Econometric Theory, **33**, 2017, 1387–1417.
doi:10.1017/S0266466616000451

UNIFORM CONVERGENCE RATES OVER MAXIMAL DOMAINS IN STRUCTURAL NONPARAMETRIC COINTEGRATING REGRESSION

JAMES A. DUFFY
University of Oxford

This paper presents uniform convergence rates for kernel regression estimators, in the setting of a structural nonlinear cointegrating regression model. We generalise the existing literature in three ways. First, the domain to which these rates apply is much wider than the domains that have been considered in the existing literature, and can be chosen so as to contain as large a fraction of the sample as desired in the limit. Second, our results allow the regression disturbance to be serially correlated, and cross-correlated with the regressor; previous work on this problem (of obtaining uniform rates) having been confined entirely to the setting of an exogenous regressor. Third, we permit the bandwidth to be data-dependent, requiring it to satisfy only certain weak asymptotic shrinkage conditions. Our assumptions on the regressor process are consistent with a very broad range of departures from the standard unit root autoregressive model, allowing the regressor to be fractionally integrated, and to have an infinite variance (and even infinite lower-order moments).

1. INTRODUCTION

Whereas data on a stationary regressor will lie, with high probability, within a fixed bounded interval of sufficient width, the randomly wandering nature of an integrated process prevents it from being contained within any such interval, no matter how wide. Consequently, the global nonparametric estimation of a regression function taking the latter as an argument is considerably more difficult, as it requires one to approximate the regression function on an ever-expanding domain—widening probabilistically at rate $n^{1/2}$, in the unit root case. The inherent randomness of the limiting occupation density associated with the (standardised) regressor process poses a further challenge, complicating the identification of domains on which observations may be guaranteed to accumulate, in a manner that seems not to have any parallel in the case of a stationary regressor.

The author thanks Xiaohong Chen, Bent Nielsen and Peter Phillips for helpful comments on this paper, and the earlier work. The manuscript was prepared with LyX 2.1.3 and JabRef 2.7b. Address correspondence to James A. Duffy, Institute for New Economic Thinking, Oxford Martin School; and Economics Department, University of Oxford, Oxford, UK; e-mail: james.duffy@economics.ox.ac.uk.

© Cambridge University Press 2017

https://doi.org/10.1017/S0266466616000451 Published online by Cambridge University Press

This paper considers kernel nonparametric estimators of m_0 in the nonlinear cointegrating model

$$y_t = m_0(x_t) + u_t, \tag{1.1}$$

where $x_t = \sum_{s=1}^{t} v_t$ is the partial sum of a linear process $\{v_t\}$, and $\{u_t\}$ is an unobserved disturbance process. Regarding the pointwise consistency and asymptotic normality of these estimators, we refer in particular to Karlsen, Myklebust, and Tjøstheim (2007) and Wang and Phillips (2009a,b). Our assumptions on the mechanism generating $\{x_t\}$ are very general, not only permitting fractional integration of order $d \in (-\frac{1}{2}, \frac{3}{2})$—where $d = 1$ corresponds to the familiar unit root autoregressive model—but also including the case where the variance (and even lower-order moments) of $\{v_t\}$ do not exist.

We obtain rates of uniform convergence for our estimators on a sample-dependent sequence of domains, which correspond as nearly as possible to the entire empirical support of $\{x_t\}_{t=1}^{n}$ in the sense that they may be chosen so as to contain as large a fraction of the data as desired in the limit. These domains are thus maximally wide; in contrast, previous work on uniform convergence rates in this setting has been limited to the consideration of smaller, deterministically expanding intervals, which necessarily contain an asymptotically negligible fraction of the data (see Wang and Wang, 2013; Chan and Wang, 2014; Gao, Kanaya, Li, and Tjøstheim, 2015). Being able to estimate m_0 uniformly on such wide domains should be especially useful in the context of certain semiparametric estimation problems, such as those arising when $m_0(x_t)$ in (1.1) is replaced by the more general formulation $m_0(x_t'\beta_0)$, where x_t is a vector nonstationary process, and both β_0 and m_0 are to be jointly estimated. Clearly, only observations lying in those domains on which m_0 may be (uniformly) consistently estimated would be of any use in estimating β_0; our results suggest, reassuringly, that 'almost all' the observed sample should be available for this purpose.

We further generalise previous work by permitting the regressor to be endogenous, and the bandwidth sequence to be data-dependent in a very general way. Endogeneity arises naturally in the setting of cointegrating models such as (1.1), which are so dynamically under-specified as to be plausible only as a model of the long-run equilibrium relationship between $\{y_t\}$ and $\{x_t\}$. The exogeneity assumption typically—though not universally, see Wang and Phillips (2009b, 2016)—imposed in these models is thus unlikely to be satisfied in applications. Moreover, in any application, the bandwidth used to compute a kernel regression estimator will not be determined, *a priori*, as some function of the sample size, but will instead be chosen with at least some reference to the sample at hand. To better accommodate this aspect of actual empirical work, we allow the bandwidth to be functionally dependent on the sample $\{(y_t, x_t)\}_{t=1}^{n}$, requiring it to satisfy only a weak asymptotic shrinkage condition.

The proofs of these convergence rates are facilitated by a number of new technical results. The first concerns the weak convergence of the standardised *signal* process

https://doi.org/10.1017/S0266466616000451 Published online by Cambridge University Press

$$\mathcal{L}_n(a) := \frac{1}{e_n h_n} \sum_{t=1}^{n} K\left(\frac{x_t - d_n a}{h_n}\right) \rightsquigarrow \mathcal{L}(a) \tag{1.2}$$

in $\ell_{\mathrm{ucc}}(\mathbb{R})$, where: $\mathcal{L}$ denotes the occupation density associated to the finite-dimensional limit of $X_n(r) := d_n^{-1} x_{\lfloor nr \rfloor}$; $\{e_n\}$ is a norming sequence; $K \in L^1(\mathbb{R})$ is a mean-zero kernel density function; and $h_n = o_p(1)$ is a smoothing (bandwidth) sequence. This type of result is proved in Duffy (2016) and is reproduced as Proposition 2.1 below. (1.2) may be loosely regarded as the nonstationary process counterpart of the uniform convergence of the signal to the corresponding invariant density that obtains when $\{x_t\}$ is stationary. In the present setting, $\mathcal{L}_n$ arises as the denominator of the Nadaraya–Watson estimator, and so, when combined with a suitable characterisation of the (random) support of $\mathcal{L}$, (1.2) allows us to identify a sequence of domains on which the signal must accumulate at a certain probabilistic rate.

The second class of relevant technical results supplies uniform order estimates for

$$\frac{1}{(e_n h_n)^{1/2}} \sum_{t=1}^{n} f\left(\frac{x_t - d_n a}{h_n}\right) u_t \qquad \frac{1}{(e_n h_n)^{1/2}} \sum_{t=1}^{n} g\left(\frac{x_t - d_n a}{h_n}\right),$$

where $f, g \in L^1(\mathbb{R})$, $\int g = 0$, and $\{u_t\}$ is weakly dependent. These are referred to as the *covariance* and *zero energy* processes, and are, respectively, relevant for a determination of the uniform order of the variance and bias of a kernel regression estimator. The estimates obtained here (Theorem 2.1 below) appear to be new to the literature. While Chan and Wang (2014) provide an estimate for the covariance process when $\{x_t\}$ is exogenous, our estimate holds even when $\{u_t\}$ is correlated with $\{x_t\}$, and is within a $\log^{1/2} n$ factor of theirs. In consequence, endogeneity of the regressor seems to penalise the rate of convergence of a kernel regression estimator by merely a factor of $\log^{1/2} n$.

Notation.[1] For deterministic sequences $\{a_n\}$ and $\{b_n\}$, we write $a_n \sim b_n$ if $\lim_{n\to\infty} a_n/b_n = 1$, and $a_n \asymp b_n$ if $\lim_{n\to\infty} a_n/b_n \in (-\infty, \infty)\backslash\{0\}$; for random sequences, $a_n \lesssim_p b_n$ denotes $a_n = O_p(b_n)$. $X_n \rightsquigarrow X$ denotes weak convergence in the sense of van der Vaart and Wellner (1996), and $X_n \rightsquigarrow_{\mathrm{fdd}} X$ the convergence of finite-dimensional distributions. For a metric space (Q, d), $\ell_\infty(Q)$ denotes the space of uniformly bounded functions on Q, equipped with the topology of uniform convergence; while $\ell_{\mathrm{ucc}}(Q)$ denotes the space of functions that are uniformly bounded on compact subsets of Q, and is equipped with the topology of uniform convergence on compacta. For $p \geq 1$, X a random variable, and $f : \mathbb{R} \to \mathbb{R}$, $\|X\|_p := (\mathbb{E}|X|^p)^{1/p}$, and $\|f\|_p := (\int_{\mathbb{R}} |f|^p)^{1/p}$. BI denotes the space of bounded and Lebesgue integrable functions on $\mathbb{R}$. $\lfloor \cdot \rfloor$ and $\lceil \cdot \rceil$, respectively, denote the floor and ceiling functions. C denotes a generic constant that may take different values even at different places in the same proof; $a \lesssim b$ denotes $a \leq Cb$.

https://doi.org/10.1017/S0266466616000451 Published online by Cambridge University Press

2. DISCUSSION OF RESULTS

2.1. Model and Assumptions

We are concerned with the estimation of m_0 in the model,

$$y_t = m_0(x_t) + u_t, \tag{2.1}$$

where (x_t, u_t) satisfies

Assumption 1.

(i) $\{x_t\}$ is generated according to

$$x_t := \sum_{s=1}^{t} v_s \qquad\qquad v_t := \sum_{k=0}^{\infty} \phi_k \epsilon_{t-k},$$

and either

(a) $\alpha \in (1,2]$, $\sum_{k=0}^{\infty} |\phi_k| < \infty$ and $\phi := \sum_{k=0}^{\infty} \phi_k \neq 0$; or

$\phi_k \sim k^{H-1-1/\alpha} \pi_k$ for some $\{\pi_k\}_{k\geq 0}$ strictly positive and slowly varying at infinity, with

(b) $H > 1/\alpha$; or

(c) $H < 1/\alpha$ and $\sum_{k=0}^{\infty} \phi_k = 0$.

In both cases (b) and (c), $H \in (\frac{1}{3}, 1)$.

(ii) $\{u_t\}$, the regression disturbance, is the linear process

$$u_t := \sum_{k=0}^{\infty} \theta_k \eta_{t-k}$$

with $\sum_{k=0}^{\infty} |\theta_k| k^{7/6} < \infty$.

(iii) $\{(\epsilon_t, \eta_t)\}$ is a bivariate i.i.d. sequence. ϵ_0 lies in the domain of attraction of a strictly stable distribution with index $\alpha \in (0,2]$, and has characteristic function $\psi(\lambda) := \mathbb{E}e^{i\lambda\epsilon_0}$ satisfying $\psi \in L^{p_0}$ for some $p_0 \geq 1$. $\mathbb{E}\eta_0 = 0$, $\mathbb{E}|\epsilon_0 \eta_0| < \infty$, and $\mathbb{E}|\eta_0|^{q_0} < \infty$ for some $q_0 > 2$.

