}\right)^{1/2} q\alpha\left(\left[\frac{T}{2q}\right]\right), \quad \text{(A.1)}$$

where $v^2(q) = 2\sigma^2(q)/p^2 + B\epsilon/2$ with $p = \frac{T}{2q}$ and

$$\sigma^2(q) = \max_{1 \leq j \leq 2q-1} \mathsf{E}\ \left(([jp] + 1 - jp)Z_{[jp]+1} + Z_{[jp]+2} + \cdots + Z_{[(j+1)p]}\right.$$
$$\left. +((j+1)p - [(j+1)p])Z_{[(j+1)p]+1}\right)^2.$$

Let $Y_t^{(m)}$ be defined as in Definition 1. We establish the result on the moment of $\Psi(Y_t^{(m)})$ in the following lemma.

LEMMA A.2 *Suppose that $\mathsf{E}|Y_t|^s < \infty$, $s \geq 1$, the function $\Psi(\cdot)$ has continuous derivatives up to the q_0–th order and*

$$\mathsf{E}\left|\Psi^{(k)}(Y_t)\right|^s < \infty, \ \ 1 < k < q_0, \ \ \sup_{y\in\mathbb{R}}\left|\Psi^{(q_0)}(y)\right| < \infty.$$

Then, if $\{Y_t\}$ is NED in L_{ν^*} ($\nu^* = (q_0 - 1)s/(s - r) \vee q_0 r$), for each $m \geq 1$,

$$\mathsf{E}\left|\Psi(Y_t^{(m)})\right|^r < \infty, \ \ 1 \leq r < s. \quad \text{(A.2)}$$

PROOF. Note that $\Psi(Y_t^{(m)}) = \Psi(Y_t) + \Psi(Y_t^{(m)}) - \Psi(Y_t)$. By applying the C_r–inequality, it is easy to check that for $r \geq 1$,

$$\mathsf{E}\left|\Psi(Y_t^{(m)})\right|^r \leq 2^{r-1}\left(\mathsf{E}\left|\Psi(Y_t)\right|^r + \mathsf{E}\left|\Psi(Y_t^{(m)}) - \Psi(Y_t)\right|^r\right). \quad \text{(A.3)}$$

22

By Taylor expansion,

$$\Psi(Y_t^{(m)}) = \Psi(Y_t) + \sum_{k=1}^{q_0-1} \Psi^{(k)}(Y_t)(Y_t^{(m)} - Y_t)^k/k! + \Psi^{(q_0)}(Y_t^*)(Y_t^{(m)} - Y_t)^{q_0}/q_0!, \quad \text{(A.4)}$$

where $Y_t^* = Y_t + \theta^*(Y_t^{(m)} - Y_t)$, $0 < \theta^* < 1$. By the C_r–inequality and the Cauchy–Schwarz inequality, we have

$$\mathsf{E}\left|\Psi(Y_t^{(m)}) - \Psi(Y_t)\right|^r \le C_0 \left(\sum_{k=1}^{q_0-1} v_{ks/(s-r)}^{(s-r)/s}(m) + v_{q_0 r}(m)\right) < \infty, \quad \text{(A.5)}$$

where C_0 is some positive constant. Then, by (A.3), (A.5) and $\mathsf{E}|Y_t|^r < \infty$, we can show that (A.2) holds. ∎

Define

$$W_T^{(m)}(\mathbf{x}) = \frac{1}{Th^d} \sum_{t=1}^{T} \Psi(Y_t^{(m)}) K_T\left(\frac{\mathbf{X}_t^{(m)} - \mathbf{x}}{h}\right). \quad \text{(A.6)}$$

The next lemma shows that $W_T(\mathbf{x})$ can be approximated by $W_T^{(m)}(\mathbf{x})$ in probability as $m \to \infty$, which is critical for uniform weak convergence rate of $W_T(\mathbf{x})$.

LEMMA A.3 *Suppose that the conditions A5–A6 are satisfied. Then, we have*

$$\sup_{\|\mathbf{x}\| \le C_T} \left|W_T(\mathbf{x}) - W_T^{(m)}(\mathbf{x})\right| = O_P\left(h^{-d-1} v_1(m) + h^{-d}\left(\sum_{k=1}^{q_0-1} v_{k\delta_p}^{1/\delta_p}(m) + v_{q_0}(m)\right)\right), \quad \text{(A.7)}$$

where $\delta_p = p_0/(p_0 - 1)$.

