Econometric Theory, **32**, 2016, 1–29.
doi:10.1017/S026646661400070X

NONPARAMETRIC TRANSFORMATION REGRESSION WITH NONSTATIONARY DATA

OLIVER LINTON
University of Cambridge

QIYING WANG
University of Sydney

We examine a kernel regression estimator for time series that takes account of the error correlation structure as proposed by Xiao et al. (2003, *Journal of the American Statistical Association* 98, 980–992). We show that this method continues to improve estimation in the case where the regressor is a unit root or a near unit root process.

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; Johansen, 1988; Phillips, 1991). Wang and Phillips (2009a, 2009b, 2011) recently considered the nonparametric cointegrating regression. They analyze the behavior of the standard kernel estimator of the cointegrating relation/nonparametric regression when the covariate is nonstationary. They showed that under self (random) normalization, the estimator is asymptotically normal. See also Phillips and Park (1998), Karlsen and Tjøstheim (2001), Karlsen, Myklebust, and Tjøstheim (2007), Schienle (2008), and Cai, Li, and Park (2009).

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 autocovariance 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 affect the asymptotic variance of the usual nonparametric estimators, the error structure can be used to construct a more

The authors thank Professor Phillips and three referees for their very helpful comments on the original version. Linton acknowledges the European Research Council for financial support and Wang acknowledges research support from the Australian Research Council. Address correspondence to Oliver Linton, Faculty of Economics, University of Cambridge, Austin Robinson Building, Sidgwick Avenue, Cambridge CB3 9DD, United Kingdom; e-mail: obl20@cam.ac.uk.

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 (2008) proposed a type of iterated version of this procedure and showed that it attains 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 et al. (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 nonlinear cointegrating regression model

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

where $u_t = \rho u_{t-1} + \epsilon_t$ with $|\rho| < 1$ and x_t is a nonstationary 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 $K(\cdot)$ is a nonnegative real function and the bandwidth parameter $h \equiv h_n \to 0$ as $n \to \infty$. Applying the Cochrane–Orcutt transformation to (2.1), we obtain

$$y_t - \rho y_{t-1} + \rho m(x_{t-1}) = m(x_t) + \epsilon_t. \tag{2.2}$$

It is expected that a two-step estimator of $m(x)$ by using (2.2) may achieve efficiency improvements over the usual estimator $\hat{m}(x)$ that uses (2.1). The strategy to provide the two-step estimator is as follows:

Step 1: Construct an estimator of ρ by

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

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

Step 2: Construct an estimator of $m(x)$, say $\hat{m}_1(x)$, by using (2.2) and the kernel method, but replace the unknown quantity $m(x)$ in (2.2) by $\hat{m}(x)$.

We now have a two-step estimator $\hat{m}_1(x)$ of $m(x)$, defined as follows:

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

https://doi.org/10.1017/S026646661400070X Published online by Cambridge University Press

To establish our claim that $\hat{m}_1(x)$ 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 $\tau \leq 0$ being a constant and $\{\xi_j, j \geq 1\}$ is a linear process defined by

$$\xi_j = \sum_{k=0}^{\infty} \phi_k v_{j-k}, \tag{2.3}$$

where $\phi_0 \neq 0$, $\phi \equiv \sum_{k=0}^{\infty} \phi_k \neq 0$, and $\sum_{k=0}^{\infty} |\phi_k| < \infty$, and $\{v_j, -\infty < j < \infty\}$ is a sequence of i.i.d. (independent and identically distributed) random variables with $Ev_0 = 0$, $Ev_0^2 = 1$, $E|v_0|^{2+\delta} < \infty$ for some $\delta > 0$; the characteristic function $\varphi(t)$ of v_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 $\mathcal{F}_{n,t} = \sigma(\epsilon_0, \epsilon_1, \ldots, \epsilon_t, x_1, \ldots, x_n)$ and $\{\epsilon_t, \mathcal{F}_{n,t}\}_{t=1}^{n}$ forms a martingale difference sequence satisfying, as $n \to \infty$ first, and then $m \to \infty$,

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

where σ^2 is a given constant, and $\sup_{1 \leq t \leq n, n \geq 1} E(|\epsilon_t|^q | \mathcal{F}_{n,t-1}) < \infty$ a.s. (almost surely) 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 $x, y \in \mathbb{R}$, $|K(x) - K(y)| \leq C\,|x - y|$, where C is a positive constant; (c) For $p \geq 2$,

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

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

$$|m(x+y) - m(x)| \leq C\,(1 + |x|^{\alpha})\,|y|^{\beta},$$

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

We have the following main results.

THEOREM 2.1. *Suppose that Assumptions 2.1, 2.2, 2.3(a), and 2.4(a) hold. Then, we have*

$$\Big(\sum_{t=1}^{n} K[(x_t - x)/h]\Big)^{1/2} \big[\hat{m}(x) - m(x)\big] \to_D N(0, \sigma_1^2), \tag{2.4}$$

https://doi.org/10.1017/S026646661400070X Published online by Cambridge University Press

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) hold, then

$$\left(\sum_{t=1}^{n} K[(x_t - x)/h]\right)^{1/2} \left[\hat{m}(x) - m(x) - \frac{h^p m^{(p)}(x)}{p!} \int_{-\infty}^{\infty} y^p K(y)dy\right] \to_D N(0, \sigma_1^2), \tag{2.5}$$

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

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

$$\widehat{\rho} - \rho = O_P\left\{n^{\alpha/2}h^{\beta} + (nh^2)^{-1/4}\right\}, \tag{2.6}$$

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

$$\left(\sum_{t=1}^{n} K[(x_t - x)/h]\right)^{1/2} \left[\hat{m}_1(x) - m(x)\right] \to_D N(0, \sigma_2^2), \tag{2.7}$$

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) hold, then

$$\left(\sum_{t=1}^{n} K[(x_t - x)/h]\right)^{1/2} \left[\hat{m}_1(x) - m(x) - \frac{h^p m^{(p)}(x)}{p!} \int_{-\infty}^{\infty} y^p K(y)dy\right] \to_D N(0, \sigma_2^2), \tag{2.8}$$

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 certain 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 noted in previous works, the conditions on $m(x)$ to establish our results are quite weak, in particular, a wide range of regression functions $m(x)$ are included in Assumption 2.4(a), like $m(x) = 1/(1+\theta|x|^{\beta})$, $m(x) = (a+be^x)/(1+e^x)$, and $m(x) = \theta_1 + \theta_2 x + \cdots + \theta_k x^{k-1}$.