Remark 2.1. The preceding conditions may be compared with those imposed by Wang and Phillips (2009b, 2016); but quite unlike those authors, we do *not* require $\mathbb{E}\epsilon_0^2 < \infty$. Although our assumptions are consistent with substantial departures from the standard unit root model—which here coincides with (i)(a) with $\alpha = 2$—$\{x_t\}$ is in all cases a partial sum process, and this feature of the generating mechanism identifies (2.1) as a *nonlinear cointegrating regression*, in the terminology of Park and Phillips (2001). To allow the alternative forms of (i) to be more concisely referenced, we shall regard part (a) as corresponding to the case where $H = 1/\alpha$.

The arguments used in this paper could be adapted to derive our main results when $H \in (0, \frac{1}{3}]$. However, for H falling within this range, certain

https://doi.org/10.1017/S0266466616000451 Published online by Cambridge University Press

simplifications—resulting, in particular, from parts (i) and (ii) of Lemma 3.4 below—are unavailable, and the statement of our results would take a more complicated form. To keep this paper to a reasonable length, we have therefore restricted to $H \in (\frac{1}{3}, 1)$. (This restriction is also necessary for certain related results in the literature: in particular, see Theorems 4 and 5 in Jeganathan, 2008.)

Remark 2.2. Part (ii) permits the regression disturbance to be serially dependent, and cross-correlated with the regressor; (2.1) is thus a *structural* model. Wang and Phillips (2016) allow $\{u_t\}$ to be generated in a slightly more general manner, according to

$$u_t = \sum_{k=0}^{\infty} [\theta_k \eta_{t-k} + \vartheta_k g(\epsilon_{t-k})] =: u_{1t} + u_{2t}.$$

By considering separately the cases in which $u_t = u_{1t}$ and $u_t = u_{2t}$, it is easily seen that Theorems 2.1 and 2.2 below hold, without any modification, when $u_t = u_{1t} + u_{2t}$, provided that $\sum_{k=0}^{\infty} |\vartheta_k| k^{7/6} < \infty$ and g is such that $\mathbb{E}g^2(\epsilon_0) < \infty$.

Instead of (ii), we might have required $\{u_t, \tilde{\mathcal{F}}_{t+1}\}$ to be a martingale difference sequence (m.d.s.), where $\tilde{\mathcal{F}}_t := \sigma(\{x_s, u_{s-1}\}_{s \le t})$ (see, e.g., Park and Phillips, 2001, Ass. 2.1); we say that $\{u_t\}$ is an *exogenous m.d.s.* in this case. While this alternative assumption would be very convenient, it seems rather restrictive in the setting of such an 'under-specified' model as (2.1), in which any short-run dynamics affecting the relationship between y_t and x_t must be absorbed into u_t.

Remark 2.3. The requirement that $\sum_{k=0}^{\infty} |\theta_k| k^{7/6} < \infty$ is stronger is necessary to ensure the pointwise consistency and asymptotic normality of kernel regression estimators in this setting: see for example Wang and Phillips (2016), who merely assume that $\sum_{k=0}^{\infty} |\theta_k| k^{1/4} < \infty$. Were (ii) to be relaxed in this direction, then the arguments used to prove the main results of this paper could still be applied, but the rates of convergence obtained would be complicated by the presence of an additional term, the magnitude of which would depend, in a somewhat complicated manner, on the rate at which $\theta_k \to 0$ as $k \to \infty$.

We shall consider both local level (Nadaraya–Watson) and local linear estimators of m_0. For a given $x \in \mathbb{R}$, the local level estimator is defined as

$$\hat{m}(x) := \frac{\sum_{t=1}^{n} K_h(x_t - x) y_t}{\sum_{t=1}^{n} K_h(x_t - x)},$$

while the local linear estimator $\hat{m}_L(x)$ corresponds to the β_0 that solves

$$\min_{(\beta_0, \beta_1)} \sum_{t=1}^{n} K_h(x_t - x)(y_t - \beta_0 - \beta_1(x_t - x))^2.$$

Here $K_h(u) := h^{-1} K(h^{-1} u)$, for K a smoothing kernel and $h = h_n$ a bandwidth parameter, whose properties are specified in Assumption 3 below. To facilitate nonparametric estimation, we require

https://doi.org/10.1017/S0266466616000451 Published online by Cambridge University Press

Assumption 2. m_0 is twice continuously differentiable.

Remark 2.4. Assumption 2 is stronger than is necessary for a determination of the convergence rates of these estimators; our arguments would also permit a derivation of these rates when Assumption 2 is relaxed to the Hölder continuity of m_0 or of its derivatives. We have refrained from doing so only in order to permit our convergence rates to be concisely stated.

As noted above, both estimators involve the use of a smoothing kernel K, and a bandwidth sequence $\{h_n\}$. In order to state our assumptions on K, let BIL denote the set of bounded, integrable, and Lipschitz continuous functions on $\mathbb{R}$, and recall that $\{e_n\}$ is the norming sequence that appears in (1.2) above (see also (2.3) below).

Assumption 3.

(i) $K \in \text{BIL}$ is compactly supported, with $\int K = 1$ and $\int x K(x)\,\mathrm{d}x = 0$.

(ii) $h_n \in \mathscr{H}_n := [\underline{h}_n, \overline{h}]$ with probability approaching 1 (w.p.a.1), where $\overline{h} < \infty$, and $\underline{h}_n^{-1} \lesssim e_n n^{-2r_0}$ for some $r_0 > 0$.

Remark 2.5. The Lipschitz continuity and compact support of K ease some of our arguments, but are certainly not necessary for the most fundamental results that underpin our derivations. For example, Theorem 3.1 in Duffy (2016) requires only that K have one-sided Lipschitz approximants, a rather weak condition that *is* consistent with the presence of simple discontinuities.

Remark 2.6. An important feature of the present work, relative to the preceding literature on nonlinear cointegration, is that we permit the bandwidth sequence to be random, and thus data-dependent, requiring only that it take values lying in the (growing) interval $[\underline{h}_n, \overline{h}_n]$, w.p.a.1. This is of considerable utility in applications, where h_n will typically be chosen with at least some reference to the sample at hand, making the assumption that $\{h_n\}$ is a 'given' deterministic sequence quite unrealistic. In the i.i.d. regressor case, results of this kind are given in Einmahl and Mason (2005). Note that restricting to deterministic bandwidth sequences would *not* help us to obtain better rates of convergence than those that are given in Theorem 2.2 below.

Assumptions 1–3 are maintained throughout the paper, even when no explicit reference is made to them. We shall treat the parameters (including H and α) describing the data generating mechanism as 'fixed', ignoring the dependence of any constants on these.

2.2. Asymptotic Behaviour of the Regressor Process

Before proceeding to an account of our main results, we describe the limiting behaviour of the standardised regressor process $X_n(r) := d_n^{-1} x_{\lfloor nr \rfloor}$ that is entailed by our assumptions, and which is fundamental to our results. (The required norming sequence $\{d_n\}$ is given in (2.3) below.)

https://doi.org/10.1017/S0266466616000451 Published online by Cambridge University Press

Part (iii) of Assumption 1 implies that there exists a slowly varying sequence $\{\varrho_k\}$ such that

$$\frac{1}{n^{1/\alpha}\varrho_n}\sum_{t=1}^{\lfloor nr\rfloor}\epsilon_t \rightsquigarrow_{\text{fdd}} Z_\alpha(r),$$

where Z_α denotes an *α-stable Lévy motion* on $\mathbb{R}$, with $Z_\alpha(0)=0$. That is, the increments of Z_α are stationary, and for any $r_1<r_2$ the characteristic function of $Z_\alpha(r_2)-Z_\alpha(r_1)$ has the logarithm

$$-(r_2-r_1)c|\lambda|^\alpha\left[1-\mathrm{i}\beta\tan\left(\frac{\pi\alpha}{2}\right)\right],$$

where $\beta\in[-1,1]$ and $c>0$; following Jeganathan (2004, p. 1773), we impose the further restriction that $\beta=0$ when $\alpha=1$. We shall also require that $\{\varrho_k\}$ be chosen such that $c=1$ here, which provides a convenient normalisation for the scale of Z_α. (Thus when $\alpha=2$, Z_α corresponds to a Brownian motion with variance 2.) Let X denote the *linear fractional stable motion* (LFSM)

$$X(r):=\int_0^r (r-s)^{H-1/\alpha}\,\mathrm{d}Z_\alpha(s)+\int_{-\infty}^0\left[(r-s)^{H-1/\alpha}-(-s)^{H-1/\alpha}\right]\mathrm{d}Z_\alpha(s),$$

with the convention that $X=Z_\alpha$ when $H=1/\alpha$. (See Samorodnitsky and Taqqu, 1994, for a detailed discussion of the LFSM; note that when $\alpha=2$, X is a fractional Brownian motion.) Associated to X is the occupation density (*local time*) process $\mathcal{L}:=\{\mathcal{L}(a)\}_{a\in\mathbb{R}}$, a process which, almost surely, has continuous paths and satisfies

$$\int_{\mathbb{R}} f(x)\mathcal{L}(x)\,\mathrm{d}x=\int_0^1 f(X(r))\,\mathrm{d}r,\qquad \forall f \text{ bounded, measurable.} \tag{2.2}$$

(See Theorem 0 in Jeganathan, 2004.)

Now let $\{c_k\}$ denote a sequence with $c_0=1$ and

$$c_k=\begin{cases}\phi & \text{if } H=1/\alpha,\\ |H-1/\alpha|^{-1}k^{H-1/\alpha}\pi_k & \text{otherwise.}\end{cases}$$

By Karamata's theorem (Bingham, Goldie, and Teugels, 1987, Thm. 1.5.11), $\sum_{l=0}^k\phi_k\sim c_k$ as $k\to\infty$. Set

$$d_k:=k^{1/\alpha}c_k\varrho_k \qquad\qquad e_k:=kd_k^{-1}, \tag{2.3}$$

and note that the sequences $\{c_k\}$, $\{d_k\}$ and $\{e_k\}$ are regularly varying with indices $H-1/\alpha$, H, and $1-H$, respectively. The following is a special case of Proposition 2.1 and Theorem 3.1 in Duffy (2016).

PROPOSITION 2.1. *For every $f\in$ BIL,*

$$\mathcal{L}_n^f(a):=\frac{1}{e_nh_n}\sum_{t=1}^n f\left(\frac{x_t-d_na}{h_n}\right)\rightsquigarrow\mathcal{L}(a)\int_{\mathbb{R}} f$$

on $\ell_{\text{ucc}}(\mathbb{R})$, jointly with $X_n(r):=d_n^{-1}x_{\lfloor nr\rfloor}\rightsquigarrow_{\text{fdd}}X(r)$.

https://doi.org/10.1017/S0266466616000451 Published online by Cambridge University Press

2.3. Order Estimates

Our rates of convergence will be obtained with the aid of the following order estimates, for the covariance and zero energy processes, respectively, which appear to be new in the literature.

