LOCAL LINEAR FITTING UNDER NEAR EPOCH DEPENDENCE: UNIFORM CONSISTENCY WITH CONVERGENCE RATES

Author(s): Degui Li, Zudi Lu and Oliver Linton

Source: *Econometric Theory*, October 2012, Vol. 28, No. 5 (October 2012), pp. 935-958

Published by: Cambridge University Press

Stable URL: https://www.jstor.org/stable/23271468

Econometric Theory, **28**, 2012, 935–958.
doi:10.1017/S0266466612000011

LOCAL LINEAR FITTING UNDER NEAR EPOCH DEPENDENCE: UNIFORM CONSISTENCY WITH CONVERGENCE RATES

DEGUI LI
Monash University

ZUDI LU
University of Adelaide

OLIVER LINTON
University of Cambridge

Local linear fitting is a popular nonparametric method in statistical and econometric modeling. Lu and Linton (2007, *Econometric Theory* 23, 37–70) established the pointwise asymptotic distribution for the local linear estimator of a nonparametric regression function under the condition of near epoch dependence. In this paper, we further investigate the uniform consistency of this estimator. The uniform strong and weak consistencies with convergence rates for the local linear fitting are established under mild conditions. Furthermore, general results regarding uniform convergence rates for nonparametric kernel-based estimators are provided. The results of this paper will be of wide potential interest in time series semiparametric modeling.

1. INTRODUCTION

Local linear fitting is a popular nonparametric method in nonlinear statistical and econometric modeling. 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 covers 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 (NED) processes. The results of this paper will be of wide potential interest in time series semiparametric modeling (see, e.g., Andrews, 1995) and structured nonparametric modeling (see, e.g., Linton and Mammen, 2005).

The authors thank the co-editor Yoon-Jae Whang and two referees for their valuable comments, which substantially improved the earlier version of this paper. This research was partially supported by Discovery Project and Future Fellowship grants from the Australian Research Council, and the European Research Council. Address correspondence to Oliver Linton, Faculty of Economics, Cambridge, CB3 9DD, United Kingdom; e-mail: obl20@cam.ac.uk.

Uniform consistency results of nonparametric kernel-based estimators have been studied by many authors, as they are useful in many applications such as semiparametric estimation and specification testing. For recent developments, the reader is referred to Liebscher (1996), Masry (1996), Bosq (1998), Fan and Yao (2003), Hansen (2008), and Kristensen (2009) and the references therein. A rather obvious feature of the preceding literature is that the observed time series are generally assumed to be α-mixing (i.e., strongly mixing). α-mixing dependence has been one of the most popular dependence conditions in statistics and econometrics. Indeed, the stationary solutions of many linear and nonlinear time series models are α-mixing under some suitable conditions; see, for example, Withers (1981), Tjøstheim (1990), Tong (1990), Masry and Tjøstheim (1995), Lu (1998), and Cline and Pu (1999).

However, from a practical point of view, α-mixing dependence suffers from many undesirable features. As pointed out by Davidson (1994) and Lu (2001), the α-mixing condition is difficult to verify in practice, especially in the case of compound processes. For example, the autoregressive conditional heteroskedasticity (ARCH) model and its generalized version GARCH have been proved to be α-mixing under some mild conditions (Bollerslev, 1986; Lu, 1996a, 1996b; Carrasco and Chen, 2002). But for compound processes such as autoregressive moving average process with ARCH or GARCH errors, it is still difficult to show 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 cases. Andrews (1984) showed that the stationary solution to a simple linear AR(1) model of the form

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

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 NED or stable process, which includes the α-mixing process as a special case. One can allow some types of nonstationarity, but this complicates the notation considerably, so we do not formally consider this but discuss subsequently 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

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

where $\mathbf{X}^\top$ 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*

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

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

The concept of NED process dates back to Ibragimov (1962) and was further developed by Billingsley (1968), McLeish (1975a, 1975b, 1977), and Lin (2004). Basically, most of these authors assumed that $\{\varepsilon_t\}$ is a martingale difference or is 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 processes that are not α-mixing.

There has been some literature on estimation and testing issues for 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. In particular, we obtain the uniform rate over expanding subsets of the covariate support. We also provide new results on estimation of a countable number of regression functions, for example, $g_j(\mathbf{x}) = \mathbb{E}(Y_t|\mathbf{X}_{t-j} = \mathbf{x})$, $j = 1, 2, \ldots$. This application occurs naturally in a number of time series settings (Hong, 2000; 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 also.

The proofs for the main results are different from those in Andrews (1995), which may be the only existing uniform convergence results for nonparametric kernel estimation under the NED assumption. Andrews (1995) made use of a Fourier transformation of the kernel and obtained a number of uniform consistency results for the nonparametric density and regression estimators based on the local constant approximation. In this paper, we will use the local linear approach and then establish the uniform consistency results by approximating the NED process by an α-mixing process and applying some effective ways such as finite covering and truncation methods in the proofs. The rate we obtain is constrained by the amount of dependence but does not explicitly depend on it, as it does in Andrews (1995), thereby yielding faster convergence rates in general. This means that in some special cases our convergence rate is optimal (see Stone, 1980).

We remark that an alternative extension of dependence beyond mixing can also be found in Nze, Bühlmann, and Doukhan (2002) and Nze and Doukhan (2004). These authors investigated a class of dependent processes they call "weakly dependent," the definition of which is quite involved. They established the asymptotic normality and uniform consistency of the local constant nonparametric regression estimator under some conditions, which included a fixed compact support.

The rest of the paper is organized as follows. The local linear fitting and the uniform convergence rates of the proposed 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. Application of our results in estimation of a countable number of conditional expectations is given in Section 4. The technical lemmas and the proofs of the main results are collected in two Appendixes.

2. UNIFORM CONVERGENCE RATES OF LOCAL LINEAR FITTING

In this section, we study the local linear estimator of the conditional mean regression function defined by

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

Local linear fitting is a widely used nonparametric estimation method, and we refer to Fan and Gijbels (1996) for a detailed account of 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

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

Locally, this suggests estimating $(a_0, \mathbf{a}_1^\top) = (g(\mathbf{x}), (g'(\mathbf{x}))^\top)$ by

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

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

There has been rich literature on the uniform convergence rates for the local linear estimators under mixing conditions; see, for example, Masry (1996), Fan and

Yao (2003), and Hansen (2008). Lu and Linton (2007) established the pointwise asymptotic distribution for the local linear estimators under the NED condition. In this section, we provide the uniform convergence rates for $\hat{g}(\mathbf{x})$ over expanding sets. The distribution of the covariates plays a role in determining the rate at which the set considered may expand, and such set is defined by

$$\{\mathbf{x} : \|\mathbf{x}\| \le C_T\}, \quad \text{where } C_T = (\log T)^{\tau_*} T^{\tau_0}, \qquad \tau_* \ge 0, \qquad \tau_0 \ge 0. \tag{2.3}$$

Define

$$a_T(f) := \inf_{\|\mathbf{x}\| \le C_T} f(\mathbf{x}) > 0, \tag{2.4}$$

where $f(\cdot)$ is the density function of $\{\mathbf{X}_t\}$. Let

$$M_T = \left[\left\{ \frac{T^{1-2/p_0} h^d}{\log T} \right\}^{1/2} \right], \qquad \rho_T = \left\{ \frac{\log T}{T h^d} \right\}^{1/2}, \tag{2.5}$$

where $p_0 = 2 + \varepsilon_*$ and $\varepsilon_* > 0$ and $[a]$ stands for the integer part of a real number a.

