NON-PARAMETRIC COINTEGRATING REGRESSION WITH NNH ERRORS

QIYING WANG AND YING XIANG RACHEL WANG

THE UNIVERSITY OF SYDNEY

THIS VERSION: DECEMBER 22, 2010

Abstract. This paper studies a non-linear cointegrating regression model with non-linear nonstationary heteroskedastic error processes. We establish uniform consistency for the conventional kernel estimate of the unknown regression function and develop a two-stage approach for the estimation of the heterogeneity generating function.

Key words and phrases: Cointegration, non-parametric regression, kernel estimate, non-stationarity, nonlinearity, heterogeneity.
JEL Classification: C14, C22.

1. Introduction

It is now well-known that the linear, time-invariant structure of the traditional cointegration model is often too restrictive and simplistic for modeling purposes. In most cases, even if a simple linear cointegrating relationship exists, it is unrealistic to assume this is not subject to any changes over the time span in which the variables are recorded. To address this issue, there is a large literature on parameter instability and time-varying coefficient models, see e.g. Park and Hahn (1999), Cai et al. (2009) and Xiao (2009), whilst Quintos and Phillips (1993) and Sun et al. (2008) considered the problem of testing for parameter constancy. It is worth noting that despite the various modifications, these models are linear in spirit.

An alternative to time-varying coefficient models are nonlinear models. Granger and Teräsvirta (1993) offers many empirical examples where nonlinear models are desirable. Indeed, given the prevalence of nonlinear relationships in economics, it is expected that nonlinear cointegration captures the features of many long-run relationships in a more realistic manner. Typical non-linear cointegrationg regression model has the form

$$y_t = f(x_t) + \epsilon_t, \qquad t = 1, 2, ..., n, \tag{1.1}$$

where $\{\epsilon_t\}$ is a zero mean equilibrium error, x_t is a non-stationary regressor and $f(\cdot)$ is an unknown function to be estimated with the observed data $\{y_t, x_t\}_{t=1}^n$.

With given observations (x_t, y_t) which may include non-stationary components, the issues on the estimation and inference of the unknown $f(\cdot)$ have been becoming increasing interests in past decade. In this regard, Phillips and Park (1998) studied nonparametric autoregression in the context of a random walk. Karlsen and Thostheim (2001) and Guerre (2004) studied nonparametric estimation for certain nonstationary processes in the framework of recurrent Markov chains. Karlsen, et al. (2007) developed an asymptotic theory for nonparametric estimation of a time series regression equation involving stochastically nonstationary time series. Karlsen, et al. (2007) address the function estimation problem for a possibly nonlinear cointegrating relation, providing an asymptotic theory of estimation and inference for nonparametric forms of cointegration. Under similar conditions and using related Markov chain methods, Schienle (2008) investigated additive nonlinear versions of (1.1) and obtained a limit theory for nonparametric regressions under smooth backfitting. More recently, Wang and Phillips (2009a, 2010) and Cai, et al. (2009) considered an alternative treatment by making use of local time limit theory and, instead of recurrent Markov chains, worked with partial sum representations of the type $x_t = \sum_{j=1}^t \xi_j$ where ξ_j is a general linear process. In another paper, Wang and Phillips (2009b) considered the errors u_t to be serially dependent and cross correlated with the regressor x_t for small lags. Other related current works, we refer to Kasparis and Phillips (2009), Park and Phillips (1999, 2001), Bandi (2004), Gao, et al (2009a, b), Choi and Saikkonen (2004, 2009) and Marmer (2008).

Whilst an extensive literature exists for the estimation theory of $f(\cdot)$, less attention has been paid to studying the the error structure, which belongs to a subclass of stochastic volatility processes. Just as it is unreasonable to assume the cointegrating relationship remains linear and unchanged throughout time, the long-run equilibrium error is unlikely to be homoskedastic over a long time span. It is well known that stock markets have variances that change substantially with time, rendering them much more volatile than many economic variables that can be used to explain their movements. In these situations, it makes sense to incorporate time heterogeneity into error term and hence motivates us to investigate the model (1.1) with error structure:

$$\epsilon_t = \sigma(x_t)\, u_t, \tag{1.2}$$

where $\sigma(\cdot)$ is a heterogeneity generating function (HGF) as called in Park (2002) and for a filtration $\mathcal{F}_t$ to which x_{t+1} is adapted, $\{u_t,\ \mathcal{F}_t\}$ forms a martingale difference.

This incorporation idea is not new in stationary regression models. Various error structures that generate heteroskedasticity can be found in e.g. Robinson (1987), Harvey and Robinson (1988) and Hansen (1995). On the nonstationary front, Kim and Park (2010) investigated a linear cointegration model in which the conditional heteroskedasticity of the error is generated by a smooth deterministic function of time. The error process considered in this paper is known as a nonlinear nonstationary heteroskedastic (NNH) process first proposed by Park (2002), where the author examined the asymptotic behaviour of the sample statistics of NNH processes generated by integrable and asymptotically homogeneous functions and pointed out that NNH models offer an attractive alternative to ARCH type volatility models because they not only demonstrate the desirable properties of volatility clustering and leptokurtosis, which are common traits in financial and economic series, but also allow for the possibility of explaining the source of volatility in terms of economic variables. In the related works, the NNH error structure is incorporated into a stationary regression model and a linear cointegration model by Chung and Park (2007). Without involving the cointegrating relationship, Han and Park (2008) handled with the NNH model with an ARCH component, Han and Zhang (2009) investigated the asymptotics for the kernel estimator of $\sigma(x)$.

The main purpose of the present paper is to estimate the HGF in the model given by (1.1) and (1.2). We consider a kernel-type estimator $\hat{\sigma}(x)$ of $\sigma(x)$. It is shown that $\hat{\sigma}(x)$ converges to $\sigma(x)$ in probability uniformly over a compact set, and limit distribution is normal under appropriate normalization. New technical mechanisms are developed to investigate the asymptotics for $\hat{\sigma}(x)$. In particular, we establish uniform consistency for the conventional kernel estimate of the unknown regression function and consider the lower bound for a class of functionals of non-stationary data over a compact set. These results are interesting in the right of themselves and will be useful in other related research fields.

We finally remark that the two-step kernel-type estimator $\hat{\sigma}(x)$ considered in this paper is quite natural and simple. Our main result provides a theoretical framework which allows the extension from the stationary time series to near $I(1)$ process. Uniform consistency is important not only in estimation theory, but also useful in other fields like model specification testing. Existing studies mainly focus on the observations coming from a stationary data set. A recent work in this regard can be found in Hansen (2008), where readers can also find a sequence of the related references. Gao, Li and Tjostheim (2010) currently investigated uninform consistency of non-parametric kernel estimators for non-stationary time series. In their work, the authors considered the situation that the regressor is a recurrent Markov chain, and assumed that the error term is independent of the regressor. The setting on the model in this paper is different. We consider a near integrated process with innovations being a linear process, and we remove the restriction on the independence between the regressor and the error terms.

The remaining of the paper is organized as follows. Section 2 explains the model in more details and defines some variables that are central to our development of the asymptotics. Section 3 outlines the asymptotic properties of the kernel estimator for both f and σ. Section 4 makes a conclusion. All the mathematical proofs are contained in Section 5.

2. The model and assumptions

To estimate the HGF $\sigma(x)$, we re-organize the model driven by (1.1) and (1.2) as follows:

$$\begin{aligned} y_t &= f(x_t) + \sigma(x_t)\, u_t, && (2.1)\\ \big[y_t - f(x_t)\big]^2 &= \sigma^2(x_t) + \sigma^2(x_t)(u_t^2 - 1). && (2.2) \end{aligned}$$

As $u_t^2 - 1$ may form a martingale difference, from the observations on x_t and $y_t - \hat{f}(x_t)$, where $\hat{f}(x)$ is the first stage estimate of $f(x)$ in model (2.1), the conventional Kernel estimate of $\sigma^2(x)$ in model (2.2) is given by

$$\hat{\sigma}^2(x) = \frac{\sum_{t=1}^n [y_t - \hat{f}(x_t)]^2 K[(x_t - x)/h]}{\sum_{t=1}^n K[(x_t - x)/h]}, \tag{2.3}$$

where $K(x)$ is a nonnegative real function and the bandwidth parameter $h \equiv h_n \to 0$ as $n \to \infty$.

The limit behavior of $\hat{\sigma}(x)$ depends on the choice of $\hat{f}(x)$. The present paper adopts the Nadaraya-Watson estimator defined by

$$\hat{f}(x) = \frac{\sum_{s=1}^n y_s\, K[(x_s - x)/h]}{\sum_{s=1}^n K[(x_s - x)/h]}, \tag{2.4}$$

where the kernel $K(x)$ and the bandwidth h are chosen to be the same as in (2.3) for the technical convenience.

Throughout the paper, let $\{\xi_j, j \geq 1\}$ be a linear process defined by

$$\xi_j = \sum_{k=0}^{\infty} \phi_k\, \epsilon_{j-k}, \tag{2.5}$$

where $\{\epsilon_j, -\infty < j < \infty\}$ is a sequence of iid random variables with $E\epsilon_0 = 0$, $E\epsilon_0^2 = 1$ and characteristic function $\varphi(t)$ of ϵ_0 satisfying $\int_{-\infty}^{\infty} |\varphi(t)| dt < \infty$. The coefficients $\phi_k, k \geq 0$, are assumed to satisfy $\sum_{k=0}^{\infty} |\phi_k| < \infty$ and $\phi \equiv \sum_{k=0}^{\infty} \phi_k \neq 0$. The absolute summable condition is commonly used literature. In the simplest case, it includes the example in which x_t is a random walk on iid random variables. We also use the following assumptions in the asymptotic development.

Assumption 1. $x_t = \rho\, x_{t-1} + \xi_t$, $(x_0 \equiv 0)$, where $\rho = 1 + \tau/n$ with τ being a constant.

Assumption 2. (i) $\{u_t, \mathcal{F}_t, 1 \le t \le n\}$ is a martingale difference sequence satisfying $E(u_t^2|\mathcal{F}_{t-1}) = 1$, $E(u_t^4|\mathcal{F}_{t-1}) \to \Lambda^2$, a.s., where $\Lambda^2 > 1$ is a constant, and $\sup_{1\le t\le n} E(|u_t|^{4\nu}|\mathcal{F}_{t-1}) < \infty$ a.s. for some $\nu > 1$. (ii) x_t is adapted to $\mathcal{F}_{t-1}$, $t = 1, 2, ..., n$.

Assumption 3. The kernel K satisfies that $\int_{-\infty}^{\infty} K(s)ds = 1$, $K(\cdot)$ has a compact support and for any $x, y \in R$,

$$|K(x) - K(y)| \le C\,|x - y|,$$

where C is a positive constant.