PROOF. Observe that

$$
\begin{aligned}
W_T(\mathbf{x}) - W_T^{(m)}(\mathbf{x}) &= \frac{1}{Th^d} \sum_{t=1}^{T} \left(\Psi(Y_t) K_T\left(\frac{\mathbf{X}_t - \mathbf{x}}{h}\right) - \Psi(Y_t^{(m)}) K_T\left(\frac{\mathbf{X}_t^{(m)} - \mathbf{x}}{h}\right)\right) \\
&= \frac{1}{Th^d} \sum_{t=1}^{T} \left(\Psi(Y_t) - \Psi(Y_t^{(m)})\right) K_T\left(\frac{\mathbf{X}_t - \mathbf{x}_t}{h}\right) \\
&\quad + \frac{1}{Th^d} \sum_{t=1}^{T} \Psi(Y_t^{(m)}) \left(K_T\left(\frac{\mathbf{X}_t - \mathbf{x}}{h}\right) - K_T\left(\frac{\mathbf{X}_t^{(m)} - \mathbf{x}}{h}\right)\right) \\
&=: I_{T,1}(\mathbf{x}) + I_{T,2}(\mathbf{x}).
\end{aligned}
\quad \text{(A.8)}
$$

We first consider $I_{T,1}(\mathbf{x})$. By Taylor expansion (A.4) and similar to the proof of (A.5),

23

we have

$$
\begin{aligned}
\mathsf{E}\left|\Psi(Y_t^{(m)}) - \Psi(Y_t)\right| &\leq C\left(\sum_{k=1}^{q_0-1} \mathsf{E}\left|\Psi^{(k)}(Y_t)(Y_t^{(m)} - Y_t)^k\right|/k! + \mathsf{E}\left|\Psi^{(q_0)}(Y_t^*)(Y_t^{(m)} - Y_t)^{q_0}\right|/q_0!\right) \\
&\leq C\left(\sum_{k=1}^{q_0-1}\left(\mathsf{E}\left|\Psi^{(k)}(Y_t)\right|^{p_0}\right)^{1/p_0} v_{p_0 k/(p_0-1)}^{(p_0-1)/p_0}(m) + v_{q_0}(m)\right) \\
&= O\left(\sum_{k=1}^{q_0-1} v_{\delta_p k}^{1/\delta_p}(m) + v_{q_0}(m)\right).
\end{aligned}
$$

(A.9)

By the boundness condition on $K_T(\cdot)$ (see A5 in Section 3) and (A.9), we have

$$
\begin{aligned}
\sup_{\|\mathbf{x}\|\leq C_T}|I_{T,1}(\mathbf{x})| &\leq h^{-d}\sup_{\|\mathbf{x}\|\leq C_T}\left|K_T(\tfrac{\mathbf{X}_t-\mathbf{x}}{h})\right|\left|\Psi(Y_t^{(m)}) - \Psi(Y_t)\right| \\
&= O_P\left(h^{-d}\left(\sum_{k=1}^{q_0-1} v_{\delta_p k}^{1/\delta_p}(m) + v_{q_0}(m)\right)\right).
\end{aligned}
$$

(A.10)

For $I_{T,2}(\mathbf{x})$, note that

$$
\begin{aligned}
I_{T,2}(\mathbf{x}) &= \frac{1}{Th^d}\sum_{t=1}^{T}\Psi(Y_t)\left(K_T\left(\frac{\mathbf{X}_t-\mathbf{x}}{h}\right) - K_T\left(\frac{\mathbf{X}_t^{(m)}-\mathbf{x}}{h}\right)\right) \\
&\quad + \frac{1}{Th^d}\sum_{t=1}^{T}\left(\Psi(Y_t^{(m)}) - \Psi(Y_t)\right)\left(K_T\left(\frac{\mathbf{X}_t-\mathbf{x}}{h}\right) - K_T\left(\frac{\mathbf{X}_t^{(m)}-\mathbf{x}}{h}\right)\right) \\
&=: I_{T,3}(\mathbf{x}) + I_{T,4}(\mathbf{x}).
\end{aligned}
$$

(A.11)

By the Lipschitz continuity of $K_T(\cdot)$, we have

$$
\left|K_T\left(\frac{\mathbf{X}_t-\mathbf{x}}{h}\right) - K_T\left(\frac{\mathbf{X}_t^{(m)}-\mathbf{x}}{h}\right)\right| = O_P(v_1(m)/h).
$$

(A.12)

By (A.12), we have

$$
\sup_{\|\mathbf{x}\|\leq C_T}|I_{T,3}(\mathbf{x})| = O_P\left(h^{-d-1}v_1(m)\right).
$$

(A.13)