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

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

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

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

A2.

(a) The density function $f(\cdot)$ is continuous on $\mathbb{R}^d$. Furthermore, the joint density function $f_{0j}(\cdot,\cdot)$ of $(\mathbf{X}_0, \mathbf{X}_j)$ exists and satisfies that for some positive integer j^* and all $j \ge j^*$, $f_{0j}(\mathbf{x}_1, \mathbf{x}_2) < C_f$ for all $(\mathbf{x}_1, \mathbf{x}_2) \in \mathbb{R}^{2d}$, $0 < C_f < \infty$.

(b) The regression function $g(\cdot)$ has continuous derivatives up to the second order over $\mathbb{R}^d$.

A3.

(a) $\{Y_t, \mathbf{X}_t\}$ is stationary NED in L_{p_0}-norm with respect to a stationary α-mixing process $\{\varepsilon_t\}$, $\mathbb{E}|Y_t|^{p_0} < \infty$, where $p_0 = 2 + \varepsilon_*$.

(b) The mixing coefficient α_t of the stationary α-mixing $\{\varepsilon_t\}$ satisfies $\alpha_t \le C_\alpha t^{-\theta_0}$, $0 < C_\alpha < \infty$, $\theta_0 > \beta_1$, $\beta_1 = \Big((3p_0 + 6)/4p_0 + (\tfrac{1}{2} + \tau_0)d\Big) \Big/ \Big(\tfrac{1}{2} - (1/p_0)\Big)$, where τ_0 is defined in C_T of (2.3).

A4.

(a) There exist two sequences of positive integers m_T and M_T^*, which satisfy $m_T \to \infty$, $m_T = o(M_T^*)$,

$$M_T^* = o(M_T), \qquad M_T^* h^d = o(1), \qquad \left(M_T^*\right)^{-(\theta_0 \varepsilon_*/(4+\varepsilon_*))+1} h^{-d} = o(1), \tag{2.6}$$

where θ_0 is defined in A3(b). Furthermore,

$$h^{-d-1} \rho_T^{-1} v_1(m_T) = O(1),$$

$$h^{-2d} \left(v_2^{1/2}(m) + h^{-(\varepsilon_*/(2+\varepsilon_*))} v_1^{(\varepsilon_*/2+\varepsilon_*)}(m_T) \right) = o(1).$$

(b) The bandwidth h satisfies, as $T \to \infty$,

$$h \to 0, \qquad (\log T)^{(\theta_0/2)+\beta_3} h^{-(\theta_0 d/2)-\beta_2} T^{(\beta_1-\theta_0)((1/2)-(1/p_0))} \to 0, \tag{2.7}$$

where $\beta_2 = \frac{(7+2d)d}{4}$ and $\beta_3 = \frac{(2\tau_*-1)d}{2} + \frac{1}{4}$.

Remark 2.1. A1 is a mild condition on the kernel function $K(\cdot)$, and some commonly used kernel functions such as the standard normal probability density function can be shown to satisfy A1. By contrast, Masry (1996) required kernels that have compact support. A2(a) and (b) 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 $\{\varepsilon_t\}$. There is a trade-off between the moment condition and dependence, and we work in the special case with at least two moments because the case with fewer moments requires different techniques; see, for example, Lu and Cheng (1997), who considered pointwise strong consistency of kernel regression estimators, and Kanaya (2010) for uniform convergence under weaker moment conditions.

A4(a) is on the stability coefficient defined by (1.3) in Section 1 and can be satisfied by some interesting time series models under mild conditions (see, e.g., Lu and Linton, 2007 Sect. 4.1). When $p_0 = 3$, $d = 1$, and θ_0 is large enough, by letting

$$m_T = \frac{\sqrt{T^{1/3} h}}{(\log T)^2}, \qquad M_T^* = \frac{\sqrt{T^{1/3} h}}{\log T}, \qquad h \propto T^{-1/5},$$

we can show that (2.6) is satisfied. The crucial assumption A4(b) allows for slow decay in general, but it can be simplified in some special cases. For example, if

$\theta_0 \to \infty$ (α-mixing process decays with the exponential rate), the second term in (2.7) can be rewritten as

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

As β_1, β_2, and β_3 are constants, this means that

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

Hence, for the case of $\theta_0 \to \infty$, the second term in (2.7) is just slightly stronger than

$$\frac{T^{(1/2)-(1/p_0)}h^{d/2}}{(\log T)^{1/2}} \to \infty,$$

which is comparable to condition (12) in Hansen (2008) and is slightly stronger than the condition $Th^d/\log T \to \infty$ as $p_0 \to \infty$.

As the NED condition (with respect to the α-mixing $\{\varepsilon_t\}$) is more general than the mixing condition in Hansen (2008), to obtain the same convergence rates in this paper, we need some technical assumptions on the mixing coefficient and stability coefficients that are a bit more involved. However, the moment condition on $\{Y_t\}$ in A3(a) is the same as the corresponding moment condition in Hansen (2008).

We first give the uniform convergence rate of the local linear estimator $\hat{g}(\mathbf{x})$ in probability. Denote $b_T(g) = \sup_{\|\mathbf{x}\|\le C_T} \left\|f(\mathbf{x})g''(\mathbf{x})\right\|$, where $g''(\mathbf{x})$ denotes the $d \times d$ matrix of second partial derivatives of the function $g(\cdot)$ and the norm here is the matrix euclidean norm $||A|| = \text{tr}(A^\top A)^{1/2}$ for matrix A.

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

$$\sup_{\|\mathbf{x}\|\le C_T} \left|\hat{g}(\mathbf{x}) - g(\mathbf{x})\right| = O_P\left(\frac{\rho_T + b_T(g)h^2}{a_T(f)}\right), \tag{2.8}$$

where $a_T(f)$ and ρ_T are defined in (2.4) and (2.5), respectively.