THEOREM 2.1. *Suppose* $f \in \mathrm{BIL}$. *Then*

$$\frac{1}{(e_n h_n)^{1/2}} \sup_{a\in\mathbb{R}} \left| \sum_{t=1}^{n} f\left(\frac{x_t - d_n a}{h_n}\right) u_t \right| \lesssim_p (1 + n^{1/q_0 - r_0}) \log n; \tag{2.4}$$

and if additionally $\int_{\mathbb{R}} f = 0$ *and* $\int |f(x)x| \,\mathrm{d}x < \infty$,

$$\frac{1}{(e_n h_n)^{1/2}} \sup_{a\in\mathbb{R}} \left| \sum_{t=1}^{n} f\left(\frac{x_t - d_n a}{h_n}\right) \right| \lesssim_p \log n. \tag{2.5}$$

See Section 5 for the proof. In both cases the assumed smoothness of f permits the supremum over $\mathbb{R}$ to be effectively reduced to a maximum over a sequence of finite sets, $\{\mathscr{F}_n\}$. For the zero energy process, the requisite bound over $\mathscr{F}_n$ is provided (essentially) by Proposition 4.2 in Duffy (2016). It turns out that a counterpart of this result is available for the covariance process, but its application requires that a truncation be first applied to $\{u_t\}$. In order to state the result, let $\eta_t^{(\le)}$ denote an appropriately truncated (and centred) version of η_t (see (5.1) below), such that $\mathbb{E}\eta_t^{(\le)} = 0$ and $\|\eta\|_n := \|\eta_0^{(\le)}\|_\infty < \infty$. For $u_t^{(\le)} := \sum_{k=0}^{n} \theta_k \eta_{t-k}^{(\le)}$, define

$$\mathcal{S}_n f := \sum_{t=1}^{n} f(x_t) u_t^{(\le)}, \tag{2.6}$$

and for $\mathscr{F} \subset \mathrm{BI}$, set

$$\delta_n(\mathscr{F}) := \|\eta\|_n \|\mathscr{F}\|_\infty + [\|\eta\|_n + e_n^{1/2}](\|\mathscr{F}\|_1 + \|\mathscr{F}\|_2), \tag{2.7}$$

where $\|\mathscr{F}\| := \sup_{f\in\mathscr{F}} \|f\|$.

PROPOSITION 2.2. *Suppose* $\mathscr{F}_n \subset \mathrm{BI}$ *with* $\#\mathscr{F}_n \lesssim n^C$. *Then*

$$\max_{f\in\mathscr{F}_n} |\mathcal{S}_n f| \lesssim_p \delta_n(\mathscr{F}_n) \log n.$$

The proof of this result appears in Section 4.

Remark 2.7. If $\{u_t\}$ is an exogenous m.d.s., then $\{f(x_t)u_t\}$ itself forms a m.d.s., and an application of Freedman's (1975, Thm. 1.6) inequality permits the $\log n$ factor on the right side of (2.4) to be reduced to $\log^{1/2} n$; see Wang and Chan (2014, Thm. 2.1). In the present case, however, $\{f(x_t)u_t\}$ is not a m.d.s., precluding a direct application of such a result. Instead, we shall rely on the combination of a suitable subgaussian tail inequality for martingales (Bercu and Touati, 2008, Thm. 2.1)—which implies Lemma 3.6 below—and a martingale decomposition of $\sum_{t=1}^{n} f(x_t)u_t$.

https://doi.org/10.1017/S0266466616000451 Published online by Cambridge University Press

2.4. Rates of Uniform Convergence

The essential features of the problem become apparent when we consider the local level estimator, which admits the decomposition

$$\hat{m}(x) - m_0(x) = \frac{\sum_{t=1}^n K_{h_n}(x_t - x)[m_0(x_t) - m_0(x)]}{\sum_{t=1}^n K_{h_n}(x_t - x)} + \frac{\sum_{t=1}^n K_{h_n}(x_t - x)u_t}{\sum_{t=1}^n K_{h_n}(x_t - x)} =: \frac{\Psi_{1n}(x)}{\Psi_{3n}(x)} + \frac{\Psi_{2n}(x)}{\Psi_{3n}(x)}.$$

We shall now examine each of Ψ_{1n}, Ψ_{2n}, and Ψ_{3n}, in turn: compared with the stationary regressor case, the treatment of the denominator poses some unique challenges here, and so we turn to it first.

Denominator. Set $\mathcal{L}_n(a) := e_n^{-1} \sum_{t=1}^n K_{h_n}(x_t - d_n a)$, and define

$$A_n^\varepsilon := \{x \in \mathbb{R} \mid \mathcal{L}_n(d_n^{-1}x) \geq \varepsilon\}.$$

Noting the different standardisations of $\mathcal{L}_n$ and Ψ_{3n}, we see that

$$\sup_{x \in A_n^\varepsilon} \Psi_{3n}^{-1}(x) = \left(\inf_{x \in A_n^\varepsilon} \Psi_{3n}(x) \right)^{-1} \leq \varepsilon^{-1} e_n^{-1}. \tag{2.8}$$

Thus A_n^ε describes a subset of $\mathbb{R}$—which depends on the trajectory of $\{x_t\}_{t=1}^n$—on which the order of $\Psi_{3n}^{-1}(x)$ may be uniformly controlled. Importantly, $\varepsilon > 0$ may be chosen (sufficiently small) such that A_n^ε contains as large a fraction of the sample as desired, in the limit as $n \to \infty$, in the sense that

$$\limsup_{n \to \infty} \mathbb{P}\left\{ \frac{1}{n} \sum_{t=1}^n \mathbf{1}\{x_t \notin A_n^\varepsilon\} \geq \delta \right\} \leq \delta \tag{2.9}$$

for any given $\delta > 0$: see the arguments used to verify (5.4) in Duffy (2016).

In general, A_n^ε will be a union of disjoint (closed) intervals, even for large n. This is necessarily the case when $H < 1/\alpha$, since in this case X has discontinuous sample paths (Samorodnitsky and Taqqu, 1994, Example 10.2.5), and so the support of $\mathcal{L}$, which is contained in the range of X, will typically contain gaps. However, in the special case where $\mathbb{E}\epsilon_0^2 < \infty$ (implying $\alpha = 2$) and $H = \frac{1}{2}$, it is possible to replace A_n^ε by a sequence of (connected) intervals. In order to state a result to this effect, let $R_n^\varepsilon := [x_{(1)}, x_{(n)}]_\varepsilon$, where $[a, b]_\varepsilon := [(1-\varepsilon)a, (1-\varepsilon)b]$ and let $\{x_{(i)}\}_{i=1}^n$ denote the order statistics of the sample $\{x_t\}_{t=1}^n$.

PROPOSITION 2.3. *Suppose $H = \frac{1}{2}$ and $\mathbb{E}\epsilon_0^2 < \infty$, and let $\delta > 0$ be given. Then for every $\varepsilon > 0$, (2.8) holds with R_n^ε in place of A_n^ε; and for $\varepsilon > 0$ sufficiently small, (2.9) holds with R_n^ε in place of A_n^ε.*

Remark 2.8. This result is essentially a consequence of Ray's (1963) theorem, which implies that the local time $\mathcal{L}$ of a diffusion J is strictly positive on the interior of the range of J. In the setting of the present paper, this seems to be

https://doi.org/10.1017/S0266466616000451 Published online by Cambridge University Press

applicable only in the case where X is a Brownian motion. However, it also applies in the important case where

$$X(r) = \int_0^r \mathrm{e}^{\kappa(r-s)}\,\mathrm{d}B(s)$$

for $\kappa \in \mathbb{R}$ and B a Brownian motion. Such a process arises as the weak limit of X_n, under the hypotheses of Proposition 2.3, if $\{x_t\}$ is generated according to $x_t = x_{t-1} + \rho_n v_t$, where $\rho_n = 1 + \frac{\kappa}{n}$: see Wang and Phillips (2009b). (To extend our results to this case would require very little modification to our arguments; indeed, we explicitly considered such a data generating mechanism in an earlier version of this paper.) Whether such a characterisation of the support of $\mathcal{L}$ is available in other cases where X has continuous sample paths, most notably when $\alpha = 2$ and $H \neq 1/2$—i.e., when X is a fractional Brownian motion—seems to be an open question.

Remark 2.9. In view of (2.9), the volumes of both A_n^ε and R_n^ε must expand, probabilistically, at rate d_n as $n \to \infty$. However, even under the hypotheses of Proposition 2.3, A_n^ε could *not* be replaced by a sequence of *deterministic* intervals $[-a_n, a_n]$ whose endpoints diverge at rate d_n. Indeed, suppose that $a_n = C_0 n^{1/2}$ for some $C_0 > 0$. Then by $X_n \rightsquigarrow X$ and the reflection principle (Revuz and Yor, 1999, Prop. III.3.7),

$$\mathbb{P}\left\{\max_{1\le t\le n} x_t \le \frac{C_0 n^{1/2}}{2}\right\} \to \mathbb{P}\left\{\sup_{r\in[0,1]} X(r) \le \frac{C_0}{2}\right\} = 1 - 2\mathbb{P}\left\{X(1) > \frac{C_0}{2}\right\}$$
$$= 2\Phi_{\mathfrak{g}}\left(\frac{C_0}{2}\right) - 1 > 0$$

for every $C_0 > 0$, no matter how small; here $\Phi_{\mathfrak{g}}$ denotes the standard normal c.d.f. With nonzero probability, $\{x_t\}_{t=1}^n$ *never* visits $[\frac{1}{2}C_0 n^{1/2}, C_0 n^{1/2}]$, and so the signal is forever negligible within this range. This accounts for why earlier work on this problem (e.g., Wang and Wang, 2013; Chan and Wang, 2014; Gao et al., 2015), which considered deterministic intervals of this form, has been restricted to domains whose volume grows at a rate strictly slower than d_n, which necessarily contain a vanishingly small fraction of the observed $\{x_t\}_{t=1}^n$ as $n \to \infty$.