If we are only interested in the asymptotics of $\hat{m}(x)$, recent research has shown that the information set $\mathcal{F}_{n,t} = \sigma(\epsilon_0, \epsilon_1, \ldots, \epsilon_t, x_1, \ldots, x_n)$ given in Assumption 2.2 can be reduced to $\mathcal{F}_t = \sigma(\epsilon_0, \epsilon_1, \ldots, \epsilon_t, x_1, \ldots, x_t)$, a more natural condition in the framework of cointegration. We refer to Wang and Phillips (2009a) and Wang (2014) for current developments. However, due to the involvement of $x_1, \ldots, x_n$ in the definition of $\hat{m}_1(x)$, it is difficult to weaken this restriction for the asymptotics of $\hat{m}_1(x)$, which is provided in Theorem 2.2. Theorem 2.1 of this paper is only stated to make a comparison with Theorem 2.2, which does not provide a general result in this direction.

https://doi.org/10.1017/S026646661400070X Published online by Cambridge University Press

Remark 2. As $|\rho| < 1$, Theorem 2.2 confirms the claim that $\hat{m}_1(x)$ 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) implies that $0 \le \alpha < \beta$, which in turn requires that the rate at which $m(x)$ diverges to ∞ in the tail is not faster than $|x|^{1+\beta}$. In comparison with Theorem 2.1, this is a little bit restrictive but it is reasonable, due to the fact that the consistency result (2.6) heavily depends on the following uniform convergence

$$\frac{1}{n}\sum_{t=1}^{n}\left[\hat{m}(x_t) - m(x_t)\right]^2 = O_P(r_n), \tag{2.9}$$

where $0 < r_n \to 0$ is a sequence of constants. As $x_t \sim \sqrt{t}$ under our model, it is natural for the restriction on the tail of $m(x)$ to enable (2.9). The result (2.9) is a consequence of Theorem 3.1 in the next section, which provides a strong approximation result on the convergence to a local time process.

One referee argued that the condition $n^{\alpha}h^{2\beta} \to 0$ can be weakened if the local linear estimator $\tilde{m}(x)$ is used in the definition of $\hat{m}_1(x)$ instead of $\hat{m}(x)$. Since this improvement requires new limit theorems, we leave the topic for future work.

Remark 3. Consider model (2.1) with AR(k) errors, i.e., u_t is assumed to be strictly stationary satisfying

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \cdots + \rho_k u_{t-k} + \epsilon_t, \tag{2.10}$$

where $\max_{1\le j\le k}|\rho_j| < 1$. In this situation, similar to (2.2), we have

$$y_t - \sum_{j=1}^{k}\rho_j[y_{t-j} - m(x_{t-j})] = m(x_t) + \epsilon_t,$$

and, as in Step 2, we may construct a two-step estimator $\hat{m}_2(x)$ of $m(x)$ as

$$\hat{m}_2(x) = \frac{\sum_{t=1}^{n}\left\{y_t - \sum_{j=1}^{k}\hat{\rho}_j[y_{t-j} - \hat{m}(x_{t-j})]\right\}K[(x_t - x)/h]}{\sum_{t=1}^{n}K[(x_t - x)/h]},$$

where $(\hat{\rho}_1, \ldots, \hat{\rho}_k)^{\tau}$ is a Least Squares (LS) estimator of $\hat{u}_t$ on $\hat{u}_{t-1}, \hat{u}_{t-2}, \ldots, \hat{u}_{t-k}$, where $\hat{u}_t = y_t - \hat{m}(x_t)$.

If $\epsilon_j, j \in Z$, are assumed to be i.i.d. random variables with $E\epsilon_0 = 0$ and $E\epsilon_0^2 = \sigma^2$ and ϵ_j are independent of x_t, (2.4) and (2.5) hold true except that σ_1^2 is replaced by

$$\sigma_1^{*2} = (\sigma^2 + \rho^{\tau}\gamma)\int_{-\infty}^{\infty}K^2(s)dt,$$

https://doi.org/10.1017/S026646661400070X Published online by Cambridge University Press

where $\rho = (\rho_1, \ldots, \rho_k)^\tau$ and $\gamma = (\gamma_1, \ldots, \gamma_k)^\tau$ with $\gamma_j = \mathrm{cov}(u_1, u_{1+j}), 1 \le j \le k$. Furthermore, (2.7) and (2.8) hold for $\hat{m}_1(x)$ being replaced by $\hat{m}_2(x)$. Since we generally have $\rho^\tau \gamma > 0$, $\hat{m}_2(x)$ achieves efficiency improvements over $\hat{m}(x)$ under model (2.1) with $AR(k)$ errors (2.10). The proof of this claim is similar to Theorems 2.1 and 2.2, but involves some complicated calculations. The details are hence omitted.

3. STRONG APPROXIMATION TO LOCAL TIME

This section investigates the strong approximation to a local time process, which provides a technical tool in the development of the uniform convergence such as (2.9) for the kernel estimator $\hat{m}(x)$. As the conditions imposed are 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 limiting behavior of the statistic $S_n(t)$, defined by

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

where c_n is a certain sequence of positive constants and g is a real integrable function on $\mathbb{R}$. As noticed in the 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 limiting behavior plays a key role in 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 $\mathbb{R}$.

