Optimal bandwidth selection in nonlinear cointegrating regression*

Qiying Wang

School of Mathematics and Statistics
The University of Sydney

Peter C. B. Phillips

Yale University, University of Auckland
University of Southampton & Singapore Management University

August 9, 2020

Abstract

We study optimal bandwidth selection in nonparametric cointegrating regression where the regressor is a stochastic trend process driven by short or long memory innovations. Unlike stationary regression, the optimal bandwidth is found to be a random sequence which depends on the sojourn time of the process. All random sequences h_n that lie within a wide band of rates as the sample size $n \to \infty$ have the property that local level and local linear kernel estimates are asymptotically normal, which enables inference and conveniently corresponds to limit theory in the stationary regression case. This finding reinforces the distinctive flexibility of data-based nonparametric regression procedures for nonstationary nonparametric regression. The present results are obtained under exogenous regressor conditions, which are restrictive but which enable flexible data-based methods of practical implementation in nonparametric predictive regressions within that environment.

JEL Classification: C13, C22.

Key words and phrases: Cointegration, nonlinear regressions, consistency, limit distribution, nonstationarity, nonlinearity, exogeneity.

1 Introduction

Extensions of cointegrating regression techniques to include nonlinear response functions have become available through a substantial body of recent work on both parametric

and nonparametric nonstationary kernel regression. Much of this literature makes use of certain foundational results concerning the asymptotic behavior of various nonlinear functions of integrated processes and standardized forms of integrated processes. Early papers in this literature by Park and Phillips (1999, 2000, 2001), de Jong (2004) and Pötscher (2004) provided a groundwork of methods and results that have assisted the development of this research. Pötscher's work introduced a boundedness assumption on the density of a standardized form of an integrated process that has proved particularly useful in establishing limit theory for parametric and nonparametric estimators of nonlinear functions of nonstationary processes. Many authors have taken advantage of this approach in advancing research in the field.

Following this research, it is now known that standard kernel methods can be employed to estimate and conduct valid nonparametric cointegrating estimation and inference with unit root, local unit root, and long memory regressors, as well as endogenous regressors and weakly dependent structural equation errors (see Wang and Phillips, 2009a&b, 2011, 2016; hereafter WP). Remarkably, nonparametric t-statistics enjoy standard Gaussian limit behavior in these environments precisely as they do in the conventional stationary exogenous regressor setting. It is further known that nonparametric kernel estimators are uniformly consistent over very wide regions with nonstationary data (Chan and Wang, 2014, 2015; Duffy, 2017a), a property that is particularly useful given the typical random wandering nature of such data.

These findings have brought estimation and inference in bivariate nonparametric cointegrating regression to a level of generality comparable to linear cointegrating regression but with the unexpected advantages of (i) not requiring endogeneity or serial correlation bias corrections, (ii) none of the difficulties of the ill-posedness that arise in the stationary nonparametric context with endogenous regressors, and (iii) simple Gaussian inferential methods that facilitate application. The methods have been found to be especially useful in predictive regression with nonstationary predictors (Kasparis et al., 2015) where nonparametric methods show effective size control and good power for a wide class of regressors. In that context, Duffy (2017a, 2017b) has further demonstrated that kernel density estimates satisfy a unified theory of limit behavior that includes both stationary and persistent processes of the integrated and mildly integrated type (Phillips and Magdalinos, 2007). In effect, in predictive regression, the limit distributions of self normalized kernel regression statistics are Gaussian and this property is unaffected by persistence in

2

the regressor, uniformly in the parameters that characterize persistence.

As in other applications of nonparametric methods, implementation requires a rule for bandwidth selection. In stationary and cross section regressions, bandwidth selection analysis is a heavily worked area where operational methods that deliver optimal rates of convergence have long been available and much experience has been accumulated through simulation and empirical practice. In the nonstationary case, while rate conditions are known, there has been little work on optimal selection or formal justification for data-based methods, although simulation evidence is available from past work (WP, 2009a&b) and some results have been recently obtained for β recurrent Markov chains (Bandi et al, 2012).

The contribution of the present paper is to provide an optimal bandwidth selection rule for use in kernel-based nonlinear cointegrating regression. It is found that the optimal bandwidth is delivered by a random sequence, unlike the deterministic function rule that is familiar in cross section and stationary kernel regression. We show that for this bandwidth and for all bandwidth sequences h_n that lie within a certain band of rates as the sample size $n \to \infty$, centred and self standardized local level and local linear kernel estimates are asymptotically standard normal, which enables convenient use in inference and corresponds to standard limit theory in the stationary regression case. This finding reinforces the distinctive flexibility of data-based nonparametric regression procedures for nonstationary nonparametric regression. The present results are obtained under exogenous regressor conditions and enable conditional data-based methods of practical implementation in environments such as certain nonparametric predictive regressions. Exogeneity is restrictive but is a useful starting point that enables the use of existing methods to gain traction on the challenging problem of bandwidth selection in nonparametric nonstationary regression.