Numerator. To provide a measure of the 'regularity' of m_0 over a given domain, we associate to m_0 the mappings $\overline{m}_1, \overline{m}_2 : \mathcal{P}(\mathbb{R}) \to \mathbb{R}_+ \cup \{\infty\}$, defined by

$$\overline{m}_1(A) := \sup_{x\in A} |m_0'(x)| \qquad\qquad \overline{m}_2(A) := \sup_{x\in A} |m_0''(x)|.$$

Let $\tilde{A}_n^\varepsilon := \{x \in \mathbb{R} \mid d(x, A_n^\varepsilon) \le c_K h\}$, where c_K is chosen such that the support of K is contained in $[-c_K, c_K]$. Then a Taylor series expansion of m_0 around x_t, for each $t \in \{1, \dots, n\}$, yields the estimate

$$\sup_{x\in A_n^\varepsilon} |\Psi_{1n}| \le h_n \overline{m}_1(\tilde{A}_n^\varepsilon) \sup_{x\in\mathbb{R}} \left|\sum_{t=1}^n K_{h_n}^{[1]}(x_t - x)\right| + h_n^2 \overline{m}_2(\tilde{A}_n^\varepsilon) \sup_{x\in\mathbb{R}} \sum_{t=1}^n |K_{h_n}^{[2]}(x_t - x)|,$$

https://doi.org/10.1017/S0266466616000451 Published online by Cambridge University Press

where $f^{[p]}$ denotes $x \mapsto x^p f(x)$. Applying Theorem 2.1 and Proposition 2.1 to the first and second terms on the right, respectively, then gives

$$\frac{1}{e_n} \sup_{x \in A_n^\varepsilon} |\Psi_{1n}| \lesssim_p h_n^2 \left[\overline{m}_1(\tilde{A}_n^\varepsilon) \frac{\log n}{h_n^{3/2} e_n^{1/2}} + \overline{m}_2(\tilde{A}_n^\varepsilon) \right]. \tag{2.10}$$

Similarly, we have by Theorem 2.1 that

$$\frac{1}{e_n} \sup_{x \in A_n^\varepsilon} |\Psi_{2n}| \lesssim_p (1 + n^{1/q_0 - r_0}) \frac{\log n}{(e_n h_n)^{1/2}}. \tag{2.11}$$

In view of (2.8), the rate at which $\hat{m}(x)$ converges uniformly to $m_0(x)$ on A_n^ε is given by the sum of the right sides of (2.10) and (2.11). In giving a formal statement of our results below, we assume that q_0 and r_0 are such that $n^{1/q_0 - r_0} \lesssim 1$, so that the right side of (2.11) takes a simplified form.

THEOREM 2.2. *Suppose* $r_0 \geq q_0^{-1}$. *Then for every* $\varepsilon > 0$,

$$\varepsilon \cdot \sup_{x \in A_n^\varepsilon} |\hat{m}(x) - m_0(x)| \lesssim_p h_n^2 \left[\overline{m}_1(\tilde{A}_n^\varepsilon) \frac{\log n}{h_n^{3/2} e_n^{1/2}} + \overline{m}_2(\tilde{A}_n^\varepsilon) \right] + \frac{\log n}{(e_n h_n)^{1/2}} \tag{2.12}$$

and

$$\varepsilon \cdot \sup_{x \in A_n^\varepsilon} |\hat{m}_L(x) - m_0(x)| \lesssim_p h_n^2 \overline{m}_2(\tilde{A}_n^\varepsilon) + \frac{\log n}{(e_n h_n)^{1/2}}. \tag{2.13}$$

Remark 2.10. These uniform convergence rates agree almost exactly with their pointwise counterparts (see, e.g., Wang and Phillips, 2011), except for

(i) the presence of the $\log n$ factors; and

(ii) the dependence of the bias terms on the *suprema* of m_0' and m_0'' over $\tilde{A}_n^\varepsilon$.

Thus, as the support of $\{x_t\}$—to which A_n^ε is an approximation—expands over the line, these 'uniform' bias terms may shrink less rapidly than their pointwise counterparts, depending on the tail behaviour of the derivatives of m_0.

Remark 2.11. After the manuscript of this paper had been completed, we obtained a copy of an unpublished manuscript by Liu, Chan, and Wang (2014), who determine the uniform rate of convergence of $\hat{m}_L$ when $\{x_t\}$ is exogenous ($\{u_t\}$ is a heteroskedastic m.d.s.) and $\{h_n\}$ is a deterministic sequence. (Rather than a sequence of random domains such as $\{A_n^\varepsilon\}$, they consider only a deterministic sequence of intervals, with the consequences discussed in Remark 2.9 above.) Due to the assumed exogeneity, the $\log n$ factor appearing in the second terms on the right sides of (2.12) and (2.13) can be improved to $\log^{1/2} n$, as per Remark 2.7 above. The presence of endogeneity would thus seem to penalise the rate of convergence of these estimators by at worst a factor of $\log^{1/2} n$.

Underpinning these authors' derivations is an analogue of our Proposition 2.1 (their Theorem 2.1), which is worked out under quite different assumptions on the regressor process than are imposed here. In this regard, we may note particularly

https://doi.org/10.1017/S0266466616000451 Published online by Cambridge University Press

their requirement that there exist a sequence of processes $\{X_n^*\}$ with $X_n^* =_d X$, and a $\delta > 0$ such that

$$\sup_{r\in[0,1]} |X_n(r) - X_n^*(r)| = o_{\text{a.s.}}(n^{-\delta}),$$

a condition which excludes a large portion of the processes considered in this paper, for which merely $X_n \rightsquigarrow_{\text{fdd}} X$ is available (this is particularly true when $H < 1/\alpha$ and $\alpha \in (0,2)$, since in this case the sample paths of X are unbounded: see Samorodnitsky and Taqqu, 1994, Example 10.2.5). On the other hand, our results do *not* subsume those of Liu et al. (2014), since those authors do not require $\{v_t\}$ to be a linear process; there is thus only a partial overlap between the class of processes considered in this paper, and in theirs.

Remark 2.12. Provided that $\varepsilon > 0$ is fixed, and $\overline{m}_1$ and $\overline{m}_2$ are bounded on $\mathbb{R}$—which is perfectly consistent with linear or sublinear growth in the tails of m_0—Theorem 2.2 implies that requiring convergence on a domain almost as large as the range of $\{x_t\}$ does not penalise the convergence *rate* of either estimator, relative to the rate that could be proved on an interval of fixed width. This might seem to contrast markedly with the situation when $\{x_t\}$ is stationary, where necessarily slower rates of convergence hold on domains that expand with the sample size, as the estimator is pushed into regions where $\{x_t\}$ has a progressively smaller density (see, e.g., Hansen, 2008, Thm. 8; Kristensen, 2009, Thm. 1; Li, Lu, and Linton, 2012, Thm. 2.1). However, this phenomenon would re-emerge here if we were to let $\varepsilon = \varepsilon_n \to 0$ as $n \to \infty$: indeed, it is immediate from (2.12) and (2.13) that the rates of uniform convergence of our estimators, over the domains $\{A_n^{\varepsilon_n}\}$, would be slowed by a factor of ε_n^{-1} in this case.

Remark 2.13. In a recent paper, Chan and Wang (2014) argue that while both $\hat{m}$ and $\hat{m}_L$ enjoy similar pointwise bias properties, the latter enjoys markedly better performance than the former, so far as the uniform behaviour of their respective bias terms is concerned. This conclusion is *partly* borne out by Theorem 2.2; but the improved order estimate obtained here for the linear bias term (the first element on the right side of (2.12)) indicates that this judgement may need qualification. In particular, if both $\overline{m}_1(A_n^\varepsilon)$ and $\overline{m}_2(\tilde{A}_n^\varepsilon)$ are of a comparable magnitude, then $e_n h_n^3 \log^{-2} n \xrightarrow{p} \infty$ will ensure that the second order bias (the second term on the right side of (2.12)) dominates the linear bias term; this is scarcely less restrictive than the condition that $e_n h_n^3 \xrightarrow{p} \infty$ that is required for this conclusion when only the *pointwise* performance of these estimators is in issue (see Wang and Phillips, 2011).

See Section 6 for the proofs of Proposition 2.3 and Theorem 2.2.

2.5. An Alternative Perspective on Our Results

There is another way of viewing the estimation problem considered in this paper, which highlights the connections between our results, and those which are

https://doi.org/10.1017/S0266466616000451 Published online by Cambridge University Press

obtained when $\{x_t\}$ is stationary. Defining a sequence of regression functions $m_n(x) := m_0(d_n x)$, the model (2.1) can be rewritten as

$$y_t = m_0(x_t) + u_t = m_n(d_n^{-1}x_t) + u_t = m_n(x_{nt}) + u_t,$$

where $x_{nt} := d_n^{-1}x_t$. Taking $b_n = d_n^{-1}h_n$, we see that

$$\frac{1}{nb_n}\sum_{t=1}^{n} K\left(\frac{x_{nt}-x}{b_n}\right) = \frac{1}{e_n h_n}\sum_{t=1}^{n} K\left(\frac{x_t - d_n x}{h_n}\right) \rightsquigarrow \mathcal{L}(x) \tag{2.14}$$

in $\ell_\infty(\mathbb{R})$ by Proposition 2.1. In light of this, we might regard the $\{x_{nt}\}$'s as being drawn from a spatial distribution with marginal density $\mathcal{L}(x)$—just as stationary regressors would be drawn from a distribution with marginal density $p(x)$. (Restricting ourselves to A_n^ε or R_n^ε yields a domain on which the density $\mathcal{L}$ can be bounded away from 0, which is equally desirable in the stationary regressor case.)

Now suppose that almost nothing is known about $\{x_{nt}\}$, beyond the fact that a convergence result of the same kind as (2.14) holds (together with some knowledge of the support of $\mathcal{L}$). Is this sufficient to determine the rate at which the local linear estimator of m_n, computed from $\{y_t, x_{nt}\}_{t=1}^n$, converges uniformly to m_n? If $\{u_t\}$ is an exogenous m.d.s., then this is indeed the case. Supposing

$$\sup_{x\in\mathbb{R}} |m_n''(x)| \leq \alpha_n,$$

then as per Remark 2.11, we would have

$$\sup_{x\in A} |\hat{m}_L(x) - m_n(x)| \lesssim_p b_n^2\alpha_n + \frac{\log^{1/2} n}{n^{1/2}b_n^{1/2}} = \alpha_n^{1/5}\frac{\log^{2/5} n}{n^{2/5}}, \tag{2.15}$$

for any set A on which $\mathcal{L}$ can be bounded away from zero; the final equality follows if b_n is chosen so as to balance the order of bias and variance terms.

The rate of convergence thus depends only on how the 'complexity' of m_n—as measured here by α_n—varies with n. In the stationary setting, m_n is (typically) a *fixed* function, and so $\alpha_n = \alpha_0$, a constant. In this case, the right side of (2.15) agrees precisely with Stone's (1982) minimax optimal rate for twice-continuously differentiable functions. On the other hand, when $\{x_t\}$ is integrated (or near-integrated), the manner in which m_n is constructed from a *fixed* m_0 gives $\alpha_n = \alpha_0 d_n^2$, and the best obtainable rate is $O_p(e_n^{-2/5}\log^{2/5} n)$, as per Chan and Wang (2014). This leads us to believe that the minimax optimality properties of local polynomial regression, for the estimation of functions belonging to Hölder classes, should extend quite straightforwardly to the case of an integrated regressor.

