cemmap
centre for microdata methods and practice

Non-parametric transformation regression with non-stationary data

Oliver Linton
Qiying Wang

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP16/13

An ESRC Research Centre

Institute for Fiscal Studies, 7 Ridgmount Street, London WC1E 7AE **Tel:** 44 (0)20 7291 4800 **Email:** info@cemmap.ac.uk **www.cemmap.ac.uk**

Non-parametric transformation regression with non-stationary data

Oliver Linton[*] Qiying Wang[†]
University of Cambridge University of Sydney

April 22, 2013

Abstract

We examine a kernel regression smoother for time series that takes account of the error correlation structure as proposed by Xiao et al. (2008). We show that this method continues to improve estimation in the case where the regressor is a unit root or near unit root process.

Key words and phrases: Dependence; Efficiency; Cointegration; Non-stationarity; Non-parametric estimation.
JEL Classification: C14, C22.

1 Introduction

This paper is concerned with estimation of a nonstationary nonparametric cointegrating regression. The theory of linear cointegration is extensive and originates with the work of Engle and Granger (1987), see also Stock (1987), Phillips (1991), and Johanssen (1988). Wang and Phillips (2009a, b, 2011) recently considered the nonparametric cointegrating regression. They analyse the behaviour of the standard kernel estimator of the cointegrating relation/nonparametric regression when the covariate is nonstationary. They showed that the under self (random) normalization, the estimator is asymptotically normal. See also Phillips and Park (1998), Karlsen and Tjostheim (2001), Karlsen, et al.(2007), Schienle (2008) and Cai, et al.(2009).

[*]Address: Faculty of Economics, Austin Robinson Building, Sidgwick Avenue, Cambridge, CB3 9DD, UK. Email: obl20@cam.ac.uk. I would like to thank the European Research Council for financial support.
[†]Address: School of Mathematics and Statistics, The University of Sydney, NSW 2006, Australia. E-mail: qiying@maths.usyd.edu.au. This author thanks the Australian Research Council for financial support.

We extend this work by investigating an improved estimator in the case where there is autocorrelation in the error term. Standard kernel regression smoothers do not take account of the correlation structure in the covariate x_t or the error process u_t and estimate the regression function in the same way as if these processes were independent. Furthermore, the variance of such estimators is proportional to the short run variance of u_t, $\sigma_u^2 = \mathrm{var}(u_t)$ and does not depend on the regressor or error covariance functions $\gamma_x(j) = \mathrm{cov}(x_t, x_{t-j})$, $\gamma_u(j) = \mathrm{cov}(u_t, u_{t-j})$, $j \neq 0$. Although the time series properties do not effect the asymptotic variance of the usual estimators, the error structure can be used to construct a more efficient estimator. Xiao, Linton, Carroll, and Mammen (2003) proposed a more efficient estimator of the regression function based on a prewhitening transformation. The transform implicitly takes account of the autocorrelation structure. They obtained an improvement in terms of variance over the usual kernel smoothers. Linton and Mammen (2006) proposed a type of iterated version of this procedure and showed that it obtained higher efficiency. Both these contributions assumed that the covariate process was stationary and weakly dependent. We consider here the case where x_t is nonstationary, of the unit root or close to unit root type. We allow the error process to have some short term memory, which is certainly commonplace in the linear cointegration literature. We show that the Xiao, Linton, Carroll, and Mammen (2003) procedure can improve efficiency even in this case and one still obtains asymptotic normality for the self normalized estimator, which allows standard inference methods to be applied. In order to establish our results we require a new strong approximation result and use this to establish the L_2 convergence rate of the usual kernel estimator.

2 The model and main results

Consider a non-linear cointegrating regression model:

$$y_t = m(x_t) + u_t, \quad t = 1, 2, ..., n, \tag{2.1}$$

where $u_t = \rho u_{t-1} + \epsilon_t$ with $|\rho| < 1$ and x_t is a non-stationary regressor. The conventional kernel estimator of $m(x)$ is defined as

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

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

On the other hand, we may write the model (2.1) as

$$y_{t-1} = m(x_{t-1}) + u_{t-1}, \tag{2.2}$$

$$y_t - \rho y_{t-1} + \rho m(x_{t-1}) = m(x_t) + \epsilon_t. \tag{2.3}$$

It is expected that a two-step estimator of $m(x)$ by using models (2.2) and (2.3) may achieve efficiency improvements over the usual estimator $\hat{m}(x)$ by (2.1).

The strategy to provide the two-step estimator is as follows:

Step 1: Construct an estimator of $m(x)$, say $\hat{m}_1(x)$, by using model (2.2). This can be the conventional kernel estimator defined by

$$\hat{m}_1(x) = \frac{\sum_{s=2}^n y_{s-1} K[(x_{s-1}-x)/h]}{\sum_{s=2}^n K[(x_{s-1}-x)/h]},$$

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

Step 2: Construct an estimator of ρ by

$$\widehat{\rho} = \frac{\sum_{s=2}^n \widehat{u}_s \widehat{u}_{s-1}}{\sum_{s=2}^n \widehat{u}_{s-1}^2}.$$

Note that $\hat{\rho}$ is a LS estimator from model:

$$\widehat{u}_t = \rho \widehat{u}_{t-1} + \epsilon_t,$$

where $\widehat{u}_t = y_t - \widehat{m}_1(x_t)$.

Step 3: Construct an estimator of $m(x)$, say $\hat{m}_2(x)$, by using (2.3) and kernel method, but instead of the left hand $m(x)$ in model (2.3) by $\hat{m}_1(x)$.

We now have a two-step estimator $\hat{m}_2(x)$ of $m(x)$, defined as follows:

$$\hat{m}_2(x) = \frac{\sum_{t=1}^n \left[y_t - \widehat{\rho} y_{t-1} + \widehat{\rho}\hat{m}_1(x_{t-1})\right] K[(x_t-x)/h]}{\sum_{t=1}^n K[(x_t-x)/h]}.$$

To establish our claim, that is, $\hat{m}_2(x)$ achieves efficiency improvements over the usual estimator $\hat{m}(x)$, we make the following assumptions.

Assumption 2.1. $x_t = \lambda\, x_{t-1} + \xi_t$, $(x_0 \equiv 0)$, where $\lambda = 1 + \tau/n$ with τ being a constant and $\{\xi_j, j \ge 1\}$ is a linear process defined by

$$\xi_j = \sum_{k=0}^{\infty} \phi_k\, \nu_{j-k}, \tag{2.4}$$

where $\phi_0 \neq 0$, $\phi \equiv \sum_{k=0}^{\infty} \phi_k \neq 0$ and $\sum_{k=0}^{\infty} |\phi_k| < \infty$, and where $\{\nu_j, -\infty < j < \infty\}$ is a sequence of iid random variables with $E\nu_0 = 0$, $E\nu_0^2 = 1$, $E|\nu_0|^{2+\delta} < \infty$ for some $\delta > 0$ and characteristic function $\varphi(t)$ of ν_0 satisfies $\int_{-\infty}^{\infty}(1+|t|)\,|\varphi(t)|dt < \infty$.

Assumption 2.2. $u_t = \rho u_{t-1} + \epsilon_t$ with $|\rho| < 1$ and $\epsilon_0 = u_0 = 0$, where $\mathcal{F}_{n,t} = \sigma(\epsilon_0, \epsilon_1, ..., \epsilon_t, x_1, ..., x_n)$ and $\{\epsilon_t, \mathcal{F}_{n,t}\}_{t=1}^n$ forms a martingale difference sequence satisfying, as $n \to \infty$ first and then $m \to \infty$,

$$\max_{m \le t \le n} |E(\epsilon_t^2|\mathcal{F}_{n,t-1}) - \sigma^2| \to 0, \quad a.s.,$$

where σ^2 is a given constant, and $\sup_{\substack{1\le t\le n \\ n\ge 1}} E(|\epsilon_t|^q|\mathcal{F}_{n,t-1}) < \infty$ a.s. for some $q > 2$.

Assumption 2.3. (a) $\int_{-\infty}^{\infty} K(s)ds = 1$ and $K(\cdot)$ has a compact support; (b) For any $x, y \in R$, $|K(x) - K(y)| \le C\,|x-y|$, where C is a positive constant; (c) For $p \ge 2$,

$$\int y^p K(y)dy \neq 0, \quad \int y^i K(y)dy = 0, \quad i = 1, 2, ..., p-1.$$

Assumption 2.4. (a) There exist a $0 < \beta \le 1$ and $\alpha \ge 0$ such that

$$|m(x+y) - m(x)| \;\le\; C\,(1 + |x|^\alpha)\,|y|^\beta,$$

for any $x \in R$ and y sufficiently small, where C is a positive constant; (b) For given fixed x, $m(x)$ has a continuous $p+1$ derivatives in a small neighborhood of x, where $p \ge 2$ is defined as in Assumption 2.3(c).

We have the following main results.

THEOREM 2.1. *Under Assumptions 2.1–2.2, 2.3(a) and 2.4(a), we have*

$$\Big(\sum_{t=1}^n K[(x_t - x)/h]\Big)^{1/2}\big[\hat{m}(x) - m(x)\big] \;\to_D\; N(0, \sigma_1^2), \tag{2.5}$$

for any h satisfying $nh^2 \to \infty$ and $nh^{2+4\beta} \to 0$, where $\sigma_1^2 = (1-\rho^2)^{-1}\sigma^2 \int_{-\infty}^{\infty} K^2(s)dt$. If in addition Assumptions 2.3(c) and 2.4(b), then

$$\Big(\sum_{t=1}^n K[(x_t - x)/h]\Big)^{1/2}\Big[\hat{m}(x) - m(x) - \frac{h^p m^{(p)}(x)}{p!}\int_{-\infty}^{\infty} y^p K(y)dy\Big] \;\to_D\; N(0, \sigma_1^2), \tag{2.6}$$

for any h satisfying $nh^2 \to \infty$ and $nh^{2+4p} = 0(1)$.

THEOREM 2.2. *Under Assumptions 2.1–2.2, 2.3(a)–(b), 2.4(a) and* $\sum_{i=0}^{\infty} i|\phi_i| < \infty$, *we have*

$$|\widehat{\rho} - \rho| = O_P\big\{n^{\alpha/2}h^{\beta} + (nh^2)^{-1/4}\big\}, \tag{2.7}$$

and with $\sigma_2^2 = \sigma^2 \int_{-\infty}^{\infty} K^2(s)dt$,

$$\big(\sum_{t=1}^{n} K[(x_t - x)/h]\big)^{1/2} \big[\hat{m}_2(x) - m(x)\big] \to_D N(0, \sigma_2^2), \tag{2.8}$$

for any h *satisfying that* $nh^{2+4\beta} \to 0$, $n^{\alpha}h^{2\beta} \to 0$ *and* $n^{1-\epsilon_0}h^2 \to \infty$ *for some* $\epsilon_0 > 0$. *If in addition Assumptions 2.3(c) and 2.4(b), then*

$$\Big(\sum_{t=1}^{n} K[(x_t - x)/h]\Big)^{1/2} \Big[\hat{m}_2(x) - m(x) - \frac{h^p m^{(p)}(x)}{p!} \int_{-\infty}^{\infty} y^p K(y)dy\Big] \to_D \; N(0, \sigma_2^2), \tag{2.9}$$

for any h *satisfying that* $nh^{2+4p} = O(1)$, $n^{\alpha}h^{2\beta} \to 0$ *and* $n^{1-\epsilon_0}h^2 \to \infty$ *for some* $\epsilon_0 > 0$.

Remark 1. Theorem 2.1 generalizes the related results in previous articles. See, for instance, Wang and Phillips (2009a, 2011), where the authors investigated the asymptotics under $\rho = 0$ and $\tau = 0$. As noticed in previous works, the conditions to establish our results are quite weak, in particular, a wide range of regression function $m(x)$ is included in Assumption 2.4(a), like $m(x) = 1/(1 + \theta\,|x|^{\beta})$, $m(x) = (a + be^x)/(1 + e^x)$ and $m(x) = \theta_1 + \theta_2 x + ... + \theta_k x^{k-1}$.