We consider a nonlinear cointegrating regression model of the form

$$y_t = m(x_t) + \sigma(x_t)\, u_t, \tag{1.1}$$

where x_t is a non-stationary regressor, $m(\cdot)$ and $\sigma(\cdot)$ are unknown real functions on $\mathbb{R}$ representing the conditional mean and error standard deviation, respectively, and u_t is an equilibrium error satisfying $\mathbb{E}(u_t|x_t) = 0$. In (1.1), the conventional local linear estimator $\widehat{m}_L(x)$ of $m(x)$ is defined by

$$\widehat{m}_L(x) = \sum_{i=1}^{n} w_i(x)y_i \Big/ \sum_{i=1}^{n} w_i(x),$$

where w_i is a weight function defined by $w_i(x) = K[(x_i - x)/h]V_{n,2} - K_1[(x_i - x)/h]V_{n,1}$, employing the non-negative real continuous kernel function $K(x)$, with $K_j(x) = x^j K(x)$ and $V_{n,j} = \sum_{i=1}^n K_j[(X_i - x)/h]$, in which the bandwidth $h \equiv h_n \to 0$.

Under various conditions on the model components $m(x)$ and $\sigma(x)$, the time series x_t, and the bandwidth h, the consistency and asymptotic normality of the kernel estimator $\widehat{m}_L(x)$ have been explored in WP (2009a&b, 2011, 2016) and Wang (2014, 2015). We now address the issue of optimal bandwidth selection to aid implementation of $\widehat{m}_L(x)$ in practical work.

The paper is organized as follows. Section 2 gives the main results and attendant discussion, Section 3 concludes and proofs are provided in Section 4.

2   Main results

In classical nonparametric kernel regression with a stationary regressor, a bandwidth selection rule for the choice of h in $\widehat{m}_L(x)$ is typically based on the asymptotic behavior of a mean squared error (MSE) criterion such as $\mathbb{E}\big[(\widehat{m}_L(x) - m(x))^2\big]$. Such a criterion is appropriate in a stationary setting where the criterion is well defined. In the nonstationary case, as will become apparent, such a criterion is not well suited because the expectation is undefined. Our approach in the present work is therefore based on a conditional MSE criterion

$$\mathbb{E}\big[(\widehat{m}_L(x) - m(x))^2 \mid x_1, ..., x_n\big] \tag{2.1}$$

or, more generally, a conditional weighted average mean squared error (WMSE) such as

$$\int_{-\infty}^{\infty} \mathbb{E}\big[(\widehat{m}_L(x) - m(x))^2 \mid x_1, ..., x_n\big] W(x)dx, \tag{2.2}$$

where $W(x)$ is a weight function having a compact support.

To fix ideas in what follows, we make precise the assumptions employed for the time series (x_t, u_t) in (1.1). Let $\eta_j, j = 0, \pm 1, \pm 2, ...$ be a sequence of iid random variables with $\mathbb{E}\eta_0 = 0$, $\mathbb{E}\eta_0^2 = 1$ and $|\mathbb{E}e^{it\eta_0}| \le t^{-\delta}$ for some $\delta > 0$. Let $\xi_j, j \ge 1$, be a linear process defined by

$$\xi_j = \sum_{k=0}^{\infty} \phi_k \eta_{j-k},$$

where the coefficients $\phi_k, k \ge 0$, satisfy one of the following two conditions that allow for long memory (LM) and short memory (SM) in ξ_j:

LM. $\phi_k \sim k^{-\mu}$, where $1/2 < \mu < 1$;

SM. $\sum_{k=0}^{\infty} |\phi_k| < \infty$ and $\phi \equiv \sum_{k=0}^{\infty} \phi_k \neq 0$.

Let $d = \left(\frac{3}{2} - \mu\right) \mathbf{1}\{(\xi_j) \in \mathbf{LM}\} + \left(\frac{1}{2}\right) \mathbf{1}\{(\xi_j) \in \mathbf{SM}\}$. For some $0 < \delta_0 < \min\{d, 1-d\}$, let $\Omega \subset \{a : |a| \leq n^{d-\delta_0}\}$ be a subset of $\mathbb{R}$ that expands with n at a rate that depends on d. Define the potential bandwidth region $\mathcal{H}_n = \{d_n : \epsilon_n^{-1} n^{d-1+\delta_0} \leq d_n \leq \epsilon_n\}$, where $\log^{-1} n \leq \epsilon_n \to 0$ is a sequence of constants, so that if $h_n \in \mathcal{H}_n$ then $h_n \to 0$ and $n^{1-d-\delta_0} h_n \to \infty$, ensuring that the bandwidth tends to zero but not as fast as $1/n^{1-d}$.

To investigate the asymptotic properties of the conditional MSE and WMSE criteria in (2.1) and (2.2), we employ the following assumptions. All time series are defined in a probability space with filtration $(\mathcal{F}_t)$.

A1. (i) $x_t = \sum_{j=1}^{t} \xi_j$; (ii) $\{u_t, \mathcal{F}_t\}_{t \geq 1}$ forms a martingale difference with $\mathbb{E}(u_t^2 \mid \mathcal{F}_{t-1}) = 1$; and (iii) $x_1, x_2, \cdots, x_n$ are $\mathcal{F}_{t-1}$ measurable for any $1 \leq t \leq n$ and $n \geq 1$.