Assumption 3.2. On a rich probability space, there exists a continuous semimartingale $G(t)$ having a local time $L_G(t,s)$ [1] such that, for some $\xi > 0$,

$$|L_G(1,s) - L_G(1,t)| \le C\,|s - t|^\xi \quad a.s., \tag{3.2}$$

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

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

for some $0 < \delta_0 < 1$.

https://doi.org/10.1017/S026646661400070X Published online by Cambridge University Press

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

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

The following is our main result.

THEOREM 3.1. *Suppose that Assumptions 3.1–3.4 hold. Then, 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* $L_{nt} = \lim_{\epsilon\to 0} \frac{1}{2\epsilon} \int_0^1 I(|G_n(s) - G_n(t)| \le \epsilon) ds$.

The rate obtained in (3.3) may not be optimal. We conjecture that the optimal rate should have the form $n^{-\delta_1}$, where $\delta_1 > 0$ is related to $\delta_0 > 0$ given in Assumption 3.2. We are not able to establish the optimal rate due to the technical difficulty. However, by noting $\{L_{nt}; 0 \le t \le 1\} =_D \{L_G(1, G(t)); 0 \le t \le 1\}$ 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 (2.9) 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) and (b) hold. Then, for any $\eta > 0$, there exist $M_1 > 0$ and $n_0 > 0$ such that*

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

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

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

that is, (2.9) holds true if in addition $n^{\alpha}h^{2\beta} \to 0$.

https://doi.org/10.1017/S026646661400070X Published online by Cambridge University Press

4. EXTENSION

We next propose another estimator that potentially can improve efficiency further, following the method of Linton and Mammen (2008). Note that we can write

$$m(x) = \frac{1}{1+\rho^2} E\left[Z_t^-(\rho) - \rho Z_{t+1}^+(\rho) | x_t = x\right],$$

where: $Z_t^-(\rho) = y_t - \rho y_{t-1} + \rho m(x_{t-1})$ and $Z_t^+(\rho) = y_t - \rho y_{t-1} - m(x_t)$. Let $\widehat{m}(.), \widehat{\rho}$ be initial consistent estimators of $m(.)$ and ρ, and let $\widehat{y}_t = \widehat{Z}_t^-(\widehat{\rho}) - \widehat{\rho}\widehat{Z}_{t+1}^+(\widehat{\rho})$, where: $\widehat{Z}_t^-(\rho) = y_t - \widehat{\rho} y_{t-1} + \widehat{\rho}\widehat{m}(x_{t-1})$ and $\widehat{Z}_t^+(\rho) = y_t - \widehat{\rho} y_{t-1} - \widehat{m}(x_t)$. Then let

$$\widehat{m}_{eff}(x) = \frac{1}{1+\widehat{\rho}^2} \frac{\sum_{s=1}^{n-1} \widehat{y}_s K[(x_s - x)/h]}{\sum_{s=1}^{n-1} K[(x_s - x)/h]}.$$

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

THEOREM 4.1. *Suppose in addition to Assumptions 2.1, 2.2, 2.3(a) and (b), and 2.4(a), that* $\sum_{i=0}^{\infty} i|\phi_i| < \infty$ *holds. Then, for any* h *satisfying* $nh^{2+4\beta} \to 0$, $n^{\alpha} h^{2\beta} \to 0$ *and* $n^{1-\epsilon_0} h^2 \to \infty$ *for some* $\epsilon_0 > 0$, *we have*

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

where $\sigma_3^2 = (1+\rho^2)^{-1} \int_{-\infty}^{\infty} K^2(s) dt$.

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 that 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 $\varepsilon_t \sim N(0,1)$; the two errors are mutually independent. We used the Epanechnikov kernel $K(u) = 0.75(1-u^2)1(|u| \le 1)$ with the bandwidth $h = n^{-bc}$. We examine a range of values of $\rho_0 \in \{-1, -0.9, \ldots, 0.9, 0.95, 1\}$ (although we only show the results for nonnegative values of ρ_0) and the bandwidth constant $bc \in (10/18, 1/2, 1/3, 1/5, 1/10\}$. We consider $n = 500$,

https://doi.org/10.1017/S026646661400070X Published online by Cambridge University Press

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
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. $n = 1{,}000$ 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
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. $E\widehat{\rho},\ m(x) = \sin(x)$

	$n = 1{,}000$						$n = 500$					
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
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. $\mathrm{stdc}(\widehat{\rho}), m(x) = \sin(x)$

	$n = 1{,}000$						$n = 500$					
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
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

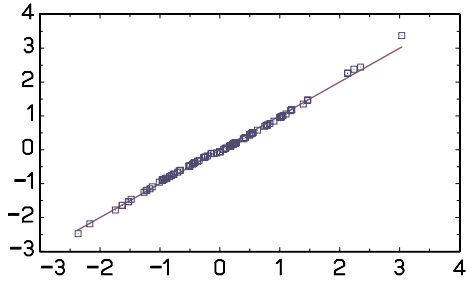

FIGURE 1. The QQ plot of the standardized estimator $\widehat{m}_2$.

1,000 and take $ns = 1{,}000$ replications throughout. We report the performance measure

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

where $K = 101$ and $x_k = \{-1, -0.98, \ldots, 1\}$. The results for the linear case are given in Tables 1 and 2. The results show that there is an improvement when going from $n = 500$ to $n = 1{,}000$ and when going from $\hat{m}$ 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.

We show in Tables 3 and 4 the performance of the estimator of ρ for $n = 500$ and $n = 1{,}000$. 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 $m(x) = \sin(x)$, $\rho = 0.95$, $n = 1{,}000$, and $bc = 1/10$.