Remark 2.2. The preceding theorem can be regarded as an extension of Theorem 10 in Hansen (2008) from α-mixing process to NED process. Hansen (2008) used the slightly different condition that the second derivatives of $g(\mathbf{x})f(\mathbf{x})$ are bounded, whereas we allow that $b_T(g) = \sup_{\|\mathbf{x}\|\le C_T} \left\|f(\mathbf{x})g''(\mathbf{x})\right\|$ increases with T. If the second-order derivatives of $g(\mathbf{x})$ and $f(\mathbf{x})$ are uniformly bounded, $b_T(g) < C_g$ for some $0 < C_g < \infty$. Then (2.8) would become

$$\sup_{\|\mathbf{x}\|\le C_T} \left|\hat{g}(\mathbf{x}) - g(\mathbf{x})\right| = O_P\left(\frac{\rho_T + h^2}{a_T(f)}\right). \tag{2.9}$$

Furthermore, if we let $C_T = C$ and $a_T(f) > c_0 > 0$, (2.9) becomes

$$\sup_{\|\mathbf{x}\| \le C} \left|\hat{g}(\mathbf{x}) - g(\mathbf{x})\right| = O_P\left(\rho_T + h^2\right). \tag{2.10}$$

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 independent and identically distributed case (see, e.g., Stone, 1980). This bandwidth is consistent with A4 under some restrictions on p_0, d, θ_0, τ_0 and the stability coefficients v_j, $j = 1, 2$. Equation (2.10) can be regarded as the extension of some existing results under the mixing dependence assumption such as Theorem 6.5 in Fan and Yao (2003).

Remark 2.3. We next briefly discuss some nonstationary extensions. There has been a lot of work recently on nonparametric regression with nonstationary covariates; see, for example, Wang and Phillips (2009) and included references. One particularly tractable type of nonstationarity is that of local stationarity; see, for example, Dahlhaus (1997). Suppose the data come from a triangular array $Z_{t,T} = \{Y_{t,T}, \mathbf{X}_{t,T},\ t = 1,\dots,T\}$. The stochastic process $\{Z_{t,T}\}$ is called locally stationary if there exists a stationary stochastic process $\{\widetilde{Z}_{u,t}\}$, $u \in [0,1]$, such that

$$\mathsf{P}\left\{\max_{1\le t\le T}\left|Z_{t,T} - \widetilde{Z}_{t/T,t}\right| \le D_T T^{-1/2}\right\} = 1 \tag{2.11}$$

for all T, where $\{D_T\}$ is a well-defined positive process satisfying for some $\eta > 0$, $\mathbb{E}\left(|D_T|^{4+\eta}\right) < \infty$; see Koo and Linton (2010). For locally stationary processes, our results will go through provided all conditions are made on $\widetilde{Z}_{u,t}$ to hold uniformly over $u \in [0,1]$.

Remark 2.4. Our C_T defined in (2.3) is quite general to cover different situations in applications of Theorem 2.1. For example, if taking $C_T = (\log T)^{1/d} T^{1/\tau_0}$ as in Hansen (2008), the uniform convergence rate on the right-hand side of (2.8) would become inapplicable when the regressor is gaussian, by noticing that when $\{X_t\}$ is real-valued gaussian, it is easy to check that

$$\inf_{|x|\le c_T} f(x) \propto \exp\left\{-\frac{c_T^2}{2}\right\}, \qquad c_T \to \infty,$$

which implies that

$$a_T(f) \propto \exp\left\{-\frac{C_T^2}{2}\right\} \propto \exp\left\{-\frac{T^{2/\tau_0}(\log T)^{2/d}}{2}\right\},$$

and the convergence rate on the right-hand side of (2.8) would tend to infinity. Hence, it is more sensible for us to consider the uniform convergence rate of the local linear estimator with gaussian regressors by letting $\tau_0 = 0$ in C_T (i.e., $C_T = (\log T)^{\tau_*}$) defined in (2.3). Hence, our results are more widely applicable than the results of Hansen (2008), who only considered the form of $C_T = (\log)^{1/d} T^{1/\tau_0}$.

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

THEOREM 2.2. *Suppose that the conditions in Theorem 2.1 are satisfied,* $\mathbb{E}|Y_t|^{s_1} < \infty$, $s_1 > 2p_0$,

$$Th^{-(d+1)}v_1(m_T)\rho_T^{-1} = O\left((\log T)^{-(1+\varsigma)}\right), \qquad \varsigma > 0, \tag{2.12}$$

and

$$(\log T)^{(\theta_0/2)+\beta_3}h^{-(\theta_0 d/2)-\beta_2}T^{1+(\beta_1-\theta_0)((1/2)-(1/p_0))} = O\left((\log T)^{-(1+\varsigma)}\right). \tag{2.13}$$

Then, we have

$$\sup_{\|\mathbf{x}\|\le C_T} \left|\hat{g}(\mathbf{x}) - g(\mathbf{x})\right| = O\left(\frac{\rho_T + b_T(g)h^2}{a_T(f)}\right) \quad a.s. \tag{2.14}$$

3. GENERAL RESULTS

Let $\{Y_t, \mathbf{X}_t\}$ be a stationary NED sequence defined in Section 1. We next consider the weighted average form

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

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 in the form of (3.1). In this section, we provide some general results for uniform convergence rates of W_T under our NED assumption, from which we can derive the two theorems in Section 2 conveniently. Hansen (2008) established the weak and strong uniform convergence rate of $W_T(\mathbf{x})$ for stationary α-mixing process. We will provide the uniform convergence rate for $W_T(\mathbf{x})$ when the α-mixing dependence is replaced by the NED condition.

To establish the uniform convergence rate of $W_T(\mathbf{x})$, we need the following regularity condition on $K_T(\cdot)$.

A5. The kernel-based weight function $K_T(\cdot)$ 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.

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

THEOREM 3.1. *Suppose that the conditions A2(a) and A3–A5 are satisfied. Then we have*

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

where ρ_T is defined in (2.5).

THEOREM 3.2. *Suppose that the conditions in Theorem 3.1, (2.12), and (2.13) are satisfied, $\mathbb{E}|Y_t|^{s_1} < \infty$, $s_1 > 2p_0$. Then, we have*

$$\sup_{\|\mathbf{x}\| \le C_T} |W_T(\mathbf{x}) - \mathbb{E}[W_T(\mathbf{x})]| = O(\rho_T) \quad a.s., \tag{3.3}$$

where ρ_T is defined in (2.5).

Remark 3.1. The preceding theorems establish the weak and strong convergence rates 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. Letting $Q_1 > Q_0 > 1$, $\mathbb{E}|Y_t|^{Q_1} < \infty$ and the mixing coefficient

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

Then, applying Theorem 4.1 in Shao and Yu (1996) and following the proofs of Lemmas A.2 and A.5 in Appendix A, we can show that, if $\{(\mathbf{X}_t, Y_t)\}$ is NED with the stable coefficient decaying at a geometric rate,

$$\sup_{\|\mathbf{x}\| \le C_T} \left(\mathbb{E}|W_T(\mathbf{x}) - \mathbb{E}[W_T(\mathbf{x})]|^{Q_0}\right)^{1/Q_0} = O\left(T^{-1/2} h^{((1-Q_1)d)/Q_1}\right) \tag{3.4}$$

under mild conditions.

Remark 3.2. It is of interest to consider the uniform consistency over the set $\{\mathbf{x} : f(\mathbf{x}) \ge d_T\}$, $d_T \to 0$, similarly to Andrews (1995). Under some conditions on $f(\cdot)$ and d_T, we conjecture that the uniform convergence rates obtained in this paper also hold over the set $\{\mathbf{x} : f(\mathbf{x}) \ge d_T\}$. We will consider this in future study.