Assumption 4. For a compact set $\Omega \subset R$, there exist real constants $\epsilon > 0$, $0 < \alpha, \beta \le 1$ and $C_1, C_2 > 0$ such that, $\sigma(x)$ is non-negative and

$$|f(y) - f(x)| \;\le\; C_1|y - x|^{\alpha}, \quad |\sigma(y) - \sigma(x)| \;\le\; C_2|y - x|^{\beta},$$

for any $x, y \in \Omega_\epsilon$, where $\Omega_\epsilon = \{y : |y - x| \le \epsilon, \text{where } x \in \Omega\}$.

Assumption 1 allows for both non-stationary ($\tau = 0$) and near non-stationary ($\tau \ne 0$) regressor, and is standard in the near integration regression framework. See, e.g., Phillips (1987, 1988), Chan and Wei (1987) and Wang and Phillips (2009b).

Assumption 2 is standard as in the stationary situation in which we impose a martingale structure so that $cov(u_{t+1}, x_t) = E[x_t E(u_{t+1} \mid \mathcal{F}_t)] = 0$. The restrictions that $E(u_t^2|\mathcal{F}_{t-1}) = 1$ and $E(u_t^4|\mathcal{F}_{t-1}) \to \Lambda^2$, a.s., are required because of the model structure (2.2). These restrictions are not necessary in the investigation of the asymptotics related to $\hat{f}(x)$, even on the uniform convergence. In the later situation, we use a less restricted Assumption 2*. See Theorem 3.2 in next section.

Assumption 3 is a standard condition on $K(x)$ as in the stationary situation. The Lipschitz condition on $K(x)$ is not necessary if we only investigate the point-wise asymptotics. See Remark 3.3 for further details.

Assumption 4 requires a Lipschitz-type condition in a small neighborhood of the targeted compact set for the functionals to be estimated. This condition is quite weak, which may host a wide set of functionals. Typical examples include that $f(x) = \theta_1 + \theta_2 x + ... + \theta_k x^{k-1}$; $f(x) = \alpha + \beta\, x^{\gamma}$; $f(x)) = x(1+\theta x)^{-1} I(x \ge 0)$; $f(x) = (\alpha + \beta\, e^x)/(1 + e^x)$. $\sigma(x)$ can have the same form as in $f(x)$, but we require that the x and the parameters are restricted on a space so that $\sigma(x)$ is non-negative. Also notice that Assumption 4 implies that

$$\sup_{x\in\Omega_\epsilon}|f(x)|<\infty \quad \text{and} \quad \sup_{x\in\Omega_\epsilon}\sigma(x)<\infty.$$

This fact will be repeatedly used in the proofs of main results without further explanation.

3. Main results and outline of the proofs

We have the following asymptotic results for the estimation of the HGF $\sigma(x)$ given in the model (2.1) and (2.2). Except explicitly mentioned, we denote by $[a]$ the integer part of a positive constant a in the following.

Theorem 3.1. *Under Assumptions 1-4, for any h satisfying $h\to 0$ and $nh^{2+4/[\nu]}\to\infty$, we have*

$$\sup_{x\in\Omega}|\hat\sigma^2(x)-\sigma^2(x)| \;=\; O_P\Big\{\big(nh^{2+4/[\nu]}\big)^{-[\nu]/[2(2[\nu]+1)]}+h^{\min\{\alpha,\beta\}}\Big\}, \tag{3.1}$$

where Ω is defined as in Assumption 4. Furthermore, if in addition u_t are independent of x_t, then for any h satisfying $nh^2/\log^4 n\to\infty$ and $nh^{2+4\min\{2\alpha,\beta\}}\to 0$,

$$(T_n(z_1),T_n(z_2),\ldots,T_n(z_k))\to_d N(0,\Sigma), \tag{3.2}$$

where $z_1,...,z_k\in\Omega$ all are different,

$$T_n(z_j)=\big\{\sum_{t=1}^n K\big[(x_t-z_j)/h\big]\big\}^{1/2}(\hat\sigma^2(z_j)-\sigma^2(z_j)),$$

and

$$\Sigma=diag\big\{\sigma^4(z_1),\ldots,\sigma^4(z_k)\big\}\,(\Lambda^2-1)\int_{-\infty}^{\infty}K^2(s)ds.$$

Remark 3.1. *The result (3.1) shows that $\hat\sigma^2(x)$ is a consistent estimator of $\sigma^2(x)$ uniformly over a compact set. If the error term u_t is assumed to be bounded by a constant, the restriction on the choice of h can be relaxed and convergence rate in the result (3.1) can be improved. Indeed, if in addition $|u_t|\le C$, then for any h satisfying $h\to 0$ and $nh^2/\log^4 n\to\infty$,*

$$\sup_{x\in\Omega}|\hat\sigma^2(x)-\sigma^2(x)| \;=\; O_P\big[(nh^2)^{-1/4}\log n+h^{\min\{\alpha,\beta\}}\big]. \tag{3.3}$$

The result (3.2) provides a distributional property of the estimator $\hat\sigma^2(x)$, by imposing the independence between u_t and x_t. Note that $\hat\sigma^2(x)$ involves the cross terms in related to u_t and x_s for all $1\le s\le n$ (see (3.4) below) by through $\hat f(x_t)$. It is not clear at the moment *for how to provide an accurate estimation for these cross terms, which are required in the investigation of distribution convergence, without the independence assumption.*

Remark 3.2. *To outline the essentials of the arguments in the proof of Theorem 3.1, we split* $\hat\sigma^2(x) - \sigma^2(x)$ *as*

$$\begin{aligned} & \hat\sigma^2(x) - \sigma^2(x) \\ = \; & \frac{\sum_{t=1}^n \sigma^2(x)\,(u_t^2-1)\,K[(x_t-x)/h]}{\sum_{t=1}^n K[(x_t-x)/h]} + \frac{\sum_{t=1}^n \left[\sigma^2(x_t)) - \sigma^2(x)\right] u_t^2\, K[(x_t-x)/h]}{\sum_{t=1}^n K[(x_t-x)/h]} \\ & + \frac{\sum_{t=1}^n (f(x_t) - \hat f(x_t))^2 K[(x_t-x)/h]}{\sum_{t=1}^n K[(x_t-x)/h]} + \frac{2\sum_{t=1}^n \sigma(x_t) u_t (f(x_t) - \hat f(x_t)) K[(x_t-x)/h]}{\sum_{t=1}^n K[(x_t-x)/h]} \\ := \; & I_{1n}(x) + I_{2n}(x) + I_{3n}(x) + I_{4n}(x). \end{aligned} \tag{3.4}$$

When $I_{2n}(x)$ *is easy to handle with, the estimates of other terms over a compact set requires new mechanisms. In particular, we need uniform asymptotics of* $\hat f(x)$ *on* Ω_ϵ *for a small* ϵ*. The result in this regards is presented in the following theorem, which is interesting in itself.*

The following assumption is weaker than Assumption 2.

Assumption 2*. (i) $\{u_t, \mathcal{F}_t, 1 \le t \le n\}$ is a martingale difference sequence satisfying $\sup_{1\le t\le n} E(|u_t|^{2p}|\mathcal{F}_{t-1}) < \infty$ a.s. for some integer $p \ge 1$. (ii) x_t is adapted to $\mathcal{F}_{t-1}$, $t = 1, 2, ..., n$.

Theorem 3.2. *Under Assumptions 1, 2*, 3 and 4, for any* h *satisfying* $h \to 0$ *and* $nh^{2+4/p} \to \infty$*, we have*

$$\sup_{x\in\Omega_\epsilon} |\hat f(x) - f(x)| = O_P\left\{ \left(nh^{2+4/p}\right)^{-p/[2(2p+1)]} + h^{\min\{\alpha,\beta\}} \right\}, \tag{3.5}$$

where Ω_ϵ *is defined as in Assumption 4. If in addition* $|u_t| \le C$*, then, for any* h *satisfying* $h \to 0$ *and* $nh^2/\log^4 n \to \infty$*,*

$$\sup_{x\in\Omega_\epsilon} |\hat f(x) - f(x)| = O_P\left\{ (nh^2)^{-1/4}\log n + h^{\min\{\alpha,\beta\}} \right\}. \tag{3.6}$$

Remark 3.3. *A better result can be obtained if we are only interested in the point-wise asymptotics for* $\hat f(x)$*. Indeed, as in Wang and Phillips (2009a, b) with minor modification, we may show that, for each fixed* $x \in \Omega$*,*

$$\hat f(x) - f(x) \;=\; O_P\left\{ (nh^2)^{-1/4} + h^{\min\{\alpha,\beta\}} \right\}. \tag{3.7}$$

Furthermore $\hat{f}(x)$ *has an asymptotic distribution that is mixing normal, under minor additional conditions to Assumption* 2^*. *More details are referred to Wang and Phillips (2009a, b).*

It is interesting to note that, in point-wise situation, the term $h^{\min\{\alpha,\beta\}}$ *can be improved if we put a bias term in the left hand of (3.7). It is not clear if there are similar properties in related to uniform consistency.*

Remark 3.4. *It is interesting to notice that, in stationary situation, the sharp rate of convergence in (3.6) is* $O_P[(\log n/nh)^{1/2}]$. *See Hansen (2008), for instance. There is an essential difference for the rate of convergence between stationary and non-stationary time series. The reason behind the difference is mainly because, in non-stationary case, the amount of time spent by the process around any specific point is of order* $\sqrt{n}$ *rather than* n. *More explanation can be found in Remark 3.3 of Wang and Phillips (2009a).*

The proof of Theorem 3.2 is quite technical. For the sake of reading convenience, we separate the key steps into the following propositions, which are interesting in themselves.