6. CONCLUSION

We have shown that the main results of Xiao et al. (2003) regarding efficiency improvements can be extended to the case where the covariate process is nonstationary. In practice it may be important to take account of error autocorrelation when conducting inference about nonparametric cointegrating regressions. The idea of using the transformed model and the results obtained in terms of efficiency improvements are closely related to the augmented regression findings in Wang and Phillips (2009b), wherein it is shown that carrying the impact of endogeneity via the conditional mean in an augmented regression reduces the asymptotic variance of the kernel regression estimator.

NOTE

1. Here and below, 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 process $G(s)$, whenever it exists.

https://doi.org/10.1017/S026646661400070X Published online by Cambridge University Press

REFERENCES

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

Chan, N. & Q. Wang (2014) Uniform convergence for Nadaraya-Watson estimators with nonstationary data. *Econometric Theory* 30, 1110–1133.

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

Csörgö, M. & P. Révész (1981) *Strong Approximations in Probability and Statistics*. *Probability and Mathematical Statistics*. Academic Press, Inc.

Engle, R.F. & C.W.J. Granger (1987) Cointegration and error correction: Representation, estimation, and testing. *Econometrica* 55, 251–276.

Hall, P. & C.C. Heyde (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, 231–254.

Karlsen, H.A. & D. Tjøstheim (2001) Nonparametric estimation in null recurrent time series. *Annals of Statistics* 29, 372–416.

Karlsen, H.A., T. Myklebust, & D. Tjøstheim (2007) Nonparametric estimation in a nonlinear cointegration model. *Annals of Statistics* 35, 252–299.

Linton, O.B. & E. Mammen (2008) Nonparametric transformation to white noise. *Journal of Econometrics* 141, 241–264.

Liu, W., N. Chan, & Q. Wang (2014) Uniform approximation to local time with applications in non-linear co-integrating regression. Preprint, School of Mathematics and Statistics, The University of Sydney. Available on http://www.maths.usyd.edu.au/u/pubs/publist/preprints/2014/liu-18.pdf.

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

Phillips, P.C.B. (1991) Optimal inference in cointegrated systems. *Econometrica* 59(2), 283–306.

Phillips, P.C.B. & J.Y. Park (1998) Nonstationary Density Estimation and Kernel Autoregression. Cowles Foundation Discussion paper no. 1181.

Revuz, D. & M. Yor (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. (2014) Martingale limit theorem revisited and nonlinear cointegrating regression. *Econometric Theory* 30, 509–535.

Wang, Q. & P.C.B. Phillips (2009a) Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory* 25, 710–738.

Wang, Q. & P.C.B. Phillips (2009b) Structural nonparametric cointegrating regression. *Econometrica* 77, 1901–1948.

Wang, Q. & P.C.B. Phillips (2011) Asymptotic theory for zero energy functionals with nonparametric regression applications. *Econometric Theory* 27, 235–259.

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

Xiao, Z., O. Linton, R.J. Carroll, & E. Mammen (2003) More efficient local polynomial estimation in nonparametric regression with autocorrelated errors. *Journal of the American Statistical Association* 98, 980–992.

APPENDIX: Proofs

Appendix A.1 provides several preliminary lemmas. Some of them are of independent interest. The proofs of main theorems will be given in Appendixes A.2–A.4. Throughout the Appendix, we denote by $C_0, C, C_1, ...$ positive constants, which might be different at each appearance.

A.1. Preliminary Lemmas

First note that

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

where

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

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. Routine calculations show that, whenever n is sufficiently large,

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

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

$$\gamma_0 \le \inf_{n \ge k \ge m} |\sum_{j=0}^{k} \lambda^{-j} \phi_j| \le \gamma_1, \tag{A.3}$$

whenever n, m are sufficiently large. By virtue of (A.2) and (A.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,

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

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 $\int (1+|u|)|Ee^{iuv_0}|dt < \infty$. See the proof of Corollary 2.2 in Wang and Phillips (2009a) and/or (7.14) and Proposition 7.2 (p. 1934) of Wang and Phillips (2009b) with a minor modification. Hence, conditional on $\mathcal{F}_k = \sigma(v_j, -\infty < j \le k)$,

$$(x_t - x_s)/\sqrt{t-s} \text{ has a density } h^*_{s,t}(x) \equiv h_{s,t}(x - x^*_{s,t}/\sqrt{t-s}), \tag{A.4}$$

https://doi.org/10.1017/S026646661400070X Published online by Cambridge University Press

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

$$\begin{aligned}
\sup_x \left|h^*_{s,t}(x+u) - h^*_{s,t}(x)\right| &\le \sup_x |h_{s,t}(x+u) - h_{s,t}(x)| \\
&\le C \left| \int_{-\infty}^{\infty} \left(e^{-iv(x+u)} - e^{-ivx}\right) \varphi_{s,t}(v) dv \right| \\
&\le C \min\{|u|, 1\} \int_{-\infty}^{\infty} (1 + |v|)\, |\varphi_{s,t}(v)| dv \le C_1 \min\{|u|, 1\}, \qquad \textbf{(A.5)}
\end{aligned}$$

where we have used the inversion formula of a characteristic function in the calculation above.