4. ESTIMATION OF A COUNTABLE NUMBER OF CONDITIONAL EXPECTATIONS

Define the quantities $g_j(x) = \mathbb{E}(Y_t | X_{t-j} = x)$, $j = 1, 2, \ldots$, where both $\{Y_t\}$ and $\{X_t\}$ are real-valued. There are many cases of interest that require estimation of this whole family of regression functions. For example, consider the quantity

$$G(x) = \sum_{j=1}^{\infty} w_j g_j(x), \tag{4.1}$$

where w_j, $j \ge 1$, are summable weights and the sum in (4.1) is assumed to be well defined. This quantity is of interest in a number of applications, and we discuss three examples in detail here.

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 because of the curse of dimensionality. He proposed checking 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 $g_j^*(x) = \mathbb{E}(X_t|X_{t-j} = x)$, $j \geq 1$, are constant. This can be done by evaluating an empirical version of the weighted sum $\sup_x \sum_{j=1}^{\infty} w_j |g_j^*(x) - g_j|$, where w_j and g_j, $j \geq 1$, are summable weights and average values, respectively.

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

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

where $\widetilde{g}(\cdot)$ is an unknown function and the parametric family $\{\psi_j(\theta) : \theta \in \Theta, j = 1, \ldots, \infty\}$ satisfies 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 $\widetilde{g}(\cdot)$ that involves a weighted sum of the form (4.1); specifically, the quantity $g_\theta^*(x) = \sum_{j=1}^{\infty} \psi_j(\theta)\eta_j(x)$. They proposed an estimation strategy for the unknown quantities, which requires as input the estimation of $\eta_j(x) = \mathbb{E}(X_t^2|X_{t-j} = x)$ for $j = 1, 2, \ldots, J(T)$, where $J(T) = c \log T$ for some $c > 0$. They required bounding the estimation error of $\eta_j(x)$ 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 strongly mixing with geometric decay. We next give more definitive results under weaker conditions.

As a final motivation, consider the nonparametric prediction of a future value X_0 given a sample $\{X_{-1}, \ldots, X_{-T}\}$. Linton and Sancetta (2009) established consistency of estimators of $\mathbb{E}(X_0|X_{-1}, \ldots)$ 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; that is, to use $\sum_{j=1}^{J(T)} w_{T,j}\hat{g}_j(X_{-j})$ to estimate $\mathbb{E}(X_0|X_{-1}, \ldots)$, where $w_{T,j}$, $j = 1, \ldots, J(T)$, are weights such that $\sum_{j=1}^{J(T)} w_{T,j} = 1$, $J(T)$ is an increasing sequence and $\hat{g}_j(\cdot)$, $j \geq 1$, are the nonparametric regression fits.

Let

$$G(x_1, x_2, \ldots) = \mathbb{E}(X_t|X_{t-1} = x_1, X_{t-2} = x_2, \ldots)$$

be the best prediction function. Then

$$G_w(x_1, x_2, \ldots) = \sum_{j=1}^{J(T)} w_{T,j} g_j^*(x_j), \qquad g_j^*(x) = \mathbb{E}(X_t | X_{t-j} = x),$$

can be considered as an approximation to $G(x_1, x_2, \ldots)$. 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 $\hat{G}_w(x_1, x_2, \ldots)$ to $G_w(x_1, x_2, \ldots)$, where

$$\hat{G}_w(x_1, x_2, \ldots) = \sum_{j=1}^{J(T)} w_{T,j} \widetilde{g}_j(x_j)$$

and $\widetilde{g}_j(\cdot)$ is the local linear estimator of $g_j^*(\cdot)$, it suffices to control the rate for each $\widetilde{g}_j(x_j)$ 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 also. 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 4.1. *Suppose that $\{X_t\}$ is stationary NED in L_{p_0}-norm with respect to a stationary α-mixing process $\{\varepsilon_t\}$ with $\mathbb{E}|X_t|^{p_0} < \infty$, A2 is satisfied when $g(\cdot)$ is replaced by $g_j^*(\cdot)$, and the remaining conditions of Theorem 2.1 are satisfied. Furthermore, suppose that*

$$J(T) h^{-(d+1)} \upsilon_1(m_T) \rho_T^{-1} = O(1)$$

and

$$J(T)(\log T)^{(\theta_0/2)+\beta_3} h^{-(\theta_0 d/2)-\beta_2} T^{(\beta_1-\theta_0)((1/2)-(1/p_0))} = o(1).$$

Then we have

$$\max_{1 \le j \le J(T)} \sup_{|x| \le C_T} \left| \widetilde{g}_j(x) - g_j^*(x) \right| = O_P\left(\frac{\rho_T + b_T h^2}{a_T(f)} \right), \tag{4.2}$$

where $b_T = \max_{1 \le j \le J(T)} b_T(g_j^*)$ and $b_T(g_j^*)$ is defined as $b_T(g)$ in Section 2.

Remark 4.1. In the preceding result, we establish the weak convergence rate for $g_j^*(x)$ uniformly over j and x. The strong uniform convergence rate result for $g_j^*(x)$ can also be established by applying proofs similar to those of Theorems 2.2 and 3.2.

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-Verlag.

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

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 ed. Lecture Notes in Statistics 110. Springer-Verlag.

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

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

Dahlhaus, R. (1997) Fitting time series models to nonstationary processes. *Annals of Statistics* 25, 1–37.

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

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

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* 62, 557–574.

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

Kanaya, S. (2010) Uniform Convergence Rates of Kernel-Based Nonparametric Estimators for Continuous Time Diffusion Processes: A Damping Function Approach. Manuscript, Department of Economics, University of Oxford.

Koo, B. & O. Linton (2010) Semiparametric Estimation of Locally Stationary Diffusion Processes. Manuscript, LSE.

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

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

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

Linton, O.B. & 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 (digest).

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. & P. Cheng (1997) Distribution-free strong consistency for nonparametric kernel regression involving nonlinear time series. *Journal of Statistical Planning and Inference* 65, 67–86.
Lu, Z. & O. Linton (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. & 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 für Wahrscheinlichkeitstheorie 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., 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.
Shao, Q. & H. Yu (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.
Wang, Q. & P.C.B. Phillips (2009) Structural nonparametric cointegrating regression. *Econometrica* 77, 1901–1948.
Withers, C.S. (1981) Conditions for linear processes to be strong-mixing. *Probability Theory and Related Fields* 57, 477–480.

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 $\mathsf{P}(|Z_t| \le B) = 1$ for all $t \ge 1$. Then for each integer $q \in [1, \frac{T}{2}]$ and each $\epsilon > 0$, we have*

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

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

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

Letting $Y_t^{(m)}$ be defined as in Definition 1, we establish the result on the moment of $Y_t^{(m)}$ in Lemma A.2, which follows.