Proposition 3.1. *Under Assumptions 1, 2* and 3, for any compact set* Ω^* *and* h *satisfying* $h \to 0$ *and* $nh^{2+4/p} \to \infty$,

$$\sup_{x\in\Omega^*} \frac{\sum_{t=1}^n u_t\, K[(x_t-x)/h]}{\sum_{t=1}^n K[(x_t-x)/h]} = O_P\big[\big(nh^{2+4/p}\big)^{-p/[2(2p+1)]}\big] \tag{3.8}$$

If in addition $|u_j| \le C$ *and* $nh^2/\log^4 n \to \infty$, *the result (3.8) can be improved to*

$$\sup_{x\in\Omega^*} \frac{\sum_{t=1}^n u_t\, K[(x_t-x)/h]}{\sum_{t=1}^n K[(x_t-x)/h]} \;=\; o_P\big[(nh^2)^{-1/4}\log n\big]. \tag{3.9}$$

Proposition 3.2. *Under Assumptions 1 and 3, for any compact set* Ω^* *and* $\eta > 0$, *there exist* $n_0 > 0$ *and* $M_0 > 0$ *such that, for all* $n \ge n_0$, $M \ge M_0$ *and* h *satisfying* $h \to 0$ *and* $nh^2 \log^{-4} n \to \infty$,

$$P\Big(\inf_{x\in\Omega^*} \sum_{t=1}^n K[(x_t-x)/h] \ge \sqrt{n}h\,/M\Big) \;\ge\; 1-\eta. \tag{3.10}$$

Remark 3.5. *Gao, Li and Tjostheim (2009) investigated a similar uninform consistency as in Theorem 3.2 in the situation that the regressor* x_t *is a recurrent Markov chain and the error term* u_t *is independent of* x_t. *Not only the setting on the model, the techniques used in the current paper are also different from Gao, Li and Tjostheim (2009). We develop our own techniques by establishing the key result like (3.10). These key results may be useful in other related fields. As in Gao, Li and Tjostheim (2009), it is possible to extend our result from the compact set* Ω_ϵ *to a range that* $|x| \le T_n$, *where* $T_n \to \infty$.

However, this kind of extensions may require quite different techniques. Since Theorem 3.2 is enough for the purpose of this paper, we leave the extensions for future research.

4. Conclusion

In this paper, we consider a non-linear cointegrating regression model with non-linear nonstationary heteroskedastic (NNH) error processes. A two-stage approach is proposed to find an estimator $\hat{\sigma}(x)$ of the heterogeneity generating function $\sigma(x)$. For a wide class of $\sigma(x)$, it is shown that $\hat{\sigma}(x)$ is consistent uniformly over a compact set and has an asymptotic distribution that is normal under appropriate normalization. New technical mechanisms are developed to investigate the asymptotics for $\hat{\sigma}(x)$. In particular, we establish uniform consistency for the conventional kernel estimate of the unknown regression function and consider the lower bound for a class of functionals of non-stationary data over a compact set. These results are interesting in the right of themselves and will be useful in other related research fields.

5. Proofs of main results

This section provides proofs of the main results. We start with Proposition 3.2. The techniques used in the proof of this proposition also provide some insights to other main results. Throughout this section, we denote constants by $C, C_1, C_2, ...$, which may be different at each appearance.

Proof of Proposition 3.2. Without loss of generality, assume $\phi_0 \neq 0$. Otherwise the proof follows from a routine modification. Let $a < b$ be integers such that $\Omega^* \subseteq [a, b]$ and

$$y_j = a + j\,(b-a)m_n^{-1}, \quad j = 0, 1, 2, ..., m_n, \tag{5.1}$$

where $m_n = [(nh^2)^{1/4}h^{-2}]$. Write $\mathcal{F}_t^* = \sigma(\epsilon_s, s \leq t)$ for $t \geq 1$. Note that, for any $M > 0$,

$$\inf_{x\in\Omega^*} \sum_{t=1}^n E\big(K[(x_t - x)/h] \mid \mathcal{F}_{t-1}^*\big) \leq \inf_{x\in\Omega^*} \sum_{t=1}^n K[(x_t - x)/h] + \sup_{x\in[a,b]} \Big| \sum_{t=1}^n \big\{K[(x_t - x)/h] - E\big(K[(x_t - x)/h] \mid \mathcal{F}_{t-1}^*\big)\big\}\Big|, \tag{5.2}$$

and the second term in right hand of (5.2) is less than

$$\begin{aligned} & \max_{0\le j\le m_n-1} \sup_{x\in[y_j,y_{j+1}]} \sum_{t=1}^n \Big\{\big|K[(x_t-x)/h]-K[(x_t-y_j)/h]\big| \\ & \qquad\qquad +E\Big(\big|K[(x_t-x)/h]-K[(x_t-y_j)/h]\big| \mid \mathcal{F}^*_{t-1}\Big)\Big\} \\ & + \max_{0\le j\le m_n} \Big|\sum_{t=1}^n \big\{K[(x_t-y_j)/h]-E\big(K[(x_t-y_j)/h] \mid \mathcal{F}^*_{t-1}\big)\big\}\Big| \\ := & \ \lambda_{1n}+\lambda_{2n}. \end{aligned}$$

It is readily seen from $nh^2\log^{-4} n\to\infty$ that (3.10) will follow if we prove

$$\lambda_{1n} = O_P\big[(nh^2)^{1/4}\log n\big], \tag{5.3}$$

$$\lambda_{2n} = O_P\big[(nh^2)^{1/4}\log n\big], \tag{5.4}$$

and for any $\eta>0$, there exist $n_0>0$ and $M_0>0$ such that, for all $n\ge n_0$ and $M\ge M_0$,

$$P\big(\inf_{x\in\Omega^*}\sum_{t=1}^n E\big(K[(x_t-x)/h] \mid \mathcal{F}^*_{t-1}\big)\ge \sqrt{n}h/M\big) \ \ge \ 1-\eta. \tag{5.5}$$

To establish (5.3)-(5.5), we start with some preliminaries. First notice that

$$\begin{aligned} x_t &= \sum_{j=1}^t \rho^{t-j}\,\xi_j=\sum_{j=1}^t \rho^{t-j}\sum_{i=-\infty}^{j}\epsilon_i\phi_{j-i} \\ &= x_t^*+x_t'+\phi_0\,\epsilon_0, \end{aligned} \tag{5.6}$$

where $x_t^*=\sum_{j=1}^t\rho^{t-j}\sum_{i=-\infty}^0\epsilon_i\phi_{j-i}$ depends only on $(...,\epsilon_{-1},\epsilon_0)$ and

$$x_t'=\sum_{j=1}^t\rho^{t-j}\sum_{i=1}^j\epsilon_i\phi_{j-i}-\phi_0\,\epsilon_0=\sum_{i=1}^{t-1}\epsilon_i\sum_{j=0}^{t-i}\rho^{t-j-i}\,\phi_j.$$

Write $a_{t,i}=\rho^{t-i}\sum_{j=0}^{t-i}\rho^{-j}\,\phi_j$ and $d_t^2=\sum_{i=1}^{t-1}a_{t,i}^2=E(x_t')^2$. The result (7.14) in Wang and Phillips (2009b) shows that $a_{t,i}\ge e^{-|\tau|}\,|\phi|/4$ whenever $1\le i\le t/2$ and $t\le n$ is sufficiently large. This implies that $d_t^2\asymp t$, where the notation $a_n\asymp b_n$ denotes that there exist constants $c_1,c_2>0$ such that $c_1\le a_n/b_n\le c_2$, and as in the proof of Corollary 2.2. of Wang and Phillips (2009a),

$$\int_{-\infty}^{\infty}|Ee^{i\lambda x_t/\sqrt{t}}|d\lambda \ \le \ C\int_{-\infty}^{\infty}|Ee^{i\lambda x_t'/\sqrt{t}}|d\lambda<\infty, \tag{5.7}$$

$$\int_{-\infty}^{\infty}|Ee^{i\lambda (x_t^*+x_t')/\sqrt{t}}|d\lambda \ \le \ C\int_{-\infty}^{\infty}|Ee^{i\lambda x_t'/\sqrt{t}}|d\lambda<\infty, \tag{5.8}$$

uniformly for $t\ge 1$. The results (5.7) and (5.8) imply that,

F. $x_t/\sqrt{t}$ and $(x_t^* + x_t')/\sqrt{t}$ have densities $\nu_{1t}(x)$ and $\nu_{2t}(x)$ respectively, and both $\nu_{1t}(x)$ and $\nu_{2t}(x)$ are uniformly bounded on t and x by a constant C.

See, e.g., Lukács (1970, Thm 3.2.2). Furthermore, by defining $\sum_k^j = 0$ if $j < k$,

$$I_1(\lambda) = E\exp\Big\{i\lambda \sum_{q=s}^{t-1} \epsilon_q\, a_{t,q}\Big\}, \quad I_2(\lambda,\lambda_1) = E\exp\Big\{i\sum_{q=1}^{s-1}\epsilon_q\,\big(\lambda\, a_{t,q} - \lambda_1\, a_{s,q}\big)\Big\}, \quad (5.9)$$

for $s \le t$, the same arguments as in the proof of Lemma 7.2 of Wang and Phillips (2009b) yield that there exist $\gamma_1, \gamma_2 > 0$ such that

$$I_1(\lambda) \le \begin{cases} e^{-\gamma_1\,(t-s)} & \text{if } |\lambda| \ge 1, \\ e^{-\gamma_2\lambda^2(t-s)} & \text{if } |\lambda| \le 1, \end{cases} \quad (5.10)$$

$$I_2(\lambda,\lambda_1) \le \exp\Big\{-\gamma_1\#(\Omega_1) - \gamma_2\, B_1\,(\lambda_1 - \lambda B_2/B_1)^2\Big\}, \quad (5.11)$$

where Ω_1 (Ω_2, respectively) denotes the set of $1 \le q \le s/2$ such that $|\lambda\, a_{t,q} - \lambda_1\, a_{s,q}| \ge 1$ ($|\lambda\, a_{t,q} - \lambda_1\, a_{s,q}| < 1$, respectively), and

$$B_1 = \sum_{q\in\Omega_2} a_{s,q}^2 \quad \text{and} \quad B_2 = \sum_{q\in\Omega_2} a_{t,q}a_{s,q}.$$

Note that $\Omega_1 + \Omega_2 = s/2$ and recall that $a_{s,q} \ge e^{-|\tau|}\,|\phi|/4$ for all sufficiently large s. It is readily seen from (5.11) that

$$I_2(\lambda,\lambda_1) \le e^{-\gamma_1\sqrt{s}} I_{(\#(\Omega_1)\ge\sqrt{s})} + e^{-\gamma_2'\,s\,(\lambda_1-\lambda B_2/B_1)^2} I_{(\#(\Omega_1)\le\sqrt{s})}, \quad (5.12)$$

for some $\gamma' > 0$.

We are now ready to prove (5.3)-(5.5).