Remark 2. As $|\rho| < 1$, Theorem 2.2 confirms the claim that $\hat{m}_2(x)$ achieves efficiency improvements over the usual estimator $\hat{m}(x)$ under certain additional conditions on $m(x)$ and the bandwidth h. Among these additional conditions, the requirement on the bandwidth h (that is, $n^{\alpha}h^{2\beta} \to 0$ and $n^{1-\epsilon_0}h^2 \to \infty$, where ϵ_0 can be sufficiently small) imply that $0 \le \alpha < \beta$, which in turn requires that the rate of $m(x)$ divergence to ∞ on the tail is not fast than $|x|^{1+\beta}$. In comparison to Theorem 2.1, this is a little bit restrictive but it is reasonable, due to the fact that the consistency result (2.7) heavily depend on the following convergence for the kernel estimator $\hat{m}(x)$:

$$\frac{1}{n}\sum_{t=1}^{n} \big[\hat{m}(x_t) - m(x_t)\big]^2 \tag{2.10}$$

as $n \to \infty$. As $x_t \sim \sqrt{t}$ under our model, it is natural for the restriction on the tail of $m(x)$ to enable (2.10). The result (2.10) is a consequence of Theorem 3.1 in next section, which provides a strong approximation result on the convergence to a local time process.

3 Strong approximation to local time

This section investigates strong approximation to a local time process which essentially provides a technical tool in the development of the uniform convergence such as (2.10) for the kernel estimator $\hat{m}(x)$. As the condition imposed is different, this section can be read separately.

Let $x_{k,n}, 1 \le k \le n, n \ge 1$ be a triangular array, constructed from some underlying nonstationary time series and assume that there is a continuous limiting Gaussian process $G(t), 0 \le t \le 1$, to which $x_{[nt],n}$ converges weakly, where $[a]$ denotes the integer part of a. In many applications, we let $x_{k,n} = d_n^{-1} x_k$ where x_k is a nonstationary time series, such as a unit root or long memory process, and d_n is an appropriate standardization factor. This section is concerned with the limit behaviour of the statistic $S_n(t)$, defined by

$$S_n(t) = \frac{c_n}{n} \sum_{k=1}^{n} g[c_n \left(x_{k,n} - x_{[nt],n}\right)], \quad t \in [0,1], \tag{3.1}$$

where c_n is a certain sequence of positive constants and g is a real integrable function on R. As noticed in last section and previous research [see, e.g., Wang and Phillips (2012)], this kind of statistic appears in the inference for the unknown regression function $m(x)$ and its limit behaviour plays a key role in the related research fields.

The aim of this section is to provide a strong approximation result for the target statistic. To achieve our aim, we make use of the following assumptions.

Assumption 3.1. $\sup_x |x|^\gamma |g(x)| < \infty$ for some $\gamma > 1$, $\int_{-\infty}^{\infty} |g(x)| dx < \infty$ *and* $|g(x) - g(y)| \le C|x-y|$ whenever $|x-y|$ is sufficiently small on R.

Assumption 3.2. On a rich probability space, there exist a continuous local martingale $G(t)$ having a local time $L_G(t,s)$ and a sequence of stochastic processes $G_n(t)$ such that $\{G_n(t); 0 \le t \le 1\} =_D \{G(t); 0 \le t \le 1\}$ for each $n \ge 1$ and

$$\sup_{0 \le t \le 1} |x_{[nt],n} - G_n(t)| \quad = \quad o_{a.s.}(n^{-\delta_0}). \tag{3.2}$$

for some $0 < \delta_0 < 1$.

Assumption 3.3. For all $0 \le j < k \le n, n \ge 1$, there exist a sequence of σ*-fields* $\mathcal{F}_{k,n}$ (define $\mathcal{F}_{0,n} = \sigma\{\phi, \Omega\}$, the trivial σ-field) such that,

(i) $x_{j,n}$ are adapted to $\mathcal{F}_{j,n}$ and, conditional on $\mathcal{F}_{j,n}$, $[n/(k-j)]^d (x_{k,n} - x_{j,n})$ where $0 < d < 1$, has a density $h_{k,j,n}(x)$ satisfying that $h_{k,j,n}(x)$ is uniformly bounded by a constant K and

(ii) $\sup_{u\in R} \big|h_{k,j,n}(u+t)-h_{k,j,n}(u)\big| \le C\, \min\{|t|,1\}$, whenever n and $k-j$ are sufficiently large and $t \in R$.

Assumption 3.4. There is a $\epsilon_0 > 0$ such that $c_n \to \infty$ and $n^{-1+\epsilon_0}c_n \to 0$.

The following is our main result.

THEOREM 3.1. *Suppose Assumptions 3.1–3.4 hold. On the same probability space as in Assumption 3.2, for any $l > 0$, we have*

$$\sup_{0\le t\le 1} |S_n(t) - \tau\, L_{nt}| \;=\; o_P(\log^{-l} n) \tag{3.3}$$

where $\tau = \int_{-\infty}^{\infty} g(t)dt$ and $L_{nt} = \lim_{\epsilon\to 0} \frac{1}{2\epsilon}\int_0^1 I(|G_n(s) - G_n(t)| \le \epsilon)ds$.

Due to technical difficulty, the rates in (3.3) may not be optimal, which, in our guess, should have the form $n^{-\delta_1}$, where $\delta_1 > 0$ is related to $\delta_0 > 0$ given in Assumption 3.2. However, by noting $\{L_{nt}; 0 \le t \le 1\} =_D \{L_G(1, G(t)); 0 \le t \le 1\}$ [1] due to $\{G_n(t); 0 \le t \le 1\} =_D \{G(t); 0 \le t \le 1\}$, the result (3.3) is enough in many applications. To illustrate, we have the following theorem which provides the lower bound of $S_n(t)$ over $t \in [0,1]$. As a consequence, we establish the result (2.10) when x_t satisfies Assumption 2.1.

THEOREM 3.2. *Let x_t be defined as in Assumption 2.1 with $\sum_{k=0}^{\infty} k|\phi_k| < \infty$. Let Assumptions 2.3(a)–(b) hold. Then, for any $\eta > 0$ and fixed $M_0 > 0$, there exist $M_1 > 0$ and $n_0 > 0$ such that*

$$P\Big(\inf_{s=1,2,\ldots,n} \sum_{t=1}^{n} K[(x_t - x_s)/h] \ge \sqrt{n}h/M_1\Big) \;\ge\; 1-\eta, \tag{3.4}$$

for all $n \ge n_0$ and h satisfying that $h \to 0$ and $n^{1-\epsilon_0}h^2 \to \infty$ for some $\epsilon_0 > 0$. Consequently, we have

$$V_n \;:=\; \frac{1}{n}\sum_{t=1}^{n} \big[\hat m(x_t) - m(x_t)\big]^2 = O_P\big\{n^{\alpha}h^{2\beta} + (nh^2)^{-1/2}\big\}, \tag{3.5}$$

that is, (2.10) holds true if in addition $n^{\alpha}h^{2\beta} \to 0$.

---

[1] Here and belew, we define $L_G(1,x) = \lim_{\epsilon\to 0}\frac{1}{2\epsilon}\int_0^1 I(|G(s) - x| \le \epsilon)ds$, a local time process of the $G(s)$ whenever it exists.

4 Extension

We next propose another estimator that potentially can improve the efficiency even more. Following Linton and Mammen (2008), we obtain

$$\begin{aligned} m(x) &= \frac{1}{1+\rho^2}\left[E(Z_t^-(\rho)|x_t=x)-\rho E(Z_t^+(\rho)|x_{t-1}=x)\right] \\ &= \frac{1}{1+\rho^2}E\left[Z_t^-(\rho)-\rho Z_{t+1}^+(\rho)|x_t=x\right] \\ Z_t^-(\rho) &= y_t-\rho y_{t-1}+\rho m(x_{t-1}) \\ Z_t^+(\rho) &= y_t-\rho y_{t-1}-m(x_t). \end{aligned}$$

Let $\widehat{m}(.),\widehat{\rho}$ be initial consistent estimators, and let

$$\begin{aligned} \widehat{Z}_t^-(\rho) &= y_t-\widehat{\rho}y_{t-1}+\widehat{\rho}m(x_{t-1}) \\ \widehat{Z}_t^+(\rho) &= y_t-\widehat{\rho}y_{t-1}-\widehat{m}(x_t). \end{aligned}$$

Then let

$$\widehat{m}_{eff}(x)=\frac{1}{1+\widehat{\rho}^2}\widehat{E}\left[\widehat{Z}_t^-(\widehat{\rho})-\widehat{\rho}\widehat{Z}_{t+1}^+(\widehat{\rho})|x_t=x\right].$$

We claim that the following result holds. The proof is similar to earlier results and is ommitted.

THEOREM 4.1. *If in addition to Assumptions 2.1-2.4, $\sum_{i=0}^{\infty}i|\phi_i|<\infty$. Then, for any h satisfying $nh^2\log^{-4}n\to\infty$ and $nh^{2+4\beta}\to 0$, we have*

$$\Big(\sum_{t=1}^{n}K[(x_t-x)/h]\Big)^{1/2}\left[\widehat{m}_2(x)-m(x)\right]\to_D N(0,\sigma_3^2)$$

where $\sigma_2^2=(1+\rho^2)^{-1}\int_{-\infty}^{\infty}K^2(s)dt$.

We have $\sigma_3^2\le\sigma_2^2\le\sigma_1^2$, and so $\widehat{m}_{eff}(x)$ is more efficient (according to asymptotic variance) than $\widehat{m}_2(x)$, which itself is more efficient than $\widehat{m}(x)$.

5 Monte Carlo Simulation

We investigate the performance of our procedure on simulated data. We chose a similar design to Wang and Phillips (2009b) except we focus on error autocorrelation rather than contemporaneous endogeneity. We suppose that

$$y_t=m(x_t)+\sigma u_t,\ u_t=\rho_0 u_{t-1}+\varepsilon_t$$

with $m(x) = x$ and $m(x) = \sin(x)$, where $x_t = x_{t-1} + \eta_t$, with $\eta_t \sim N(0,1)$, $\sigma = 0.2$, and $\varepsilon_t \sim N(0,1)$. We used the Epanechnikov kernel for estimation with the bandwidth n^{-bc}. We examine a range of values of ρ_0 and the bandwidth constant bc, which are given below. We consider $n = 500, 1000$ and take $ns = 1000$ replications. We report the performance measure

$$\frac{1}{K}\sum_{k=1}^{K} \left|\widehat{m}(x_k) - m(x_k)\right|^2,$$

where $K = 101$ and $x_k = \{-1, -0.98, \ldots, 1\}$. The results for the linear case are given below in Tables 1-2. The results show that there is an improvement when going from $n = 500$ to $n = 1000$ and when going from $\widehat{m}_1$ to $\widehat{m}_2$. In the linear case, the bigger the bandwidth the better. In the cubic case (not shown), smaller bandwidths do better as the bias issue is much more severe in this case.

*** Tables 1-2 here ***

We show in Tables 3 and 4 the performance of the estimator of ρ for $n = 500$ and $n = 1000$. This varies with bandwidth and is generally quite poor, although improves with sample size. Finally, we give some indication of the distributional approximation. In Figure 1 we show the QQ plot for our (standardized) estimator $\widehat{m}_2$ in the case where $m(x) = \sin(x)$, $\rho = 0.95$, $n = 1000$, and $bc = 1/10$.

6 Proofs

Section 6.1 provides several preliminary lemmas. Some of them are are of independent interests. The proofs of main theorems will be given in Sections 6.2-6.4.