LEMMA A.2. *Suppose that the sequence $\{Y_t\}$ is NED in L_s with $\mathbb{E}|Y_t|^s < \infty$ for $s \geq 1$. Then we have $\mathbb{E}\left|Y_t^{(m)}\right|^s < \infty$.*

Proof. Note that $Y_t^{(m)} = Y_t + Y_t^{(m)} - Y_t$. By applying the C_r inequality and (1.3) in Definition 1, we can prove that $\mathbb{E}\left|Y_t^{(m)}\right|^s < \infty$. ■

Define

$$W_T^{(m)}(\mathbf{x}) = \frac{1}{Th^d}\sum_{t=1}^{T} Y_t^{(m)} K_T\left(\frac{\mathbf{X}_t^{(m)} - \mathbf{x}}{h}\right). \tag{A.2}$$

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 of Theorem 3.1 are satisfied. Then, we have*

$$\sup_{\|\mathbf{x}\| \leq C_T}\left|W_T(\mathbf{x}) - W_T^{(m)}(\mathbf{x})\right| = O_P\left(h^{-d-1}\upsilon_1(m)\right). \tag{A.3}$$

Proof. Observe that

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

We first consider $I_{T,2}(\mathbf{x})$. Noting that $\mathbb{E}\left|Y_t^{(m)} - Y_t\right| = \upsilon_1(m)$ and by the boundedness condition on $K_T(\cdot)$ (see A5 in Section 3), we have

$$\sup_{\|\mathbf{x}\| \leq C_T}\left|I_{T,2}(\mathbf{x})\right| \leq h^{-d}\sup_{\|\mathbf{x}\| \leq C_T}\left|K_T\left(\tfrac{\mathbf{X}_t - \mathbf{x}}{h}\right)\right|\left|Y_t^{(m)} - Y_t\right| = O_P\left(h^{-d}\upsilon_1(m)\right). \tag{A.5}$$

For $I_{T,1}(\mathbf{x})$, note that

$$\begin{aligned} I_{T,1}(\mathbf{x}) &= \frac{1}{Th^d}\sum_{t=1}^{T} Y_t\left(K_T\left(\frac{\mathbf{X}_t - \mathbf{x}}{h}\right) - K_T\left(\frac{\mathbf{X}_t^{(m)} - \mathbf{x}}{h}\right)\right) \\ &\quad + \frac{1}{Th^d}\sum_{t=1}^{T}\left(Y_t^{(m)} - Y_t\right)\left(K_T\left(\frac{\mathbf{X}_t - \mathbf{x}}{h}\right) - K_T\left(\frac{\mathbf{X}_t^{(m)} - \mathbf{x}}{h}\right)\right) \\ &=: I_{T,3}(\mathbf{x}) + I_{T,4}(\mathbf{x}). \end{aligned} \tag{A.6}$$

By the Lipschitz continuity of $K_T(\cdot)$, we have uniformly for $\|\mathbf{x}\| \le C_T$,

$$\left| K_T\left(\frac{\mathbf{X}_t - \mathbf{x}}{h}\right) - K_T\left(\frac{\mathbf{X}_t^{(m)} - \mathbf{x}}{h}\right) \right| = O_P(v_1(m)/h). \tag{A.7}$$

By (A.7), we have

$$\sup_{\|\mathbf{x}\| \le C_T} \left| I_{T,3}(\mathbf{x}) \right| = O_P\left(h^{-d-1} v_1(m)\right). \tag{A.8}$$

On the other hand, we have

$$\sup_{\|\mathbf{x}\| \le C_T} \left| I_{T,4}(\mathbf{x}) \right| = O_P\left(h^{-d-1} v_1^2(m)\right). \tag{A.9}$$

In view of (A.4)–(A.6), (A.8), and (A.9), we can show that (A.3) holds. ■

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

$$\sup_{\|\mathbf{x}\| \le C_T} \left| W_T(\mathbf{x}) - W_T^{(m_T)}(\mathbf{x}) \right| = O\left(\rho_T\right) \quad a.s., \tag{A.10}$$

where m_T satisfies the condition A4(a) and ρ_T is defined in (2.5).

Proof. Let $I_{T,1}(\mathbf{x})$ and $I_{T,2}(\mathbf{x})$ be defined as in (A.4). By (2.12) and the Markov inequality, we have

$$\begin{aligned} \sum_{T=1}^{\infty} \mathsf{P}\left(\left|Y_t^{(m_T)} - Y_t\right| > \rho_T h^d\right) &\le \sum_{T=1}^{\infty} \rho_T^{-1} h^{-d} \mathbb{E}\left|Y_t^{(m_T)} - Y_t\right| \\ &\le C \sum_{T=1}^{\infty} \rho_T^{-1} h^{-d} v_1(m_T) = C \sum_{T=1}^{\infty} \frac{1}{T \log^{1+\varsigma} T} < \infty. \end{aligned} \tag{A.11}$$

By the boundedness condition on $K_T(\cdot)$ and (A.11), we have

$$\begin{aligned} \sup_{\|\mathbf{x}\| \le C_T} \left| I_{T,2}(\mathbf{x}) \right| &\le \frac{1}{Th^d} \sum_{t=1}^{T} \left|Y_t^{(m_T)} - Y_t\right| \sup_{\|\mathbf{x}\| \le C_T} \left|K_T\left(\frac{\mathbf{X}_t - \mathbf{x}}{h}\right)\right| \\ &= O(\rho_T) \quad \text{a.s.} \end{aligned} \tag{A.12}$$

Analogously, we can show that $\sup_{\|\mathbf{x}\| \le C_T} \left|I_{T,1}(\mathbf{x})\right| = O(\rho_T)$ a.s., which together with (A.12) implies that (A.10) holds. ■

LEMMA A.5. *Let r_T be a sequence of positive integers such that $m_T/r_T = o(1)$ and*

$$U_r(\mathbf{x}) = \frac{1}{h^d} \sum_{t=1}^{r_T} Y_t^{(m_T)} K_T\left(\frac{\mathbf{X}_t^{(m_T)} - \mathbf{x}}{h}\right). \tag{A.13}$$

Suppose that the conditions of Theorem 3.1 are satisfied. Then, we have

$$\operatorname{Var}[U_r(\mathbf{x})] = O(r_T h^{-d}). \tag{A.14}$$

Proof. For simplicity, we let $m = m_T$ and $r = r_T$ in this proof. Observe that

$$\mathrm{Var}[U_r(\mathbf{x})] = \Xi(1) + \Xi(2), \tag{A.15}$$

where

$$\Xi(1) = \frac{1}{h^{2d}} \sum_{t=1}^{r} \mathrm{Var}\left[Y_t^{(m)} K_T\left(\frac{\mathbf{X}_t^{(m)} - \mathbf{x}}{h}\right)\right],$$

$$\Xi(2) = \frac{1}{h^{2d}} \sum_{t=1}^{r} \sum_{s\neq t} \mathrm{Cov}\left[Y_t^{(m)} K_T\left(\frac{\mathbf{X}_t^{(m)} - \mathbf{x}}{h}\right),\ Y_s^{(m)} K_T\left(\frac{\mathbf{X}_s^{(m)} - \mathbf{x}}{h}\right)\right].$$