Start with (5.3). Recall that $K(x)$ has a compact support. Without loss of generality, suppose $K(x) = 0$ when $x \notin [c,d]$. Since $x_t/\sqrt{t}$ has a uniformly bounded density $\nu_{1t}(x)$ by the fact **F**, it is readily seen from Assumption 3 that

$$\begin{aligned} E\lambda_{1n} &\le 2\sum_{t=1}^n E \max_{0\le j\le m_n-1}\ \sup_{x\in[y_j,y_{j+1}]} \big|K[(x_t-x)/h] - K[(x_t-y_j)/h]\big| \\ &\le 2\sum_{t=1}^n \int_{-\infty}^{\infty} \max_{0\le j\le m_n-1}\ \sup_{x\in[y_j,y_{j+1}]} \big|K[(\sqrt{t}\,y-x)/h] - K[(\sqrt{t}y-y_j)/h]\big|\nu_{1t}(y)dy \\ &\le 2\sum_{t=1}^n \frac{h}{\sqrt{t}} \int_{-|a|/h-|c|}^{|b|/h+|d|} \max_{0\le j\le m_n-1}\ \sup_{x\in[y_j,y_{j+1}]} \big|K(y-x/h) - K(y-y_j/h)\big|\nu_{1t}(yh/t)dy \\ &\le C\sum_{t=1}^n \frac{h^{-1}}{\sqrt{t}} \max_{0\le j\le m_n-1} |y_{j+1}-y_j| \\ &\le C\sqrt{n}\,m_n^{-1}\,h^{-1} \le C\,(nh^2)^{1/4}. \end{aligned} \quad (5.13)$$

Hence, $P(\lambda_{1n} \geq (nh^2)^{1/4} \log n) \leq C/\log n$, which yields the required (5.3).

To prove (5.4), write $Z_{jt} = K[(x_t - y_j)/h] - E\big(K[(x_t - y_j)/h] \mid \mathcal{F}_{t-1}^*\big)$. We first claim that

$$\begin{aligned}\max_{0\leq j\leq m_n} \sum_{t=1}^n E\big(Z_{jt}^2 \mid \mathcal{F}_{t-1}^*\big) &\leq 2\sum_{t=1}^n \max_{0\leq j\leq m_n} E\big(K^2[(x_t - y_j)/h] \mid \mathcal{F}_{t-1}^*\big) \\ &= O_P(\sqrt{n}h). \qquad (5.14)\end{aligned}$$

In fact, by recalling that $\int_{-\infty}^{\infty} |Ee^{it\epsilon_0}|dt < \infty$, the $\phi_0\epsilon_0$ has a density, say, $d(x)$. This, together with (5.6) and the independence between ϵ_0 and $x_{1t} = x_t^* + x_t'$, yields that

$$\sup_{x\in\Omega^*} E\big(K^2[(x_t - x)/h] \mid \mathcal{F}_{t-1}^*\big) \leq C \int_{-\infty}^{\infty} K[(x_{1t} + y)/h] \sup_{x\in\Omega^*} d(y + x)\, dy. \qquad (5.15)$$

Hence, by noting

$$EK[(x_{1t} + y)/h] = \int_{-\infty}^{\infty} K\big(s\sqrt{t}/h + y/h\big)\nu_{2t}(s)ds \leq C\, h/\sqrt{t},$$

uniformly on $y \in R$ because of the fact **F**, it follows that

$$\begin{aligned}E\Big[\max_{0\leq j\leq m_n} \sum_{t=1}^n E\big(Z_{jt}^2 \mid \mathcal{F}_{t-1}^*\big)\Big] &\leq 2\sum_{t=1}^n \int_{-\infty}^{\infty} EK[(x_{1t} + y)/h] \sup_{x\in\Omega^*} d(y + x)\, dy \\ &\leq C\, h \sum_{t=1}^n t^{-1/2} \leq C_1 \sqrt{n}\, h, \qquad (5.16)\end{aligned}$$

where we have used the fact that $\int_{-\infty}^{\infty} \sup_{x\in\Omega^*} d(y + x)\, dy < \infty$, as Ω^* is a compact set. The claim (5.14) follows from (5.16) by Markov's inequality..

By virtue of (5.14), it follows from $|Z_{jt}| \leq C$, $nh^2 \log^{-4} n \to \infty$ and an exponential inequality for martingale given by de la Pena (1999) that, for any $0 < \delta < 1$,

$$\begin{aligned}&P(\lambda_{2n} \geq (\sqrt{n}h)^{1/2} \log n) \\ \leq\ & P\Big[\lambda_{2n} \geq (\sqrt{n}h)^{1/2} \log n, \quad \max_{0\leq j\leq m_n} \sum_{t=1}^n E\big(Z_{jt}^2 \mid \mathcal{F}_{t-1}^*\big) \leq \sqrt{n}h \log^\delta n\Big] + o(1) \\ \leq\ & \sum_{j=0}^{m_n} P\Big[\sum_{t=1}^n Z_{jt} \geq (\sqrt{n}h)^{1/2} \log n, \quad \sum_{t=1}^n E\big(Z_{jt}^2 \mid \mathcal{F}_{t-1}^*\big) \leq \sqrt{n}h \log^\delta n\Big] + o(1) \\ \leq\ & C(nh^2)^{1/4} h^{-2} \exp\Big\{ - \frac{\sqrt{n}h \log^2 n}{2\sqrt{n}h \log^\delta n + C(\sqrt{n}h)^{1/2} \log n}\Big\} + o(1) \\ \leq\ & C(nh^2)^{1/4} h^{-2}\, e^{-C_1 \log^{1-\delta} n} + o(1) = o(1), \qquad (5.17)\end{aligned}$$

which implies the required (5.4).

The proof of (5.5) is more laborious. First note that, as in the proof of (5.15),

$$\begin{aligned}\inf_{x\in\Omega^*} E\big(K[(x_t-x)/h] \mid \mathcal{F}^*_{t-1}\big) &= \inf_{x\in\Omega^*}\int_{-\infty}^{\infty} K[(x_{1t}+y-x)/h]d(y)\,dy \\ &\geq \int_{|y|\leq A_0} K[(x_{1t}+y)/h]\ \inf_{x\in\Omega^*} d(y+x)\,dy,\end{aligned}$$

where A_0 is a constant chosen such that $I := \int_{|y|\leq A_0} \inf_{x\in\Omega^*} d(y+x)\,dy > 0$. Since

$$P\Big(\inf_{x\in\Omega^*}\sum_{t=1}^n E\big(K[(x_t-x)/h] \mid \mathcal{F}^*_{t-1}\big) \geq \sqrt{n}h\,/M\Big) \ \geq\ P(\Gamma_n \geq \sqrt{n}h\,/M),$$

where

$$\Gamma_n = \int_{|y|\leq A_0}\sum_{t=1}^n K[(x_{1t}+y)/h]\ \inf_{x\in\Omega^*} d(y+x)\,dy,$$

(5.5) will follow if we prove

$$\frac{\Gamma_n}{\sqrt{n}h} \to_D \Gamma, \tag{5.18}$$

where Γ is a random variable satisfying $P(\Gamma > 0) = 1$. To prove (5.18), we first assume that

Con: $K(x)$ is continuous and $\hat{K}(x)$ has a compact support, where $\hat{K}(x) = \int_{-\infty}^{\infty} e^{ixt}K(t)dt$. (5.19)

This restriction will be removed later. Under condition (5.19), we have $K(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i\lambda x}\hat{K}(-\lambda)d\lambda$. This implies that, for each $A > 0$,

$$\begin{aligned}\Gamma_n &= \int_{|y|\leq A_0}\big[V_{1yA}+V_{2yA}\big]\ \inf_{x\in\Omega^*} d(y+x)\,dy \\ &:= \Gamma_{1nA}+\Gamma_{2nA},\end{aligned} \tag{5.20}$$

where, with $c_n = \sqrt{n}/h$,

$$\begin{aligned}V_{1yA} &= \frac{1}{2\pi\, c_n}\sum_{t=1}^n\int_{|\lambda|\geq A} e^{i\lambda(x_{1t}+y)/\sqrt{n}}\hat{K}(-\lambda/c_n)d\lambda, \\ V_{2yA} &= \frac{1}{2\pi\, c_n}\sum_{t=1}^n\int_{|\lambda|< A} e^{i\lambda(x_{1t}+y)/\sqrt{n}}\hat{K}(-\lambda/c_n)d\lambda\end{aligned}$$

Recall that $0 < I = \int_{|y|\le A_0} \inf_{x\in\Omega^*} d(y+x)\,dy \le \int \inf_{x\in\Omega^*} d(y+x)\,dy < \infty$, since Ω^* is a compact set. It is readily seen that, to prove (5.18), it suffices to show that

$$\sup_{|y|\le A_0} E|V_{1yA}| = o(\sqrt{n}\,h), \tag{5.21}$$

and on a richer probability space which also holds a standard Brownian motion $W(t)$ such that

$$\sup_{|y|\le A_0} \Big|\frac{V_{2yA}}{\sqrt{n}h} - \frac{1}{\phi}\,L_G(1,0)\Big| \;\;=\;\; o_P(1), \tag{5.22}$$

as $n\to\infty$ first and then $A\to\infty$, where $G(t) = W(t) + \kappa\int_0^t e^{\kappa(t-s)}W(s)ds$ and $L_G(r,x)$ is a local time of the Gaussian process $G(t)$ expressed as

$$L_G(t,x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\,ux}\int_0^t e^{iuG(s)}ds\,du. \tag{5.23}$$

Here and below the right hand of (5.23) is understood as a limitation in L_2 of $L_G^N(t,x)$, as $N\to\infty$, defined as

$$L_G^N(t,x) = \frac{1}{2\pi}\int_{-N}^{N} e^{-i\,ux}\int_0^t e^{iuG(s)}ds\,du.$$

The process $\{L_\zeta(t,s), t\ge 0, s\in R\}$ is said to be the local time of a measurable process $\{\zeta(t), t\ge 0\}$ if, for any locally integrable function $T(x)$,

$$\int_0^t T[\zeta(s)]ds = \int_{-\infty}^{\infty} T(s)L_\zeta(t,s)ds, \quad \text{all } t\in R, \tag{5.24}$$

with probability one. For more details in these regards, we refer to Berman (1969).