On the other hand, by (A.9) and (A.12), we have

$$
\begin{aligned}
\sup_{\|\mathbf{x}\|\leq C_T}|I_{T,4}(\mathbf{x})| &= O_P\left(h^{-d-1}v_1(m)\left(\sum_{k=1}^{q_0-1} v_{\delta_p k}^{1/\delta_p}(m) + v_{q_0}(m)\right)\right) \\
&= O_P\left(h^{-d-1}v_1(m)\right).
\end{aligned}
$$

(A.14)

In view of (A.8), (A.10), (A.11), (A.13) and (A.14), we can show that (A.7) holds. ∎

LEMMA A.4 *Suppose that the conditions of Theorem 3.2 are satisfied. Then, we have*

$$
\sup_{\|\mathbf{x}\|\leq C_T}\left|W_T(\mathbf{x}) - W_T^{(m_T)}(\mathbf{x})\right| = O\left(\rho_T\right), \quad a.s.,
$$

(A.15)

where m_T is defined in A4 (i).

PROOF. Let $I_{T,1}(\mathbf{x})$ and $I_{T,2}(\mathbf{x})$ be defined as in (A.8). By (3.3) and (A.9), we have

$$
\begin{aligned}
& \sum_{T=1}^{\infty} \mathsf{P}\left(\left|\Psi(Y_t^{(m)}) - \Psi(Y_t)\right| > \rho_T h^d\right) \\
\leq\ & \sum_{T=1}^{\infty} \rho_T^{-1} h^{-d} \mathsf{E}\left|\Psi(Y_t^{(m)}) - \Psi(Y_t)\right| \\
\leq\ & C \sum_{T=1}^{\infty} \rho_T^{-1} h^{-d}\left(\sum_{k=1}^{q_0-1} v_{\delta_p k}^{1/\delta_p}(m) + v_{q_0}(m)\right) \\
=\ & C \sum_{T=1}^{\infty} \frac{1}{T \log^{1+\varsigma} T} < \infty.
\end{aligned}
\tag{A.16}
$$

By the boundness condition on $K_T(\cdot)$ and (A.16), we have

$$
\begin{aligned}
\sup_{\|\mathbf{x}\| \leq C_T} |I_{T,1}(\mathbf{x})| & \leq h^{-d} \sup_{\|\mathbf{x}\| \leq C_T}\left|K_T(\tfrac{\mathbf{X}_t - \mathbf{x}}{h})\right|\left|\Psi(Y_t^{(m)}) - \Psi(Y_t)\right| \\
& = O(\rho_T)\ \ a.s.
\end{aligned}
\tag{A.17}
$$

Analogously, we can show that

$$
\sup_{\|\mathbf{x}\| \leq C_T} |I_{T,2}(\mathbf{x})| = O(\rho_T)\ \ a.s.
\tag{A.18}
$$

In view of (A.17) and (A.18), we can show that (A.15) holds. $\blacksquare$

Appendix B: Proofs of the main results

We first prove Theorems 3.1 and 3.2 and then provide the proofs of the uniform convergence rate results in Sections 2 and 4. In fact, the results in Sections 2 and 4 can be obtained as applications of Theorems 3.1 and 3.2.

PROOF OF THEOREM 3.1. Note that

$$
\begin{aligned}
\sup_{\|\mathbf{x}\| \leq C_T} |W_T(\mathbf{x}) - \mathsf{E}[W_T(\mathbf{x})]| & \leq \sup_{\|\mathbf{x}\| \leq C_T}\left|W_T^{(m)}(\mathbf{x}) - \mathsf{E}[W_T^{(m)}(\mathbf{x})]\right| \\
& \quad + \sup_{\|\mathbf{x}\| \leq C_T}\left|W_T(\mathbf{x}) - W_T^{(m)}(\mathbf{x})\right| \\
& \quad + \sup_{\|\mathbf{x}\| \leq C_T}\left|\mathsf{E}[W_T(\mathbf{x})] - \mathsf{E}[W_T^{(m)}(\mathbf{x})]\right| \\
& =: \Pi_{T,1} + \Pi_{T,2} + \Pi_{T,3}.
\end{aligned}
\tag{B.1}
$$

25

By Lemma A.3, we have

$$\Pi_{T,2} = O_P\left(h^{-d-1}v_1(m) + h^{-d}\left(\sum_{k=1}^{q_0-1} v_{k\delta_p}^{1/\delta_p}(m_T) + v_{q_0}(m_T)\right)\right)$$

by taking $m = m_T$, where m_T is defined in A4 (i). Furthermore, by A6, we have

$$\Pi_{T,2} = O_P(\rho_T). \tag{B.2}$$

Analogously,

$$\Pi_{T,3} = O(\rho_T). \tag{B.3}$$

By (B.1)–(B.3), to prove (3.2), we need only to show that

$$\Pi_{T,1} = O_P(\rho_T). \tag{B.4}$$

It is easy to check that $\{Y_t^{(m)}, \mathbf{X}_t^{(m)}\}$ is an α–mixing process with mixing coefficient

$$\alpha_m(t) \leq \begin{cases} \alpha_{t-m}, & t \geq m+1; \\ 1, & t \leq m. \end{cases}$$