We also have the following representation for x_t:

$$\begin{aligned}
x_t &= \sum_{j=1}^{t} \lambda^{t-j} \xi_j = \sum_{j=1}^{t} \lambda^{t-j} \left(\sum_{i=0}^{j-1} + \sum_{i=j}^{\infty} \right) \phi_i \nu_{j-i} \\
&= \sum_{i=0}^{t-1} \phi_i\, \lambda^{t-i} \sum_{j=1}^{t} \lambda^{-j} \nu_j - \sum_{i=0}^{t-1} \phi_i\, \lambda^{t-i} \sum_{j=t-i+1}^{t} \lambda^{-j} \nu_j + \sum_{j=1}^{t} \lambda^{t-j} \sum_{i=0}^{\infty} \phi_{i+j} \nu_{-i} \\
&= a_t\, x'_t - x''_t + x'''_t, \quad \text{say,} \qquad \textbf{(A.6)}
\end{aligned}$$

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

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

for some constant $C_0 > 0$. Indeed, using (A.2) and the strong law, we obtain that, for some constant $C_0 > 0$,

$$\begin{aligned}
|x''_t| &\le 2e^{|\tau|} \sum_{i=0}^{t-1} |\phi_i| \sum_{j=t-i+1}^{t} |\nu_j| \le 2e^{|\tau|} \max_{1 \le j \le t} |\nu_j| \sum_{i=0}^{t-1} i |\phi_i| \\
&\le C\, t^{1/(2+\delta)} \left(\frac{1}{t} \sum_{j=1}^{t} |\nu_j|^{2+\delta} \right)^{1/(2+\delta)} \le C_0 t^{1/(2+\delta)}, \quad a.s., \qquad \textbf{(A.7)}
\end{aligned}$$

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

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

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 (A.2) again and the Kronecker lemma that

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

This proves (A.6).

https://doi.org/10.1017/S026646661400070X Published online by Cambridge University Press

We are now ready to provide several preliminary lemmas.

LEMMA A.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*

$$E\left(|p(x_t/h)| \,\middle|\, \mathcal{F}_s\right) \le \frac{C_0 h}{\sqrt{t-s}} \int_{-\infty}^{\infty} |p(x + x_s/h)|dx$$

$$= \frac{C_0 h}{\sqrt{t-s}} \int_{-\infty}^{\infty} |p(x)|dx, \quad a.s., \tag{A.8}$$

where $\mathcal{F}_s = \sigma\{v_s, v_{s-1}, \ldots\}$.

Proof. Recall (A.1), (A.4), and the independence of v_k. The result (A.8) follows from a routine calculation and hence the details are omitted. ■

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

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

where $G(t) = W(t) + \tau \int_0^t e^{\tau(t-s)} W(s)ds$ 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 of the conditions is similar to Proposition 7.2 of Wang and Phillips (2009b). We omit the details. ■

LEMMA A.3. *Suppose that Assumptions 2.1, 2.2, and 2.3(a) hold. Then, for any $h \to 0$ and $nh^2 \to \infty$,*

$$\sum_{t=1}^{n} u_t Z_{nt} \to_D N(0, \sigma_1^2), \tag{A.9}$$

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

Proof. For notational convenience, we assume that $\sigma^2 = 1$ 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

$$\sum_{k=1}^{n} Z_{nk}^2 \to_P \int_{-\infty}^{\infty} K^2(x)dx, \tag{A.10}$$

$$\sum_{k=1}^{n} Z_{nk}^{*2} \to_P (1-\rho^2)^{-1} \int_{-\infty}^{\infty} K^2(x)dx. \tag{A.11}$$

https://doi.org/10.1017/S026646661400070X Published online by Cambridge University Press

The proof of (A.10) is simple by applying Lemma A.2. To see (A.11), note that

$$\sum_{k=1}^{n} Z_{nk}^{*2} = \Lambda_n^{-1} \sum_{k=1}^{n} \left(\sum_{t=k}^{n} \rho^{t-k} K[(x_t - x)/h] \right)^2$$
$$= \Lambda_n^{-1} \sum_{k=1}^{n} \sum_{t=k}^{n} \rho^{2(t-k)} K^2[(x_t - x)/h] + \Lambda_n^{-1} \Gamma_{1n}$$
$$= (1-\rho^2)^{-1} \sum_{k=1}^{n} Z_{nk}^2 + \Lambda_n^{-1} (\Gamma_{1n} - \Gamma_{2n})$$
$$= (1-\rho^2)^{-1} \int_{-\infty}^{\infty} K^2(x) dx + \Lambda_n^{-1} (\Gamma_{1n} - \Gamma_{2n}) + o_P(1), \qquad \textbf{(A.12)}$$

by (A.10), 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 A.2. The result (A.11) will follow if we prove

$$\Gamma_{1n} + \Gamma_{2n} = o_P[(nh^2)^{1/2}]. \qquad \textbf{(A.13)}$$

Recalling that $K(x) \le C$ and $|\rho| < 1$, it is readily seen that $\Gamma_{2n} \le C$. On the other hand, by applying Lemma A.1, for any $t > s$, we have

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

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 $\Gamma_{1n} = O_P(h^2\sqrt{n})$. Hence, (A.13) follows due to $h \to 0$ and $nh^2 \to \infty$. This also completes the proof of (A.11).

We now turn to the proof of (A.9). 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)] in Hall and Heyde (1980) with $\delta = q/2 - 1$ that

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

https://doi.org/10.1017/S026646661400070X Published online by Cambridge University Press

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

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

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

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

by Assumption 2.3 and Lemma A.2. Routine calculations, together with (A.11), show that

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

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

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

This completes the proof of Lemma A.3. ■

LEMMA A.4. *Suppose that Assumptions 2.3(a) and 2.4(a) hold. Then, for any $x \in \mathbb{R}$, we have*

$$\left|\Lambda_n(x)-m(x)\right| \le C(1+|x|^\alpha)h^\beta, \tag{A.14}$$

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) holds, we have

$$\left|\Lambda_n(x)-m(x)-\frac{h^p m^{(p)}(x)}{p!}\int_{-\infty}^{\infty}y^p K(y)dy\right| = o_P[(nh^2)^{-1/4}], \tag{A.15}$$