6.1 Preliminary lemmas

First note that

$$\begin{aligned} x_t &= \sum_{j=1}^{t} \lambda^{t-j}\, \xi_j = \sum_{j=1}^{t} \lambda^{t-j} \sum_{i=-\infty}^{j} \nu_i \phi_{j-i} \\ &= \lambda^{t-s}\, x_s + \sum_{j=s+1}^{t} \lambda^{t-j} \sum_{i=-\infty}^{s} \nu_i \phi_{j-i} + \sum_{j=s+1}^{t} \lambda^{t-j} \sum_{i=s+1}^{j} \nu_i \phi_{j-i} \\ &:= \lambda^{t-s}\, x_s + \Delta_{s,t} + x'_{s,t}, \end{aligned} \tag{6.1}$$

where

$$x'_{s,t} = \sum_{j=1}^{t-s} \lambda^{t-j-s} \sum_{i=1}^{j} \nu_{i+s}\phi_{j-i} = \sum_{i=s+1}^{t} \nu_i \sum_{j=0}^{t-i} \lambda^{t-j-i}\, \phi_j.$$

Write $d_{s,t}^2 = \sum_{i=s+1}^{t} \lambda^{2(t-i)} (\sum_{j=0}^{t-i} \lambda^{-j}\phi_j)^2 = E(x'_{s,t})^2$. Recall $\lim_{n\to\infty} \lambda^n = e^{\tau}$ and $\lim_{n\to\infty} \lambda^m = 1$ for any fixed m. The routine calculations show that, whenever n is sufficiently large,

$$e^{-|\tau|}/2 \le \lambda^k \le 2e^{|\tau|}, \quad \text{for all } -n \le k \le n \tag{6.2}$$

and there exist $\gamma_1, \gamma_0 > 0$ such that

$$\gamma_0 \le \inf_{n\ge k\ge m} |\sum_{j=0}^{k} \lambda^{-j}\phi_j| \le \gamma_1, \tag{6.3}$$

whenever n, m are sufficiently large. By virtue of (6.2)-(6.3), it is readily seen that $d_{s,t} \neq 0$ for all $0 \le s < t \le n$ because $\phi = \sum_{j=0}^{\infty} \phi_j \neq 0$ and $C_1(t-s) \le d_{s,t}^2 \le C_2(t-s)$. Consequently,

$$\tfrac{1}{\sqrt{t-s}} x'_{s,t} \text{ has a density } h_{s,t}(x), \tag{6.4}$$

which is uniformly bounded by a constant C_0 and $\int_{-\infty}^{\infty}(1+|u|)|\varphi_{s,t}(u)|du < \infty$ uniformly for $0 \le s < t \le n$, where $\varphi_{s,t}(u) = Ee^{iux'_{s,t}/\sqrt{t-s}}$, due to $\int(1+|u|)|Ee^{iu\nu_0}|dt < \infty$. See the proof of Corollary 2.2 in Wang and Phillips (2009a) and/or (7.14) and Proposition 7.2 (page 1934 there) of Wang and Phillips (2009b) with a minor modification. Hence, conditional on $\mathcal{F}_k = \sigma(\nu_j, -\infty < j \le k)$,

$$(x_t - x_s)/\sqrt{t-s} \text{ has a density } h_{s,t}^*(x) \equiv h_{s,t}(x - x_{s,t}^*/\sqrt{t-s}) \tag{6.5}$$

where $x_{s,t}^* = (\lambda^{t-s} - 1)\, x_s + \Delta_{s,t}$, satisfying, for any $u \in R$,

$$\begin{aligned} & \sup_x \left|h_{s,t}^*(x+u) - h_{s,t}^*(x)\right| \le \sup_x |h_{s,t}(x+u) - h_{s,t}(x)| \\ \le\ & C\Big|\int_{-\infty}^{\infty} (e^{-iv(x+u)} - e^{-ivx})\varphi_{s,t}(v)dv\Big| \\ \le\ & C \min\{|u|, 1\} \int_{-\infty}^{\infty} (1+|v|)\, |\varphi_{s,t}(v)|dv \le C_1 \min\{|u|, 1\}, \end{aligned} \tag{6.6}$$

where in the last second inequality of (6.6), we have used the inverse formula of the characteristic function.

We also have the following presentation for the x_t:

$$
\begin{aligned}
x_t &= \sum_{j=1}^{t} \lambda^{t-j}\,\xi_j = \sum_{j=1}^{t} \lambda^{t-j}\,\Big(\sum_{i=0}^{j-1}+\sum_{i=j}^{\infty}\Big)\phi_i \nu_{j-i} \\
&= \sum_{i=0}^{t-1}\phi_i\,\lambda^{t-i}\sum_{j=1}^{t}\lambda^{-j}\nu_j - \sum_{i=0}^{t-1}\phi_i\,\lambda^{t-i}\sum_{j=t-i+1}^{t}\lambda^{-j}\nu_j + \sum_{j=1}^{t}\lambda^{t-j}\sum_{i=0}^{\infty}\phi_{i+j}\nu_{-i} \\
&= a_t\,x_t' - x_t'' + x_t''', \quad \text{say,} \qquad (6.7)
\end{aligned}
$$

where $a_t = \sum_{i=0}^{t-1}\phi_i\,\lambda^{-i}$, $x_t' = \sum_{j=1}^{t}\lambda^{t-j}\nu_j$ and

$$
|x_t''| + |x_t'''| \le C_0 t^{1/(2+\delta)}, \quad a.s. \qquad (6.8)
$$

for some constant $C_0 > 0$. Indeed, using (6.2) and strong law, we obtain that, for some constant $C_0 > 0$,

$$
\begin{aligned}
|x_t''| &\le 2e^{|\tau|}\sum_{i=0}^{t-1}|\phi_i|\sum_{j=t-i+1}^{t}|\nu_j| \le 2e^{|\tau|}\max_{1\le j\le t}|\nu_j|\sum_{i=0}^{t-1} i|\phi_i| \\
&\le C\,t^{1/(2+\delta)}\,\Big(\frac{1}{t}\sum_{j=1}^{t}|\nu_j|^{2+\delta}\Big)^{1/(2+\delta)} \le C_0 t^{1/(2+\delta)}, \quad a.s., \qquad (6.9)
\end{aligned}
$$

since $E|\nu_1|^{2+\delta} < \infty$ and $\sum_{i=0}^{\infty} i|\phi_i| < \infty$. Note that

$$
\begin{aligned}
\sum_{j=1}^{\infty} j^{-1/(2+\delta)}E\Big|\sum_{i=0}^{\infty}\phi_{i+j}\nu_{-i}\Big| &\le \sum_{j=1}^{\infty} j^{-1/(2+\delta)}\Big(\sum_{i=j}^{\infty}\phi_i^2\Big)^{1/2} \\
&\le C\sum_{j=1}^{\infty} j^{-1-1/(2+\delta)}\Big(\sum_{i=j}^{\infty} i|\phi_i|\Big)^{1/2} < \infty,
\end{aligned}
$$

which yields that $\sum_{j=1}^{\infty} j^{-1/(2+\delta)}|\sum_{i=0}^{\infty}\phi_{i+j}\nu_{-i}| < \infty, a.s.$ It follows from (6.2) again and the Kronecker lemma that

$$
|x_t'''| \le C\sum_{j=1}^{t}\Big|\sum_{i=0}^{\infty}\phi_{i+j}\nu_{-i}\Big| = o(t^{1/(2+\delta)}), \quad a.s. \qquad (6.10)
$$

This proves (6.8).

We are now ready to provide several preliminary lemmas.

LEMMA 6.1. *Suppose that $p(x)$ satisfies $\int |p(x)|dx < \infty$ and Assumption 2.1 holds. Then, for any $h \to 0$ and all $0 \le s < t \le n$, we have*

$$
\begin{aligned}
E\big(|p(x_t/h)|\,\big|\mathcal{F}_s\big) &\le \frac{C_0 h}{\sqrt{t-s}}\int_{-\infty}^{\infty}|p(x + x_s/h)|dx \\
&= \frac{C_0 h}{\sqrt{t-s}}\int_{-\infty}^{\infty}|p(x)|dx, \quad a.s., \qquad (6.11)
\end{aligned}
$$

where $\mathcal{F}_s = \sigma\{\nu_s, \nu_{s-1}, ...\}$.

Proof. Recall (6.1), (6.5) and the independence of ν_k. The result (6.11) follows from a routine calculation and hence the details are omitted. □

LEMMA 6.2. *Suppose that $p(x)$ satisfies $\int[|p(x)| + p^2(x)]dx < \infty$ and $\int p(x)dx \neq 0$. Suppose Assumption 2.1 holds. Then, for any $h \to 0$ and $nh^2 \to \infty$,*

$$\frac{\phi}{\sqrt{n}h} \sum_{t=1}^{n} p\big[(x_t - x)/h\big] \quad \to_D \quad \int_{-\infty}^{\infty} p(x)dx \, L_G(1,0), \tag{6.12}$$

where $G(t) = W(t) + \kappa \int_0^t e^{\kappa(t-s)} W(s)ds$ with $W(s)$ being a standard Brownian motion and $L_G(r,x)$ is a local time of the Gaussian process $G(t)$.

Proof. This is a corollary of Theorem 3.1 of Wang and Phillips (2009a). The inspection on the conditions is similar to Proposition 7.2 of Wang and Phillips (2009b). We omit the details. □

LEMMA 6.3. *Suppose Assumptions 2.1-2.2 and 2.3(a) hold. Then, for any $h \to 0$ and $nh^2 \to \infty$,*

$$\sum_{t=1}^{n} u_t Z_{nt} \quad \to_D \quad N(0, \sigma_1^2), \tag{6.13}$$

where $Z_{nt} = K[(x_t - x)/h] / \big(\sum_{t=1}^n K[(x_t - x)/h]\big)^{1/2}$ and $\sigma_1^2 = (1-\rho^2)^{-1}\sigma^2 \int_{-\infty}^{\infty} K^2(x)dx$.

Proof. For the notation convenience, we assume $\sigma^2 = 1$ in the following proof. Note that $u_t = \sum_{k=1}^t \rho^{t-k}\epsilon_k$. We have $\sum_{t=1}^n u_t Z_{nt} = \sum_{k=1}^n \epsilon_k Z_{nk}^*$, where $Z_{nk}^* = \sum_{t=k}^n \rho^{t-k} Z_{nt}$. We first claim that

$$\sum_{k=1}^{n} Z_{nk}^2 \quad \to_P \quad \int_{-\infty}^{\infty} K^2(x)dx, \tag{6.14}$$

$$\sum_{k=1}^{n} Z_{nk}^{*2} \quad \to_P \quad (1-\rho^2)^{-1} \int_{-\infty}^{\infty} K^2(x)dx. \tag{6.15}$$

The proof of (6.14) is simple by applying Lemma 6.2. To see (6.15), note that

$$\begin{aligned}
\sum_{k=1}^{n} Z_{nk}^{*2} &= \Lambda_n^{-1} \sum_{k=1}^{n} \Big(\sum_{t=k}^{n} \rho^{t-k} K[(x_t - x)/h]\Big)^2 \\
&= \Lambda_n^{-1} \sum_{k=1}^{n} \sum_{t=k}^{n} \rho^{2(t-k)} K^2[(x_t - x)/h] + \Lambda_n^{-1} \Gamma_{1n} \\
&= (1-\rho^2)^{-1} \sum_{k=1}^{n} Z_{nk}^2 + \Lambda_n^{-1} (\Gamma_{1n} - \Gamma_{2n}) \quad \text{[by (6.14)]} \\
&= (1-\rho^2)^{-1} \int_{-\infty}^{\infty} K^2(x)dx + \Lambda_n^{-1} (\Gamma_{1n} - \Gamma_{2n}) + o_P(1), \qquad (6.16)
\end{aligned}$$

where $\Lambda_n = \sum_{t=1}^n K[(x_t - x)/h]$,

$$\Gamma_{1n} = 2 \sum_{k=1}^n \sum_{k \le s < t \le n} \rho^{s-k} \rho^{t-k} K[(x_s - x)/h] K[(x_t - x)/h],$$
$$\Gamma_{2n} = (1-\rho^2)^{-1} \sum_{t=1}^n K^2[(x_t - x)/h] \rho^{2t}.$$

Note that $\Lambda_n/(\sqrt{n}h) \to_D \phi^{-1} L_G(1,0)$ by Lemma 6.2. The result (6.15) will follow if we prove

$$\Gamma_{1n} + \Gamma_{2n} = o_P[(nh^2)^{1/2}]. \tag{6.17}$$

Recall $K(x) \le C$ and $|\rho| < 1$. It is readily seen that $\Gamma_{2n} \le C = o_P[(nh^2)^{1/2}]$, due to $nh^2 \to \infty$. On the other hand, by applying Lemma 6.1, for any $t > s$, we have

$$E\big\{K[(x_s - x)/h] K[(x_t - x)/h]\big\} \le E\Big[K[(x_s - x)/h] E\big\{K[(x_t - x)/h] \mid \mathcal{F}_s\big\}\Big]$$
$$\le \frac{Ch}{\sqrt{t-s}} \frac{h}{\sqrt{s}}.$$

It follows that

$$E\Gamma_{1n} \le C h^2 \sum_{k=1}^n \sum_{k \le s < t \le n} \rho^{s-k} \rho^{t-k} \frac{1}{\sqrt{t-s}} \frac{1}{\sqrt{s}}$$
$$\le C_1 h^2 \sum_{k=1}^n \frac{1}{\sqrt{k}} \le C h^2 \sqrt{n}.$$

which implies that $\Gamma_{1n} = O_P(h^2\sqrt{n}) = o_P[(nh^2)^{1/2}]$. This proves (6.17) and hence the claim (6.15).