A2. On Ω, (i) $m(x)$ is bounded and twice continuously differentiable, and (ii) $\sigma(x)$ is bounded and continuous.

A3. $K(x)$ has finite support, $\int_{-\infty}^{\infty} K(x)dx = 1$, $\int_{-\infty}^{\infty} xK(x)dx = 0$ and $|K(x) - K(y)| \leq C|x - y|$ whenever $|x - y|$ is sufficiently small.

The first result describes the asymptotic behavior of the conditional MSE and conditional WMSE.

THEOREM 2.1. *Suppose* ***A1***–***A3*** *hold. For any* $h = h_n(x_1, ..., x_n)$ *such that* $\lim_{n \to \infty} P(h \in \mathcal{H}_n) = 1$, *we have*

$$
\mathbb{E}\big[(\widehat{m}_L(x) - m(x))^2 \mid x_1, ..., x_n\big]
$$
$$
= \frac{1}{4} \tau^2 h^4 [m''(x)]^2 + (h A_n)^{-1} \sigma^2(x) \int_{-\infty}^{\infty} K^2(t)dt
$$
$$
+ o_P\Big[\big(n^{1-d}h\big)^{-1} + h^4\Big], \tag{2.3}
$$

uniformly in $x \in \Omega$, *where* $\tau = \int_{-\infty}^{\infty} t^2 K(t)dt$ *and* $A_n = \sum_{k=1}^{n} K(x_k)$. *Moreover, for any weight function* $W(x)$ *having a compact support that is covered by* Ω, *we have*

$$
\int_{-\infty}^{\infty} \mathbb{E}\big[(\widehat{m}_L(x) - m(x))^2 \mid x_1, ..., x_n\big] W(x)dx
$$
$$
= \frac{1}{4} \tau^2 h^4 \int_{-\infty}^{\infty} [m''(x)]^2 W(x)dx + (hA_n)^{-1} \int_{-\infty}^{\infty} \sigma^2(x) W(x)dx \int_{-\infty}^{\infty} K^2(t)dt
$$
$$
+ o_P\Big[\big(n^{1-d}h\big)^{-1} + h^4\Big]. \tag{2.4}
$$

Remark 1. Based on (2.3), for any $x \in \Omega$, the optimal pointwise bandwidth is taken to be

$$h_{opt} = \Big[\frac{\sigma^2(x) \int_{-\infty}^{\infty} K^2(t)dt}{[\tau\, m''(x)]^2} \Big]^{1/5} A_n^{-1/5}. \tag{2.5}$$

Similarly, based on (2.4), the optimal weighted bandwidth is

$$h_{opt} = \Big[\frac{\int_{-\infty}^{\infty} \sigma^2(x)\, W(x)dx \int K^2(x)dx}{\int_{-\infty}^{\infty} [m''(x)]^2 W(x)dx\, \tau^2} \Big]^{1/5} A_n^{-1/5}. \tag{2.6}$$

Note that $(s_n/n)A_n \to_D L_G(1,0)$, where $s_n^2 = var(x_n)$ and $L_G(t,a)$ is the local time at spatial location a at time t of the (fractional) Brownian motion limit process G_t with index d for which $s_n^{-1} x_{\lfloor nt \rfloor} \Rightarrow G_t$ (see Lemma 4.2 and equation (4.14) below), where $\lfloor \cdot \rfloor$ is the floor function. Unlike the stationary time series case where the optimal bandwidth is a deterministic sequence, the optimal bandwidth here is a random sequence involving A_n which upon normalization has the random limit $L_G(1,0)$. Due to the fact that $\mathbb{E}L_G(1,0)^{-1} = \infty$, the conventional mean squared error criterion based on the unconditional mean squared error $\mathbb{E}(\widehat{m}_L(x) - m(x))^2$ cannot be used as a selection rule for the bandwidth. On the other hand, we do have the following uniform convergence result: for any $h = h_n(x_1, ..., x_n)$ such that $\lim_{n\to\infty} P(h \in \mathcal{H}_n) = 1$,

$$\sup_{x\in\Omega} |\widehat{m}_L(x) - m(x)| = O_P\{(n^{1-d}h)^{-1/2} \log^{1/2} n + h^2\}. \tag{2.7}$$

The proof of (2.7) is similar to that in Section 5.1.4 of Wang (2015). We omit the details.[1]

Remark 2. Using Theorem 2.1, an explicit presentation of the optimal bandwidth h is provided that depends on the sojourn time of the process. It is clear from the proof that results (2.3) and (2.4) still hold if A_n is replaced by $A_{n,h} := \frac{1}{h}\sum_{k=1}^{n} K\big[(x_k - x)/h\big]$. In consequence, the optimal pointwise bandwidth can be taken to be

$$\hat{h}_{opt} = \arg\min_{h} \Big\{ \frac{1}{4}\tau^2 h^4\, [m''(x)]^2 + (h\, A_{n,h})^{-1}\, \sigma^2(x) \int_{-\infty}^{\infty} K^2(t)dt \Big\},$$

and, similarly, the optimal weighted bandwidth can be taken as

$$\hat{h}_{opt} = \arg\min_{h} \Big\{ \frac{1}{4}\tau^2 h^4 \int_{-\infty}^{\infty} \big[m''(x)\big]^2 W(x)dx$$
$$+ (hA_{n,h})^{-1} \int_{-\infty}^{\infty} \sigma^2(x)\, W(x)dx \int_{-\infty}^{\infty} K^2(t)dt \Big\}.$$