We first consider $\Xi(1)$. It is easy to check that

$$\begin{aligned}
\Xi(1) \le\ & \frac{1}{h^{2d}} \sum_{t=1}^{r} \mathbb{E}\left[(Y_t^{(m)})^2 K_T^2\left(\frac{\mathbf{X}_t^{(m)} - \mathbf{x}}{h}\right)\right] \\
=\ & \frac{1}{h^{2d}} \sum_{t=1}^{r} \mathbb{E}\left[Y_t^2 K_T^2\left(\frac{\mathbf{X}_t - \mathbf{x}}{h}\right)\right] \\
& + \frac{1}{h^{2d}} \sum_{t=1}^{r} \mathbb{E}\left[\left((Y_t^{(m)})^2 - Y_t^2\right) K_T^2\left(\frac{\mathbf{X}_t^{(m)} - \mathbf{x}}{h}\right)\right] \\
& + \frac{1}{h^{2d}} \sum_{t=1}^{r} \mathbb{E}\left[Y_t^2 \left(K_T^2\left(\frac{\mathbf{X}_t^{(m)} - \mathbf{x}}{h}\right) - K_T^2\left(\frac{\mathbf{X}_t - \mathbf{x}}{h}\right)\right)\right] \\
=:\ & \Xi(3) + \Xi(4) + \Xi(5).
\end{aligned} \tag{A.16}$$

By the condition A4(a) and standard but tedious calculation similar to that in the proof of Lemma A.3, we have

$$\Xi(3) = O(rh^{-d}), \tag{A.17}$$

$$\Xi(4) = O\left(rh^{-2d} v_2^{1/2}(m)\right) = o(rh^{-d}). \tag{A.18}$$

Letting $B_T = (h/v_1(m))^{1/(2+\varepsilon_*)} = (h/v_1(m))^{1/p_0}$, we have

$$\begin{aligned}
& \mathbb{E}\left[Y_t^2 \left(K_T^2\left(\frac{\mathbf{X}_t^{(m)} - \mathbf{x}}{h}\right) - K_T^2\left(\frac{\mathbf{X}_t - \mathbf{x}}{h}\right)\right)\right] \\
&\quad = \mathbb{E}\left[Y_t^2 I(|Y_t| \le B_T) \left(K_T^2\left(\frac{\mathbf{X}_t^{(m)} - \mathbf{x}}{h}\right) - K_T^2\left(\frac{\mathbf{X}_t - \mathbf{x}}{h}\right)\right)\right] \\
&\quad\quad + \mathbb{E}\left[Y_t^2 I(|Y_t| > B_T) \left(K_T^2\left(\frac{\mathbf{X}_t^{(m)} - \mathbf{x}}{h}\right) - K_T^2\left(\frac{\mathbf{X}_t - \mathbf{x}}{h}\right)\right)\right] \\
&\quad = O\left(B_T^2 v_1(m) h^{-1} + B_T^{-\varepsilon_*}\right) \\
&\quad = O\left((h/v_1(m))^{-\varepsilon_*/p_0}\right),
\end{aligned}$$

where $I(\cdot)$ is the indicator function. Then it is easy to check that

$$\Xi(5) = O\left(rh^{-2d}(h/\upsilon_1(m))^{-\varepsilon_*/p_0}\right) = o(rh^{-d}), \tag{A.19}$$

as $h^{-d-(\varepsilon_*/(2+\varepsilon_*))}\upsilon_1^{\varepsilon_*/(2+\varepsilon_*)}(m) = o(1)$ in A4(a).

Then, by (A.17)–(A.19), we have

$$\Xi(1) = O(rh^{-d}). \tag{A.20}$$

We next turn to the calculation of $\Xi(2)$. Note that

$$\begin{aligned}\Xi(2) = {} & \frac{1}{h^{2d}}\sum_{t=1}^{r}\sum_{|s-t|\le M_T^*}\text{Cov}\left[Y_t^{(m)}K_T\left(\frac{\mathbf{X}_t^{(m)}-\mathbf{x}}{h}\right),\ Y_s^{(m)}K_T\left(\frac{\mathbf{X}_s^{(m)}-\mathbf{x}}{h}\right)\right]\\ & +\frac{1}{h^{2d}}\sum_{t=1}^{r}\sum_{|s-t|> M_T^*}\text{Cov}\left[Y_t^{(m)}K_T\left(\frac{\mathbf{X}_t^{(m)}-\mathbf{x}}{h}\right),\ Y_s^{(m)}K_T\left(\frac{\mathbf{X}_s^{(m)}-\mathbf{x}}{h}\right)\right]\\ & =: \Xi(6)+\Xi(7),\end{aligned} \tag{A.21}$$

where M_T^* is defined in the condition A4(a).

By standard calculation, we have

$$\begin{aligned}\Xi(6) = {} & \frac{1}{h^{2d}}\sum_{t=1}^{r}\sum_{|s-t|\le M_T^*}\text{Cov}\left[Y_tK_T\left(\frac{\mathbf{X}_t-\mathbf{x}}{h}\right),\ Y_sK_T\left(\frac{\mathbf{X}_s-\mathbf{x}}{h}\right)\right]\\ & +O\left(rM_T^*h^{-2d}\left(\upsilon_2^{1/2}(m)+(h/\upsilon_1(m))^{-\varepsilon_*/p_0}\right)\right),\end{aligned} \tag{A.22}$$

which together with $h^{-2d}\left(\upsilon_2^{1/2}(m)+(h/\upsilon_1(m))^{-\varepsilon_*/p_0}\right) = o(1)$ in A4(a) and the fact that $M_T^* = o(h^{-d})$, implies that

$$\Xi(6) = O(rM_T^*) = o(rh^{-d}). \tag{A.23}$$

On the other hand, noting that $\{Y_t^{(m)}, \mathbf{X}_t^{(m)}\}$ is an α-mixing process with mixing coefficient

$$\alpha_m(t) \le \begin{cases}\alpha_{t-m}, & t\ge m+1;\\ 1, & t\le m,\end{cases}$$

we have

$$\begin{aligned}\Xi(7) & \le \frac{C}{h^{2d}}\sum_{t=1}^{r}\sum_{|s-t|> M_T^*}\alpha_{|s-t|-m}^{\varepsilon_*/(4+\varepsilon_*)}\left\{\mathbb{E}\left[\left|Y_t^{(m)}K_T\left(\frac{\mathbf{X}_t^{(m)}-\mathbf{x}}{h}\right)\right|^{2+(\varepsilon_*/2)}\right]\right\}^{4/(4+\varepsilon_*)}\\ & \le \frac{C}{h^{2d}}\sum_{t=1}^{r}\sum_{|s-t|> M_T^*}(|s-t|-m)^{-\theta_0\varepsilon_*/(4+\varepsilon_*)}\\ & \quad\times\left\{\mathbb{E}\left[\left|Y_t^{(m)}K_T\left(\frac{\mathbf{X}_t^{(m)}-\mathbf{x}}{h}\right)\right|^{2+(\varepsilon_*/2)}\right]\right\}^{4/(4+\varepsilon_*)}.\end{aligned} \tag{A.24}$$