By recalling $x_{1t} = x_t^* + x_t'$ and the independence between x_t^* and x_t', we have that

$$\begin{aligned} E|V_{1yA}|^2 \;&\le\; \frac{Ch^2}{n}\sum_{s,t=1}^n \int_{|\lambda|\ge A}\int_{|\lambda_1|\ge A} |\hat K(-\lambda/c_n)|\,|\hat K(-\lambda_1/c_n)|\,\big|Ee^{i(\lambda\,x_{1t}-\lambda_1\,x_{1s})/\sqrt{n}}\big|\,d\lambda\,d\lambda_1 \\ &\le\; \frac{Ch^2}{n}\sum_{s,t=1}^n \int_{|\lambda|\ge A}\int_{|\lambda_1|\ge A} |\hat K(-\lambda/c_n)|\,|\hat K(-\lambda_1/c_n)|\,\big|Ee^{i(\lambda\,x_t'-\lambda_1\,x_s')/\sqrt{n}}\big|\,d\lambda\,d\lambda_1 \\ &\le\; \frac{Ch^2}{n}\sum_{s\le t} \int_{|\lambda|\ge A} I_1\big(\frac{\lambda}{\sqrt{n}}\big)\,|\hat K(-\lambda/c_n)|d\lambda \int_{|\lambda_1|\ge A} I_2\big(\frac{\lambda}{\sqrt{n}},\frac{\lambda_1}{\sqrt{n}}\big)\,|\hat K(-\lambda_1/c_n)|d\lambda_1, \end{aligned} \tag{5.25}$$

where $I_1(s)$ and $I_2(s,t)$ are defined as in (5.9). It follows from (5.12) that

$$\begin{aligned} &\int_{|\lambda_1|\geq A} I_2\big(\frac{\lambda}{\sqrt{n}}, \frac{\lambda_1}{\sqrt{n}}\big)\, |\hat{K}(-\lambda_1/c_n)| d\lambda_1 \\ \leq\ & e^{-\gamma_1\sqrt{s}} \int |\hat{K}(-\lambda_1/c_n)| d\lambda_1 + C\int e^{-\gamma_2' s\lambda_1^2/n} d\lambda_1 \\ \leq\ & C\big[c_n e^{-\gamma_1\sqrt{s}} + (n/s)^{1/2}\big]. \end{aligned}$$

It follows from (5.10) that

$$\begin{aligned} &\int_{|\lambda|\geq A} I_1\big(\frac{\lambda}{\sqrt{n}}\big)\, |\hat{K}(-\lambda/c_n)| d\lambda \\ \leq\ & e^{-\gamma_1(t-s)} \int_{|\lambda|\geq\sqrt{n}} |\hat{K}(-\lambda/c_n)| d\lambda + C\int_{A\leq|\lambda|\leq\sqrt{n}} e^{-\gamma_2\lambda^2(t-s)/n} d\lambda \\ \leq\ & C\big[c_n e^{-\gamma_1(t-s)} + \int_{A\leq|\lambda|\leq\sqrt{n}} e^{-\gamma_2\lambda^2(t-s)/n} d\lambda\big]. \end{aligned}$$

Taking these estimates into (5.25), simple calculations show that

$$\begin{aligned} \sup_{x\in\omega^*} E|V_{1yA}|^2 \ &\leq\ \frac{Ch^2}{n} \sum_{t-s=0}^{n} \Big[c_n e^{-\gamma_1(t-s)} + \int_{A\leq|\lambda|\leq\sqrt{n}} e^{-\gamma_2\lambda^2(t-s)/n} d\lambda\Big] \\ &\qquad \sum_{s=1}^{n} \big[c_n e^{-\gamma_1\sqrt{s}} + (n/s)^{1/2}\big] \\ &\leq\ Ch^2\Big[c_n + \int_{A\leq|\lambda|\leq\sqrt{n}} \sum_{t-s=0}^{n} e^{-\gamma_2\lambda^2(t-s)/n} d\lambda\Big] \\ &\leq\ C\sqrt{n}h + nh^2 \int_{|\lambda|\geq A}\int_0^1 e^{-\gamma_2\lambda^2 x}\, dx\, d\lambda \\ &=\ o(nh^2), \end{aligned}$$

as $n\to\infty$ first and then $A\to\infty$, which yields the required (5.21).

To establish (5.22), first notice that $\{\epsilon_j, j\in Z\}$ can be redefined on a richer probability space which also holds a standard Brownian motion $W(t)$ such that

$$\sup_{0\leq t\leq 1} |x_{[nt],n} - G(t)| \ =\ o_P(1). \tag{5.26}$$

where $x_{k,n} = x_k/(\sqrt{n}\phi)$ and $G(t) = W(t) + \kappa\int_0^t e^{\kappa(t-s)} W(s) ds$. Indeed, on a richer probability space,

$$\sup_{0\leq t\leq 1} \Big|\frac{1}{\sqrt{n}} \sum_{j=1}^{[nt]} \epsilon_j - W_1(t)\Big| = o_P(1).$$

See, e.g., Csörgö and Révész (1981). Using this result in replacement of the fact that $\frac{1}{\sqrt{n}}\sum_{j=1}^{[nt]}\epsilon_j \Rightarrow W(t)$ on $D[0,1]$, the same technique as in the proof of Phillips (1987) [also see Chan and Wei (1987)] yields that

$$\sup_{0\le t\le 1}\Big|\frac{1}{\sqrt{n}}\sum_{j=1}^{[nt]}\rho^{[nt]-j}\epsilon_j - G(t)\Big| \;=\; o_P(1).$$

The result (5.26) can now be obtained by the same argument, with minor modifications, as in the proof of Proposition 7.1 in Wang and Phillips (2009b). On the other hand, simple calculations show that

$$\sup_{|\lambda|\le A, |y|\le A_0}\Big|e^{i\lambda y/\sqrt{n}}\hat{K}(-\lambda/c_n)-\hat{K}(0)\Big| = o(1)$$

as $c_n\to\infty$. This, together with (5.26) and the fact that Γ_{2nA} can be rewritten as

$$\begin{aligned}\frac{\Gamma_{2nA}}{\sqrt{n}h} &= \frac{1}{2\pi}\int_{|\lambda|<A} e^{i\lambda y/\sqrt{n}}\hat{K}(-\lambda/c_n)\frac{1}{n}\sum_{t=1}^{n}e^{i\lambda\phi x_{k,n}}\,d\lambda\\ &= \frac{1}{2\pi}\int_{|\lambda|<A} e^{i\lambda y/\sqrt{n}}\hat{K}(-\lambda/c_n)\int_0^1 e^{i\lambda\,\phi\, x_{[tn],n}}dt\,d\lambda\end{aligned}$$

yields that, for any $A>0$,

$$\begin{aligned}&\sup_{|y|\le A_0}\Big|\frac{V_{2yA}}{\sqrt{n}h}-\frac{\hat{K}(0)}{2\pi}\int_{|\lambda|\le A}\int_0^1 e^{i\lambda\phi\,G(t)}dt\,d\lambda\Big|\\ &\le\; C\sup_{|\lambda|\le A, |y|_0}\Big|e^{i\lambda y/\sqrt{n}}\hat{K}(-\lambda/c_n)-\hat{K}(0)\Big| + \frac{|\hat{K}(0)|}{2\pi}\int_{|\lambda|\le A}\int_0^1\Big|e^{i\lambda\,\phi\,(x_{[nt],n}-G(t))}-1\Big|dtd\lambda\\ &\le\; o(1)+O(1)\sup_{0\le t\le 1}|x_{[nt],n}-G(t)| = o_P(1).\end{aligned}$$

This proves (5.22) since $\hat{K}(0)=\int K(x)dx = 1$ and

$$\frac{1}{2\pi}\int_{|\lambda|\le A}\int_0^1 e^{i\lambda\phi\,G(t)}dt\,d\lambda \to_P \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_0^1 e^{i\lambda\phi\,G(t)}dt\,d\lambda = \frac{1}{\phi}L_G(1,0),$$

by the expression (5.23).

Finally, we remove the restriction (5.19). This only need to notice that, for any $\epsilon>0$, there exist $K_{\delta_0}^{+}(x)$ and $K_{\delta_0}^{-}(x)$ such that $K_{\delta_0}^{-}(x)\le K(x)\le K_{\delta_0}^{+}(x)$, both $K_{\delta_0}^{+}(x)$ and $K_{\delta_0}^{-}(x)$ satisfy (5.19), $\int_{-\infty}^{\infty}(|K_{\delta_0}^{+}(x)|+|K_{\delta_0}^{-}(x)|)dx<\infty$ and

$$\int_{-\infty}^{\infty}\big[K_{\delta_0}^{+}(x)-K_{\delta_0}^{-}(x)\big]dx<\epsilon. \tag{5.27}$$

The constructions of $K_{\delta_0}^{+}(x)$ and $K_{\delta_0}^{-}(x)$ are similar to the proof of Theorem 4.2.1 in Borodin and Ibragimov (1995). We omit the details.

This completes the proof of (5.5), and hence also for Proposition 3.2. □

Proof of Proposition 3.1. By virtue of (3.10), it only need to prove that

$$\sup_{x\in\Omega^*}\sum_{t=1}^n u_t\,K[(x_t-x)/h]=O_P\big[(nh^2)^{\frac{p+1}{2(2p+1)}}\,h^{-\frac{2}{2p+1}}\big], \tag{5.28}$$

and if addition $|u_t|\le C$, then, for some $\delta>0$,

$$\sup_{x\in\Omega^*}\sum_{t=1}^n u_t\,K[(x_t-x)/h]=o_P\big[(nh^2)^{1/4}\,\log^{1-\delta} n\big]. \tag{5.29}$$

To prove (5.28), we adopt the same partition as in (5.1), but with

$$m_n=[(nh^2)^{\frac{p}{2(2p+1)}}\,h^{-\frac{4p}{2p+1}}].$$

It follows that

$$\begin{aligned}
&\sup_{x\in\Omega^*}\big|\sum_{t=1}^n u_t\,K[(x_t-x)/h]\big| \\
\le\;& \max_{0\le j\le m_n-1}\sup_{x\in[y_j,y_{j+1}]}\sum_{t=1}^n|u_t|\,\big|K[(x_t-x)/h]-K[(x_t-y_j)/h]\big| \\
&\quad+\max_{0\le j\le m_n}\big|\sum_{t=1}^n u_t\,K[(x_t-y_j)/h]\big| \\
:=\;& \lambda_{3n}+\lambda_{4n}. \qquad (5.30)
\end{aligned}$$

Recall that x_t is adapted to $\mathcal{F}_{t-1}$ and $\sup_{1\le t\le n}E(|u_t|\mid\mathcal{F}_{t-1})<\infty$. As in the proof of (5.13), we have

$$\begin{aligned}
E\lambda_{3n} \;&\le\; \sum_{t=1}^n\sup_{1\le t\le n}E(|u_t|\mid\mathcal{F}_{t-1})\,E\max_{0\le j\le m_n-1}\sup_{x\in[y_j,y_{j+1}]}\big|K[(x_t-x)/h]-K[(x_t-y_j)/h]\big| \\
&\le\; C\sqrt{n}m_n^{-1}h^{-1}. \qquad (5.31)
\end{aligned}$$