---

[1]Another approach, which we do not pursue here, is to consider the median of the conditional mean squared error $\mathbb{E}\big[(\widehat{m}_L(x) - m(x))^2 \mid x_1, ..., x_n\big]$ rather than its mean. Since the density of the local time is known in certain cases such as the local time of Brownian motion, the median of the reciprocal of the local time $L_G(1,0)$ may be deduced in those cases and a sample approximation constructed in terms of a function of A_n.

Use of these alternative formulations of the optimal bandwidth may have some finite sample benefit in performance at the cost of more complex calculation.

Remark 3. As noticed by a reviewer, the optimal bandwidth would ideally be defined with respect to a criterion that would remain meaningful if $\{x_t\}_{t=1}^n$ were merely predetermined. The challenge in this nonstationary case is that the criteria (2.3) and (2.4) depend on the random quantity $A_n \sim_a \frac{n}{s_n} L_G(1,0)$ that relies on the sojourn time $L_G(1,0)$ of the limit process associated with x_t. This dependence ensures that a small sojourn time implies the need for a large bandwidth in optimal estimation of $m(x)$, thereby compensating for the fact that the data is less informative about the function in the immediate vicinity of this location. This characteristic feature of the problem of bandwidth selection is particular to the nonstationary case and should persist when the nonstationary regressor is predetermined. The criterion for selection in the present paper does not meet the requirement of demonstrating this feature analytically but the result itself is suggestive and provides some indications that will be useful in future development of this line of research.

Let $\mathcal{H}_{1n} = \{d_n : \epsilon_n^{-1} n^{d-1+\delta_0} \le d_n \le \epsilon_n n^{(d-1)/7}\}$, where $\delta_0 > 0$ is chosen as small as required and $\log^{-1} n \le \epsilon_n \to 0$.

THEOREM 2.2. *In addition to **A1**–**A3**, suppose that $m^{(3)}(x)$ is bounded on Ω and $\sup_{k\ge 1} \mathbb{E}|u_k|^{2+\delta} < \infty$, for some $\delta > 0$. Then, for any $h = h_n(x_1, ..., x_n)$ such that $\lim_{n\to\infty} P(h \in \mathcal{H}_{1n}) = 1$, we have*

$$\left(\sum_{k=1}^n K\big[(x_k - x)/h\big]\right)^{1/2} \big[\widehat{m}_L(x) - m(x) - \frac{h^2}{2} m''(x)\tau\big] \to_D \mathcal{N}\left(0,\ \sigma^2(x) \int_{-\infty}^{\infty} K^2(t)dt\right),$$
(2.8)

for any $x \in \Omega$, where $\tau = \int_{-\infty}^{\infty} t^2 K(t)dt$.

Remark 4. Theorem 2.2 shows that the self normalized bias corrected estimation error $\widehat{m}_L(x) - m(x) - \frac{1}{2} h^2 m''(x)\tau$ has the same standard normal limit distribution for all choices of bandwidth that lie within the region $\mathcal{H}_{1n}$ as $n \to \infty$. Since $P(h_{opt} \in \mathcal{H}_{1n}) \to 1$, Theorem 2.2 indicates that result (2.8) applies for the optimal bandwidth h_{opt}. The optimal bandwidth formula

$$h_{opt} = h\left(\sigma^2(x), m''(x)\right) = \left[\frac{\sigma^2(x) \int_{-\infty}^{\infty} K^2(t)dt}{\big[\tau\, m''(x)\big]^2}\right]^{1/5} A_n^{-1/5}$$
(2.9)

is infeasible as it depends on $\sigma^2(x)$ and $m''(x)$, analogous to the optimal bandwidth formula in the usual stationary case. Both $\sigma^2(x)$ and $m''(x)$ are consistently estimable

and the resulting plug-in feasible version of the optimal bandwidth $\hat{h}_{opt} = h\left(\hat{\sigma}^2(x), \hat{m}''(x)\right)$ with such consistent estimates continues to lie within the region $\mathcal{H}_{1n}$ as $n \to \infty$. Nonetheless, as pointed out by a referee, the proof of Theorem 2.2 relies on the martingale array property of the sample covariance $\sum_{k=1}^{n} w_k(x)\sigma(x_k)\,u_k$, which fails when a feasible bandwidth such as $\hat{h}_{opt}$ is used because of the resulting dependence on $\{y_1, ..., y_n\}$ that is introduced to the kernel weights and estimation of $\sigma^2(x)$ and $m''(x)$. Thus, Theorem 2.2 is no longer established when a feasible bandwidth choice such as $\hat{h}_{opt}$ is used in estimation. This problem of dependence is common in kernel nonparametric estimation and arises, for instance, in stationary nonparametric regression when a plug-in optimal bandwidth is employed, so it is not confined to the present nonlinear cointegrating regression model.