It remains to be seen how this analogy might be further extended to the case where $\{x_{nt}\}$ is endogenous. Heuristically, estimation of m_n must be possible, in the integrated case, because the joint dependence between u_t and x_{nt} becomes progressively weaker, as $n, t \to \infty$. Although it is not immediately clear how this notion should be made precise—let alone what would be a suitable analogue of it in the stationary setting—the 'location shift' model of Phillips and Su (2011) may be counted as an important effort in this direction.

https://doi.org/10.1017/S0266466616000451 Published online by Cambridge University Press

3. PRELIMINARIES

Preliminary to the proofs of our main results, this section collects some auxiliary lemmas, proofs of which are given in the Supplementary Material.[2] We shall rely heavily on the use of the inverse Fourier transform to analyse objects of the form $\mathbb{E}_t f(x_{t+k})$, similarly to Borodin and Ibragimov (1995), Jeganathan (2004, 2008), and Wang and Phillips (2009b, 2011). The following result permits the use of the 'usual' inversion formula, even in cases where $\hat{f} \notin L^1$, for $\hat{f}(\lambda) := \int f(x)\mathrm{e}^{\mathrm{i}\lambda x}\,\mathrm{d}x$.

LEMMA 3.1. *Suppose $Y = Y_1 + Y_2$, where Y_1 is independent of (Y_2, Z), and Y_i has integrable characteristic function ψ_{Y_i}. Then, for every $f \in \mathrm{BI}$, $y_0 \in \mathbb{R}$, and $\mathbb{E}|g(Z)| < \infty$,*

$$\mathbb{E}f(y_0+Y)g(Z) = \frac{1}{2\pi}\int_{\mathbb{R}} \hat{f}(\lambda)\mathrm{e}^{-\mathrm{i}\lambda y_0}\mathbb{E}[\mathrm{e}^{-\mathrm{i}\lambda Y}g(Z)]\,\mathrm{d}\lambda. \tag{3.1}$$

Let $\mathcal{F}_s^t := \sigma(\{\epsilon_r\}_{r=s}^t)$, noting that $\mathcal{F}_{s_1}^{s_2}$ and $\mathcal{F}_{s_3}^{s_4}$ are independent whenever $s_1 \leq s_2 < s_3 \leq s_4$. We shall have frequent recourse to the following decomposition,

$$\begin{aligned} x_t = \sum_{k=1}^{t} v_t = \sum_{k=1}^{t}\sum_{l=0}^{\infty}\phi_l\epsilon_{k-l} = &\left[\sum_{i=0}^{\infty}\epsilon_{-i}\sum_{j=i+1}^{i+t}\phi_j + \sum_{i=t-s+1}^{t-1}\epsilon_{t-i}\sum_{j=0}^{i}\phi_j\right] \\ &+ \sum_{i=0}^{t-s}\epsilon_{t-i}\sum_{j=0}^{i}\phi_j =: x^*_{s-1,t} + x'_{s,t,t}, \end{aligned} \tag{3.2}$$

for $1 \leq s \leq t$, where $x^*_{s-1,t}$ and $x'_{s,t,t}$ are independent, and $x^*_{s-1,t}$ is $\mathcal{F}_{-\infty}^{s-1}$-measurable. Defining $a_i := \sum_{j=0}^{i}\phi_j$, we may further decompose $x'_{s,t,t}$ as

$$x'_{s,t,t} = \sum_{i=s}^{t} a_{t-i}\epsilon_i = \sum_{i=s}^{r} a_{t-i}\epsilon_i + \sum_{i=r+1}^{t} a_{t-i}\epsilon_i =: x'_{s,r,t} + x'_{r+1,t,t}, \tag{3.3}$$

where $x'_{s,r,t}$ is $\mathcal{F}_s^r$-measurable, and $x'_{r+1,t,t}$ is $\mathcal{F}_{r+1}^t$-measurable. The following property of the coefficients $\{a_i\}$ is particularly important: there exist $0 < \underline{a} \leq \overline{a} < \infty$, and a $k_0 \in \mathbb{N}$ such that

$$\underline{a} \leq \inf_{k_0+1\leq k}\inf_{\lfloor k/2\rfloor \leq l \leq k} c_k^{-1}|a_l| \leq \sup_{k_0+1\leq k}\sup_{\lfloor k/2\rfloor \leq l \leq k} c_k^{-1}|a_l| \leq \overline{a}. \tag{3.4}$$

This is an easy consequence of Karamata's theorem (see Section F of the Supplement to Duffy (2016) for a proof). Throughout the remainder of the paper, k_0 refers to the object of (3.4); it is also implicitly maintained $k_0 \geq 8p_0$ for p_0 as in Assumption 1(iii).

Having decomposed x_t into a sum of independent components, we shall proceed to control such objects as the right side of (3.1) with the aid of the following lemma, which provides bounds on integrals involving the characteristic functions of some of those components of x_t. (This lemma summarises and refines

https://doi.org/10.1017/S0266466616000451 Published online by Cambridge University Press

some of the calculations presented on pp. 15–21 of Jeganathan, 2008.) In order to state this result, we first note that Assumption 1(iii) is equivalent to the statement that

$$\log \psi(\lambda) = -|\lambda|^{\alpha} G(\lambda)\left[1 - \mathrm{i}\beta \tan\left(\frac{\pi\alpha}{2}\right)\right]$$

for all λ in a neighbourhood of the origin, where G is slowly varying at zero (see Ibragimov and Linnik, 1971, Thm. 2.6.5). (Here, as throughout the remainder of this paper, a slowly varying (or regularly varying) function is understood to take only strictly positive values, and have the property that $G(\lambda) = G(|\lambda|)$ for every $\lambda \in \mathbb{R}$.)

LEMMA 3.2. *Let $p \in [0,5]$, $q \in [1,2]$, and $z_1, z_2 \in \mathbb{R}_+$. Then*

(*i*) *there exists a $\gamma_1 > 0$ such that for every $t \geq 0$, $k \geq k_0 + 1$, and $m \in \{0, \ldots, k-1\}$,*

$$\int_{\mathbb{R}} (z_1|\lambda|^p \wedge z_2)|\mathbb{E}\eta_{t+k-m} \mathrm{e}^{-\mathrm{i}\lambda x'_{t+1,t+k,t+k}}|^q \,\mathrm{d}\lambda \lesssim z_1 c_m^q d_k^{-(1+p+q)} + z_2 \mathrm{e}^{-\gamma_1 k}; \tag{3.5}$$

and if $F(u) \asymp G^{p/\alpha}(u)$ as $u \to 0$,

$$\int_{\mathbb{R}} (z_1|a_k|^p|\lambda|^p F(a_k\lambda) \wedge z_2)|\mathbb{E}\eta_{t+k-m} \mathrm{e}^{-\mathrm{i}\lambda x'_{t+1,t+k,t+k}}|^q \,\mathrm{d}\lambda \lesssim z_1 c_m^q k^{-p/\alpha} d_k^{-(1+q)} + z_2 \mathrm{e}^{-\gamma_1 k}; \tag{3.6}$$

(*ii*) *for every $t \geq 1$, $k \geq k_0 + 1$ and $s \in \{k_0 + 1, \ldots, t\}$,*

$$\int_{\mathbb{R}} |\mathbb{E}\mathrm{e}^{-\mathrm{i}\lambda x'_{t-s+1,t-1,t+k}}| \,\mathrm{d}\lambda \lesssim \frac{c_s}{c_{k+s}} d_s^{-1}.$$

We note here, for future reference, that the preceding continues to hold when η_t is replaced by $\eta_t^{(\leq)}$ as defined in (5.1) below. Let $\mathbb{E}_t[\cdot] := \mathbb{E}[\cdot \mid \mathcal{F}_{-\infty}^t]$. The following is an easy consequence of the preceding.

LEMMA 3.3. *Suppose $f \in$ BI. Then*

(*i*) *for every $t \geq 0$ and $k \geq k_0 + 1$,*

$$\mathbb{E}_t|f(x_{t+k})\eta_{t+k-m}| \lesssim d_k^{-1}\|f\|_1 \cdot \begin{cases} 1 & \text{if } m \in \{0, \ldots, k-1\}, \\ |\eta_{t+k-m}| & \text{if } m \geq k; \end{cases}$$

(*ii*) *and, if in addition $m \in \{0, \ldots, k-1\}$,*

$$|\mathbb{E}_t f(x_{t+k})\eta_{t+k-m}| \lesssim c_m d_k^{-2}\|f\|_1.$$

Recall the definitions of $\{d_k\}$ and $\{e_k\}$ given in (2.3) above. The following is a straightforward consequence of Karamata's theorem.

https://doi.org/10.1017/S0266466616000451 Published online by Cambridge University Press

LEMMA 3.4.

(*i*) $\sum_{t=1}^{n} d_t^{-2} \lesssim e_n^{1/2}$;

(*ii*) $\sum_{k=1}^{\infty} k^{-1/2} d_k^{-3/2} < \infty$;

(*iii*) $\sum_{m=0}^{\infty} |\theta_m|(c_m + m^{1/2} e_m) < \infty$.

For the reader's convenience, Lemmas 9.4 and 7.1 from Duffy (2016) are reproduced below; see that paper for the proofs. For the first of these, define

$$\vartheta(z_1, z_2) := \mathbb{E}\big[\mathrm{e}^{-\mathrm{i}z_1\epsilon_0} - \mathbb{E}\mathrm{e}^{-\mathrm{i}z_1\epsilon_0}\big]\big[\mathrm{e}^{-\mathrm{i}z_2\epsilon_0} - \mathbb{E}\mathrm{e}^{-\mathrm{i}z_2\epsilon_0}\big].$$

LEMMA 3.5. *Uniformly over* $z_1, z_2 \in \mathbb{R}$,

$$|\vartheta(z_1, z_2)| \lesssim [|z_1|^{\alpha} \tilde{G}(z_1) \wedge 1]^{1/2} [|z_2|^{\alpha} \tilde{G}(z_2) \wedge 1]^{1/2},$$

where $\tilde{G}(u) \asymp G(u)$ *as* $u \to 0$.

Let $\|\cdot\|_{\tau_1}$ denote the Orlicz norm associated to $\tau_1(x) := \mathrm{e}^x - 1$. (See van der Vaart and Wellner, 1996, p. 95 for the definition of an Orlicz norm.) For a martingale $M := \{M_t\}_{t=0}^n$ with associated filtration $\mathcal{G} := \{\mathcal{G}_t\}_{t=0}^n$, define

$$[M] := \sum_{t=1}^{n} (M_t - M_{t-1})^2 \qquad \langle M \rangle := \sum_{t=1}^{n} \mathbb{E}[(M_t - M_{t-1})^2 \mid \mathcal{G}_{t-1}].$$

We say that M is *initialised at zero* if $M_0 = 0$. The next result is a straightforward consequence of Theorem 2.1 in Bercu and Touati (2008).