We now turn to the proof of (6.13). Since, given $\{x_1, x_2, ..., x_n\}$, the sequence $(Z_{nk}^* \epsilon_k, k = 1, 2, ..., n)$ still forms a martingale difference by Assumption 2.2, it follows from Theorem 3.9 [(3.75) there] in Hall and Heyde (1980) with $\delta = q/2 - 1$ that

$$\sup_x \big|P\big(\sum_{t=1}^n \epsilon_t Z_{nt}^* \le x\sigma_1 \mid x_1, x_2, ..., x_n\big) - \Phi(x)\big| \le A(\delta)\, \mathcal{L}_n^{1/(1+q)}, \quad a.s.,$$

where $A(\delta)$ is a constant depending only on $q > 2$ and

$$\mathcal{L}_n = \frac{1}{\sigma_1^q} \sum_{k=1}^n |Z_{nk}^*|^q E(|\epsilon_k|^q \mid x_1, ..., x_n) + E\Big\{\Big|\frac{1}{\sigma_1^2} \sum_{k=1}^n Z_{nk}^{*2} \big[E(\epsilon_k^2 | \mathcal{F}_{k-1}) - 1\big]\Big|^{q/2} \mid x_1, ..., x_n\Big\}.$$

Recall that $K(x)$ is uniformly bounded and

$$\max_{1 \le k \le n} |Z_{nk}^*| \le C \max_{1 \le k \le n} |Z_{nk}| \le C/\big(\sum_{t=1}^n K[(x_t - x)/h]\big)^{1/2} = o_P(1),$$

by Assumption 2.3 and Lemma 6.2. Routine calculations, together with (6.15), show that

$$\mathcal{L}_n = o_P(1),$$

since $q > 2$. Therefore the dominate convergence theorem yields that

$$\begin{aligned} & \sup_x \big|P\big(\sum_{t=1}^n u_t Z_{nt} \le x\sigma_1\big) - \Phi(x)\big| \\ \le \quad & E\Big[\sup_x \big|P\big(\sum_{t=1}^n \epsilon_t Z_{nt}^* \le x\sigma_1 \mid x_1, x_2, ..., x_n\big) - \Phi(x)\big|\Big] \to 0. \end{aligned}$$

This completes the proof of Lemma 6.3. □

LEMMA 6.4. *Under Assumptions 2.3(a) and 2.4(a), for any $x \in R$, we have*

$$\big|\Lambda_n(x) - m(x)\big| \le C\,(1 + |x|^\alpha) h^\beta, \tag{6.18}$$

where $\Lambda_n(x) = \frac{\sum_{t=1}^n m(x_t)\, K[(x_t - x)/h]}{\sum_{t=1}^n K[(x_t - x)/h]}$. If in addition Assumption 2.4(b), we have

$$\big|\Lambda_n(x) - m(x) - \frac{h^p m^{(p)}(x)}{p!} \int_{-\infty}^{\infty} y^p K(y) dy\big| = o_P[(nh^2)^{-1/4}], \tag{6.19}$$

whenever $nh^2 \to \infty$ and $nh^{2+4p} = O(1)$, for any fixed x.

Proof. By Assumption 2.4(a) and the compactness of $K(x)$, the result (6.18) is simple. The proof of (6.19) is the same as in the proof of Theorem 2.2 in Wang and Phillips (2011). We omit the details. □

LEMMA 6.5. *For any $s, t \in R$ and $\xi > 0$, there exists a constant C such that*

$$|L_G(1, s) - L_G(1, t)| \le C\,|s - t|^{1/2-\xi} \quad a.s., \tag{6.20}$$

where $G(x)$ is a continuous local martingale.

Proof. See Corollary 1.8 of Revuz and Yor (1994, p. 226). □

LEMMA 6.6. *Suppose that Assumptions 3.1-3.4 hold. Then, for any $l > 0$, we have*

$$I_n := \sup_{t\in R} \sup_{s:|s-t|\le\epsilon_n} \Big|\frac{c_n}{n} \sum_{k=1}^n f_{t,s}(x_{k,n})\Big| = O_{a.s.}(\log^{-l} n) \tag{6.21}$$

where $\epsilon_n \le c_n n^{-l_1}$ for some $l_1 > 0$ and $f_{t,s}(x) = g(c_n x + t) - g(c_n x + s)$.

Proof. See Lemma 3.5 of Liu, Chan and Wang (2013). □

6.2 Proofs of Theorem 2.1 and 2.2

We only prove Theorem 2.2. Using Lemmas 6.3 and 6.4, the proof of Theorem 2.1 is standard [see, e.g., Wang and Phillips (2011)], and hence the details are omitted.

Start with (2.7). Recall that $\hat{u}_t = y_t - \hat{m}_1(x_t) = u_t + m(x_t) - \hat{m}_1(x_t)$. Simple calculations show that

$$\begin{aligned}
\hat{\rho} - \rho &= \frac{\sum_{s=2}^{n}(\widehat{u}_s - \rho\widehat{u}_{s-1})\widehat{u}_{s-1}}{\sum_{s=2}^{n}\widehat{u}_{s-1}^2} \\
&= \frac{\sum_{s=2}^{n}\epsilon_s\,\widehat{u}_{s-1}}{\sum_{s=2}^{n}\widehat{u}_{s-1}^2} + \frac{\sum_{s=2}^{n}\widehat{u}_{s-1}\big[m(x_s) - \hat{m}_1(x_s) + \rho\,\{\hat{m}_1(x_{s-1}) - m(x_{s-1})\}\big]}{\sum_{s=2}^{n}\widehat{u}_{s-1}^2} \\
&:= R_{1n} + R_{2n}. \qquad (6.22)
\end{aligned}$$

As $V_n = \frac{1}{n}\sum_{t=1}^{n}\big[\hat{m}_1(x_t) - m(x_t)\big]^2 = O_P\big\{n^\alpha h^{2\beta} + (nh^2)^{-1/2}\big\}$ by (3.5) of Theorem 3.2, it follows from $\frac{1}{n}\sum_{t=1}^{n}u_t^2 \to (1-\rho^2)^{-1}\sigma^2, a.s.$, that $\frac{1}{n}\sum_{s=2}^{n}\widehat{u}_{s-1}^2 \to_P (1-\rho^2)^{-1}\sigma^2$, whenever $n^\alpha h^{2\beta} \to 0$ and $nh^2 \to \infty$. This, together with Hölder's inequality, yields that

$$|R_{2n}| \;\leq\; 2(1+\rho^2)^{1/2}V_n^{1/2}/\big(\sum_{s=2}^{n}\widehat{u}_{s-1}^2\big)^{1/2} = O_P\big\{n^{\alpha/2}h^\beta + (nh^2)^{-1/4}\big\}.$$

On the other hand, by recalling Assumption 2.2, it is readily seen that $R_{1n} = O_P(n^{-1/2})$. Taking these facts into (6.22), we obtain (2.7).

We next prove (2.8). We may write

$$\begin{aligned}
\hat{m}_2(x) - m(x) &= \frac{\widehat{\rho}\,\sum_{t=1}^{n}\big[\hat{m}_1(x_{t-1}) - m(x_{t-1})\big]\,K[(x_t - x)/h]}{\sum_{t=1}^{n}K[(x_t - x)/h]} \\
&\quad + \frac{(\rho - \widehat{\rho})\,\sum_{t=1}^{n}u_{t-1}\,K[(x_t - x)/h]}{\sum_{t=1}^{n}K[(x_t - x)/h]} \\
&\quad + \frac{\sum_{t=1}^{n}\big[m(x_t) - m(x)\big]\,K[(x_t - x)/h]}{\sum_{t=1}^{n}K[(x_t - x)/h]} + \frac{\sum_{t=1}^{n}\epsilon_t\,K[(x_t - x)/h]}{\sum_{t=1}^{n}K[(x_t - x)/h]} \\
&:= \widehat{\rho}\,I_{1n} + (\rho - \widehat{\rho})\,I_{2n} + I_{3n} + I_{4n}.
\end{aligned}$$

Furthermore we may divide I_{1n} into

$$\begin{aligned}
I_{1n} &= \frac{1}{\sum_{t=1}^{n}K[(x_t - x)/h]}\Big\{\sum_{t=1}^{n}K[(x_t - x)/h]\,\frac{\sum_{s=2}^{n}\big[m(x_{s-1}) - m(x_{t-1})\big]\,K[(x_{s-1} - x_{t-1})/h]}{\sum_{s=2}^{n}K[(x_{s-1} - x_{t-1})/h]} \\
&\quad + \sum_{t=1}^{n}K[(x_t - x)/h]\,\frac{\sum_{s=2}^{n}u_{s-1}\,K[(x_{s-1} - x_{t-1})/h]}{\sum_{s=2}^{n}K[(x_{s-1} - x_{t-1})/h]}\Big\} \\
&:= I_{1n1} + I_{1n2}.
\end{aligned}$$

Using Lemma 6.3 and (2.7), we have

$$(\rho - \widehat{\rho})\,\big(\sum_{t=1}^{n}K[(x_t - x)/h]\big)^{1/2}\,I_{2n} \to_P 0$$

and (with $\rho = 0$ in Lemma 6.3)

$$\Big(\sum_{t=1}^{n} K[(x_t - x)/h]\Big)^{1/2} I_{4n} \to_D N(0, \sigma_2^2).$$

Using (6.18) and $\frac{1}{\sqrt{n}h}\sum_{t=1}^{n} K[(x_t - x)/h] \to_D \phi^{-1} L_G(1,0)$, we have

$$\begin{aligned}
& \Big(\sum_{t=1}^{n} K[(x_t - x)/h]\Big)^{1/2} |I_{3n}| \\
\le\ & C\,(nh^{2+4\beta})^{1/4} \Big(\frac{1}{\sqrt{n}h}\sum_{t=1}^{n} K[(x_t - x)/h]\Big)^{1/2} = o_P(1).
\end{aligned}$$

Similarly, by recalling K has a compact support, we obtain

$$\begin{aligned}
& \Big(\sum_{t=1}^{n} K[(x_t - x)/h]\Big)^{1/2} |I_{1n1}| \\
\le\ & Ch^{\beta} \Big(\sum_{t=1}^{n} K[(x_t - x)/h]\Big)^{-1/2} \sum_{t=1}^{n} (1 + |x_t|^{\alpha})\, K[(x_t - x)/h] \\
\le\ & C\,(nh^{2+4\beta})^{1/4} \Big(\frac{1}{\sqrt{n}h}\sum_{t=1}^{n} K[(x_t - x)/h]\Big)^{1/2} = o_P(1).
\end{aligned}$$