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

Proof. By Assumption 2.4(a) and the fact that $K(x)$ has a compact support, the result (A.14) is simple. The proof of (A.15) is the same as in the proof of Theorem 2.2 in Wang and Phillips (2011). We omit the details. ■

LEMMA A.5. *Suppose that Assumptions 3.1–3.4 hold. Then, for any $l > 0$, we have*

$$I_n := \sup_{t\in R}\sup_{s:|s-t|\le\epsilon_n}\left|\frac{c_n}{n}\sum_{k=1}^{n}f_{t,s}(x_{k,n})\right| = o(\log^{-l}n), \quad a.s., \tag{A.16}$$

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 (2014). ■

https://doi.org/10.1017/S026646661400070X Published online by Cambridge University Press

A.2. Proofs of Theorems 2.1 and 2.2

We only prove Theorem 2.2. Using Lemmas A.3 and A.4, the proof of Theorem 2.1 is standard (see, e.g., Wang and Phillips, 2011), and hence the details are omitted.

Start with (2.6). Recall that $\hat{u}_t = y_t - \hat{m}(x_t) = u_t + m(x_t) - \hat{m}(x_t)$. Simple calculations show that

$$\begin{aligned}\hat{\rho} - \rho &= \frac{\sum_{s=2}^n (\widehat{u}_s - \rho \widehat{u}_{s-1})\widehat{u}_{s-1}}{\sum_{s=2}^n \widehat{u}_{s-1}^2} \\ &= \frac{\sum_{s=2}^n \epsilon_s \widehat{u}_{s-1}}{\sum_{s=2}^n \widehat{u}_{s-1}^2} + \frac{\sum_{s=2}^n \widehat{u}_{s-1}\left[m(x_s) - \hat{m}(x_s) + \rho\{\hat{m}(x_{s-1}) - m(x_{s-1})\}\right]}{\sum_{s=2}^n \widehat{u}_{s-1}^2} \\ &=: R_{1n} + R_{2n}. \end{aligned} \tag{A.17}$$

As $V_n = \frac{1}{n}\sum_{t=1}^n \left[\hat{m}(x_t) - m(x_t)\right]^2 = O_P\left\{n^\alpha h^{2\beta} + (nh^2)^{-1/2}\right\}$ 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

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

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 (A.17), we obtain (2.6).

We next prove (2.7). We may write

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

Furthermore, we may divide I_{1n} into

$$\begin{aligned}I_{1n} &= \frac{1}{\sum_{t=1}^n K[(x_t - x)/h]} \\ &\quad \times \left\{ \sum_{t=1}^n K[(x_t - x)/h] \frac{\sum_{s=1}^n [m(x_s) - m(x_{t-1})] K[(x_s - x_{t-1})/h]}{\sum_{s=1}^n K[(x_s - x_{t-1})/h]} \right. \\ &\quad \left. + \sum_{t=1}^n K[(x_t - x)/h] \frac{\sum_{s=1}^n u_s K[(x_s - x_{t-1})/h]}{\sum_{s=1}^n K[(x_s - x_{t-1})/h]} \right\} \\ &:= I_{1n1} + I_{1n2}.\end{aligned}$$

https://doi.org/10.1017/S026646661400070X Published online by Cambridge University Press

Using Lemma A.3 and (2.6), we have

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

and (with $\rho = 0$ in Lemma A.3)

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

Using (A.14) and $\frac{1}{\sqrt{nh}} \sum_{t=1}^{n} K[(x_t - x)/h] \to_D \phi^{-1} L_G(1, 0)$, we have

$$\left(\sum_{t=1}^{n} K[(x_t - x)/h] \right)^{1/2} |I_{3n}|$$
$$\leq C\,(nh^{2+4\beta})^{1/4} \left(\frac{1}{\sqrt{nh}} \sum_{t=1}^{n} K[(x_t - x)/h] \right)^{1/2} = o_P(1).$$

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

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

Combining all the above facts, to prove (2.7), it suffices to show that

$$I_{1n2} = o_P[(nh^2)^{-1/4}]. \tag{A.18}$$

To this end, for each fixed x, write

$$\Omega_{1n} = \left\{ \omega : \sum_{s=1}^{n} K[(x_s - x)/h] \geq \sqrt{n} h \delta_n \right\},$$

$$\Omega_{2n} = \left\{ \omega : \inf_{t=1,2,\ldots,n} \sum_{s=1}^{n} K[(x_s - x_t)/h] \geq \sqrt{n} h \delta_n \right\},$$

where $\delta_n \downarrow 0$ is chosen later and ω denotes the sample points. As $P(\bar{\Omega}_{1n} \cup \bar{\Omega}_{2n}) \to 0$ by (3.4) and Lemma A.2, the result (A.18) will follow if we prove

$$I_{\Omega_{1n}}(\omega)\, I_{\Omega_{2n}}(\omega)\, I_{1n2} = o_P[(nh^2)^{-1/4}], \tag{A.19}$$

where $I_A(\omega)$ denotes the indicator function.

https://doi.org/10.1017/S026646661400070X Published online by Cambridge University Press