Under $E|u_t|^{2p}<\infty$, the well-known martingale moment inequality yields that

$$\begin{aligned} E\lambda_{4n}^{2p} &\le \sum_{j=0}^{m_n} E\,\big|\sum_{t=1}^n u_t\,K[(x_t-y_j)/h]\big|^{2p} \\ &\le C\sum_{j=0}^{m_n} E\,\big\{\sum_{t=1}^n K^2[(x_t-y_j)/h]\big\}^p \\ &\le C\,m_n \max_{0\le j\le m_n}\Big\{\sum_{t=1}^n K[(x_t-y_j)/h] + \sum_{1\le s<t\le n} K[(x_s-y_j)/h]\,K[(x_t-y_j)/h] \\ &\quad + \sum_{1\le t_1<t_2<...<t_p\le n} K[(x_{t_1}-y_j)/h]\,K[(x_{t_2}-y_j)/h]...K[(x_{t_p}-y_j)/h]\Big\}, \quad (5.32) \end{aligned}$$

where we have used $K(x)\le C$. Note that, given on $\mathcal{F}_s$, $(x_t-x_s)/\sqrt{t-s}$ has a density $h_{s,t}(x)$ which is uniformly bounded by a constant C. See the proof of Proposition 7.2 of Wang and Phillips (2009b, page 1934 there). Simple calculations show that, for $1\le j\le p$,

$$\begin{aligned} E\big\{K[(x_{t_j}-y_j)/h]\mid\mathcal{F}_{t_{j-1}}\big\} &= \int_{-\infty}^{\infty} K[\sqrt{t_j-t_{j-1}}y/h+(x_{t_{j-1}}-y_j)/h]h_{t_{j-1},t_j}(y)dy \\ &\le \frac{Ch}{\sqrt{t_j-t_{j-1}}}\int_{-\infty}^{\infty} K[y+(x_{t_{j-1}}-y_j)/h]dy \\ &\le C_1h/\sqrt{t_j-t_{j-1}}. \end{aligned}$$

Taking his estimate into (5.32) and using the conditional arguments repeatedly, we obtain

$$\begin{aligned} E\lambda_{4n}^{2p} &\le C\,m_n\Big\{\sum_{t=1}^n \frac{h}{\sqrt{t}} + \sum_{1\le s<t\le n}\frac{h^2}{\sqrt{s}}\frac{1}{\sqrt{t-s}} \\ &\quad + \sum_{1\le t_1<t_2<...<t_p\le n}\frac{h^p}{\sqrt{t_1}}\,\frac{1}{\sqrt{t_2-t_1}}\cdots\frac{1}{\sqrt{t_p-t_{p-1}}}\Big\} \\ &\le C\,m_n\,(nh^2)^{p/2}, \quad (5.33) \end{aligned}$$

as $nh^2\to\infty$. It follows from (5.31) and (5.33) that

$$\lambda_{3n}+\lambda_{4n} = O_P\big[\sqrt{n}m_n^{-1}h^{-1}+m_n^{1/(2p)}\,(nh^2)^{1/4}\big] = O_P\big[(nh^2)^{\frac{p+1}{2(2p+1)}}\,h^{-\frac{2}{2p+1}}\big],$$

which yields (5.28).

The proof of (5.29) is similar to (5.28), but with

$$m_n=\Big[\frac{(nh^2)^{1/4}h^{-2}}{\log^{1-\delta/2}n}\Big].$$

In this case, we still have (5.30), and by (5.31)

$$\lambda_{3n} \;=\; O_P(\sqrt{n}m_n^{-1}h^{-1}) = o_P\big[(nh^2)^{1/4}\,\log^{1-\delta} n\big].$$

So it only need to prove, for any $0<\delta<1$,

$$P\big[\lambda_{4n} \geq (nh^2)^{1/4}\,\log^{1-\delta/2} n\big] \;\to\; 0. \tag{5.34}$$

This follows from the same arguments as in the proof of (5.17). In fact, as in (5.16), we have

$$\begin{aligned} E\Big\{\max_{0\le j\le m_n}\sum_{t=1}^n K^2[(x_t-y_j)/h]\Big\} &\le E\Big[\max_{0\le j\le m_n}\sum_{t=1}^n E\big(K^2[(x_t-y_j)/h]\mid \mathcal{F}_{t-1}^*\big)\Big] \\ &\le C\sqrt{n}\,h, \end{aligned}$$

which implies that

$$\max_{0\le j\le m_n}\sum_{t=1}^n K^2[(x_t-y_j)/h] = O_P(\sqrt{n}\,h). \tag{5.35}$$

Hence, by recalling $nh^2/\log^4 n\to\infty$,

$$\begin{aligned} & P(\lambda_{4n}\ge (\sqrt{n}h)^{1/2}\log^{1-\delta/2} n) \\ \le\;& P\Big[\lambda_{4n}\ge ((\sqrt{n}h)^{1/2}\log^{1-\delta/2} n, \quad \max_{0\le j\le m_n}\sum_{t=1}^n K^2[(x_t-y_j)/h]\le \sqrt{n}h\log^{\delta/2} n\Big]+o(1) \\ \le\;& \sum_{j=0}^{m_n} P\Big[\sum_{t=1}^n u_jK[(x_t-y_j)/h]\ge (\sqrt{n}h)^{1/2}\log^{1-\delta/2} n, \\ &\qquad\qquad \sum_{t=1}^n K^2[(x_t-y_j)/h]\le \sqrt{n}h\log^{\delta/2} n\Big]+o(1) \\ \le\;& m_n\,\exp\Big\{-\frac{\sqrt{n}h\log^{2-\delta} n}{2\sqrt{n}h\log^{\delta/2} n + C(\sqrt{n}h)^{1/2}\log^{1-\delta/2} n}\Big\}+o(1)=o(1), \end{aligned}$$

which yields (5.34). □

Proof of Theorem 3.2. We may write $\hat f(x)-f(x)$ as

$$\begin{aligned} \hat f(x)-f(x) &= \frac{\sum_{t=1}^n \sigma(x)u_tK[(x_t-x)/h]}{\sum_{t=1}^n K[(x_t-x)/h]} + \frac{\sum_{t=1}^n [\sigma(x_t)-\sigma(x)]\,u_tK[(x_t-x)/h]}{\sum_{t=1}^n K[(x_t-x)/h]} \\ &\quad + \frac{\sum_{t=1}^n [f(x_t)-f(x)]\,K[(x_t-x)/h]}{\sum_{t=1}^n K[(x_t-x)/h]} \\ &:= \Theta_{1n}(x)+\Theta_{2n}(x)+\Theta_{3n}(x). \end{aligned} \tag{5.36}$$

It follows from (3.8) that

$$\sup_{x\in\Omega_\epsilon} |\Theta_{1n}(x)| \;=\; O_P\big[\big(nh^{2+4/p}\big)^{-p/[2(2p+1)]}\big],$$

since $\sup_{x\in\Omega_\epsilon}|\sigma(x)| \le C$, by Assumption 4. Note that, for any $x\in\Omega_\epsilon$, there exists a $C_0>0$ such that $K[(x_t-x)/h]=0$ if $|x_t-x|\ge hC_0$. It is readily seen from Assumption 4 that, whenever n is sufficiently large,

$$\sup_{x\in\Omega_\epsilon} |\Theta_{3n}(x)| \;\le\; \frac{C_1\,\sum_{t=1}^n |x_t-x|^\alpha\, K[(x_t-x)/h]}{\sum_{t=1}^n K[(x_t-x)/h]} \le Ch^\alpha.$$

Similarly, by noting that $|u_t| - E(|u_t| \mid \mathcal{F}_{t-1})$ forms a martingale difference satisfying Assumption 2*, we have

$$\begin{aligned}
\sup_{x\in\Omega_\epsilon} |\Theta_{2n}(x)| \;&\le\; Ch^\beta \sup_{x\in\Omega_\epsilon} \frac{\sum_{t=1}^n |u_t| K[(x_t-x)/h]}{\sum_{t=1}^n K[(x_t-x)/h]} \\
&\le\; Ch^\beta \Big\{ \sup_{t\ge 1} E(|u_t| \mid \mathcal{F}_{t-1}) + \sup_{x\in\Omega_\epsilon} \frac{\sum_{t=1}^n \big[|u_t| - E(|u_t| \mid \mathcal{F}_{t-1})\big] K[(x_t-x)/h]}{\sum_{t=1}^n K[(x_t-x)/h]} \Big\} \\
&\le\; C_1 h^\beta. \qquad (5.37)
\end{aligned}$$

Taking these estimates into (5.36), we derive (3.5). The proof of (3.6) is similar except using (3.9) instead of (3.8). We omit the details. □

Proof of Theorem 3.1. First for (3.1), which is similar to that of (3.5). We still use the decomposition in the (3.4). Recall $\sup_{1\le t\le n} E(|u_t^2|^{2[\nu]}|\mathcal{F}_{t-1}) < \infty$ a.s., where $[\nu]\ge 1$. It follows from (3.8) that

$$\sup_{x\in\Omega} |I_{1n}(x)| \;=\; O_P\big[\big(nh^{2+4/[\nu]}\big)^{-[\nu]/[2(2[\nu]+1)]}\big],$$

since $\sup_{x\in\Omega_\epsilon}|\sigma(x)| \le C$, by Assumption 4. Since Assumption 4 also implies that, for any $x,y\in\Omega_\epsilon$,

$$|\sigma^2(y) - \sigma^2(x)| \le 2|\sigma(y)-\sigma(x)| \sup_{x\in\Omega_\epsilon} |\sigma(x)| \le C\,|y-x|^\beta,$$

similarly to the estimate of (5.37), we have

$$\sup_{x\in\Omega} |I_{2n}(x)| \;\le\; Ch^\beta \frac{\sum_{t=1}^n u_t^2\, K[(x_t-x)/h]}{\sum_{t=1}^n K[(x_t-x)/h]} = O_P(h^\beta). \qquad (5.38)$$

Note that $h \to 0$ and, for any $x \in \Omega$, there exists a $C_0 > 0$ such that $K[(x_t - x)/h] = 0$ if $|x_t - x| \geq hC_0$. This, together with (3.5), implies that, whenever n is sufficiently large,

$$\begin{aligned} \sup_{x\in\Omega} |I_{3n}(x)| &\leq C \sup_{x\in\Omega_\epsilon} |\hat{f}(x) - f(x)|^2 \\ &= O_P\left\{ \left(nh^{2+4/[\nu]}\right)^{-[\nu]/(2[\nu]+1)} + h^{2\min\{\alpha,\beta\}} \right\}. \end{aligned} \tag{5.39}$$

Similarly, we have

$$\begin{aligned} \sup_{x\in\Omega} |I_{4n}(x)| &\leq C \sup_{x\in\Omega_\epsilon} |\hat{f}(x) - f(x)| \sup_{x\in\Omega_\epsilon} |\sigma(x)| \frac{\sum_{t=1}^n |u_t| K[(x_t - x)/h]}{\sum_{t=1}^n K[(x_t - x)/h]} \\ &= O_P\left\{ \left(nh^{2+4/[\nu]}\right)^{-[\nu]/[2(2[\nu]+1)]} \delta_n^{-1} + h^{\min\{\alpha,\beta\}} \right\}. \end{aligned} \tag{5.40}$$

Taking these estimates into (3.4), we obtain the required (3.1).