Combining all above facts, to prove (2.8), it suffices to show that

$$I_{1n2} \quad = \quad o_P[(nh^2)^{-1/4}]. \tag{6.23}$$

To this end, for each fixed x, write

$$\begin{aligned}
\Omega_{1n} &= \Big\{\omega : \sum_{s=1}^{n} K[(x_s - x)/h] \ge \sqrt{n}h\delta_n\Big\}, \\
\Omega_{2n} &= \Big\{\omega : \inf_{t=1,2,\dots,n} \sum_{s=2}^{n} K[(x_{s-1} - x_t)/h] \ge \sqrt{n}h\delta_n\Big\}
\end{aligned}$$

where $\delta_n \downarrow 0$ is chosen later and ω denotes the sample points. As $P(\Omega_{1n} \cup \Omega_{2n}) \to 0$ by (3.4) and Lemma 6.2, the result (6.23) will follow if we prove

$$I(\Omega_{1n}\,\Omega_{2n})\, I_{1n2} \quad = \quad o_P[(nh^2)^{-1/4}]. \tag{6.24}$$

Recall $u_s = \sum_{k=1}^{s} \rho^{s-k}\epsilon_k$. I_{1n2} can be rewritten as

$$\begin{aligned}
I_{1n2} &= \frac{1}{\sum_{t=1}^{n} K[(x_t - x)/h]} \sum_{s=2}^{n} u_{s-1}\, J_{ns} \\
&= \frac{1}{\sum_{t=1}^{n} K[(x_t - x)/h]} \sum_{k=1}^{n-1} \epsilon_k \sum_{s=k+1}^{n} \rho^{s-1-k}\, J_{ns},
\end{aligned}$$

where

$$J_{ns} \;=\; \sum_{t=1}^{n} \frac{K[(x_{s-1}-x_{t-1})/h]\,K[(x_t-x)/h]}{\sum_{s=2}^{n} K[(x_{s-1}-x_{t-1})/h]}.$$

It follows easily from the the conditional arguments and Holder's inequality that

$$\begin{aligned} E\big[I(\Omega_{1n}\,\Omega_{2n})\,I_{2n2}\big]^2 &\le \frac{C}{(nh^2\delta_n)^2}\sum_{k=1}^{n-1} E\Big(\sum_{s=k+1}^{n}\rho^{s-1-k}\,J^*_{ns}\Big)^2 \\ &\le \frac{C}{(nh^2\delta_n)^2}\sum_{k=1}^{n-1}\sum_{s=k+1}^{n}\rho^{s-1-k}\sum_{s=k+1}^{n}\rho^{s-1-k}\,EJ^{*2}_{ns} \\ &\le \frac{nC}{(nh^2\delta_n)^2}\max_{1\le s\le n} EJ^{*2}_{ns}. \end{aligned} \tag{6.25}$$

where

$$J^*_{ns} \;=\; \sum_{t=1}^{n} K[(x_{s-1}-x_{t-1})/h]\,K[(x_t-x)/h].$$

Simple calculations show that, by letting $\sum_{k=i}^{j}=0$ if $j<i$,

$$\begin{aligned} EJ^{*2}_{ns} &= E\Big(\sum_{t=1}^{n} K[(x_{s-1}-x_{t-1})/h]\,K[(x_t-x)/h]\Big)^2 \\ &\le 2E\Big\{\sum_{t=1}^{s-2} K[(x_{s-1}-x_{t-1})/h]\,K[(x_t-x)/h]\Big\}^2 \\ &\quad + 2E\Big\{\sum_{t=s}^{n} K[(x_{s-1}-x_{t-1})/h]\,K[(x_t-x)/h]\Big\}^2 \\ &\quad + 2EK^2[(x_{s-1}-x_{s-2})/h]\,K^2[(x_{s-1}-x)/h] \\ &:= T_{1n}+T_{2n}+T_{3n}. \end{aligned}$$

Assume $t_1<t_2\le s-2$. Recall $K(x)$ has a compact support. It follows from Lemma 6.1 with $p(x)=K(x-x_{t_1-1}/h)K(x-x_{t_2-1}/h)$ that

$$\begin{aligned} &E\Big\{K[(x_{s-1}-x_{t_1-1})/h]K[(x_{s-1}-x_{t_2-1})/h]\mid \mathcal{F}_{t_2}\Big\} \\ &\le \frac{Ch}{\sqrt{s-t_2}}\int_{-\infty}^{\infty} K[y+(x_{t_2}-x_{t_1-1})/h]K[y+(x_{t_2}-x_{t_2-1})/h]\,dy \\ &\le \frac{Ch}{\sqrt{s-t_2}}\int_{-\infty}^{\infty} K(y)K[y+(x_{t_2-1}-x_{t_1-1})/h]\,dy, \quad a.s. \end{aligned}$$

This, together with the repeatedly similar utilization of Lemma 6.1, yields that, for $t_1<$

$t_2 \le s-2$,

$$\begin{aligned}
\Psi_{s,t_1,t_2} &:= E\Big\{K[(x_{s-1}-x_{t_1-1})/h]K[(x_{s-1}-x_{t_2-1})/h]K[(x_{t_1}-x)/h]K[(x_{t_2}-x)/h]\Big\} \\
&\le \frac{Ch}{\sqrt{s-t_2}}\int_{-\infty}^{\infty} E\Big\{K[y+(x_{t_2-1}-x_{t_1-1})/h]K[(x_{t_1}-x)/h]K[(x_{t_2}-x)/h]\Big\}K(y)dy \\
&\le \frac{Ch^2}{\sqrt{s-t_2}}\int_{-\infty}^{\infty} E\Big\{K[y+(x_{t_2-1}-x_{t_1-1})/h]K[(x_{t_1}-x)/h]\Big\}K(y)dy \\
&\quad .... \\
&\le \frac{Ch^4}{\sqrt{s-t_2}}\frac{1}{\sqrt{t_2-t_1}}\frac{1}{\sqrt{t_1}}.
\end{aligned}$$

Similarly, for $t_1 = t_2 < s-2$,

$$\Psi_{s,t_1,t_1} = E\Big\{K^2[(x_{s-1}-x_{t_1-1})/h]K^2[(x_{t_1}-x)/h]\Big\} \le \frac{Ch^2}{\sqrt{s-t_1}}\frac{1}{\sqrt{t_1}}.$$

We now obtain, for any $1 \le s \le n$,

$$\begin{aligned}
T_{1n} &= \sum_{t_1=1}^{s-2}\Psi_{s,t_1,t_1} + 2\sum_{1\le t_1<t_2\le s-2}\Psi_{s,t_1,t_2} \\
&\le \sum_{t_1=1}^{s-2}\frac{Ch^2}{\sqrt{s-t_1}}\frac{1}{\sqrt{t_1}} + \sum_{1\le t_1<t_2\le s-2}\frac{Ch^4}{\sqrt{s-t_2}}\frac{1}{\sqrt{t_2-t_1}}\frac{1}{\sqrt{t_1}} \\
&\le Ch^2(1+\sqrt{n}h^2).
\end{aligned}$$

Similarly we may prove

$$T_{2n}+T_{3n} \le Ch^2(1+\sqrt{n}h^2).$$

Combining all these estimates, it follows from (6.25) that

$$E\big[I(\Omega_{1n}\,\Omega_{2n})\,I_{2n2}\big]^2 \le C\delta_n^{-2}\big[(nh^2)^{-1}+n^{-1/2}\big] = o[(nh^2)^{-1/2}],$$

by choosing $\delta_n = \min\{(nh^2)^{1/8}, h^{-1/4}\} \to \infty$, whenever $h \to 0$ and $nh^2 \to \infty$. This proves (6.23) and also completes the proof of Theorem 2.2. $\Box$

6.3 Proof of Theorem 3.1.

The idea for the proof of (3.3) is similar to that of Liu, Chan and Wang (2013), but there are some essential difference in details. We restate here for the convenience of reading. Without loss of generality, we assume $\tau = \int g(x)dx = 1$. Define $\bar{g}(x) = g(x)I\{|x| \le n^{\zeta}/2\}$, where $0 < \zeta < 1-\delta_0/\gamma$ is small enough such that $n^{\zeta}/c_n \le n^{-\delta_0}$, where γ and δ_0

are given in Assumption 3.1 and 3.2 respectively. Further let $\varepsilon = n^{-\alpha}$ with $0 < \alpha < \delta_0/2$ and define a triangular function

$$g_\epsilon(y) = \begin{cases} 0, & |y-\varepsilon| > \epsilon, \\ \frac{y+\varepsilon}{\varepsilon^2}, & -\epsilon \le y \le 0, \\ \frac{\varepsilon-y}{\varepsilon^2}, & 0 \le y \le \epsilon, \end{cases}$$

It suffices to show that

$$\Phi_{1n} := \max_{1\le k\le n} \Big| \frac{c_n}{n} \sum_{j=1}^{n} \{ g[c_n(x_{j,n}-x_{k,n})] - \bar g[c_n(x_{j,n}-x_{k,n})] \} \Big| = o_{a.s.}(\log^{-l} n), \tag{6.26}$$

$$\Phi_{2n} := \max_{1\le k\le n} \Big| \frac{c_n}{n} \sum_{j=1}^{n} \bar g[c_n(x_{j,n}-x_{k,n})] - \frac{1}{n}\sum_{j=1}^{n} g_\epsilon(x_{j,n}-x_{k,n}) \Big| = o_{a.s.}(\log^{-l} n), \tag{6.27}$$

$$\Phi_{3n} := \sup_{0\le t\le 1} \Big| \frac{1}{n} \sum_{j=1}^{n} g_\epsilon(x_{j,n}-x_{[nt],n}) - L_{nt} \Big| = o_P(\log^{-l} n). \tag{6.28}$$

The proof of (6.26) is simple. Indeed, by recalling $\sup_x |x|^\gamma |g(x)| < \infty$, it follows that

$$\Phi_{1n} \;\le\; c_n \sup_{|x|\ge n^\tau/2} |g(x)| I\{|x| > n^\zeta/2\} \le C\, n^{-\zeta\gamma} c_n = o(\log^{-l} n).$$

as $n^\zeta/c_n \le n^{-\delta_0}$ and $\gamma > \delta_0/(1-\zeta)$.

We next prove (6.28). Recalling $\int_{-\infty}^{\infty} g_\epsilon(y)dy = 1$, it follows from the definition of occupation time and Lemma 6.5 that

$$\begin{aligned} & \Big| \int_0^1 g_\epsilon[G(s)-G(t)]ds - L_G(1,G(t)) \Big| \\ = & \Big| \int_{-\infty}^{\infty} g_\epsilon[y-G(t)] L_G(1,y)dy - L_G(1,G(t)) \Big| \\ \le & \int_{-\infty}^{\infty} g_\epsilon(y) |L_G(1,y+G(t)) - L_G(1,G(t))| dy \\ \le & C\varepsilon^{1/2-\xi} \quad a.s. \end{aligned}$$

for any $\xi > 0$, uniformly for $t \in [0,1]$. Hence, by taking $\xi = 1/4$ and noting

$$\int_0^1 g_\epsilon[G_n(s)-G_n(t)]ds - L_{nt} =_D \int_0^1 g_\epsilon[G(s)-G(t)]ds - L_G(1,G(t)),$$

due to $\{G_n(t); 0\le t\le 1\} =_D \{G(t); 0 \le t \le 1\}, n \ge 1$, we have

$$\begin{aligned} & P\Big(\Big| \int_0^1 g_\epsilon[G_n(s)-G_n(t)]ds - L_{nt} \Big| \ge \log^{-l-1} n \Big) \\ = & P\Big(\Big| \int_0^1 g_\epsilon[G(s)-G(t)]ds - L_G(1,G(t)) \Big| \ge \log^{-l-1} n \Big) = o(1), \end{aligned} \tag{6.29}$$

as $n \to \infty$. This, together with Assumption 3.2 and the fact that $|g_\epsilon(y) - g_\epsilon(z)| \le \varepsilon^{-2}|y-z|$, implies that

$$\begin{aligned}
& \Big|\frac{1}{n}\sum_{j=1}^n g_\epsilon(x_{j,n} - x_{[nt],n}) - L_{nt}\Big| \\
\le & \ \Big|\int_0^1 g_\epsilon(x_{[ns],n} - x_{[nt],n})ds - \int_0^1 g_\epsilon[G_n(s) - G_n(t)]ds\Big| + 2/(\varepsilon n) \\
& + \Big|\int_0^1 g_\epsilon[G_n(s) - G_n(t)]ds - L_{nt}\Big| \\
= & \ O_{a.s}(\varepsilon^{-2} n^{-\delta_0}) + 2/(\varepsilon n) + O_P(\log^{-l-1} n) \\
\le & \ O_P(n^{2\alpha - \delta_0} + \log^{-l-1} n) = o_P(\log^{-l} n),
\end{aligned}$$

uniformly for $t \in [0,1]$, as $\alpha < \delta_0/2$. This yields (6.28).