By Lemma A.2, we have

$$\mathbb{E}\left[\left|Y_t^{(m)} K_T\left(\frac{\mathbf{X}_t^{(m)}-\mathbf{x}}{h}\right)\right|^{2+(\varepsilon_*/2)}\right] < \infty. \tag{A.25}$$

By the condition A4(a), (A.24), and (A.25), we have

$$\begin{aligned}\Xi(7) &\leq \frac{C}{h^{2d}}\sum_{t=1}^{r}\sum_{s>M_T^*/2} s^{-\theta_0\varepsilon_*/(4+\varepsilon_*)}\left\{\mathbb{E}\left[\left|\Psi(Y_t^{(m)})K_T\left(\frac{\mathbf{X}_t^{(m)}-\mathbf{x}}{h}\right)\right|^{2+(\varepsilon_*/2)}\right]\right\}^{4/(4+\varepsilon_*)}\\ &\leq \frac{C}{h^{2d}}\sum_{t=1}^{r}\sum_{s>M_T^*/2} s^{-\theta_0\varepsilon_*/(4+\varepsilon_*)} \leq Crh^{-d}\left((M_T^*)^{-\theta_0\varepsilon_*/(4+\varepsilon_*)+1}h^{-d}\right) = o(rh^{-d}).\end{aligned} \tag{A.26}$$

By (A.21), (A.23), and (A.26), we have $\Xi(2) = o(rh^{-d})$, which together with (A.15) and (A.20) implies that (A.14) holds. ∎

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. As in the proof of Lemma A.5, we let $m = m_T$ throughout this Appendix.

Proof of Theorem 3.1. Note that

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

By Lemma A.3, we have

$$\Pi_{T,2} = O_P\left(h^{-d-1}v_1(m)\right) = O_P(\rho_T), \qquad \Pi_{T,3} = O(\rho_T). \tag{B.2}$$

By (B.1) and (B.2), to prove (3.2), we need only to show that $\Pi_{T,1} = O_P(\rho_T)$. Recall 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 cover the set $\{\mathbf{x} : \|\mathbf{x}\| \leq C_T\}$ by a finite number of subsets S_k, $k = 1, \ldots, N_T$, which are centered at s_k with radius $\rho_T h^{d+1}$. Observe that

$$\begin{aligned}\Pi_{T,1} \le & \max_{1\le k\le N_T}\sup_{\|\mathbf{x}\|\in S_k}\left|W_T^{(m)}(\mathbf{x})-W_T^{(m)}(s_k)\right| \\ & +\max_{1\le k\le N_T}\sup_{\|\mathbf{x}\|\in S_k}\left|\mathbb{E}[W_T^{(m)}(\mathbf{x})]-\mathbb{E}[W_T^{(m)}(s_k)]\right| \\ & +\max_{1\le k\le N_T}\left|W_T^{(m)}(s_k)-\mathbb{E}[W_T^{(m)}(s_k)]\right| \\ =:& \Pi_{T,4}+\Pi_{T,5}+\Pi_{T,6}. \end{aligned} \tag{B.3}$$

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

$$\max_{T\ge 1}\max_{1\le k\le N_T}\sup_{\mathbf{x}\in S_k}\left|K_T\left(\frac{\mathbf{X}_t-\mathbf{x}}{h}\right)-K_T\left(\frac{\mathbf{X}_t-s_k}{h}\right)\right|\le \max_{1\le k\le N_T}\sup_{\mathbf{x}\in S_k}\left\|\frac{\mathbf{x}-s_k}{h}\right\|. \tag{B.4}$$

By (B.4) and noting that $\mathbb{E}\left|Y_t^{(m)}\right|<\infty$ by Lemma A.2, we have

$$\Pi_{T,4}=O_P\left(\frac{\rho_T h^{d+1}}{h^{d+1}}\right)=O_P(\rho_T),\qquad \Pi_{T,5}=O(\rho_T). \tag{B.5}$$

By (B.3) and (B.5), to prove $\Pi_{T,1}=O_P(\rho_T)$, we need only to show that $\Pi_{T,6}=O_P(\rho_T)$. Let $\Delta_T=T^{1/p_0}$,

$$\begin{aligned} &\overline{Y}_t^{(m)}=Y_t^{(m)}I\left(|Y_t^{(m)}|\le\Delta_T\right),\qquad \widetilde{Y}_t^{(m)}=Y_t^{(m)}I\left(|Y_t^{(m)}|>\Delta_T\right),\\ &\overline{W}_T^{(m)}(\mathbf{x})=\frac{1}{Th^d}\sum_{t=1}^T\overline{Y}_t^{(m)}K_T\left(\frac{\mathbf{X}_t^{(m)}-\mathbf{x}}{h}\right),\\ &\widetilde{W}_T^{(m)}(\mathbf{x})=\frac{1}{Th^d}\sum_{t=1}^T\widetilde{Y}_t^{(m)}K_T\left(\frac{\mathbf{X}_t^{(m)}-\mathbf{x}}{h}\right). \end{aligned}$$

It is easy to check that

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

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

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

where $p_0<\lambda_0<s_0$. Hence, we have

$$\Pi_{T,8}=o_P(\rho_T). \tag{B.7}$$

Letting

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

in Lemma A.1 and by Lemma A.5, we have

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

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(b), we have for η large enough, $\mathsf{P}\left(\Pi_{T,7} > \eta\rho_T\right) = o(1)$, which implies that

$$\Pi_{T,7} = O_P(\rho_T). \tag{B.8}$$

By (B.6)–(B.8), we can show that $\Pi_{T,6} = O_P(\rho_T)$. Then, the proof of Theorem 3.1 is completed. ■

Proof of Theorem 3.2. By Lemma A.4 and following the proof of Theorem 3.1, we need only to show that $\Pi_{T,6} = O(\rho_T)$ a.s., where $\Pi_{T,6}$ is defined in (B.3).

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

$$\begin{aligned}
\sum_{T=1}^{\infty} \mathsf{P}\left(\Pi_{T,8} > \eta\rho_T\right) &\le \sum_{T=1}^{\infty}\sum_{t=1}^{T} \frac{\mathbb{E}\left|Y_t^{(m)}\right|^{s_1}}{\Delta_T^{s_1}} \\
&\le C\sum_{T=1}^{\infty} T^{1-s_1/p_0} < \infty,
\end{aligned}$$

as $s_1 > 2p_0$. Hence, we have

$$\Pi_{T,8} = o(\rho_T) \quad \text{a.s.} \tag{B.9}$$

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 (2.13) and Lemma A.5, we have

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

by taking $\eta > 0$ large enough. Hence, we have

$$\Pi_{T,7} = O(\rho_T) \quad \text{a.s.} \tag{B.10}$$

By (B.9) and (B.10), we have $\Pi_{T,6} = O(\rho_T)$ a.s. 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),

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

where

$$w_{T,t}(x) = \frac{K\left(\frac{X_t - x}{h}\right)\left(S_{T,2}(x) - \left(\frac{X_t - x}{h}\right) S_{T,1}(x)\right)}{Th\left(S_{T,0}(x) S_{T,2}(x) - S_{T,1}^2(x)\right)},$$

$$S_{T,j}(x) = \frac{1}{Th}\sum_{t=1}^{T}\left(\frac{X_t - x}{h}\right)^j K\left(\frac{X_t - x}{h}\right), \qquad j = 0, 1, 2.$$