LEMMA 3.6. *Let* $\{\Theta_n\}$ *denote a sequence of index sets, and* $\{K_n\}$ *a real sequence such that* $\#\Theta_n + K_n \lesssim n^C$. *Suppose that for each* $n \in \mathbb{N}$, $k \in \{1, \dots, K_n\}$ *and* $\theta \in \Theta_n$, $M_{nk}(\theta)$ *is a martingale, initialised at zero, for which*

$$\omega_{nk}^2 := \max_{\theta \in \Theta_n} \{\|[M_{nk}(\theta)]\|_{\tau_1} \vee \|\langle M_{nk}(\theta)\rangle\|_{\tau_1}\} < \infty.$$

Then

$$\max_{\theta \in \Theta_n} \left| \sum_{k=1}^{K_n} M_{nk}(\theta) \right| \lesssim_p \left(\sum_{k=1}^{K_n} \omega_{nk} \right) \log n.$$

4. CONTROLLING THE TRUNCATED COVARIANCE PROCESS

We turn first to the proof of Proposition 2.2. For this section only, we shall denote $\eta_t^{(\leq)}$ by simply η_t. Then recalling (2.6) above, we may write

$$\mathcal{S}_n f = \sum_{t=1}^{n} f(x_t) u_t^{(\leq)} = \sum_{m=0}^{n} \theta_m \sum_{t=1}^{n} f(x_t) \eta_{t-m} =: \sum_{m=0}^{n} \theta_m \mathcal{S}_{nm} f. \tag{4.1}$$

For each $m \in \{0, \dots, n\}$, by following a procedure identical to that described in Section 7 in Duffy (2016), the process $\mathcal{S}_{nm} f$ may be decomposed as

https://doi.org/10.1017/S0266466616000451 Published online by Cambridge University Press

$$\mathcal{S}_{nm} f = \mathcal{N}_{nm} f + \sum_{k=0}^{n-1} \mathcal{M}_{nmk} f, \tag{4.2}$$

where

$$\mathcal{N}_{nm} f := \sum_{t=1}^{n} \mathbb{E}_0 f(x_t) \eta_{t-m} \qquad \mathcal{M}_{nmk} f := \sum_{t=1}^{n-k} \xi_{mkt} f$$

$$\xi_{mkt} f := \mathbb{E}_t f(x_{t+k}) \eta_{t+k-m} - \mathbb{E}_{t-1} f(x_{t+k}) \eta_{t+k-m},$$

and we have defined $\mathbb{E}_t[\cdot] := \mathbb{E}[\cdot \mid \mathcal{F}^t_{-\infty}]$.

A suitable bound for $\|\mathcal{N}_{nm} f\|_\infty$ is provided by Lemma 3.3(ii). By construction, $\{\xi_{mkt}, \mathcal{F}^t_{-\infty}\}_{t=1}^{n-k}$ forms a martingale difference sequence for each (m,k), and so control over each of the martingale 'pieces' $\mathcal{M}_{nmk} f$ may be obtained via control over

$$\mathcal{U}_{nmk} f := [\mathcal{M}_{nmk} f] = \sum_{t=1}^{n-k} \xi^2_{mkt} f,$$

$$\mathcal{V}_{nmk} f := \langle \mathcal{M}_{nmk} f \rangle = \sum_{t=1}^{n-k} \mathbb{E}_{t-1} \xi^2_{mkt} f,$$

in combination with Lemma 3.6. Defining

$$\varsigma_{nm}(f) := \|\eta\|_n \|f\|_\infty + (e_m \|\eta\|_n + c_m e_n^{1/2}) \|f\|_1 \tag{4.3}$$

and

$$\sigma^2_{nmk}(f) := \begin{cases} \|\eta\|_n^2 \|f\|_\infty^2 + (e_m \|\eta\|_n^2 + e_n) \|f\|_2^2 & \text{if } k \in \{0, \ldots, k_0\}, \\ d_k^{-1} (e_m \|\eta\|_n^2 + e_n) \|f\|_1^2 & \text{if } k \in \{k_0 + 1, \ldots, m\}, \\ (k^{-1} d_k^{-3} c_m^2 + \mathrm{e}^{-\gamma_1 k}) e_n \|f\|_1^2 & \text{if } k \in \{k_0 \vee m + 1, \ldots, n-1\}, \end{cases} \tag{4.4}$$

our first result is

LEMMA 4.1. *For all* $m \in \{0, \ldots, n\}$

$$\|\mathcal{N}_{nm} f\|_\infty \lesssim \varsigma_{nm}(f), \tag{4.5}$$

and all $0 \le k \le n-1$,

$$\|\mathcal{U}_{nmk} f\|_{\tau_1} \vee \|\mathcal{V}_{nmk} f\|_{\tau_1} \lesssim \sigma^2_{nmk}(f). \tag{4.6}$$

The proof of (4.6), in turn, relies upon

LEMMA 4.2. *For every* $m \in \{0, \ldots, n\}$, $k \in \{0, \ldots, n-1\}$ *and* $t \in \{1, \ldots, n-k\}$,

$$\|\xi^2_{mkt} f\|_\infty + \sum_{s=1}^{n-k-t} \|\mathbb{E}_t \xi^2_{m,k,t+s} f\|_\infty \lesssim \sigma^2_{nmk}(f).$$

For the next result, recall the definition of $\delta_n(\mathscr{G})$ given in (2.7) above.

https://doi.org/10.1017/S0266466616000451 Published online by Cambridge University Press

LEMMA 4.3. *If* $\mathscr{G} \subset \mathrm{BI}$, *then*

$$\sum_{m=0}^{n} |\theta_m| \sup_{f\in\mathscr{G}} \varsigma_{nm}(f) + \sum_{m=0}^{n}\sum_{k=0}^{n-1} |\theta_m| \sup_{f\in\mathscr{G}} \sigma_{nmk}(f) \lesssim \delta_n(\mathscr{G}).$$

The proofs of these results are given below. We first turn to the

Proof of Proposition 2.2. The proof is almost identical to the proof of Proposition 4.2 in Duffy (2016). In view of Lemmas 4.1 and 4.3, we have immediately that

$$\max_{f\in\mathscr{F}_n} \left| \sum_{m=0}^{n} \theta_m \mathcal{N}_{nm} f \right| \lesssim \sum_{m=0}^{n} |\theta_m| \max_{f\in\mathscr{F}_n} \varsigma_{nm}(f) \lesssim \delta_n(\mathscr{F}_n),$$

and through an application of Lemma 3.6, that

$$\max_{f\in\mathscr{F}_n} \left| \sum_{m=0}^{n}\sum_{k=0}^{n-1} \theta_m \mathcal{M}_{nmk} f \right| \lesssim_p \delta_n(\mathscr{F}_n) \log n,$$

whence the result follows from (4.1) and (4.2). ∎

Proof of Lemma 4.1. For (4.5), note that by Lemma 3.3,

$$\|\mathbb{E}_0 f(x_t)\eta_{t-m}\|_\infty \lesssim \begin{cases} \|\eta\|_n \|f\|_\infty & \text{if } t \in \{1,\dots,k_0\}, \\ d_t^{-1}\|\eta\|_n \|f\|_1 & \text{if } t \in \{k_0+1,\dots,m\}, \\ d_t^{-2} c_m \|f\|_1 & \text{if } t \in \{k_0 \vee m + 1, \dots n\}, \end{cases}$$

whence, by Karamata's theorem and Lemma 3.4(i),

$$\begin{aligned} \|\mathcal{N}_{nm} f\|_\infty &\le \left(\sum_{t=1}^{k_0} + \sum_{t=k_0+1}^{m} + \sum_{t=k_0\vee m+1}^{n} \right) \|\mathbb{E}_0 f(x_t)\eta_{t-m}\|_\infty \\ &\lesssim k_0 \|\eta\|_n \|f\|_\infty + \left(\|\eta\|_n \sum_{t=k_0+1}^{m} d_t^{-1} + c_m \sum_{t=k_0\vee m+1}^{n} d_t^{-2} \right) \|f\|_1 \\ &\lesssim \|\eta\|_n \|f\|_\infty + \left[e_m \|\eta\|_n + c_m e_n^{1/2} \right] \|f\|_1. \end{aligned}$$

(4.6) follows from Lemma 4.2 in exactly the manner described in the proof of Lemma 7.3 in Duffy (2016). ∎

Proof of Lemma 4.2. The proof is similar to that of Lemma 7.4 in Duffy (2016). Let $m_0 := k_0 \vee m$. We shall obtain the requisite bound for $\mathbb{E}_t \xi^2_{m,k,t+s} f$ by providing a bound for $\mathbb{E}_{t-s} \xi^2_{mkt} f$ (for $s \in \{1,\dots,t\}$) that depends only on m, k, and s (and *not* t), separately considering the cases where

(i) $k \in \{m_0+1,\dots,n-t\}$;
(ii) $k \in \{k_0,\dots,m_0\}$; and
(iii) $k \in \{0,\dots,k_0\}$.

https://doi.org/10.1017/S0266466616000451 Published online by Cambridge University Press

(i) Recall the decomposition given in (3.2) and (3.3) above, applied here to reduce x_{t+k} to a sum of independent pieces,

$$\begin{aligned} x_{t+k} &= x^*_{0,t+k} + x'_{1,t-1,t+k} + x'_{t,t,t+k} + x'_{t+1,t+k,t+k} \\ &= x^*_{0,t+k} + x'_{1,t-1,t+k} + a_k\epsilon_t + x'_{t+1,t+k,t+k} \end{aligned}$$

with the convention that $x'_{1,t-1,t+k} = 0$ if $t = 1$, so that by Fourier inversion (Lemma 3.1),

$$\begin{aligned} \xi_{mkt} f &= \mathbb{E}_t f(x_{t+k})\eta_{t+k-m} - \mathbb{E}_{t-1} f(x_{t+k})\eta_{t+k-m} \\ &= \frac{1}{2\pi}\int \hat f(\lambda) \mathrm{e}^{-\mathrm{i}\lambda x^*_{0,t+k}} \mathrm{e}^{-\mathrm{i}\lambda x'_{1,t-1,t+k}} \\ &\qquad \cdot \left[\mathrm{e}^{-\mathrm{i}\lambda a_k\epsilon_t} - \mathbb{E}\mathrm{e}^{-\mathrm{i}\lambda a_k\epsilon_t}\right] \mathbb{E}\eta_{t+k-m} \mathrm{e}^{-\mathrm{i}\lambda x'_{t+1,t+k,t+k}}\, \mathrm{d}\lambda. \end{aligned} \tag{4.7}$$

Thence

$$\begin{aligned} \xi^2_{mkt} f = \frac{1}{(2\pi)^2}\iint & \hat f(\lambda_1)\hat f(\lambda_2) \mathrm{e}^{-\mathrm{i}(\lambda_1+\lambda_2)x^*_{0,t+k}} \mathrm{e}^{-\mathrm{i}(\lambda_1+\lambda_2)x'_{1,t-1,t+k}} \\ & \cdot \left[\mathrm{e}^{-\mathrm{i}\lambda_1 a_k\epsilon_t} - \mathbb{E}\mathrm{e}^{-\mathrm{i}\lambda_1 a_k\epsilon_t}\right]\left[\mathrm{e}^{-\mathrm{i}\lambda_2 a_k\epsilon_t} - \mathbb{E}\mathrm{e}^{-\mathrm{i}\lambda_2 a_k\epsilon_t}\right] \\ & \cdot \mathbb{E}\eta_{t+k-m}\mathrm{e}^{-\mathrm{i}\lambda_1 x'_{t+1,t+k,t+k}} \mathbb{E}\eta_{t+k-m}\mathrm{e}^{-\mathrm{i}\lambda_2 x'_{t+1,t+k,t+k}}\, \mathrm{d}\lambda_1\, \mathrm{d}\lambda_2. \end{aligned} \tag{4.8}$$

(Note that $\xi_{mkt} f$ is real-valued, so $\xi^2_{mkt} f = |\xi_{mkt} f|^2 = \xi_{mkt} f \cdot \overline{\xi_{mkt} f} = \xi_{mkt} f \cdot \xi_{mkt} f$.)