Remark 5. Due to the nonstationarity of the regressor x_t condition **A1** (iii) plays a significant role in the proof of Theorem 2.2, wherein the extended martingale limit theorem given by Wang (2014) is employed. The condition **A1** (iii) essentially requires independence between the regressor and the error process, thereby excluding endogenous and predetermined regressors. Relaxation of this condition is technically difficult and seems unlikely to be possible using the present approach and techniques.

Remark 6. Let $\widehat{m}(x)$ be the conventional local level kernel estimator defined by

$$\widehat{m}(x) = \frac{\sum_{k=1}^{n} K\big[(x_k - x)/h\big] y_k}{\sum_{k=1}^{n} K\big[(x_k - x)/h\big]}.$$

Let $\mathcal{H}_{2n} = \{d_n : n^{(d-1)/3}\log n \leq d_n \leq n^{(d-1)/7}/\log n\}$. Results (2.3)–(2.8) remain true when $\widehat{m}_L(x)$ is replaced by $\widehat{m}(x)$ if $h = h_n(x_1, ..., x_n)$ is a random bandwidth sequence satisfying $\lim_{n\to\infty} P(h \in \mathcal{H}_{1n}) = 1$. This additional rate control condition on the bandwidth is used to remove the first order biases which have no impact on the choice of the optimal bandwidth.

3 Conclusion

Nonparametric methods in cointegrated systems have presented technical challenges to the development of limit theory for kernel estimators and theoretical underpinnings of automated methods of implementation. This paper, in conjunction with other recent work discussed in the Introduction, helps to advance the available limit theory for these systems to enable practical implementation in empirical work. The optimal bandwidth selection rules given in Section 2 enable conditional data-based implementation of kernel

techniques in environments such as pure cointegrating regression and nonparametric predictive regression that are valid under exogenous regressors. Several technical challenges remain. We need to extend the present results, or some version of them, to models in which the regressor is endogenous or predetermined. This is a demanding task that seems to require new methods for the reasons explained above. And, just as in the stationary regression case, it will also be useful in practical work to have suitable mechanisms for plug in estimation of the quantities involved to develop empirical versions of the optimal bandwidth formulae (2.5) and (2.6). While this paper does not address those challenges, it does demonstrate some progress towards automated data-based methods of inference in nonstationary nonparametric regression and prediction.

4   Proofs

For $0 < t_1 < t_2 < \infty$, let $\mathcal{L}_n = \{(t,a) : t_1 n^{d-1+\delta_0/2} \leq t \leq t_2, |a| \leq n^{d-\delta_0}\}$. Suppose $g(x)$ is a bounded real function satisfying $\int_{-\infty}^{\infty} |g(x)| dx < \infty$. We start with the following lemmas, which play key roles in the main proofs.

LEMMA 4.1. *Suppose $g(x)$ has finite support and satisfies a Lipschitz condition. Then*

$$\sup_{(h,x)\in\mathcal{L}_n} \Big| \sum_{k=1}^{n} \Big\{ \frac{1}{h} g\big[(x_k - x)/h\big] - g(x_k) \Big\} \Big| = O_{a.s.}(n^{1-d-\gamma\delta_0}) \tag{4.1}$$

for some $\gamma > 0$.

Proof. Set $t_1 n^{d-1+\delta_0/2} = h_1 < ... < h_{q_{n1}} = t_2$ and $-n^{d-\delta_0} = t_1 < ... < t_{q_{n2}} = n^{d-\delta_0}$ with $h_i - h_{i-1} \sim n^{-7}$ and $t_i - t_{i-1} \sim n^{-10}$. Due to $\mathbb{E}x_n^2 \asymp n^d$, we have $x_n = o_{a.s}(n^2)$. Standard arguments show that to prove (4.1) it suffices to show

$$\max_{1\leq i\leq q_{n1}} \max_{1\leq j\leq q_{n2}} \Big| \sum_{k=1}^{n} f_{h_i,t_j}(x_k) \Big| = O_{a.s.}(n^{1-d-\gamma\delta_0}), \tag{4.2}$$

where $f_{h_i,t_j}(y) = h_i^{-1} g[(y-t_j)/h_i] - g(y)$. Note that $\int_{-\infty}^{\infty} f_{h_i,t_j}(y) dy = 0$, $\int_{-\infty}^{\infty} |f_{h_i,t_j}(y)| dy < \infty$, $\sup_y |f_{h_i,t_j}(y)| \leq Cn^{1-d-\delta_0}$ and

$$
\begin{aligned}
\inf_t \int_{-\infty}^{\infty} |f_{h_i,t_j}(y-t)||y| dy &\leq \int_{-\infty}^{\infty} |y| \, |g(y)| dy + \int_{-\infty}^{\infty} |g(y)|(t_j + h_i|y|) dy \\
&\leq C n^{d-\delta_0}.
\end{aligned}
$$

Result (4.2) is then a direct corollary of Theorem 2.30 in Wang (2015). $\square$