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

$$I_{1n2} = \frac{1}{\sum_{t=1}^{n} K[(x_t - x)/h]} \sum_{s=1}^{n} u_s J_{ns}$$
$$= \frac{1}{\sum_{t=1}^{n} K[(x_t - x)/h]} \sum_{k=1}^{n} \epsilon_k \sum_{s=k}^{n} \rho^{s-k} J_{ns},$$

where

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

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

$$E\left[I_{\Omega_{1n}}(\omega) I_{\Omega_{2n}}(\omega) I_{2n2}\right]^2 \le \frac{C}{(nh^2\delta_n)^2} \sum_{k=1}^{n} E\left(\sum_{s=k}^{n} \rho^{s-k} J_{ns}^*\right)^2$$
$$\le \frac{C}{(nh^2\delta_n)^2} \sum_{k=1}^{n} \sum_{s=k}^{n} \rho^{s-k} \sum_{s=k}^{n} \rho^{s-k} E J_{ns}^{*2}$$
$$\le \frac{nC}{(nh^2\delta_n)^2} \max_{1\le s\le n} E J_{ns}^{*2}, \tag{A.20}$$

where

$$J_{ns}^* = \sum_{t=1}^{n} K[(x_s - x_{t-1})/h] K[(x_t - x)/h].$$

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

$$E J_{ns}^{*2} = E\left(\sum_{t=1}^{n} K[(x_s - x_{t-1})/h] K[(x_t - x)/h]\right)^2$$
$$\le 4E\left\{\sum_{t=1}^{s-1} K[(x_s - x_{t-1})/h] K[(x_t - x)/h]\right\}^2$$
$$+ 4E\left\{\sum_{t=s+1}^{n} K[(x_s - x_{t-1})/h] K[(x_t - x)/h]\right\}^2$$
$$+ 4E K^2[(x_s - x_{s-1})/h] K^2[(x_s - x)/h]$$
$$:= T_{1n} + T_{2n} + T_{3n}.$$

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

https://doi.org/10.1017/S026646661400070X Published online by Cambridge University Press

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

This, together with the repeatedly similar utilization of Lemma A.1, yields that, for $t_1 < t_2 \le s-1$,

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

Similarly, for $t_1 = t_2 < s-1$,

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

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 (A.20) that

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

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 (A.18) and also completes the proof of Theorem 2.2. ■

A.3. Proof of Theorem 3.1

The idea for the proof of (3.3) is similar to that of Liu et al. (2014), but there are some essential differences in details. We restate here for the convenience of reading. Without

https://doi.org/10.1017/S026646661400070X Published online by Cambridge University Press

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 $n^{\zeta}/c_n \le \min\{n^{-\delta_0}, n^{-\epsilon_0/2}\}$, where γ and δ_0 are given in Assumptions 3.1 and 3.2, respectively. Further let $\epsilon = n^{-\alpha}$ with $0 < \alpha < \delta_0/2$ and define a triangular function

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

It suffices to show that:

$$\Phi_{1n} := \max_{1\le k\le n} \Big| \frac{c_n}{n} \sum_{j=1}^{n} \big\{ g[c_n(x_{j,n} - x_{k,n})] - \bar{g}[c_n(x_{j,n} - x_{k,n})] \big\} \Big|$$
$$= o_{a.s.}(\log^{-l} n), \tag{A.21}$$

$$\Phi_{2n} := \max_{1\le k\le n} \Big| \frac{c_n}{n} \sum_{j=1}^{n} \bar{g}[c_n(x_{j,n} - x_{k,n})] - \frac{1}{n} \sum_{j=1}^{n} g_\epsilon(x_{j,n} - x_{k,n}) \Big|$$
$$= o_{a.s.}(\log^{-l} n), \tag{A.22}$$

$$\Phi_{3n} := \sup_{0\le t\le 1} \Big| \frac{1}{n} \sum_{j=1}^{n} g_\epsilon(x_{j,n} - x_{[nt],n}) - L_{nt} \Big| = o_P(\log^{-l} n). \tag{A.23}$$

The proof of (A.21) is simple. Indeed, by recalling $\sup_x |x|^{\gamma} |g(x)| < \infty$, it follows that

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

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

We next prove (A.23). Recalling $\int_{-\infty}^{\infty} g_\epsilon(y)dy = 1$, it follows from the definition of occupation time and (3.2) in Assumption 3.2 that

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

for some $\xi > 0$, uniformly for $t \in [0, 1]$. Hence, by noting

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

https://doi.org/10.1017/S026646661400070X Published online by Cambridge University Press

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

$$P\left(\left|\int_0^1 g_\epsilon[G_n(s) - G_n(t)]ds - L_{nt}\right| \ge \log^{-l-1} n\right)$$
$$= P\left(\left|\int_0^1 g_\epsilon[G(s) - G(t)]ds - L_G(1, G(t))\right| \ge \log^{-l-1} n\right) = o(1), \quad \textbf{(A.24)}$$

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

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

uniformly for $t \in [0, 1]$, as $\alpha < \delta_0/2$. This yields (A.23).

We finally prove (A.22), let $\bar{g}_{\epsilon n}(z)$ 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 $\bar{g}_j(y) = \bar{g}(c_n(x_{j,n} - x_{k,n}) - y)$):

$$\Delta_{1n}(k) := \left|\frac{1}{n}\sum_{j=1}^n g_\epsilon(x_{j,n} - x_{k,n}) - \frac{1}{n}\sum_{j=1}^n \bar{g}_{\epsilon n}(x_{j,n} - x_{k,n}) \int_{-\infty}^{\infty} \bar{g}_j(y)dy\right|$$
$$= o_{a.s.}(\log^{-l} n), \quad \textbf{(A.25)}$$

$$\Delta_{2n}(k) := \left|\frac{1}{n}\sum_{j=1}^n \bar{g}_{\epsilon n}(x_{j,n} - x_{k,n}) \int_{-\infty}^{\infty} \bar{g}_j(y)dy - \int_{-\infty}^{\infty} \frac{1}{n}\sum_{j=1}^n g_\epsilon(y/c_n)\bar{g}_j(y)dy\right|$$
$$= o_{a.s.}(\log^{-l} n), \quad \textbf{(A.26)}$$

$$\Delta_{3n}(k) := \left|\int_{-\infty}^{\infty} \frac{1}{n}\sum_{j=1}^n g_\epsilon(y/c_n)\bar{g}_j(y)dy - \frac{c_n}{n}\sum_{j=1}^n \bar{g}[c_n(x_{j,n} - x_{k,n})]\right|$$
$$= o_{a.s.}(\log^{-l} n). \quad \textbf{(A.27)}$$