Now suppose $s \in \{k+1, \dots, t\}$. Taking conditional expectations on both sides of (4.8) gives

$$\begin{aligned} \mathbb{E}_{t-s}\xi^2_{mkt} f = \frac{1}{(2\pi)^2}\iint & \hat f(\lambda_1)\hat f(\lambda_2) \mathrm{e}^{-\mathrm{i}(\lambda_1+\lambda_2)x^*_{0,t+k}} \mathrm{e}^{-\mathrm{i}(\lambda_1+\lambda_2)x'_{1,t-s,t+k}} \\ & \cdot \vartheta(\lambda_1 a_k, \lambda_2 a_k) \mathbb{E}\mathrm{e}^{-\mathrm{i}(\lambda_1+\lambda_2)x'_{t-s+1,t-1,t+k}} \\ & \cdot \mathbb{E}\eta_{t+k-m}\mathrm{e}^{-\mathrm{i}\lambda_1 x'_{t+1,t+k,t+k}} \mathbb{E}\eta_{t+k-m}\mathrm{e}^{-\mathrm{i}\lambda_2 x'_{t+1,t+k,t+k}}\, \mathrm{d}\lambda_1\, \mathrm{d}\lambda_2, \end{aligned}$$

where we have defined

$$\vartheta(z_1, z_2) := \mathbb{E}\left[\mathrm{e}^{-\mathrm{i}z_1\epsilon_0} - \mathbb{E}\mathrm{e}^{-\mathrm{i}z_1\epsilon_0}\right]\left[\mathrm{e}^{-\mathrm{i}z_1\epsilon_0} - \mathbb{E}\mathrm{e}^{-\mathrm{i}z_2\epsilon_0}\right]$$

for $z_1, z_2 \in \mathbb{R}$, and made the further decomposition

$$x'_{1,t-1,t+k} = x'_{1,t-s,t+k} + x'_{t-s+1,t-1,t+k}$$

with the convention that $x'_{1,t-s,t+k} = 0$ if $s = t$. Thence, using (3.4) and Lemma 3.5, and the inequalities $|\hat f(\lambda)| \le \|f\|_1$ and $|ab| \lesssim |a|^2 + |b|^2$, we obtain

https://doi.org/10.1017/S0266466616000451 Published online by Cambridge University Press

$$\mathbb{E}_{t-s}\xi_{mkt}^2 f \lesssim \iint |\hat{f}(\lambda_1)\hat{f}(\lambda_2)||\mathbb{E}e^{-i(\lambda_1+\lambda_2)x'_{t-s+1,t-1,t+k}}| \tag{4.9}$$
$$\cdot (|a_k\lambda_1|^\alpha \tilde{G}(a_k\lambda_1)\wedge 1)^{1/2}(|a_k\lambda_2|^\alpha \tilde{G}(a_k\lambda_1)\wedge 1)^{1/2}$$
$$\cdot |\mathbb{E}\eta_{t+k-m}e^{-i\lambda_1 x'_{t+1,t+k,t+k}}||\mathbb{E}\eta_{t+k-m}e^{-i\lambda_2 x'_{t+1,t+k,t+k}}|\,d\lambda_1\,d\lambda_2$$
$$\lesssim \|f\|_1^2 \int (|a_k\lambda_1|^\alpha \tilde{G}(a_k\lambda_1)\wedge 1)|\mathbb{E}\eta_{t+k-m}e^{-i\lambda_1 x'_{t+1,t+k,t+k}}|^2 \tag{4.10}$$
$$\int |\mathbb{E}e^{-i(\lambda_1+\lambda_2)x'_{t-s+1,t-1,t+k}}|\,d\lambda_2\,d\lambda_1,$$

where we have appealed to symmetry (in λ_1 and λ_2) to reduce the final bound to a single term. By a change of variables and Lemma 3.2(ii),

$$\int |\mathbb{E}e^{-i(\lambda_1+\lambda_2)x'_{t-s+1,t-1,t+k}}|\,d\lambda_2 = \int |\mathbb{E}e^{-i\lambda x'_{t-s+1,t-1,t+k}}|\,d\lambda \lesssim \frac{c_s}{c_{k+s}}d_s^{-1}, \tag{4.11}$$

while Lemma 3.2(i) gives

$$\int (|a_k\lambda_1|^\alpha \tilde{G}(a_k\lambda_1)\wedge 1)|\mathbb{E}\eta_{t+k-m}e^{-i\lambda_1 x'_{t+1,t+k,t+k}}|^2\,d\lambda \lesssim k^{-1}d_k^{-3}c_m^2 + e^{-\gamma_1 k}. \tag{4.12}$$

Together, (4.10)–(4.12) yield

$$\mathbb{E}_{t-s}\xi_{mkt}^2 f \lesssim \frac{c_s}{c_{k+s}}d_s^{-1}(k^{-1}d_k^{-3}c_m^2 + e^{-\gamma_1 k})\|f\|_1^2. \tag{4.13}$$

When $s \in \{1,\dots,k\}$, (4.9) continues to hold, whence

$$\mathbb{E}_{t-s}\xi_{mkt}^2 f \lesssim \left(\int |\hat{f}(\lambda)|(|a_k\lambda_1|^\alpha \tilde{G}(a_k\lambda_1)\wedge 1)|\mathbb{E}\eta_{t+k-m}e^{-i\lambda x'_{t+1,t+k,t+k}}|\,d\lambda\right)^2$$
$$\lesssim \|f\|_1^2(k^{-1/2}d_k^{-2}c_m + e^{-\gamma_1 k})^2$$
$$\lesssim d_s^{-1}(k^{-1}d_k^{-3}c_m^2 + e^{-\gamma_1 k})\|f\|_1^2 \tag{4.14}$$

by Lemma 3.2(i); the replacement of a d_k^{-1} by d_s^{-1} in the final bound is justified because $s \le k$. Since $\{c_k\}$ is regularly varying and $k \ge k_0+1$, it follows from Potter's inequality (Bingham et al., 1987, Thm. 1.5.6(iii)) that

$$\sum_{s=1}^{k} d_s^{-1} + \sum_{s=k+1}^{n} \frac{c_s}{c_{k+s}}d_s^{-1} \lesssim \sum_{s=1}^{n} d_s^{-1} \lesssim n d_n^{-1} = e_n,$$

with the final bound following by Karamata's theorem. As noted above, since the bounds (4.13) and (4.14) do not depend on t, they apply also to $\mathbb{E}_t\xi_{m,k,t+s}^2 f$. Hence, in view of the preceding,

$$\sum_{s=1}^{n-k-t} \mathbb{E}_t\xi_{m,k,t+s}^2 f \lesssim (k^{-1}d_k^{-3}c_m^2 + e^{-\gamma_1 k})\|f\|_1^2\left[\sum_{s=1}^{k} d_s^{-1} + \sum_{s=k+1}^{n-k-t}\frac{c_s}{c_{k+s}}d_s^{-1}\right]$$
$$\lesssim (k^{-1}d_k^{-3}c_m^2 + e^{-\gamma_1 k})e_n\|f\|_1^2.$$

https://doi.org/10.1017/S0266466616000451 Published online by Cambridge University Press

Turning now to $\|\xi^2_{mkt} f\|_\infty$, note that (4.7) still holds, with the convention that $x_{1,t-1,t+k} = 0$ if $t = 1$. Thus, again by Lemma 3.2(i),

$$\begin{aligned}\|\xi^2_{mkt} f\|_\infty &\lesssim \left(\int |\hat{f}(\lambda)||\mathbb{E}\eta_{t+k-m} \mathrm{e}^{-\mathrm{i}\lambda x'_{t+1,t+k,t+k}}|\,\mathrm{d}\lambda\right)^2 \\ &\leq \|f\|_1^2 \left(\int |\mathbb{E}\eta_{t+k-m} \mathrm{e}^{-\mathrm{i}\lambda x'_{t+1,t+k,t+k}}|\,\mathrm{d}\lambda\right)^2 \\ &\lesssim (d_k^{-2} c_m + \mathrm{e}^{-\gamma_1 k})^2 \|f\|_1^2 \\ &\lesssim (k^{-1} d_k^{-3} c_m^2 + \mathrm{e}^{-\gamma_1 k}) e_n \|f\|_1^2;\end{aligned}$$

where the final bound follows because $k \leq n$, and so $d_k^{-1} \lesssim k^{-1} e_n$.

(ii) Note that $\{k_0 + 1, \ldots, m_0\}$ can only be nonempty if $k_0 < m = m_0$; thus $k \leq m$ in this case. Using $|a+b|^2 \lesssim |a|^2 + |b|^2$ gives

$$\mathbb{E}_{t-s} \xi^2_{mkt} f \lesssim \mathbb{E}_{t-s}\left[(\mathbb{E}_t |f(x_{t+k})\eta_{t+k-m}|)^2 + (\mathbb{E}_{t-1}|f(x_{t+k})\eta_{t+k-m}|)^2\right]. \quad \textbf{(4.15)}$$

Suppose $s \in \{m-k+1, \ldots, t\}$. Then by successive applications of Lemma 3.3(i),

$$\mathbb{E}_{t-s}\,(\mathbb{E}_t |f(x_{t+k})\eta_{t+k-m}|)^2 \lesssim d_s^{-1} d_k^{-1} \|f\|_1^2,$$

and similarly for the second term on the right side of (4.15). Thus

$$\mathbb{E}_{t-s} \xi^2_{mkt} f \lesssim d_s^{-1} d_k^{-1} \|f\|_1^2. \quad \textbf{(4.16)}$$

When $s \in \{1, \ldots, m-k\}$, further applications of Lemma 3.3(i) give

$$\begin{aligned}\mathbb{E}_{t-s}\,(\mathbb{E}_t |f(x_{t+k})\eta_{t+k-m}|)^2 &\leq \|\eta\|_n^2 \mathbb{E}_{t-s}\,(\mathbb{E}_t |f(x_{t+k})|)^2 \\ &\lesssim d_s^{-1} d_k^{-1} \|\eta\|_n^2 \|f\|_1^2,\end{aligned}$$

whence

$$\mathbb{E}_{t-s} \xi_{mkt} \lesssim d_s^{-1} d_k^{-1} \|\eta\|_n^2 \|f\|_1^2. \quad \textbf{(4.17)}$$

Together, (4.16) and (4.17) give

$$\begin{aligned}\sum_{s=1}^{n-k-t} \mathbb{E}_t \xi^2_{m,k,t+s} f &= \sum_{s=1}^{m-k} \mathbb{E}_t \xi^2_{m,k,t+s} f + \sum_{s=m-k+1}^{n-k-t} \mathbb{E}_t \xi^2_{m,k,t+s} f \\ &\lesssim d_k^{-1} \left[\|\eta\|_n^2 \sum_{s=1}^{m-k} d_s^{-1} + \sum_{s=m-k+1}^{n-k-t} d_s^{-1}\right] \|f\|_1^2 \\ &\lesssim d_k^{-1} \left[e_m \|\eta\|_n^2 + e_n\right] \|f\|_1^2\end{aligned}$$

by Karamata's theorem.