We next cover the set $\{\mathbf{x} : \|\mathbf{x}\| \leq C_T\}$ by a finite number of subsets S_k, $k = 1, \cdots, N_T$, which are centered at s_k with radius $r_T = \rho_T h^{d+1}$. Observe that

$$\begin{aligned} \Pi_{T,1} &\leq \max_{1 \leq k \leq N_T} \sup_{\|\mathbf{x}\| \in S_k} \left|W_T^{(m)}(\mathbf{x}) - W_T^{(m)}(s_k)\right| \\ &\quad + \max_{1 \leq k \leq N_T} \sup_{\|\mathbf{x}\| \in S_k} \left|\mathsf{E}[W_T^{(m)}(\mathbf{x})] - \mathsf{E}[W_T^{(m)}(s_k)]\right| \\ &\quad + \max_{1 \leq k \leq N_T} \left|W_T^{(m)}(s_k) - \mathsf{E}[W_T^{(m)}(s_k)]\right| \\ &=: \Pi_{T,4} + \Pi_{T,5} + \Pi_{T,6}. \end{aligned} \tag{B.5}$$

By the Lipschitz continuity of $K_T(\cdot)$ in A5, we have

$$\max_{T \geq 1} \max_{1 \leq k \leq N_T} \sup_{\mathbf{x} \in S_k} \left|K_T\left(\frac{\mathbf{X}_t - \mathbf{x}}{h}\right) - K_T\left(\frac{\mathbf{X}_t - s_k}{h}\right)\right| \leq \max_{1 \leq k \leq N_T} \sup_{\mathbf{x} \in S_k} \left\|\frac{\mathbf{x} - s_k}{h}\right\| = O\left(\frac{r_T}{h}\right). \tag{B.6}$$

By (B.6) and noting that $\mathsf{E}\left|\Psi(Y_t^{(m)})\right| < \infty$ by Lemma A.2, we have

$$\Pi_{T,4} = O_P\left(\frac{r_T}{h^{d+1}}\right) = O_P(\rho_T). \tag{B.7}$$

26

Similarly,

$$\Pi_{T,5} = O(\rho_T). \tag{B.8}$$

By (B.5), (B.7) and (B.8), to prove (B.4), we need only to show that

$$\Pi_{T,6} = O_P(\rho_T). \tag{B.9}$$

Let $\Delta_T = T^{1/p_0}$,

$$\overline{\Psi}(Y_t^{(m)}) = \Psi(Y_t^{(m)})I\left(|\Psi(Y_t^{(m)})| \le \Delta_T\right), \quad \widetilde{\Psi}(Y_t^{(m)}) = \Psi(Y_t^{(m)})I\left(|\Psi(Y_t^{(m)})| > \Delta_T\right),$$

$$\overline{W}_T^{(m)}(\mathbf{x}) = \frac{1}{Th^d}\sum_{t=1}^{T} \overline{\Psi}(Y_t^{(m)})K_T\left(\frac{\mathbf{X}_t^{(m)} - \mathbf{x}}{h}\right),$$

$$\widetilde{W}_T^{(m)}(\mathbf{x}) = \frac{1}{Th^d}\sum_{t=1}^{T} \widetilde{\Psi}(Y_t^{(m)})K_T\left(\frac{\mathbf{X}_t^{(m)} - \mathbf{x}}{h}\right).$$

It is easy to check that

$$\begin{aligned} \Pi_{T,6} &\le \max_{1\le k\le N_T} \left|\overline{W}_T^{(m)}(s_k) - \mathsf{E}[\overline{W}_T^{(m)}(s_k)]\right| \\ &\quad + \max_{1\le k\le N_T} \left|\widetilde{W}_T^{(m)}(s_k) - \mathsf{E}[\widetilde{W}_T^{(m)}(s_k)]\right| \\ &=: \Pi_{T,7} + \Pi_{T,8}. \end{aligned} \tag{B.10}$$