Then,

$$\begin{aligned}\hat{g}(x) - g(x) &= \left(\sum_{t=1}^{T} w_{T,t}(x) g(X_t) - g(x)\right) + \sum_{t=1}^{T} w_{T,t}(x) e_t \\ &=: \Pi_{T,1}^*(x) + \Pi_{T,2}^*(x),\end{aligned} \tag{B.11}$$

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

By Theorem 3.1, for any $j \geq 1$,

$$\sup_{|x| \leq C_T} \left| S_{T,j}(x) - \mu_j f(x) \right| = o_P(1), \tag{B.12}$$

where $\mu_j = \int_{\mathbb{R}} u^j K(u) du$. By (B.12) and standard calculation, we have

$$\sup_{|x| \leq C_T} \left| \Pi_{T,1}^*(x) \right| = O_P\left(\frac{b_T(g) h^2}{a_T(f)}\right). \tag{B.13}$$

Hence, to prove (2.8), we need only to show that

$$\sup_{|x| \leq C_T} \left| \Pi_{T,2}^*(x) \right| = O_P\left(\frac{\rho_T}{a_T(f)}\right). \tag{B.14}$$

By (B.12) and the definition of $w_{T,t}(\cdot)$, to prove (B.14), we need only to show that

$$\sup_{|x| \leq C_T} \left| \frac{1}{Th}\sum_{t=1}^{T} K\left(\frac{X_t - x}{h}\right) e_t \right| = O_P\left(\rho_T\right) \tag{B.15}$$

and

$$\sup_{|x|\le C_T}\left|\frac{1}{Th}\sum_{t=1}^{T}\left(\frac{X_t-x}{h}\right)K\left(\frac{X_t-x}{h}\right)e_t\right| = O_P(\rho_T). \tag{B.16}$$

By letting $Y_t = e_t$ in Theorem 3.1, we can show that (B.15) and (B.16) hold. 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 (2.14) holds. The details are omitted here. ■

Proof of Proposition 4.1. The detailed proof is similar to the proof of Theorem 2.1. By the definition of the local linear estimators $\widetilde{g}_j(x)$, $j = 1,\dots,J(T)$, we have

$$\widetilde{g}_j(x) = \sum_{t=j+1}^{T} w_{T,j,t}(x)X_t,$$

where

$$w_{T,j,t}(x) = \frac{K\left(\frac{X_{t-j}-x}{h}\right)\left(S_{T,j,2}(x)-\left(\frac{X_{t-j}-x}{h}\right)S_{T,j,1}(x)\right)}{(T-j)h\left(S_{T,j,0}(x)S_{T,j,2}(x)-S_{T,j,1}^2(x)\right)},$$

$$S_{T,j,k}(x) = \frac{1}{(T-j)h}\sum_{t=j+1}^{T}\left(\frac{X_{t-j}-x}{h}\right)^k K\left(\frac{X_{t-j}-x}{h}\right).$$

Then,

$$\begin{aligned}\widetilde{g}_j(x)-g_j^*(x) &= \left(\sum_{t=j+1}^{T} w_{T,j,t}(x)g_j^*(X_t)-g_j^*(x)\right)+\sum_{t=j+1}^{T} w_{T,j,t}(x)\widetilde{e}_{t,j}\\ &=: \Pi_{T,j,1}(x)+\Pi_{T,j,2}(x),\end{aligned} \tag{B.17}$$

where $\widetilde{e}_{t,j} = X_t - g_j^*(X_{t-j}) = X_t - \mathbb{E}\left(X_t|X_{t-j}\right)$.

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

$$\max_{1\le j\le J(T)}\sup_{|x|\le C_T}\left|S_{T,j,k}(x)-\mu_k f(x)\right| = o_P(1), \qquad k\ge 1. \tag{B.18}$$

By (B.18), to prove

$$\max_{1\le j\le J(T)}\sup_{|x|\le C_T}\Pi_{T,j,2}(x) = O_P\left(\frac{\rho_T}{a_T(f)}\right), \tag{B.19}$$

we need only to show

$$\max_{1\le j\le J(T)}\sup_{|x|\le C_T}\left|\frac{1}{(T-j)h}\sum_{t=j+1}^{T}K\left(\frac{X_{t-j}-x}{h}\right)\widetilde{e}_{t,j}\right| = O_P\left(\frac{\rho_T}{a_T(f)}\right) \tag{B.20}$$

and

$$\max_{1\le j\le J(T)}\sup_{|x|\le C_T}\left|\frac{1}{(T-j)h}\sum_{t=j+1}^{T}\left(\frac{X_{t-j}-x}{h}\right)K\left(\frac{X_{t-j}-x}{h}\right)\widetilde{e}_{t,j}\right| = O_P\left(\frac{\rho_T}{a_T(f)}\right). \tag{B.21}$$

We only prove (B.20) as the proof of (B.21) is analogous. Let

$$\Omega_{T,j}(x) = \frac{1}{(T-j)h}\sum_{t=j+1}^{T}K\left(\frac{X_{t-j}-x}{h}\right)\widetilde{e}_{t,j},$$

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

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

$$\begin{aligned}
&\max_{1\le j\le J(T)}\sup_{|x|\le C_T}\left|\Omega_{T,j}(x)\right| \\
&\quad\le \max_{1\le j\le J(T)}\sup_{|x|\le C_T}\left|\Omega_{T,j}^{(m)}(x)-\mathbb{E}\left[\Omega_{T,j}^{(m)}(x)\right]\right| \\
&\qquad+\max_{1\le j\le J(T)}\sup_{|x|\le C_T}\left|\Omega_{T,j}(x)-\Omega_{T,j}^{(m)}(x)\right| \\
&\qquad+\max_{1\le j\le J(T)}\sup_{|x|\le C_T}\left|\mathbb{E}\left[\Omega_{T,j}^{(m)}(x)\right]-\mathbb{E}\left[\Omega_{T,j}(x)\right]\right| \\
&\quad=: \Omega_T(1)+\Omega_T(2)+\Omega_T(3).
\end{aligned} \tag{B.22}$$

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

$$\Omega_T(2)+\Omega_T(3) = O_P(\rho_T) \tag{B.23}$$

as $J(T)h^{-(d+1)}\upsilon_1(m)\rho_T^{-1} = O(1)$. On the other hand, following the proof of Theorem 3.1, we can show that

$$\Omega_T(1) = O_P(\rho_T). \tag{B.24}$$

By (B.22)–(B.24), we can show that (B.20) holds.

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

$$\max_{1\le j\le J(T)}\sup_{|x|\le C_T}\Pi_{T,j,1}(x) = O_P\left(\frac{b_T h^2}{a_T(f)}\right). \tag{B.25}$$

Then, by (B.17), (B.19), and (B.25), we can prove Proposition 4.1. ■