LEMMA 4.2. *We have*

$$\frac{s_n}{n} \sum_{k=1}^{n} g(x_k) \to_D \int_{-\infty}^{\infty} g(x)dx \, L_G(1,0), \tag{4.3}$$

where $s_n^2 = var(x_n)$ and $L_G(t,a)$ is the local time process at time t and spatial location a of the fractional Brownian motion G_t with index $H = 3/2 - \mu$ for which $s_n^{-1} x_{\lfloor nt \rfloor} \Rightarrow G_t$. If in addition $\int_{-\infty}^{\infty} g(t)dt = 0$, $\int_{-\infty}^{\infty} |t\,g(t)|dt < \infty$ and $g(t)$ satisfies the Lipschitz condition, then

$$\sup_{(h,x)\in\mathcal{L}_n} (n^{1-d}h)^{-1/2} \left| \sum_{k=1}^{n} g\big[(x_k - x)/h\big] \right| = O_P(\log n). \tag{4.4}$$

Proof. Result (4.3) follows from Wang and Phillips (2009a) and result (4.4) follows from Duffy (2017b) and Chan and Wang (2014). □

4.1 Proof of Theorem 2.1

We are now ready to prove (2.3). Result (2.4) follows from (2.3) with a minor calculation and hence the details are omitted. For convenience in the arguments that follow we introduce the notation $X_n \leq_P Y_n$ for $X_n/Y_n = O_P(1)$, and $X_n(x) = Y_n(x) + o_P(1)Z_n(x)$, uniformly in Ω, for

$$\sup_{x\in\Omega} \frac{|X_n(x) - Y_n(x)|}{Z_n(x)} = o_P(1).$$

Recalling condition **A1** and the fact that $h = h_n(x_1, ..., x_n)$ is a function only of $x_1, ..., x_n$, simple calculations show that

$$
\begin{aligned}
& \mathbb{E}\big[(\widehat{m}_L(x) - m(x))^2 \mid x_1, ..., x_n\big] \\
&= \left\{ \frac{\sum_{k=1}^{n} w_k(x)[m(x_k) - m(x)]}{\sum_{k=1}^{n} w_k(x)} \right\}^2 + \frac{\sum_{k=1}^{n} w_k^2(x)\sigma^2(x_k)}{\left[\sum_{k=1}^{n} w_k(x)\right]^2} \\
&= I_1(n) + I_2(n), \quad \text{say.}
\end{aligned} \tag{4.5}
$$

Since $K(x)$ has finite support, there exists a $C_0 > 0$ such that $w_k(x)\big[m(x_k) - m(x)\big] = 0$ if $|x_k - x|/h \geq C_0$. This, together with condition **A2** (i), ensures that

$$
\begin{aligned}
& w_k(x)\big[m(x_k) - m(x)\big] \\
&= w_k(x)\big\{m\big[x + h(x_k - x)/h\big] - m(x)\big\} \\
&= w_k(x)\big[m'(x)(x_k - x) + \frac{1}{2}m''(x)(x_k - x)^2 + o_P(1)(x_k - x)^2\big],
\end{aligned}
$$

as $h = o_P(1)$. Hence, by noting that $\sum_{k=1}^{n}(x_k - x)w_k(x) = 0$, we may write

$$I_1(n) = \frac{1}{4}h^4 \left\{\big[m''(x)\big]^2 + o_P(1)\right\} l_n^2, \tag{4.6}$$

10

uniformly in $x \in \Omega$, where, recalling that $K_j(x) = x^j K(x), j = 0, 1, 2, ...$, and $V_{n,j} = \sum_{k=1}^{n} K_j[(x_k - x)/h]$, we have

$$l_n = \frac{V_{n,2}^2 - V_{n,3}V_{n,1}}{V_{n,0}V_{n,2} - V_{n,1}^2} = \frac{V_{n,2} - V_{n,3}V_{n,1}/V_{n,2}}{V_{n,0} - V_{n,1}^2/V_{n,2}}.$$

Note that $s_n \asymp n^d$ and $P(L_G(1,0) > 0) = 1$ (e.g., see Chapter 2 of Wang, 2015). It is readily seen from (4.1) and (4.3) that

$$\inf_{x \in \Omega} V_{n,0}, \quad \inf_{x \in \Omega} V_{n,2} \geq_P h \sum_{k=1}^{n} K(x_k) \geq_P \delta_n n^{1-d} h, \tag{4.7}$$

where $\log^{-1} n < \delta_n \to 0$ is chosen as slowly as required. On the other hand by (4.4), setting $g(t) = tK(t)$ and $t^2 K(t) - \int t^2 K(t)dt\, K(t)$, we have

$$\sup_{x \in \Omega} |V_{n,1}|, \quad \sup_{x \in \Omega} |V_{n,2} - V_{n,0} \int_{-\infty}^{\infty} t^2 K(t)dt\,| \leq_P (n^{1-d}h)^{1/2} \log n. \tag{4.8}$$

Results (4.7)–(4.8) and the fact that $V_{n,3} \leq C V_{n,2}$ imply

$$\begin{aligned}
\sup_{x \in \Omega} |l_n^2 - \tau^2| &\leq \sup_{x \in \Omega} |l_n - \tau| \, |l_n + \tau| \\
&\leq_P (n^{1-d}h)^{-1/2} \log^2 n = o_P(1),
\end{aligned}$$

where $\tau = \int_{-\infty}^{\infty} t^2 K(t)dt$. Taking this estimate into (4.6) and noting that $\sup_{x \in \Omega} |m''(x)| < \infty$, we obtain

$$I_1(n) = \frac{1}{4} h^4 \left[m''(x) \right]^2 \tau^2 + o_P(h^4), \tag{4.9}$$

uniformly in $x \in \Omega$.