We next prove (3.2), and to do so by showing

$$\sum_{j=1}^k b_j T_n(z_j) \to_d N(0, \sigma_1^2), \tag{5.41}$$

for any $(b_1, \ldots, b_k) \in \mathbb{R}^k$, where $\sigma_1^2 = (\Lambda^2 - 1) \int_{-\infty}^{\infty} K^2(u) du \sum_{j=1}^k b_j^2 \sigma^4(z_j)$. Define

$$\Delta_n(x) = \sum_{t=1}^n K[(x_t - x)/h], \quad Z_{nt}(x) = \sigma^2(x) K[(x_t - x)/h] \Delta_n^{-1/2}(x)$$

and $Y_{nt} = \sum_{j=1}^k b_j Z_{nt}(z_j)$. Recall (3.4). We may write

$$\sum_{j=1}^k b_j T_n(z_j) = \sum_{t=1}^n (u_t^2 - 1) Y_{nt} + \sum_{j=1}^k b_j \Delta_n^{1/2}(z_j) \left[I_{2n}(z_j) + I_{3n}(z_j) + I_{4n}(z_j) \right].$$

Recall (5.35), that is, $\Delta_n(x) = O_P(\sqrt{n}h)$ for each fixed $x \in \Omega$. The result (5.41) will follow if we prove

$$\sum_{t=1}^n (u_t^2 - 1) Y_{nt} \to_d N(0, \sigma_1^2) \tag{5.42}$$

and for each fixed $x \in \Omega$,

$$I_{jn}(x) = o_P[(nh^2)^{-1/4}], \quad j = 2,\ 3,\ 4. \tag{5.43}$$

First prove (5.42). Write $\mathcal{F}_{nj} = \sigma(u_1, ..., u_j, x_1, ..., x_n), 1 \leq j \leq n$. Recalling that u_t is independent of x_t, it is readily seen that $\mathcal{F}_{nj} \subseteq \mathcal{F}_{n,j+1}$ for $1 \leq j \leq n, n \geq 1$, and $\{(u_t^2 - 1) Y_{nt}, \mathcal{F}_{nt}\}_{t=1}^n$ forms a martingale difference sequence. By the classical martingale limit theorem (see, e.g., Corollary 3.1 of Hall and Heyde (1980)), it suffices to show that, for any $\nu > 1$

$$\sum_{t=1}^{n} |Y_{nt}|^{2\nu} \quad \to_P \quad 0, \tag{5.44}$$

and

$$\sum_{t=1}^{n} Y_{nt}^2\, E\big[(u_t^2-1)^2 \mid \mathcal{F}_{t-1}\big] \quad \to_P \quad \sigma_1^2. \tag{5.45}$$

Note that $\inf_{x\in\Omega} \Delta_n^{-1}(x) = o_P(1)$ by (3.10). The proof of (5.44) is simple. Indeed it is readily seen that

$$\begin{aligned}
\sum_{t=1}^{n} |Y_{nt}|^{2\nu} &\leq C \sum_{j=1}^{k}\sum_{t=1}^{n} |Z_{nt}(z_j)|^{2\nu} \\
&\leq C \sum_{j=1}^{k} |\sigma(z_j)|^{2\nu} \Delta_n^{-\nu}(z_j) \sum_{t=1}^{n} K[(x_t - z_j)/h] \\
&\leq C_1 \sum_{j=1}^{k} \Delta_n^{1-\nu}(z_j) = o_P(1),
\end{aligned}$$

which yields (5.44).

In order to prove (5.45), first notice that, for any a, b and $1 \leq j \leq k$,

$$\begin{aligned}
&\frac{\phi}{\sqrt{n}h} \sum_{t=1}^{n} \big\{a\, K[(x_t - z_j)/h] + b\, K^2[(x_t - z_j)/h]\big\} \\
&\qquad \to_D \int_{-\infty}^{\infty} \big[aK(x) + bK^2(x)\big] dx\, L_G(1,0),
\end{aligned}$$

by Proposition 7.2 of Wang and Phillips (2009b)[1]. This implies that (recalling $\int K(x)dx = 1$)

$$\begin{aligned}
&\Big\{\frac{\phi}{\sqrt{n}h} \sum_{t=1}^{n} K[(x_t - z_j)/h],\ \frac{\phi}{\sqrt{n}h} \sum_{t=1}^{n} K^2[(x_t - z_j)/h]\Big\} \\
&\qquad \to_D \Big\{L_G(1,0), \int_{-\infty}^{\infty} K^2(x)dx\, L_G(1,0)\Big\},
\end{aligned}$$

and hence by the continuous mapping theorem,

$$\frac{\sum_{t=1}^{n} K^2[(x_t - z_j)/h]}{\sum_{t=1}^{n} K[(x_t - z_j)/h]} \to_P \int_{-\infty}^{\infty} K^2(x)dx.$$

[1]Proposition 7.2 of Wang and Phillips (2009b) requires $\int_{-\infty}^{\infty} |\hat{K}(t)|dt < \infty$, where $\hat{K}(t) = \int_{-\infty}^{\infty} e^{itx} K(x)dx$. But this condition is only used in the proof of tightness, not for the finite dimensional distribution convergence.

On the other hand, for fixed $z_i \neq z_j$, we have

$$K[(x_t - z_i)/h]K[(x_t - z_j)/h] = 0,$$

whenever n is sufficiently large, as $K(x)$ has a compact support. Simple calculations show that

$$\begin{aligned}\sum_{t=1}^n Y_{nt}^2 &= \sum_{j=1}^k b_j^2 \sigma^4(z_j)\, \Delta_n^{-1}(z_j) \sum_{t=1}^n K^2[(x_t - z_j)/h] \\ &\quad +2 \sum_{1\le i<j\le k} b_i b_j \sigma^2(z_i)\sigma^2(z_j)\Delta_n^{-1/2}(z_i)\Delta_n^{-1/2}(z_j) \sum_{t=1}^n K[(x_t - z_i)/h]K[(x_t - z_j)/h] \\ &\to_P \int_{-\infty}^{\infty} K^2(x)dx \sum_{j=1}^k b_j^2 \sigma^4(z_j). \qquad (5.46)\end{aligned}$$

This, together with the fact that

$$E\big[(u_t^2 - 1)^2 \mid \mathcal{F}_{t-1}\big] = E\big[u_t^4 \mid \mathcal{F}_{t-1}\big] - 1 \to_{a.s.} \Delta^2 - 1,$$

yields (5.45). The proof of (5.42) is now complete.

We next prove (5.43). The result (5.43) with $j = 2$ follows from (5.38) and $nh^{2+4\beta} \to 0$. To prove (5.43) with $j = 3$ and 4, for $x \in \Omega$, write

$$\Delta_{1n}(x) = \sum_{t=1}^n (f(x_t) - \hat{f}(x_t))^2 K[(x_t - x)/h], \qquad (5.47)$$

$$\Delta_{2n}(x) = \sum_{t=1}^n \sigma(x_t) u_t (f(x_t) - \hat{f}(x_t)) K[(x_t - x)/h]. \qquad (5.48)$$

We may rewrite $\Delta_{2n}(x)$ as

$$\begin{aligned}\Delta_{2n}(x) &= \sum_{\substack{s,t=1\\ s\neq t}}^n u_s\, u_t\, \sigma(x_s)\sigma(x_t)\, \Delta_n^{-1}(x_t) K[(x_s - x_t)/h]K[(x_t - x)/h] \\ &\quad +K(0) \sum_{s=1}^n u_s^2\, \sigma^2(x_s)\, \Delta_n^{-1}(x_s) K[(x_s - x)/h] \\ &\quad + \sum_{\substack{s,t=1\\ s\neq t}}^n u_t\, \sigma(x_t)\, [f(x_s) - f(x_t)] \Delta_n^{-1}(x_t) K[(x_s - x_t)/h]K[(x_t - x)/h] \\ &= \Delta_{2n1}(x) + \Delta_{2n2}(x) + \Delta_{2n3}(x), \quad \text{say.} \qquad (5.49)\end{aligned}$$

Let $\Xi_n = I\big\{ \inf_{x\in\Omega^*} \sum_{t=1}^n K[(x_t - x)/h] \geq \sqrt{n}h\,/M \big\}$, where $M \geq M_0$ is chosen as in (3.10). Note that, for $x \in \Omega$,

$$K[(x_s - x_t)/h]K[(x_t - x)/h] \;=\; 0, \qquad (5.50)$$

if $|x_t - x| \geq C_0 h$ or $|x_s - x_t| \geq C_0 h$ for some $C_0 > 0$. By virtue of the independence between u_t and x_t, it is readily seen that, there exists $\epsilon > 0$ such that

$$\begin{aligned} E\big\{\Delta_{2n1}^2(x)\Xi_n\big\} &= E\Big\{\Xi_n \sum_{\substack{s,t=1\\ s\neq t}}^n u_s^2\, u_t^2\, \sigma^2(x_s)\sigma^2(x_t)\, \Delta_n^{-2}(x_t) K^2[(x_s-x_t)/h]K^2[(x_t - x)/h]\Big\} \\ &\leq C \sup_{x\in\Omega_\epsilon} \sigma^4(x) E\Big\{\Xi_n \sum_{t=1}^n \Delta_n^{-1}(x_t) K^2[(x_t-x)/h]\Big\} \\ &\leq \frac{CM}{\sqrt{n}h} E\sum_{t=1}^n K^2[(x_t - x)/h] \leq C_1 M, \end{aligned} \qquad (5.51)$$

whenever n is sufficiently large. By virtue of this fact and (3.10), for any $\eta > 0$, there exists a $M_0 > 0$ such that for $M \geq M_0$,

$$\begin{aligned} & P\Big[\frac{\Delta_{2n1}(x)}{\Delta_n(x)} \geq (nh^2)^{-1/4}/M\Big] \\ \leq\;& P\big\{\inf_{x\in\Omega^*}\sum_{t=1}^n K[(x_t-x)/h] \leq \sqrt{n}h\,/M\big\} + \frac{M}{(nh^2)^{1/2}} E\big\{\Delta_{2n1}^2(x)\Xi_n\big\} \\ \leq\;& \eta + \frac{C_1 M^2}{\delta_n^2 (nh^2)^{1/2}}, \end{aligned} \qquad (5.52)$$

which yields $\frac{\Delta_{2n1}(x)}{\Delta_n(x)} = o_P[(nh^2)^{-1/4}]$. Similarly, $\frac{\Delta_{2n2}(x)}{\Delta_n(x)} = o_P[(nh^2)^{-1/4}]$. As for $\Delta_{2n3}(x)$, it follows from (5.50) and Assumption 4 that

$$\begin{aligned} E\Delta_{2n3}^2(x) &= E\Big\{\sum_{t=1}^n u_t^2\, \sigma^2(x_t)\, \Delta_n^{-2}(x_t) K^2[(x_t-x)/h] \\ &\qquad\qquad \Big(\sum_{\substack{s=1\\ s\neq t}}^n [f(x_s) - f(x_t)] K[(x_s - x_t)/h]\Big)^2\Big\} \\ &\leq C \sup_{x\in\Omega_\epsilon} \sigma^2(x)\, h^{2\alpha}\, E\sum_{t=1}^n K^2[(x_t - x)/h] \\ &\leq C_1 \sqrt{n} h^{1+2\alpha}. \end{aligned} \qquad (5.53)$$