By the Markov inequality and Lemma A.2, for any $\eta > 0$,

$$\begin{aligned} \mathsf{P}\left(\Pi_{T,8} > \eta\rho_T\right) &\le \sum_{t=1}^{T} \frac{\mathsf{E}\left|\Psi(Y_t^{(m)})\right|^{\lambda_0}}{\Delta_T^{\lambda_0}} \\ &\le CT\Delta_T^{-\lambda_0} = O(T^{1-\lambda_0/p_0}) = o(1), \end{aligned}$$

where $p_0 < \lambda_0 < s$. Hence, we have

$$\Pi_{T,8} = O_P(\rho_T). \tag{B.11}$$

Letting

$$B = \Delta_T h^{-d} = T^{1/p_0}h^{-d}, \quad \epsilon = \eta\rho_T, \quad q = T^{1+1/p_0}\rho_T$$

in Lemma A.1, we have

$$\mathsf{P}\left(\Pi_{T,7} > \eta\rho_T\right) \le \sum_{k=1}^{N_T} \mathsf{P}\left(\left|\overline{W}_T^{(m)}(s_k) - \mathsf{E}[\overline{W}_T^{(m)}(s_k)]\right| > \eta\rho_T\right)$$

$$\le N_T\exp\left\{-\frac{c\eta^2\rho_T^2 Th^d}{16}\right\} + cN_T(\log T)^{(2\theta_0+1)/4}h^{-(3+2\theta_0)d/4}T^{(3p_0+6)/4p_0+\theta_0(1/p_0-1/2)}$$

27

for some positive constant c. Noting that

$$N_T = O\left(\frac{C_T^d}{\rho_T^d h^{d^2+d}}\right),$$

by the bandwidth condition in A4 (ii), we have for η large enough

$$\mathsf{P}\left(\Pi_{T,7} > \eta\rho_T\right) = o(1),$$

which implies that

$$\Pi_{T,7} = O_P(\rho_T). \tag{B.12}$$

By (B.10)–(B.12), we can show that (B.9) holds. Then, the proof of Theorem 3.1 is completed. $\blacksquare$

PROOF OF THEOREM 3.2. By Lemma A.4 and following the proof of Theorem 3.1, we need only to show that

$$\Pi_{T,6} = O(\rho_T) \quad a.s., \tag{B.13}$$

where $\Pi_{T,6}$ is defined in (B.5)

Let $\Delta_T = T^{1/p_0}$, $\overline{\Psi}(Y_t^{(m)})$, $\widetilde{\Psi}(Y_t^{(m)})$, $\overline{W}_T^{(m)}(\mathbf{x})$, $\widetilde{W}_T^{(m)}(\mathbf{x})$, $\Pi_{T,7}$ and $\Pi_{T,8}$ be defined as in the proof of Theorem 3.1. By the Markov inequality and Lemma A.2, for any $\eta > 0$,

$$\begin{aligned}
\sum_{T=1}^{\infty} \mathsf{P}\left(\Pi_{T,8} > \eta\rho_T\right) &\leq \sum_{T=1}^{\infty}\sum_{t=1}^{T} \frac{\mathsf{E}\left|\Psi(Y_t^{(m)})\right|^s}{\Delta_T^s} \\
&\leq C\sum_{T=1}^{\infty} T^{1-s/p_0} < \infty,
\end{aligned}$$

as $s > 2p_0$. Hence, we have

$$\Pi_{T,8} = O(\rho_T) \quad a.s. \tag{B.14}$$

Letting $B = \Delta_T h^{-d} = T^{1/p_0} h^{-d}$, $\epsilon = \eta\rho_T$, $q = T^{1+1/p_0}\rho_T$ in Lemma A.1, by (2.5) and (2.10) we have

$$\begin{aligned}
\sum_{T=1}^{\infty} \mathsf{P}\left(\Pi_{T,7} > \eta\rho_T\right) &\leq \sum_{T=1}^{\infty}\sum_{k=1}^{N_T} \mathsf{P}\left(\left|\overline{W}_T^{(m)}(s_k) - \mathsf{E}[\overline{W}_T^{(m)}(s_k)]\right| > \eta\rho_T\right) \\
&\leq \sum_{T=1}^{\infty} N_T\left(\exp\left\{-\frac{\eta^2\rho_T^2 T h^d}{16}\right\} + (\log T)^{(2\theta_0+1)/4} h^{-(3+2\theta_0)d/4} T^{(3p_0+6)/4p_0+\theta_0(1/p_0-1/2)}\right) \\
&\leq C\sum_{T=1}^{\infty} \frac{1}{T\log^{1+\varsigma} T} < \infty.
\end{aligned}$$

28

Hence

$$\Pi_{T,7} = O(\rho_T) \quad a.s. \tag{B.15}$$

By (B.10), (B.14) and (B.15), we can show that (B.13) holds. Then, the proof of Theorem 3.2 is completed. ∎