We next consider $I_2(n)$. Similar arguments making use of (4.7)–(4.8) yield that

$$\frac{hV_{n,2}}{\sum_{k=1}^{n} w_k(x)} = \frac{h}{V_{n,0} - V_{n,1}^2/V_{n,2}} = A_n^{-1} + O_P(n^{d-1-\gamma\delta_0}), \tag{4.10}$$

where $A_n = \sum_{k=1}^{n} K(x_k)$, and

$$\begin{aligned}
\frac{V_{n,2}^{-1} \sum_{k=1}^{n} w_k^2(x)}{\sum_{k=1}^{n} w_k(x)} &= \frac{\sum_{k=1}^{n} K^2[(x_k - x)/h] + R_n}{V_{n,0} - V_{n,1}^2/V_{n,2}} \\
&= \int_{-\infty}^{\infty} K^2(t)dt + O_P[(n^{1-d}h)^{-1/2} \log^2 n], \tag{4.11}
\end{aligned}$$

uniformly in $x \in \Omega$, where we have used the fact that since both $K(x)$ and $K_1(x)$ are bounded,

$$|R_n| \leq 2C|V_{n,1}| V_{n,0}/V_{n,2} + CV_{n,1}^2 V_{n,0}/V_{n,2}^2 = O_P[(n^{1-d}h)^{1/2} \log n].$$

11

By virtue of (4.10) and (4.11), we have

$$
\frac{\sum_{k=1}^{n} w_k^2(x)}{\left[\sum_{k=1}^{n} w_k(x)\right]^2} = h^{-1} \frac{hV_{n,2}}{\sum_{k=1}^{n} w_k(x)} \frac{V_{n,2}^{-1} \sum_{k=1}^{n} w_k^2(x)}{\sum_{k=1}^{n} w_k(x)}
$$

$$
= \left\{ (hA_n)^{-1} + o_P\left[(n^{1-d}h)^{-1}\right] \right\} \left\{ \int_{-\infty}^{\infty} K^2(t)dt + O_P\left(n^{-\delta_0/2} \log^2 n\right) \right\}
$$

$$
= \int_{-\infty}^{\infty} K^2(t)dt \, (hA_n)^{-1} + o_P\left[(n^{1-d}h)^{-1}\right],
$$

uniformly in $x \in \Omega$. In consequence, recalling that $K(x)$ has finite support and $\sigma(x)$ is bounded and continuous on Ω, standard arguments now yield that

$$
\begin{aligned}
I_2(n) &= \frac{\left[\sigma^2(x) + o_P(1)\right] \sum_{k=1}^{n} w_k^2(x)}{\left[\sum_{k=1}^{n} w_k(x)\right]^2} \\
&= \sigma^2(x) \int_{-\infty}^{\infty} K^2(t)dt \, (hA_n)^{-1} + o_P\left[(n^{1-d}h)^{-1}\right]. \tag{4.12}
\end{aligned}
$$

Result (2.4) follows immediately from (4.9) and (4.12).

4.2   Proof of Theorem 2.2

Noting that $\sum_{k=1}^{n}(x_k - x)w_k(x) = 0$, we may write

$$
\begin{aligned}
& \widehat{m}_L(x) - m(x) - \frac{1}{2}h^2 m''(x)\tau \\
&= \frac{\sum_{k=1}^{n} w_k(x)\sigma(x_k)\, u_k}{\sum_{k=1}^{n} w_k(x)} \\
&\quad + \frac{\sum_{k=1}^{n} w_k(x)[m(x_k) - m(x) - m'(x)(x_k - x) - \frac{1}{2}m''(x)(x_k - x)^2]}{\sum_{k=1}^{n} w_k(x)} \\
&\quad + \frac{1}{2}m''(x)\left[\frac{\sum_{k=1}^{n} w_k(x)(x_k - x)^2}{\sum_{k=1}^{n} w_k(x)} - h^2 \int_{-\infty}^{\infty} t^2 K(t)dt\right] \\
&= R_{n1} + R_{n2} + R_{n3}, \quad \text{say.} \tag{4.13}
\end{aligned}
$$

Recall that $m^{(3)}(x)$ is bounded on $\mathbb{R}$. As in the proof of Theorem 2.1, for any $x \in \Omega$, we have

$$
|R_{n2}| \leq_P \frac{h^3(|V_{n3}| + |V_{n4}||V_{n1}|/V_{n2})}{|V_{n0} - V_{n1}^2/V_{n2}|} \leq_P h^3,
$$

where $K_j(x) = x^j K(x)$ and $V_{nj} = \sum_{k=1}^{n} K_j\left[(x_k - x)/h\right]$, and

$$
\begin{aligned}
|R_{n3}| &\leq_P \frac{h^2(|V_{n2} - \int_{-\infty}^{\infty} t^2 K(t)dt\, V_{n0}| + |V_{n3}||V_{n1}|/V_{n2})}{|V_{n0} - V_{n1}^2/V_{n2}|} \\
&\leq_P h^2 \, (n^{1-d}h)^{-1/2} \log n.
\end{aligned}
$$