Similar to the proof of (5.52) we have $\frac{\Delta_{2n3}(x)}{\Delta_n(x)} = o_P[(nh^2)^{-1/4}]$, since $h \to 0$. Taking these estimates into (5.49), we proves

$$I_{4n} = \frac{2\Delta_{2n}(x)}{\Delta_n(x)} = o_P[(nh^2)^{-1/4}].$$

It is quite similar to show that

$$I_{3n} = \frac{\Delta_{1n}(x)}{\Delta_n(x)} = o_P[(nh^2)^{-1/4}].$$

Indeed, we may rewrite $\Delta_{1n}(x)$ as

$$\begin{aligned}\Delta_{1n}(x) &\leq \sum_{t=1}^n \Delta_n^{-2}(x_t)\, K[(x_t - x)/h] \Big(\sum_{\substack{s=1\\ s\neq t}}^n [f(x_s) - f(x_t)] K[(x_s - x_t)/h] \Big)^2 \\ &\quad + \sum_{t=1}^n \Delta_n^{-2}(x_t)\, K[(x_t - x)/h] \Big(\sum_{s=1}^n \sigma(x_s)\, u_s\, K[(x_s - x_t)/h] \Big)^2 \\ &:= \Delta_{1n1}(x) + \Delta_{1n2}(x). \end{aligned} \tag{5.54}$$

It follows from the similar arguments as in (5.53) that

$$\Delta_{1n1}(x) \leq Ch^{2\alpha} \sum_{t=1}^n K[(x_t - x)/h] = O_P(\sqrt{n}h^{1+2\alpha}) = o_P[(nh^2)^{1/4}],$$

whenever $nh^{2+8\alpha} \to 0$, which yields $\frac{\Delta_{1n1}(x)}{\Delta_n(x)} = o_P[(nh^2)^{-1/4}]$. Note that, as in (5.51),

$$\begin{aligned} E\big\{\Delta_{1n2}(x)\Xi_n\big\} &= E\Big\{\Xi_n \sum_{\substack{s,t=1\\ s\neq t}}^n \sigma^2(x_s)\, u_s^2\, \Delta_n^{-2}(x_t)\, K^2[(x_s - x_t)/h] K[(x_t - x)/h]\Big\} \\ &\leq C \sup_{x\in\Omega_\epsilon} \sigma^2(x) E\Big\{\Xi_n \sum_{t=1}^n \Delta_n^{-1}(x_t) K[(x_t - x)/h]\Big\} \\ &\leq CM. \end{aligned}$$

As in the proof of (5.52), we have $\frac{\Delta_{1n2}(x)}{\Delta_n(x)} = o_P[(nh^2)^{-1/4}]$. Taking these estimates into (5.54), we obtain the required $I_{3n}(x) = \frac{\Delta_{2n1}(x)}{\Delta_n(x)} = o_P[(nh^2)^{-1/4}]$. The proof of (5.43) is now complete, and hence that of Theorem 3.1. □

Acknowledgments: The authors would like to thank the co-Editor Professor Saikkonen and two referees for their very helpful comments on the original version. The main idea of this paper was greatly benefited from the rich discussions with Professor Peter Phillips when the first author visited Singapore Management University in April, 2009. Wang also thanks the partial research support from the Australian research council.

REFERENCES

Bandi, F. (2004). On Persistence and nonparametric estimation (with an application to stock return predictability). *Unpublished manuscript.*

Berman, S. M. (1969). Local times and sample function properties of stationary Gaussian processes. *Trans. Amer. Math. Soc.* **137** 277–299.

Borodin, A. N.; Ibragimov, I. A. (1995). *Limit theorems for functionals of random walks.* Proc. Steklov Inst. Math., no. 2.

Cai, Z., Li, Q. and Park, J. Y. (2009). Functional-coefficient models for nonstationary time series data. *J. Econometrics* **148**, 101–113.

Chan, N. H. and Wei, C. Z. (1987). Asymptotic inference for nearly nonstationary $AR(1)$ process. *Annals of Statistics,* **15**, 1050-1063.

Choi, I. and Saikkonen, P.(2004). Testing linearity in cointegrating smooth transition regressions. *Econom. J.* **7**, 341–365.

Choi, I. and Saikkonen, P. (2009). Tests for nonlinear cointegration. Forthcoming in Econometric Theory

Chung, H. and Park, J. Y. (2007), Nonstationary nonlinear heteroskedasticity in regression, *Journal of Econometrics*, **137**, 230-259.

Csörgö, M.; Révész, P. (1981). *Strong approximations in probability and statistics.* Probability and Mathematical Statistics. Academic Press, Inc. New York-London.

Gao, J., Li, D. and Tjstheim, D. (2009). Uniform consistency for nonparametric estimates in null recurrent time series. *Working paper series No. 0085,* , The university of Adelaide, School of Economics.

Gao, J., K., Maxwell, Lu, Z., Tjstheim, D. (2009a). Nonparametric specification testing for nonlinear time series with nonstationarity. *Econometric Theory*, **25**, 1869–1892.

Gao, J., K., Maxwell, Lu, Z., Tjstheim, D. (2009b). Specification testing in nonlinear and nonstationary time series autoregression. *Ann. Statist.* **37**, 3893–3928.

Granger, C. W. J. and Teräsvirta, T (1993), *Modelling nonlinear economic relationships*, Oxford University Press, New York.

Guerre, E. (2004). Design-Adaptive pointwise nonparametric regression estimation for recurrent Markov time series. *Unpublished pamnuscript.*

Hall, P.; Heyde, C. C. (1980). *Martingale limit theory and its application,* Probability and Mathematical Statistics. Academic Press, Inc.

Hansen, B. E. (1995). Regression with nonstationary volatility, *Econometrica*, **63**, 1113-1132.

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

Harvey, A. C. and Robinson, P. M. (1987), Efficient estimation of nonstationary time series regression, *Journal of Time Series Analysis*, **9**, 201-214.

Han, H. and Park, J. Y. (2008), Time series properties of ARCH processes with persistent covariates, *Journal of Econometrics*, **146**, 275-292.

Han, H. and Zhang, S. (2009), Nonstationary semiparametric ARCH model and non-parametric cointegrating volatility model, *working paper*.

Kim, C. S. and Park, J. Y. (2010), Cointegrating regressions with time heterogeneity, *Econometric Review*, **29**, 397-438.

Karlsen, H. A. and Tjøstheim, D. (2001). Nonparametric estimation in null recurrent time series. *Ann. Statist.* **29**, no. 2, 372–416.

Karlsen, H. A., Myklebust, T. and Tjostheim, D. (2007). Nonparametric estimation in a nonlinear cointegration model, *Ann. Statist.*, to appear.

Kasparis, I. and Phillips, P.C.B. (2009). Dynamic misspecification in nonparametric cointegrating regression. Cowles foundation discussion paper No. 1700.

Lobato, I. N. and Robinson, P. M. (1998). A nonparametric test for I(0). *Rev. Econom. Stud.* **65**, 475–495.

Lukacs, E. (1970). *Characteristic functions*. Hafner Publishing Co., New York.

Marmer, V. (2008). Nonlinearity, nonstationarity, and spurious forecasts. *J. Econometrics* **142** , 1–27.

De La Pena, V. H. (1999). A general class of exponential inequalities for martingales and ratios. *Ann. Probab.* **27**, 537–564.

Park, J. Y. (2002), Nonstationary nonlinear heteroskedasticity, *Journal of Econometrics*, **110**, 383-415.

Park, J. Y. and Hahn, S. B. (1999). Cointegrating regressions with time varying coefficients, *Econometric Theory*, **15**, 664-703.

Park, J. Y. and P. C. B. Phillips (1999). Asymptitocs for nonlinear transformation of integrated time series, *Econometric Theory*, **15**, 269-298.

Park, J. Y. and P. C. B. Phillips (2001). Nonlinear regressions with integrated time series, *Econometrica*, **69**, 117-161.

Phillips, P. C. B. (1987). Towards a unified asymptotic theory for autoregression. *Biometrika*, **74**, 535-547.

Phillips, P. C. B. (1988), "Regression Theory for Near-Integrated Time Series," *Econometrica*, **56**, 1021-1044.

Phillips, P. C. B. and J. Y. Park (1998). Nonstationary Density Estimation and Kernel Autoregression. *Cowles Foundation discuss paper No. 1181.*

Quintos, C. E. and Phillips, P. C. B. (1993), Parameter constancy in cointegrating regressions, *Empirical Economics*, **18**, 675-706.

Robinson, P. M. (1987), Asymptotically efficient estimation in the presence of heteroskedasticity of unknown form, *Econometrica*, **55**, 875-891.

Schienle, M. (2008). Nonparametric Nonstationary Regression, Unpublished Ph.D. Thesis, University of Mannheim.

Sun, Y., Cai, Z. and Li, Q. (2008), Consistent nonparametric test on parametric smooth coefficient model with nonstationary data, *Manuscript*.

Tassos, M. and Phillips, P. C. B. (2009). Limit theory for cointegrated systems with moderately integrated and moderately explosive regressors. *Econometric Theory* **25** ,

482–526.

Wang, Q. and P. C. B. Phillips (2009a). Asymptotic Theory for Local Time Density Estimation and Nonparametric Cointegrating Regression, *Econometric Theory*, **25**, 710-738.

Wang, Q. and P. C. B. Phillips (2009b). Structural Nonparametric Cointegrating Regression, *Econometrica*, **77**, 1901-1948.

Wang, Q. and P. C. B. Phillips (2010). Asymptotic Theory for Zero Energy Functionals with Nonparametric Regression Applications. *Econometric Theory*, forthcoming.

Xiao, Z. (2009). Functional-coefficient cointegration models, *Journal of Econometrics*, **152**, 81-92.

Address correspondence to Qiying Wang, School of Mathematics and Statistics, The University of Sydney, NSW 2006, Australia; e-mail:qiying@maths.usyd.edu.au.