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

$$|\bar{g}_{\epsilon n}(y) - g_\epsilon(z)| \le |\bar{g}_{\epsilon n}(y) - g_\epsilon(y)| + |g_\epsilon(y) - g_\epsilon(z)|$$
$$\le C\epsilon^{-2}(n^\zeta/c_n + |y - z|), \quad \textbf{(A.28)}$$

(A.25) follows from that, uniformly for all $1 \le j, k \le n$,

https://doi.org/10.1017/S026646661400070X Published online by Cambridge University Press

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

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

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

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

By the definition of $\bar{g}_j(y)$ and (A.28) again, (A.26) follows from that, uniformly for all $1\le j,k\le n$,

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

As for (A.27), the result follows from that, by using Lemma A.5,

$$\Delta_{3n}(k)=\left|\int_{-\infty}^{\infty}\frac{1}{n}\sum_{j=1}^{n}\left\{\bar{g}[c_n(x_{j,n}-x_{k,n})-y]-\bar{g}[c_n(x_{j,n}-x_{k,n})]\right\}g_\epsilon(y/c_n)dy\right|$$
$$\le\sup_t\sup_{|s-t|\le c_n\epsilon}\left|\frac{c_n}{n}\sum_{j=1}^{n}\left\{\bar{g}(c_nx_{j,n}+s)-\bar{g}(c_nx_{j,n}+t)\right\}\right|$$
$$\times\left(\frac{1}{c_n}\int_{-\infty}^{\infty}g_\epsilon(y/c_n)dy\right)$$
$$=o_{a.s.}(\log^{-l}n)\qquad\textbf{(A.29)}$$

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

The proof of Theorem 3.1 is complete. ■

A.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 (2014), $x_{k,n}$ satisfies Assumption 3.3. $x_{k,n}$ also satisfies Assumption 3.2. Explicitly we will show later that $\{v_j,j\in Z\}$ 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

$$\sup_{0\le t\le 1}|x_{[nt],n}-G_{1n}(t)|=o(n^{-\delta_0}),\quad a.s.,\qquad\textbf{(A.30)}$$

https://doi.org/10.1017/S026646661400070X Published online by Cambridge University Press

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 semimartingale having a local time $L_{G_1}(t,x)$ that satisfies (3.2). To see this claim, it suffices to use Corollary 1.8 of Revuz and Yor (1994, p. 226) with minor modifications.

Due to these facts, it follows from Theorem 3.1 that

$$\sup_{0\le t\le 1}\left|\frac{1}{\sqrt{n}h}\sum_{k=1}^{n} K\left[(x_k - x_{[nt]})/\sqrt{n}\phi\right] - L_{nt}\right| = o_P(\log^{-l} n), \tag{A.31}$$

for any $l > 0$, $h \to 0$, and $n^{1-\epsilon_0}h^2 \to \infty$ where $\epsilon_0 > 0$ can be taken arbitrarily 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 follows from (A.31) and the well known 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 (A.30). In fact, the classical strong approximation theorem implies that, on a richer probability space,

$$\sup_{0\le t\le 1}\left|\sum_{j=1}^{[nt]} v_j - W_1(nt)\right| = o\left[n^{1/(2+\delta)}\right], \quad a.s. \tag{A.32}$$

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

$$\sup_{0\le t\le 1}\left|\sum_{j=1}^{[nt]} \lambda^{[nt]-j} v_j - G^*_{1n}(t)\right| = o\left[n^{1/(2+\delta)}\right], \quad a.s., \tag{A.33}$$

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 (A.6), it follows that

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

https://doi.org/10.1017/S026646661400070X Published online by Cambridge University Press

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 $\max_{1\le k\le n}|\sum_{j=1}^{k}\lambda^{k-j}v_j| = o(\sqrt{n\log n}), a.s.$, it follows from $\sum_{i=0}^{\infty} i|\phi_i| < \infty$ that

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

This proves (A.30) and also completes the proof of (3.4).

We finally prove (3.5). Simple calculations show that

$$V_n \le \frac{2}{n}\sum_{t=1}^{n}(J_{1t}^2 + J_{2t}^2), \tag{A.34}$$

where

$$J_{1t} = \frac{\sum_{s=1}^{n}\left[m(x_s)-m(x_t)\right]K[(x_s-x_t)/h]}{\sum_{s=1}^{n}K[(x_s-x_t)/h]},\ J_{2t} = \frac{\sum_{s=1}^{n}u_s\,K[(x_s-x_t)/h]}{\sum_{s=1}^{n}K[(x_s-x_t)/h]}.$$

Assumption 2.4(a) implies that, when $|x_s - x_t| \le Mh$ and h is sufficiently small,

$$|m(x_s)-m(x_t)| \le C\,|x_s-x_t|^{\beta}(1+|x_t|^{\alpha}) \le C_1 h^{\beta}(1+|x_t|^{\alpha}),$$

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

$$\frac{1}{n}\sum_{t=1}^{n}J_{1t}^2 \le \frac{Ch^{2\beta}}{n}\sum_{t=1}^{n}(1+|x_t|^{2\alpha}) = O_P(n^{\alpha}h^{2\beta}). \tag{A.35}$$

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

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

where

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

https://doi.org/10.1017/S026646661400070X Published online by Cambridge University Press

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

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

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

$$
\frac{1}{n} \sum_{t=1}^{n} J_{2t}^2 = O_P[(nh^2)^{-1/2}]. \tag{A.36}
$$

Combining (A.34)–(A.36), we prove (3.5) and hence complete the proof of Theorem 3.2. ■

https://doi.org/10.1017/S026646661400070X Published online by Cambridge University Press