We finally prove (6.27), let $\bar{g}_{\varepsilon n}(z)$ be the step function which takes the value $g_\epsilon(mn^\zeta/c_n)$ for $z \in [mn^\zeta/c_n, (m+1)n^\zeta/c_n)$, $m \in \mathbb{Z}$. It suffices to show that, uniformly for all $1 \le k \le n$, (letting $\bar{g}_j(y) = \bar{g}(c_n(x_{j,n} - x_{k,n}) - y)$),

$$\begin{aligned}
\Delta_{1n}(k) &:= \Big|\frac{1}{n}\sum_{j=1}^n g_\epsilon(x_{j,n} - x_{k,n}) - \frac{1}{n}\sum_{j=1}^n \bar{g}_{\varepsilon n}(x_{j,n} - x_{k,n}) \int_{-\infty}^{\infty} \bar{g}_j(y)dy\Big| \\
&= o_{a.s.}(\log^{-l} n) \qquad (6.30) \\
\Delta_{2n}(k) &:= \Big|\frac{1}{n}\sum_{j=1}^n \bar{g}_{\varepsilon n}(x_{j,n} - x_{k,n}) \int_{-\infty}^{\infty} \bar{g}_j(y)dy - \int_{-\infty}^{\infty} \frac{1}{n}\sum_{j=1}^n g_\epsilon(y/c_n)\bar{g}_j(y)dy\Big| \\
&= o_{a.s.}(\log^{-l} n), \qquad (6.31) \\
\Delta_{3n}(k) &:= \Big|\int_{-\infty}^{\infty} \frac{1}{n}\sum_{j=1}^n g_\epsilon(y/c_n)\bar{g}_j(y)dy - \frac{c_n}{n}\sum_{j=1}^n \bar{g}[c_n(x_{j,n} - x_{k,n})]\Big| \\
&= o_{a.s.}(\log^{-l} n), \qquad (6.32)
\end{aligned}$$

In fact, by noting that $|g_\epsilon(y) - g_\epsilon(z)| \le \varepsilon^{-2}|y-z|$ and

$$\begin{aligned}
|\bar{g}_{\varepsilon n}(y) - g_\epsilon(z)| &\le |\bar{g}_{\varepsilon n}(y) - g_\varepsilon(y)| + |g_\epsilon(y) - g_\epsilon(z)| \\
&\le C\varepsilon^{-2}(n^\zeta/c_n + |y-z|), \qquad (6.33)
\end{aligned}$$

(6.30) follows from that, uniformly for all $1 \le j, k \le n$,

$$\begin{aligned}
& \Big|g_\varepsilon(x_{j,n} - x_{k,n}) - \bar{g}_{\varepsilon n}(x_{j,n} - x_{k,n}) \int_{-\infty}^{\infty} \bar{g}_j(y)dy\Big| \\
& \le \Big|g_\varepsilon(x_{j,n} - x_{k,n}) - \bar{g}_{\varepsilon n}(x_{j,n} - x_{k,n})\Big| + |\bar{g}_{\varepsilon n}(x_{j,n} - x_{k,n})|\Big|1 - \int_{-\infty}^{\infty} \bar{g}_j(y)dy\Big| \\
& \le C\varepsilon^{-2} n^\zeta/c_n + C_1 \epsilon^{-1} n^{-\zeta(\gamma-1)} = o_{a.s.}(\log^{-l} n).
\end{aligned}$$

where we have used the fact that (recalling $\int g(y)dy = 1$),

$$\Big|1 - \int_{-\infty}^{\infty} \bar{g}_j(y)dy\Big| \le \Big| \int_{-\infty}^{\infty} g(y) I\{|y| > n^\zeta/2\} dy\Big| \le C\, n^{-\zeta(\gamma-1)}$$

due to $\sup_y |y|^\gamma |g(y)| < \infty$ and $\gamma > 1$.

By the definition of $\bar{g}_j(y)$ and (6.33) again, (6.31) follows from that, uniformly for all $1 \le j, k \le n$,

$$\begin{aligned}
&\int_{-\infty}^{\infty} |\bar{g}_{\varepsilon n}(x_{j,n} - x_{k,n})\bar{g}_j(y) - g_\varepsilon(y/c_n)\bar{g}_j(y)| dy \\
&\le \Big(\int_{-\infty}^{\infty} g(y)dy\Big) \Big(\sup_y \big|\bar{g}_{\varepsilon n}(x_{j,n} - x_{k,n}) - g_\varepsilon(y/c_n)\big|\, I\{|c_n(x_{j,n} - x_{k,n}) - y| \le n^\zeta/2\}\Big) \\
&\le C \sup_y \Big[\varepsilon^{-2}(n^\zeta/c_n + |x_{j,n} - x_{k,n} - y/c_n|)\, I\big\{\big|x_{j,n} - x_{k,n} - y/c_n\big| \le n^\zeta/(2c_n)\big\}\Big] \\
&\le C\varepsilon^{-2}(n^\zeta/c_n) = o_{a.s.}(\log^{-l} n).
\end{aligned}$$

As for (6.32), the result follows from that, by using Lemma 6.6,

$$\begin{aligned}
\Delta_{3n}(k) &= \Big| \int_{-\infty}^{\infty} \frac{1}{n}\sum_{j=1}^n \big\{\bar{g}[c_n(x_{j,n} - x_{k,n}) - y] - \bar{g}[c_n(x_{j,n} - x_{k,n})]\big\}\, g_\epsilon(y/c_n) dy\Big| \\
&\le \sup_t \sup_{|s-t|\le c_n\varepsilon} \Big|\frac{c_n}{n}\sum_{j=1}^n \big\{\bar{g}(c_n x_{j,n} + s) - \bar{g}(c_n x_{j,n} + t)\big\}\Big| \\
&\qquad\qquad \times \Big(\frac{1}{c_n}\int_{-\infty}^{\infty} g_\epsilon(y/c_n)dy\Big) \\
&= o_{a.s.}(\log^{-l} n). \qquad (6.34)
\end{aligned}$$

uniformly for $1 \le k \le n$.

The proof of Theorem 3.1 is complete. □

6.4 Proof of Theorem 3.2.

To prove (3.4), we make use of Theorem 3.1. First note that $K(x)$ satisfies Assumption 3.1 as it has a compact support. Let $x_{k,n} = \frac{x_k}{\sqrt{n}\phi}$, $1 \le k \le n$, where x_k satisfies Assumption 2.1 with $\sum_{i=0}^\infty i|\phi_i| < \infty$. As shown in Chan and Wang (2012), $x_{k,n}$ satisfies Assumption 3.3. $x_{k,n}$ also satisfies Assumption 3.2. Explicitly we will show later that $\{\nu_j, j \in Z\}$ can be redefined on a richer probability space which also contains a standard Brownian motion $W_1(t)$ and a sequence of stochastic processes $G_{1n}(t)$ such that $\{G_{1n}(t), 0 \le t \le 1\} =_D \{G_1(t), 0 \le t \le 1\}$ for each $n \ge 1$ and

$$\sup_{0\le t\le 1} |x_{[nt],n} - G_{1n}(t)| = o(n^{-\delta_0}), \quad a.s., \qquad (6.35)$$

for some $\delta_0 > 0$, where $G_1(t) = W_1(t) + \kappa \int_0^t e^{\kappa(t-s)} W_1(s)ds$. We remark that $G_1(t)$ is a continuous local martingale having a local time.

Due to these fact, it follows from Theorem 3.1 that

$$\sup_{0\le t\le 1} \Big| \frac{1}{\sqrt{n}h} \sum_{k=1}^n K\big[(x_k - x_{[nt]})/\sqrt{n}\phi\big] - L_{nt}\Big| = o_P(\log^{-l} n), \tag{6.36}$$

for any $l > 0$, $h \to 0$ and $n^{1-\epsilon_0}h^2 \to \infty$ where $\epsilon_0 > 0$ can be taken arbitrary small and

$$L_{nt} = \lim_{\epsilon\to 0} \frac{1}{2\epsilon} \int_0^1 I(|G_{1n}(s) - G_{1n}(t)| \le \epsilon)ds.$$

Note that, for each $n \ge 1$, $\{L_{nt}, 0 \le t \le 1\} =_D \{L_t, 0 \le t \le 1\}$ due to $\{G_{1n}(t), 0 \le t \le 1\} =_D \{G_1(t), 0 \le t \le 1\}$, where $L_t = \lim_{\epsilon\to 0} \frac{1}{2\epsilon} \int_0^1 I(|G_1(s) - G_1(t)| \le \epsilon)ds$. The result (3.4) now follow from (6.36) and the well-know fact that $P(\inf_{0\le t\le 1} L_t = 0) = 0$, due to the continuity of the process $G_1(s)$.

To end the proof of (3.4), it remains to show (6.35). In fact, the classical strong approximation theorem implies that, on a richer probability space,

$$\sup_{0\le t\le 1} \Big| \sum_{j=1}^{[nt]} \nu_j - W_1(nt)\Big| = o\big[n^{1/(2+\delta)}\big], \quad a.s. \tag{6.37}$$

See, e.g., Csörgö and Révész (1981). Taking this result into consideration, the same technique as in the proof of Phillips (1987) [see also Chan and Wei (1987)] yields

$$\sup_{0\le t\le 1} \Big| \sum_{j=1}^{[nt]} \lambda^{[nt]-j} \nu_j - G_{1n}^*(t)\Big| = o\big[n^{1/(2+\delta)}\big], \quad a.s., \tag{6.38}$$

where $G_{1n}^*(t) = W_1(nt) + \kappa \int_0^t e^{\kappa(t-s)} W_1(ns)ds$. Let $G_{1n}(t) = G_{1n}^*(t)/\sqrt{n}$. It is readily seen that $\{G_{1n}(t), 0 \le t \le 1\} =_D \{G_1(t), 0 \le t \le 1\}$ due to $\{W_1(nt)/\sqrt{n}, 0 \le t \le 1\} =_D \{W_1(t), 0 \le t \le 1\}$. Now, by virtue of (6.7)-(6.8), the result (6.35) follows from that

$$\begin{aligned}
\sup_{0\le t\le 1} |x_{[nt],n} - G_{1n}(t)| &\le \sup_{0\le t\le 1} \Big|\frac{a_{[nt]}\, x'_{[nt]}}{\sqrt{n}\phi} - G_{1n}(t)\Big| + \frac{\sup_{0\le t\le 1} |x''_{[nt]} + x'''_{[nt]}|}{\sqrt{n}\phi} \\
&\le \frac{C}{\sqrt{n}} \sup_{0\le t\le 1} \Big|\Big(\sum_{i=0}^{[nt]} \phi_i \lambda^{-i} - \phi\Big) \sum_{j=1}^{[nt]} \lambda^{[nt]-j} \nu_j\Big| \\
&\quad \frac{1}{\sqrt{n}} \sup_{0\le t\le 1} \Big| \sum_{j=1}^{[nt]} \lambda^{[nt]-j} \nu_j - G_{1n}^*(t)\Big| + O_{a.s.}(n^{-\delta/2(2+\delta)}) \\
&= o(n^{-\delta_0}), \quad a.s.,
\end{aligned}$$

for some $\delta_0 > 0$, where we have used the fact: due to $\max_{1\le i\le n}|\lambda^{-i}-1| \le Cn^{-1/2}$ and $\max_{1\le k\le n}|\sum_{j=1}^k \lambda^{k-j}\nu_j| = o(\sqrt{n\log n}), a.s.$, it follows from $\sum_{i=0}^\infty i|\phi_i| < \infty$ that

$$\begin{aligned}
&\sup_{0\le t\le 1} |\Big(\sum_{i=0}^{[nt]} \phi_i\lambda^{-i} - \phi\Big) \sum_{j=1}^{[nt]} \lambda^{[nt]-j}\nu_j| \\
\le\ & C \max_{1\le k\le \sqrt{n}} |\sum_{j=1}^{k} \lambda^{k-j}\nu_j| + O(\sqrt{n\log n}) \max_{\sqrt{n}\le k\le n} |\sum_{i=0}^{k} \phi_i\lambda^{-i} - \phi| \\
\le\ & O_P(n^{1/4}\sqrt{\log n}) + O(\sqrt{n\log n}) \Big(\big|\sum_{i=0}^{\sqrt{n}} \phi_i(\lambda^{-i}-1)\big| + \sum_{i=\sqrt{n}}^{\infty} |\phi_i|\Big) \\
=\ & O(n^{1/4}\sqrt{\log n}), \quad a.s.
\end{aligned}$$