PROOF OF THEOREM 2.1. We only consider the case of $d = 1$ as the extension to the case of $d \geq 2$ is similar. Then $\mathbf{X}_t$ and $\mathbf{x}$ become X_t and x, respectively. By the standard argument of local linear estimator as in Fan and Gijbels (1996).

$$\widehat{g}(x) = \sum_{t=1}^{T} w_{T,t}(x) Y_t,$$

where

$$
\begin{aligned}
w_{T,t}(x) &= \widetilde{K}\left(\frac{X_t - x}{h}\right) \Big/ \sum_{t=1}^{T} \widetilde{K}\left(\frac{X_t - x}{h}\right), \\
\widetilde{K}\left(\frac{X_t - x}{h}\right) &= K\left(\frac{X_t - x}{h}\right)\left(S_{T,2}(x) - \left(\frac{X_t - x}{h}\right) S_{T,1}(x)\right), \\
S_{T,j}(x) &= \frac{1}{Th} \sum_{t=1}^{T} \left(\frac{X_t - x}{h}\right)^j K\left(\frac{X_t - x}{h}\right).
\end{aligned}
$$

Then,

$$
\begin{aligned}
\widehat{g}(x) - g(x) &= \left(\sum_{t=1}^{T} w_{T,t}(x) g(X_t) - g(x) \right) + \sum_{t=1}^{T} w_{T,t}(x) e_t \\
&=: \Pi_{T,1}^*(x) + \Pi_{T,2}^*(x),
\end{aligned}
\tag{B.16}
$$

where $e_t = Y_t - g(X_t)$.

Applying Theorem 3.1 with $\Psi(y) \equiv 1$, for any $j \geq 1$,

$$\sup_{|x| \leq C_T} \left| S_{T,j}(x) - \mu_j f(x) \right| = o_P(1), \tag{B.17}$$

where $\mu_j = \int_{\mathbb{R}} u^j K(u) du$. By (B.17) and standard calculation, we have

$$\sup_{|x| \leq C_T} \left| \Pi_{T,1}^*(x) \right| = O_P\left(\frac{b_T(g) h^2}{a_T(f)} \right). \tag{B.18}$$

Hence, to prove (2.6), we need only to show that

$$\sup_{|x| \leq C_T} \left| \Pi_{T,2}^*(x) \right| = O_P\left(\frac{\rho_T}{a_T(f)} \right). \tag{B.19}$$

29

By (B.17) and the definition of $w_{T,t}(\cdot)$, to prove (B.19), we need only to show that

$$\sup_{|x|\leq C_T}\left|\frac{1}{Th}\sum_{t=1}^{T}K\left(\frac{X_t-x}{h}\right)e_t\right|=O_P\left(\rho_T\right). \tag{B.20}$$

By Theorem 3.1, we can show that (B.20) holds. Then, the proof of Theorem 2.1 is completed. ∎

PROOF OF THEOREM 2.2. Following the proofs of Theorems 2.1 and 3.2, we can show that (2.11) holds. The details are omitted here. ∎

PROOF OF PROPOSITION 4.1. As in the proof of Theorem 3.1, we only consider the case of $d=1$. Noting that $\widehat{e}_t=Y_t-g(X_t)+g(X_t)-\widehat{g}(X_t)=e_t+g(X_t)-\widehat{g}(X_t)$, $\widehat{\sigma}(x)$ can be rewritten as

$$
\begin{aligned}
\widehat{\sigma}^2(x) &= \sum_{t=1}^{T}w_{T,t}^*(x)\widehat{e}_t^2, \\
&= \sum_{t=1}^{T}w_{T,t}^*(x)e_t^2+2\sum_{t=1}^{T}w_{T,t}^*(x)e_t(g(X_t)-\widehat{g}(X_t))+\sum_{t=1}^{T}w_{T,t}^*(x)(g(X_t)-\widehat{g}(X_t))^2, \\
&=: \Xi_{T,1}(x)+\Xi_{T,2}(x)+\Xi_{T,3}(x),
\end{aligned}
$$

where

$$
\begin{aligned}
w_{T,t}^*(x) &= \widetilde{W}\left(\frac{X_t-x}{h_*}\right)\Big/\sum_{t=1}^{T}\widetilde{W}\left(\frac{X_t-x}{h_*}\right), \\
\widetilde{W}\left(\frac{X_t-x}{h_*}\right) &= W\left(\frac{X_t-x}{h}\right)\left(S_{T,2}^*(x)-\left(\frac{X_t-x}{h_*}\right)S_{T,1}^*(x)\right), \\
S_{T,j}^*(x) &= \frac{1}{Th}\sum_{t=1}^{T}\left(\frac{X_t-x}{h_*}\right)^j W\left(\frac{X_t-x}{h_*}\right).
\end{aligned}
$$