In consequence, it follows that

$$\Big(\sum_{k=1}^{n} K\big[(x_k - x)/h\big]\Big)^{1/2} \big(|R_{n2}| + |R_{n3}|\big)$$

$$\leq_P \quad (n^{1-d}h)^{1/2} \big[h^3 + h^2 (n^{1-d}h)^{-1/2} \log n\big] = o_P(1),$$

since $\lim_{n\to\infty} P(h \in \mathcal{H}_{1n}) = 1$. Hence, Theorem 2.2 will follow by continuous mapping if we prove

$$\Big\{\frac{s_n}{nh} \sum_{k=1}^{n} K\big[(x_k - x)/h\big], \ \frac{s_n}{nh} V_{n2}^{-1} \sum_{k=1}^{n} w_k(x), \ \big(\frac{s_n}{nh}\big)^{1/2} V_{n2}^{-1} \sum_{k=1}^{n} w_k(x)\sigma(x_k)\, u_k\Big\}$$

$$\to_D \quad \big\{L_G(1,0), \ L_G(1,0), \ c_0\, L_G^{1/2}(1,0)\, N\big\},$$

or equivalently,

$$\Big\{\frac{s_n}{nh} \sum_{k=1}^{n} K\big[(x_k - x)/h\big], \big(\frac{s_n}{nh}\big)^{1/2} \sum_{k=1}^{n} K\big[(x_k - x)/h\big]\sigma(x_k)\, u_k\Big\}$$

$$\to_D \quad \big\{L_G(1,0), \ c_0\, L_G^{1/2}(1,0)\, N\big\}, \tag{4.14}$$

where $c_0^2 = \int_{-\infty}^{\infty} K^2(t)dt\, \sigma^2(x)$ and N is a standard normal variate independent of $L_G(1,0)$. To prove (4.14), let $\mathcal{F}_k$ be defined as in condition **A1** and

$$X_{nk} = \Big(\frac{s_n}{nh}\Big)^{1/2} K\big[(x_k - x)/h\big]\sigma(x_k)\, u_k.$$

By recalling $h = h_n(x_1, ... x_n)$, i.e., h is $\mathcal{F}_k$-measurable for each $1 \leq k \leq n$, we have

$$\mathbb{E}(X_{nk} \mid x_1, ..., x_n) = \Big(\frac{d_n}{nh}\Big)^{1/2} K\big[(x_k - x)/h\big]\sigma(x_k)\mathbb{E}(u_k \mid x_1, ..., x_n) = 0.$$

The remainder of the proof of (4.14) follows from the same arguments as in the proof of Theorem 5.2 in Wang (2015) [see pages 196-197 there and note that the extra condition on $\sigma(x)$ is not necessary under the current condition on $K(x)$].

REFERENCES

Bandi, F. M., V. Corradi and D. Wilhelm (2012). "Data-driven bandwidth selection for nonparametric nonstationary regressions", Working Paper, Johns Hopkins.

Chan, N. and Wang, Q. (2014). Uniform convergence for non-parametric estimators with non-stationary data, *Econometric Theory*, 30, 1110–33.

Chan, N. and Wang, Q. (2015). Nonlinear regression with nonstationary time series. *Journal of Econometrics*, **185**, 182-195.

de Jong, R. M. (2004). "Addendum to: 'Asymptotics for nonlinear transformations of integrated time series'," *Econometric Theory*, 20, 627-635.

Duffy, J.A. (2017a). Uniformly valid inference in nonparametric predictive regression, Working Paper, University of Oxford.

Duffy, J.A. (2017b). Uniform convergence rates over maximal domains in structural nonparametric cointegrating regression, *Econometric Theory*, 33, 1387–1417.

Kasparis, I., E. Andreou, and P. C. B. Phillips (2015). "Nonparametric Predictive Regression", *Journal of Econometrics*, 185(2), 468–94.

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

Park, J. Y. and P. C. B. Phillips (2000). "Nonstationary binary choice,"*Econometrica*, 68, 1249-1280.

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

Phillips, P. C. B., and T. Magdalinos (2007), "Limit Theory for Moderate Deviations from a Unit Root," *Journal of Econometrics,* 136, 115-130.

Pötscher, B. M. (2004). "Nonlinear functions and convergence to Brownian motion: beyond the continuous mapping theorem, " *Econometric Theory,* 20, 1-22.

Wang, Q. (2014). Martingale limit theorem revisited and nonlinear cointegrating regression, *Econometric Theory*, 30, 509–535.

Wang, Q. (2015). *Limit theorems for nonlinear cointegrating regression.* World Scientific.

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 P. C. B. Phillips (2016). "Nonparametric Cointegrating Regression with Endogeneity and Long Memory", *Econometric Theory* , 32, 359–401.