We finally prove (3.5). Simple calculations show that

$$V_n \ \le\ \frac{2}{n}\sum_{t=1}^n (J_{1t}^2 + J_{2t}^2), \tag{6.39}$$

where

$$J_{1t} = \frac{\sum_{s=2}^n \big[m(x_{s-1}) - m(x_t)\big]\, K[(x_{s-1}-x_t)/h]}{\sum_{s=2}^n K[(x_{s-1}-x_t)/h]}, \quad J_{2t} = \frac{\sum_{s=2}^n u_{s-1}\, K[(x_{s-1}-x_t)/h]}{\sum_{s=2}^n K[(x_{s-1}-x_t)/h]}.$$

Assumption 2.4(a) implies that, when $|x_{s-1}-x_t| \le M\,h$ and h is sufficiently small,

$$|m(x_{s-1}) - m(x_t)| \ \le\ C\,|x_{s-1}-x_t|^\beta(1+|x_t|^\alpha),$$

uniformly on s,t. Using this fact and K has a compact support, it follows that

$$\frac{1}{n}\sum_{t=1}^n J_{1t}^2 \ \le\ \frac{C\,h^{2\beta}}{n}\sum_{t=1}^n (1+|x_t|^{2\alpha}) = O_P(n^\alpha h^{2\beta}). \tag{6.40}$$

As for J_{2t}, by recalling $u_s = \sum_{k=1}^s \rho^{s-k}\epsilon_k$, we have

$$\frac{1}{n}\sum_{t=1}^n J_{2t}^2 \ \le\ \Big\{\inf_{t=1,\dots,n}\sum_{s=1}^n K[(x_s-x_t)/h]\Big\}^{-1} \frac{1}{n}\sum_{t=1}^n J_{2t}^{*2},$$

where

$$J_{2t}^* = \frac{\sum_{k=1}^{n-1}\epsilon_k \sum_{s=k}^{n-1}\rho^{s-k}\, K[(x_s-x_t)/h]}{\big(\sum_{s=1}^{n-1} K[(x_s-x_t)/h]\big)^{1/2}}.$$

It follows from Assumption 2.2 that, for $1\le t\le n$,

$$\begin{aligned}
E(J_{2t}^{*2} \mid x_1,...,x_n) \ &\le\ \frac{\sum_{k=1}^{n-1} E(\epsilon_k^2 \mid x_1,...,x_n)\Big\{\sum_{s=k}^{n-1}\rho^{s-k}\, K[(x_s-x_t)/h]\Big\}^2}{\sum_{s=1}^{n-1} K[(x_s-x_t)/h]} \\
&\le\ \frac{C\,\sum_{k=1}^{n-1}\sum_{s=k}^{n-1}\rho^{s-k}\, K^2[(x_s-x_t)/h]\,\sum_{s=k}^{n-1}\rho^{s-k}}{\sum_{s=1}^{n-1} K[(x_s-x_t)/h]} \\
&\le\ C_1,
\end{aligned}$$

where we have used the fact that $\sum_{k=0}^{\infty} \rho^k = 1/(1-\rho) < \infty$. Hence $\frac{1}{n}\sum_{t=1}^{n} J_{2t}^{*2} = O_P(1)$. Due to this fact and (3.4), we get

$$\frac{1}{n}\sum_{t=1}^{n} J_{2t}^2 = O_P[(nh^2)^{-1/2}]. \tag{6.41}$$

Combining (6.39)–(6.41), we prove (3.5) and hence complete the proof of Theorem 3.2. □

REFERENCES

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

Chan, N. and Wang, Q. (2012). Uniform convergence for Nadaraya-Watson estimators with non-stationary data. *Econometric Theory*, conditionally accepted.

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

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

Engle, R.F. and C.W.J. Granger (1987). Cointegration and Error Correction: Representation, Estimation, and Testing, *Econometrica*, **55**, 251-276.

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

Johansen, S. (1988) . Statistical Analysis of Cointegrating Vectors, *Journal of Economic Dynamics and Control*, **12**, 231-54.

Linton, O.B., and E, Mammen (2008). Nonparametric Transformation to White Noise *Journal of Econometrics* 141, 241-264.

Liu, W., Chan, N. and Wang, Q. (2013). Uniform approximation to local time with applications in non-linear co-integrating regression. *Preprint.*

Karlsen, H. A.; Tjøstheim, D. (2001). Nonparametric estimation in null recurrent time series. *Ann. Statist.* **29**, 372–416.

Karlsen, H. A., Myklebust, T. and Tjøstheim, D. (2007). Nonparametric estimation in a nonlinear cointegration model, *Ann. Statist.*, **35**, 252-299.

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. (1991). Optimal Inference in Cointegrated Systems. *Econometrica*, **59(2)**, 283-306.

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

Phillips, P. C. B., and Durlauf, S. N. (1986): Multiple Time Series Regression With Integrated Processes, *Review of Economic Studies*, **53**, 473–495.

Phillips, P. C. B., and Hansen, B. E.(1990): Statistical Inference in Instrumental Variables Regression With I(1) Processes, *Review of Economic Studies*, **57**, 99–125.

Revuz, D. and Yor, M. (1994) *Continuous Martingales and Brownian Motion.* Fundamental Principles of Mathematical Sciences 293. Springer-Verlag

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

Stock, J. H. (1987): Asymptotic Properties of Least Squares Estimators of Cointegration Vectors, *Econometrica*, **55**, 1035–1056.

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 (2011). Asymptotic Theory for Zero Energy Functionals with Nonparametric Regression Applications. *Econometric Theory*, **27**, 235-259.

Wang, Q. and Phillips, P. C. B., (2012). A specification test for nonlinear nonstationary models. *The Annals of Statistics*, **40**, 727-758.

Xiao, Z., O. Linton, R. J. Carroll, and E. Mammen (2003). More Efficient Local Polynomial Estimation in Nonparametric Regression with Autocorrelated Errors *Journal of the American Statistical Association* 98, 980-992.

Table 1. $n = 500$ and $m(x) = x$

	$\widehat{m}_2$						$\widehat{m}_1$					
ρ/bc	10/18	1/2	1/3	1/5	1/6	1/10	10/18	1/2	1/3	1/5	1/6	1/10
−1.00	2.0143	1.7216	0.5675	0.1597	0.1108	0.0563	2.7782	2.7354	1.5466	0.7070	0.5572	0.3174
−0.90	0.2074	0.1729	0.0797	0.0390	0.0333	0.0263	0.2238	0.1950	0.1047	0.0547	0.0464	0.0347
−0.80	0.1813	0.1493	0.0701	0.0364	0.0315	0.0256	0.1882	0.1587	0.0811	0.0434	0.0375	0.0296
−0.70	0.1722	0.1411	0.0665	0.0354	0.0308	0.0254	0.1760	0.1462	0.0726	0.0393	0.0342	0.0277
−0.60	0.1677	0.1370	0.0647	0.0348	0.0304	0.0253	0.1699	0.1400	0.0683	0.0372	0.0325	0.0268
−0.50	0.1651	0.1346	0.0635	0.0345	0.0302	0.0252	0.1664	0.1364	0.0658	0.0360	0.0315	0.0262
−0.40	0.1635	0.1330	0.0628	0.0343	0.0301	0.0252	0.1642	0.1341	0.0641	0.0352	0.0309	0.0259
−0.30	0.1624	0.1321	0.0624	0.0342	0.0300	0.0253	0.1628	0.1326	0.0631	0.0347	0.0305	0.0257
−0.20	0.1618	0.1315	0.0621	0.0341	0.0300	0.0254	0.1619	0.1317	0.0625	0.0344	0.0302	0.0256
−0.10	0.1614	0.1312	0.0620	0.0342	0.0301	0.0255	0.1615	0.1312	0.0621	0.0342	0.0302	0.0256
0.00	0.1613	0.1311	0.0621	0.0343	0.0302	0.0257	0.1613	0.1311	0.0621	0.0343	0.0302	0.0257
0.10	0.1615	0.1313	0.0622	0.0344	0.0304	0.0260	0.1615	0.1313	0.0623	0.0345	0.0304	0.0260
0.20	0.1619	0.1317	0.0626	0.0347	0.0307	0.0263	0.1620	0.1319	0.0628	0.0349	0.0308	0.0263
0.30	0.1627	0.1325	0.0631	0.0351	0.0311	0.0268	0.1630	0.1329	0.0637	0.0355	0.0314	0.0269
0.40	0.1639	0.1337	0.0640	0.0358	0.0317	0.0275	0.1645	0.1345	0.0650	0.0365	0.0323	0.0277
0.50	0.1657	0.1355	0.0652	0.0367	0.0326	0.0284	0.1668	0.1370	0.0671	0.0380	0.0337	0.0289
0.60	0.1686	0.1383	0.0672	0.0382	0.0341	0.0299	0.1705	0.1408	0.0704	0.0404	0.0359	0.0308
0.70	0.1736	0.1432	0.0707	0.0409	0.0366	0.0324	0.1768	0.1475	0.0759	0.0445	0.0397	0.0342
0.80	0.1836	0.1530	0.0778	0.0466	0.0422	0.0379	0.1894	0.1608	0.0872	0.0530	0.0477	0.0413
0.90	0.2130	0.1823	0.1010	0.0661	0.0612	0.0564	0.2261	0.1998	0.1209	0.0796	0.0730	0.0644
0.95	0.2717	0.2423	0.1538	0.1126	0.1068	0.1011	0.2969	0.2757	0.1909	0.1378	0.1288	0.1168
1.00	3.1401	3.3402	3.3938	3.2839	3.2661	3.2429	3.3769	3.6514	3.7446	3.5232	3.4795	3.4196