By Theorem 3.1 and following the proof of Theorem 2.1, we have

$$\sup_{|x|\leq C_T}\left|\Xi_{T,1}(x)-\sigma^2(x)\right|=O_P\left(\frac{\rho_T}{a_T(f)}+\frac{b_T(\sigma^2)h^2}{a_T(f)}\right). \tag{B.21}$$

On the other hand, observe that

$$
\begin{aligned}
\Xi_{T,3}(x) &= \sum_{t=1}^{T}w_{T,t}^*(x)(g(X_t)-\widehat{g}(X_t))^2 I\left(\max_{1\leq t\leq T}|X_T|\leq C_T\right) \\
&\quad + \sum_{t=1}^{T}w_{T,t}^*(x)(g(X_t)-\widehat{g}(X_t))^2 I\left(\max_{1\leq t\leq T}|X_T|>C_T\right) \\
&=: \Xi_{T,4}(x)+\Xi_{T,5}(x).
\end{aligned}
$$

30

By Theorem 2.1, we have

$$\sup_{|x| \le C_T} |\Xi_{T,4}(x)| = O_P\left(\left(\frac{\rho_T}{a_T(f)} + \frac{b_T(g)h^2}{a_T(f)}\right)^2\right). \tag{B.22}$$

For $\Xi_{T,5}(x)$, we have for any small $\eta > 0$

$$
\begin{aligned}
\mathsf{P}\left(\sup_{|x| \le C_T} |\Xi_{T,5}(x)| > \eta\left(\frac{\rho_T}{a_T(f)} + \frac{b_T(g)h^2}{a_T(f)}\right)^2\right) &= \mathsf{P}\left(\max_{1 \le t \le T} |X_T| > C_T\right) \\
&\le \sum_{t=1}^{T} \frac{\mathsf{E}|X_t|^{2p_0}}{C_T^{2p_0}} \\
&= O(T^{1-2p_0/d}) \\
&= o(1)
\end{aligned}
$$

as $2p_0 > d$. Hence, we have

$$\sup_{|x| \le C_T} |\Xi_{T,5}(x)| = O_P\left(\left(\frac{\rho_T}{a_T(f)} + \frac{b_T(g)h^2}{a_T(f)}\right)^2\right). \tag{B.23}$$

By (B.22) and (B.23), we have

$$\sup_{|x| \le C_T} |\Xi_{T,3}(x)| = O_P\left(\left(\frac{\rho_T}{a_T(f)} + \frac{b_T(g)h^2}{a_T(f)}\right)^2\right) = o_P\left(\frac{\rho_T}{a_T(f)} + \frac{b_T(\sigma^2)h^2}{a_T(f)}\right). \tag{B.24}$$

By the Cauchy–Schwarz inequality, we have

$$\Xi_{T,2}(x) = 2\left(\sum_{t=1}^{T} w_{T,t}^*(x)e_t^2\right)^{1/2}\left(\sum_{t=1}^{T} w_{T,t}^*(x)(g(X_t) - \widehat{g}(X_t))^2\right)^{1/2}.$$

Then, by (B.21) and (B.24), we have

$$\sup_{|x| \le C_T} |\Xi_{T,2}(x)| = O_P\left(\frac{\rho_T}{a_T(f)} + \frac{b_T(\sigma^2)h^2}{a_T(f)}\right). \tag{B.25}$$

By (B.21), (B.24) and (B.25), we can show that (4.2) holds. ∎

PROOF OF PROPOSITION 4.2. From the proof of Proposition 4.1, and by Theorems 2.2 and 3.2, we can show that (4.3) holds. Details are omitted here. ∎

PROOF OF PROPOSITION 4.3. The detailed proof is similar to the proof of Theorem 2.1. By the definition of the local linear estimators $\widehat{m}_j(x)$, $j = 1, \cdots, J(T)$, we have

$$\widehat{m}_j(x) = \sum_{t=j+1}^{T} w_{T,j,t}(x) Y_t,$$

where

$$
\begin{aligned}
w_{T,j,t}(x) &= \widetilde{K}\left(\frac{X_{t-j} - x}{h}\right) \bigg/ \sum_{t=j+1}^{T} \widetilde{K}\left(\frac{X_{t-j} - x}{h}\right), \\
\widetilde{K}\left(\frac{X_{t-j} - x}{h}\right) &= K\left(\frac{X_{t-j} - x}{h}\right)\left(S_{T,j,2}(x) - \left(\frac{X_{t-j} - x}{h}\right) S_{T,j,1}(x)\right), \\
S_{T,j,k}(x) &= \frac{1}{(T-j)h} \sum_{t=j+1}^{T} \left(\frac{X_{t-j} - x}{h}\right)^k K\left(\frac{X_{t-j} - x}{h}\right).
\end{aligned}
$$

Then,

$$
\begin{aligned}
\widehat{m}_j(x) - m(x) &= \left(\sum_{t=j+1}^{T} w_{T,j,t}(x) m_j(X_t) - m_j(x)\right) + \sum_{t=j+1}^{T} w_{T,j,t}(x) \widetilde{e}_{t,j} \\
&=: \Pi_{T,j,1}(x) + \Pi_{T,j,2}(x),
\end{aligned}
\tag{B.26}
$$

where $\widetilde{e}_{t,j} = Y_t - m_j(X_t)$.

Following the proof of Theorem 3.1 with some modification, we can show that

$$\max_{1 \leq j \leq J(T)} \sup_{|x| \leq C_T} |S_{T,j,k}(x) - \mu_k f(x)| = o_P(1), \quad k \geq 1. \tag{B.27}$$

By (B.27), to prove

$$\max_{1 \leq j \leq J(T)} \sup_{|x| \leq C_T} \Pi_{T,j,2}(x) = O_P\left(\frac{\rho_T}{a_T(f)}\right), \tag{B.28}$$

we need only to show

$$\max_{1 \leq j \leq J(T)} \sup_{|x| \leq C_T} \left|\frac{1}{(T-j)h} \sum_{t=j+1}^{T} K\left(\frac{X_{t-j} - x}{h}\right) \widetilde{e}_{t,j}\right| = O_P\left(\frac{\rho_T}{a_T(f)}\right), \tag{B.29}$$

Let

$$
\begin{aligned}
\Omega_{T,j}(x) &= \frac{1}{(T-j)h} \sum_{t=j+1}^{T} K\left(\frac{X_{t-j} - x}{h}\right) \widetilde{e}_{t,j}, \\
\Omega_{T,j}^{(m)}(x) &= \frac{1}{(T-j)h} \sum_{t=j+1}^{T} K\left(\frac{X_{t-j}^{(m)} - x}{h}\right) \widetilde{e}_{t,j}^{(m)}, \quad m = m_T,
\end{aligned}
$$

32

where $X_{t-j}^{(m)}$ and $\tilde{e}_{t,j}^{(m)}$ are defined as in Definition 1. Note that $\mathsf{E}\left[\Omega_{T,j}(x)\right] = 0$ for all $j = 1, \cdots, J(T)$. Then, we have

$$
\begin{aligned}
& \max_{1 \leq j \leq J(T)} \sup_{|x| \leq C_T} \left|\Omega_{T,j}(x)\right| \\
\leq \ & \max_{1 \leq j \leq J(T)} \sup_{|x| \leq C_T} \left|\Omega_{T,j}^{(m)}(x) - \mathsf{E}\left[\Omega_{T,j}^{(m)}(x)\right]\right| \\
& + \max_{1 \leq j \leq J(T)} \sup_{|x| \leq C_T} \left|\Omega_{T,j}(x) - \Omega_{T,j}^{(m)}(x)\right| \\
& + \max_{1 \leq j \leq J(T)} \sup_{|x| \leq C_T} \left|\mathsf{E}\left[\Omega_{T,j}^{(m)}(x)\right] - \mathsf{E}\left[\Omega_{T,j}(x)\right]\right| \\
=: \ & \Omega_T(1) + \Omega_T(2) + \Omega_T(3).
\end{aligned}
\tag{B.30}
$$

Following the argument in the proof of Lemma A.3, we have

$$
\Omega_T(2) + \Omega_T(3) = O_P(\rho_T) \tag{B.31}
$$

as $J(T)h^{-(d+1)}v_1(m_T)\rho_T^{-1} = O(1)$. On the other hand, following the proof of Theorem 3.1, we can show that

$$
\Omega_T(1) = O_P(\rho_T) \tag{B.32}
$$

By (B.30)–(B.32), we can show that (B.28) holds.

By (B.27), Taylor expansion and the proof of (B.18), we can show that

$$
\max_{1 \leq j \leq J(T)} \sup_{|x| \leq C_T} \Pi_{T,j,1}(x) = O_P\left(\frac{b_T h^2}{a_T(f)}\right). \tag{B.33}
$$

Then, by (B.26), (B.28) and (B.33), we can prove Proposition 4.3. ∎