Table 2. $n = 1000$ and $m(x) = x$

	$\widehat{m}_2$						$\widehat{m}_1$					
ρ/bc	10/18	1/2	1/3	1/5	1/6	1/10	10/18	1/2	1/3	1/5	1/6	1/10
−1.00	3.9147	3.3783	0.9869	0.1960	0.1248	0.0525	5.3312	5.3166	3.0086	1.1622	0.8541	0.4341
−0.90	0.2094	0.1705	0.0693	0.0292	0.0239	0.0175	0.2256	0.1936	0.0956	0.0441	0.0359	0.0247
−0.80	0.1818	0.1457	0.0598	0.0269	0.0223	0.0169	0.1887	0.1555	0.0711	0.0334	0.0276	0.0202
−0.70	0.1724	0.1372	0.0564	0.0260	0.0217	0.0168	0.1762	0.1425	0.0626	0.0296	0.0246	0.0187
−0.60	0.1677	0.1329	0.0546	0.0255	0.0214	0.0167	0.1699	0.1361	0.0583	0.0277	0.0232	0.0179
−0.50	0.1649	0.1305	0.0536	0.0253	0.0212	0.0167	0.1663	0.1324	0.0558	0.0265	0.0223	0.0174
−0.40	0.1632	0.1289	0.0529	0.0251	0.0211	0.0167	0.1640	0.1300	0.0542	0.0258	0.0217	0.0172
−0.30	0.1621	0.1279	0.0525	0.0250	0.0211	0.0168	0.1625	0.1285	0.0532	0.0254	0.0214	0.0170
−0.20	0.1614	0.1273	0.0523	0.0250	0.0211	0.0169	0.1616	0.1275	0.0526	0.0252	0.0212	0.0170
−0.10	0.1611	0.1269	0.0522	0.0250	0.0212	0.0170	0.1611	0.1270	0.0523	0.0251	0.0212	0.0170
0.00	0.1610	0.1269	0.0522	0.0251	0.0213	0.0172	0.1610	0.1269	0.0522	0.0251	0.0213	0.0171
0.10	0.1611	0.1270	0.0524	0.0253	0.0214	0.0174	0.1612	0.1271	0.0524	0.0253	0.0215	0.0174
0.20	0.1616	0.1275	0.0527	0.0256	0.0217	0.0177	0.1617	0.1276	0.0529	0.0257	0.0218	0.0177
0.30	0.1623	0.1282	0.0532	0.0259	0.0221	0.0181	0.1626	0.1286	0.0538	0.0263	0.0224	0.0182
0.40	0.1635	0.1294	0.0540	0.0265	0.0226	0.0186	0.1642	0.1302	0.0551	0.0272	0.0232	0.0189
0.50	0.1654	0.1311	0.0552	0.0273	0.0234	0.0194	0.1665	0.1327	0.0571	0.0286	0.0245	0.0199
0.60	0.1683	0.1339	0.0570	0.0287	0.0247	0.0207	0.1703	0.1367	0.0603	0.0309	0.0266	0.0216
0.70	0.1735	0.1388	0.0602	0.0310	0.0270	0.0229	0.1768	0.1435	0.0658	0.0347	0.0301	0.0246
0.80	0.1838	0.1486	0.0668	0.0363	0.0320	0.0278	0.1899	0.1572	0.0769	0.0427	0.0375	0.0310
0.90	0.2147	0.1784	0.0887	0.0543	0.0496	0.0449	0.2287	0.1979	0.1107	0.0682	0.0615	0.0525
0.95	0.2762	0.2392	0.1386	0.0978	0.0922	0.0867	0.3041	0.2780	0.1814	0.1248	0.1153	0.1024
1.00	6.1783	6.6740	6.9673	6.8136	6.7811	6.7500	6.6138	7.2842	7.5941	7.2184	7.1329	7.0197

Table 3. $E\widehat{\rho}, \quad m(x) = \sin(x)$

	$n = 1000$						$n = 500$					
ρ/bc	10/18	1/2	1/3	1/5	1/6	1/10	10/18	1/2	1/3	1/5	1/6	1/10
−1.00	-0.4587	-0.5555	-0.8031	-0.9151	-0.9330	-0.9600	-0.4627	-0.5496	-0.7803	-0.8962	-0.9158	-0.9464
−0.90	-0.4148	-0.5021	-0.7239	-0.8239	-0.8393	-0.8601	-0.4215	-0.5002	-0.7071	-0.8083	-0.8247	-0.8475
−0.80	-0.3704	-0.4478	-0.6451	-0.7330	-0.7462	-0.7616	-0.3773	-0.4474	-0.6310	-0.7194	-0.7331	-0.7492
−0.70	-0.3259	-0.3934	-0.5661	-0.6424	-0.6534	-0.6644	-0.3326	-0.3940	-0.5547	-0.6309	-0.6421	-0.6529
−0.60	-0.2814	-0.3391	-0.4871	-0.5521	-0.5612	-0.5686	-0.2877	-0.3403	-0.4783	-0.5429	-0.5519	-0.5584
−0.50	-0.2368	-0.2849	-0.4082	-0.4622	-0.4695	-0.4740	-0.2427	-0.2867	-0.4019	-0.4554	-0.4624	-0.4657
−0.40	-0.1922	-0.2307	-0.3295	-0.3726	-0.3783	-0.3804	-0.1977	-0.2330	-0.3257	-0.3683	-0.3736	-0.3743
−0.30	-0.1476	-0.1765	-0.2508	-0.2833	-0.2874	-0.2876	-0.1527	-0.1792	-0.2496	-0.2817	-0.2854	-0.2842
−0.20	-0.1030	-0.1223	-0.1721	-0.1942	-0.1968	-0.1955	-0.1078	-0.1255	-0.1735	-0.1953	-0.1975	-0.1949
−0.10	-0.0583	-0.0681	-0.0935	-0.1052	-0.1064	-0.1038	-0.0628	-0.0718	-0.0974	-0.1091	-0.1099	-0.1062
0.00	-0.0137	-0.0138	-0.0149	-0.0163	-0.0161	-0.0124	-0.0179	-0.0181	-0.0213	-0.0230	-0.0225	-0.0179
0.10	0.0310	0.0405	0.0637	0.0726	0.0741	0.0789	0.0270	0.0356	0.0549	0.0632	0.0649	0.0702
0.20	0.0756	0.0948	0.1424	0.1616	0.1645	0.1702	0.0720	0.0893	0.1311	0.1494	0.1524	0.1585
0.30	0.1204	0.1492	0.2212	0.2507	0.2550	0.2618	0.1169	0.1432	0.2075	0.2358	0.2401	0.2471
0.40	0.1652	0.2037	0.3000	0.3400	0.3457	0.3538	0.1619	0.1971	0.2841	0.3225	0.3282	0.3363
0.50	0.2100	0.2583	0.3791	0.4296	0.4368	0.4464	0.2070	0.2511	0.3608	0.4095	0.4167	0.4263
0.60	0.2550	0.3131	0.4583	0.5196	0.5284	0.5398	0.2521	0.3052	0.4378	0.4971	0.5059	0.5173
0.70	0.3003	0.3683	0.5380	0.6101	0.6206	0.6343	0.2974	0.3596	0.5152	0.5853	0.5959	0.6098
0.80	0.3461	0.4241	0.6183	0.7015	0.7137	0.7303	0.3432	0.4144	0.5932	0.6745	0.6871	0.7043
0.90	0.3930	0.4811	0.7003	0.7946	0.8088	0.8287	0.3902	0.4702	0.6730	0.7659	0.7807	0.8017
0.95	0.4174	0.5109	0.7433	0.8430	0.8581	0.8799	0.4153	0.4996	0.7147	0.8135	0.8293	0.8523
1.00	0.4466	0.5462	0.7903	0.8935	0.9093	0.9324	0.4399	0.5296	0.7567	0.8614	0.8781	0.9033

Table 4. stdc($\widehat{\rho}$), $m(x) = \sin(x)$

	$n = 1000$						$n = 500$					
ρ/bc	10/18	1/2	1/3	1/5	1/6	1/10	10/18	1/2	1/3	1/5	1/6	1/10
−1.00	0.0805	0.0816	0.0557	0.0283	0.0229	0.0145	0.0829	0.0835	0.0611	0.0344	0.0288	0.0197
−0.90	0.0706	0.0704	0.0466	0.0265	0.0230	0.0187	0.0756	0.0733	0.0512	0.0333	0.0302	0.0261
−0.80	0.0635	0.0634	0.0429	0.0269	0.0245	0.0220	0.0704	0.0672	0.0481	0.0346	0.0327	0.0307
−0.70	0.0568	0.0568	0.0399	0.0276	0.0261	0.0247	0.0650	0.0617	0.0463	0.0364	0.0353	0.0346
−0.60	0.0505	0.0506	0.0374	0.0284	0.0274	0.0267	0.0599	0.0568	0.0453	0.0382	0.0377	0.0378
−0.50	0.0450	0.0450	0.0353	0.0291	0.0285	0.0283	0.0554	0.0527	0.0447	0.0400	0.0398	0.0402
−0.40	0.0403	0.0402	0.0336	0.0298	0.0295	0.0295	0.0518	0.0495	0.0444	0.0417	0.0417	0.0422
−0.30	0.0366	0.0363	0.0325	0.0305	0.0304	0.0305	0.0491	0.0473	0.0445	0.0432	0.0433	0.0437
−0.20	0.0342	0.0336	0.0320	0.0311	0.0311	0.0312	0.0476	0.0461	0.0450	0.0446	0.0446	0.0450
−0.10	0.0331	0.0323	0.0321	0.0318	0.0318	0.0318	0.0470	0.0459	0.0457	0.0457	0.0457	0.0459
0.00	0.0334	0.0326	0.0327	0.0325	0.0325	0.0323	0.0474	0.0466	0.0467	0.0467	0.0467	0.0467
0.10	0.0351	0.0345	0.0339	0.0333	0.0331	0.0328	0.0488	0.0481	0.0479	0.0475	0.0474	0.0472
0.20	0.0381	0.0376	0.0356	0.0340	0.0337	0.0332	0.0510	0.0504	0.0493	0.0481	0.0479	0.0476
0.30	0.0419	0.0417	0.0376	0.0346	0.0341	0.0335	0.0540	0.0534	0.0508	0.0486	0.0482	0.0477
0.40	0.0466	0.0466	0.0398	0.0351	0.0344	0.0336	0.0577	0.0569	0.0525	0.0488	0.0482	0.0476
0.50	0.0517	0.0519	0.0423	0.0354	0.0344	0.0334	0.0619	0.0609	0.0543	0.0488	0.0480	0.0472
0.60	0.0573	0.0575	0.0448	0.0354	0.0341	0.0328	0.0666	0.0654	0.0561	0.0485	0.0474	0.0463
0.70	0.0633	0.0635	0.0472	0.0350	0.0333	0.0316	0.0718	0.0701	0.0580	0.0479	0.0464	0.0449
0.80	0.0697	0.0697	0.0498	0.0343	0.0322	0.0298	0.0777	0.0753	0.0599	0.0469	0.0449	0.0426
0.90	0.0767	0.0765	0.0526	0.0337	0.0309	0.0274	0.0844	0.0820	0.0618	0.0454	0.0427	0.0395
0.95	0.0803	0.0800	0.0544	0.0339	0.0307	0.0264	0.0877	0.0863	0.0628	0.0448	0.0417	0.0377
1.00	0.0869	0.0860	0.0572	0.0347	0.0310	0.0253	0.0942	0.0929	0.0668	0.0461	0.0427	0.0377

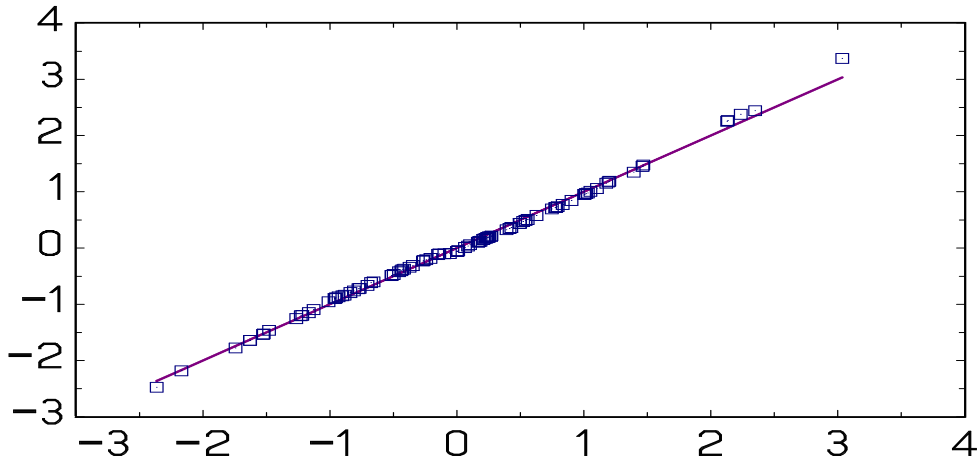

Figure 1. Shows the QQ plot of the standardized estimator $\widehat{m}_2$