Regarding $\|\xi^2_{mkt} f\|_\infty$, it follows from Lemma 3.3(i) that

$$(\mathbb{E}_t |f(x_{t+k})\eta_{t+k-m}|)^2 \leq \|\eta\|_n^2 (\mathbb{E}_t |f(x_{t+k})|)^2 \lesssim d_k^{-2} \|\eta\|_n^2 \|f\|_1^2.$$

https://doi.org/10.1017/S0266466616000451 Published online by Cambridge University Press

(iii) Suppose that $s \in \{(m-k) \vee k_0 + 1, \dots, t\}$. Using $|a+b|^2 \lesssim |a|^2 + |b|^2$, Jensen's inequality and Lemma 3.3(i), we have

$$\mathbb{E}_{t-s}\xi_{mkt}^2 f \lesssim \mathbb{E}_{t-s}|f(x_{t+k})\eta_{t+k-m}|^2 \lesssim d_s^{-1}\|f\|_2^2.$$

When $s \in \{k_0+1, \dots, m-k\}$, we have similarly

$$\mathbb{E}_{t-s}\xi_{mkt}^2 f \lesssim \|\eta\|_n^2 \mathbb{E}_{t-s}|f(x_{t+k})|^2 \lesssim d_s^{-1}\|\eta\|_n^2\|f\|_2^2,$$

and for $s \in \{1, \dots, k_0\}$, we may use the crude bound

$$\mathbb{E}_{t-s}\xi_{mkt}^2 f \leq \|\xi_{mkt}^2 f\|_\infty \leq \|\eta\|_n^2\|f\|_\infty^2. \tag{4.18}$$

Hence

$$\begin{aligned}\sum_{s=1}^{n-k-t} \mathbb{E}_t \xi_{m,k,t+s}^2 f &= \left(\sum_{s=1}^{k_0} + \sum_{s=k_0+1}^{m-k} + \sum_{s=(m-k)\vee k_0+1}^{n-k-t}\right)\mathbb{E}_t\xi_{m,k,t+s}^2 f \\ &\lesssim k_0\|\eta\|_n^2\|f\|_\infty^2 + \|f\|_2^2\left(\|\eta\|_n^2\sum_{s=k_0+1}^{m-k} d_s^{-1} + \sum_{s=(m-k)\vee k_0+1}^{n-k-t} d_s^{-1}\right) \\ &\lesssim \|\eta\|_n^2\|f\|_\infty^2 + \left[e_m\|\eta\|_n^2 + e_n\right]\|f\|_2^2\end{aligned}$$

by Karamata's theorem. The required bound for $\|\xi_{mkt}^2 f\|_\infty$ is given in (4.18). ■

Proof of Lemma 4.3. It is evident from (4.3) and Lemma 3.4(iii) that

$$\begin{aligned}\sum_{m=0}^{n}|\theta_m| \sup_{f\in\mathscr{G}} \varsigma_{nm}(f) &\leq \|\eta\|_n\|\mathscr{G}\|_\infty + \left[\|\eta\|_n\sum_{m=0}^{n}|\theta_m|e_m + e_n^{1/2}\sum_{m=0}^{n}|\theta_m|c_m\right]\|\mathscr{G}\|_1 \\ &\lesssim \|\eta\|_n\|\mathscr{G}\|_\infty + \left[\|\eta\|_n + e_n^{1/2}\right]\|\mathscr{G}\|_1. \end{aligned} \tag{4.19}$$

Considering the three parts of (4.4) separately, we first have

$$\begin{aligned}\sum_{m=0}^{n}\sum_{k=0}^{k_0}|\theta_m| \sup_{f\in\mathscr{G}} \sigma_{nmk}(f) &\lesssim \|\eta\|_n\|\mathscr{G}\|_\infty + \left[\|\eta\|_n\sum_{m=0}^{n}|\theta_m|e_m^{1/2} + e_n^{1/2}\right]\|\mathscr{G}\|_2 \\ &\lesssim \|\eta\|_n\|\mathscr{G}\|_\infty + \left[\|\eta\|_n + e_n^{1/2}\right]\|\mathscr{G}\|_2, \end{aligned} \tag{4.20}$$

since k_0 is fixed and finite. Next,

$$\begin{aligned}\sum_{m=0}^{n}\sum_{k=k_0+1}^{m}|\theta_m| \sup_{f\in\mathscr{G}} \sigma_{nmk}(f) &\lesssim \sum_{m=0}^{n}|\theta_m|\left[e_m^{1/2}\|\eta\|_n + e_n^{1/2}\right]\|\mathscr{G}\|_1\sum_{k=k_0+1}^{m} d_k^{-1/2} \\ &\lesssim \sum_{m=0}^{n}|\theta_m|m^{1/2}e_m^{1/2}\left[e_m^{1/2}\|\eta\|_n + e_n^{1/2}\right]\|\mathscr{G}\|_1 \\ &\lesssim \left[\|\eta\|_n + e_n^{1/2}\right]\|\mathscr{G}\|_1, \end{aligned} \tag{4.21}$$

https://doi.org/10.1017/S0266466616000451 Published online by Cambridge University Press

using Karamata's theorem and Lemma 3.4(iii). Finally,

$$\sum_{m=0}^{n}\sum_{k=m+1}^{n}|\theta_m|\sup_{f\in\mathscr{G}}\sigma_{nmk}(f)\lesssim\sum_{m=0}^{n}|\theta_m|c_m e_n^{1/2}\|\mathscr{G}\|_1\sum_{k=m+1}^{n}k^{-1/2}d_k^{-3/2}$$
$$\lesssim e_n^{1/2}\|\mathscr{G}\|_1 \tag{4.22}$$

by parts (ii) and (iii) of Lemma 3.4. Recalling (2.7), the result now follows from (4.19)–(4.22). ■

5. PROOFS OF THE ORDER ESTIMATES

The proof of (2.4) in Theorem 2.1 may be broken into three parts:

(a) a truncation argument permits $\{u_t\}$ to be replaced by $\{u_t^{(\le)}\}$ on the left side of (2.4);
(b) the supremum over $(a,h)\in\mathbb{R}\times\mathscr{H}_n$ is reduced to a maximum over a (growing) finite set; and
(c) an application of Proposition 2.2 yields the requisite bound over this finite set.

These steps are described below, following which we provide details of the modifications necessary for the proof of (2.5).

5.1. Truncation

Decomposing

$$u_t=\sum_{k=0}^{\infty}\theta_k\eta_{t-k}=\sum_{k=0}^{n}\theta_k\eta_{t-k}+\sum_{k=n+1}^{\infty}\theta_k\eta_{t-k}=:u_t^{(n)}+u_t^{(-n)},$$

we have by the Cauchy–Schwarz inequality that

$$\sup_{a\in\mathbb{R}}\left|\frac{1}{(e_nh_n)^{1/2}}\sum_{t=1}^{n}f\left(\frac{x_t-d_na}{h_n}\right)u_t^{(-n)}\right|$$
$$\le\left(\sup_{a\in\mathbb{R}}\frac{1}{(e_nh_n)^{1/2}}\sum_{t=1}^{n}f^2\left(\frac{x_t-d_na}{h_n}\right)\right)^{1/2}\left(\sum_{t=1}^{n}(u_t^{(-n)})^2\right)^{1/2}=o_p(1),$$

since the first term on the RHS is $O_p(1)$ by Proposition 2.1, and

$$\mathbb{E}\sum_{t=1}^{n}(u_t^{(-n)})^2\lesssim n\sum_{j=n+1}^{\infty}\theta_j^2\le\sum_{j=n+1}^{\infty}\theta_j^2\,j\to0.$$

Next, define for $-n\le t\le n$,

$$\eta_t^{(\le)}:=\eta_t\mathbf{1}\{|\eta_t|\le n^{1/q_0}\}-\mathbb{E}\eta_0\mathbf{1}\{|\eta_0|\le n^{1/q_0}\} \tag{5.1}$$

and $\eta_t^{(>)}:=\eta_t-\eta_t^{(\le)}$. Then setting

https://doi.org/10.1017/S0266466616000451 Published online by Cambridge University Press

$$u_t^{(\leq)} := \sum_{k=0}^{n} \theta_k \eta_{t-k}^{(\leq)} \qquad\qquad u_t^{(>)} := \sum_{k=0}^{n} \theta_k \eta_{t-k}^{(>)},$$

we see that

$$\mathbb{P}\left\{ \sup_{a\in\mathbb{R}} \left| \sum_{t=1}^{n} f\left(\frac{x_t - d_n a}{h_n}\right) u_t^{(>)} \right| \neq 0 \right\} \leq \mathbb{P}\left\{ \sup_{t\leq n} |u_t^{(>)}| \neq 0 \right\}$$
$$\leq \mathbb{P}\left\{ \max_{-n\leq t\leq n} |\eta_t| > n^{1/q_0} \right\} = o(1),$$

where the final equality follows by Theorem 2.12.1 in Hansen (2012), since $\{\eta_t\}$ is i.i.d. with bounded q_0th moment. (2.4) will therefore follow once we have shown that

$$\frac{1}{(e_n h_n)^{1/2}} \sup_{a\in\mathbb{R}} \left| \sum_{t=1}^{n} f\left(\frac{x_t - d_n a}{h_n}\right) u_t^{(\leq)} \right| \lesssim_p (1 + n^{1/q_0 - r_0}) \log n.$$

5.2. Reduction to the Maximum Over a Finite Set

For the remainder of the proof, we may without loss of generality take f to be bounded by unity, with a Lipschitz constant of unity. To simplify the exposition, we shall require that $h_n \in \mathscr{H}_n$ always, and take $\overline{h} = 1$; the proof in the general case (where this occurs w.p.a.1) requires no new ideas. As it is less cumbersome to work with the inverse bandwidth $b := h^{-1}$, we define

$$\mathscr{B}_n := \{h^{-1} \mid h \in \mathscr{H}_n\} = [1, \overline{b}_n],$$

where $\overline{b}_n := \underline{h}_n^{-1}$. For $(a,b) \in \mathbb{R}\times\mathbb{R}_+$, let $f_{(a,b)}(x) := b^{1/2} f[b(x - d_n a)]$. Then

$$\frac{1}{(e_n h)^{1/2}} \sum_{t=1}^{n} f\left(\frac{x_t - d_n a}{h}\right) u_t^{(\leq)} = \frac{1}{e_n^{1/2}} \sum_{t=1}^{n} f_{(a,b)}(x_t) u_t^{(\leq)} =: \mathcal{R}_n^f(a,b)$$

for $b = h^{-1}$.

Take $C_n := [-n^\gamma, n^\gamma] \times \mathscr{B}_n$, and let $\mathscr{C}_n \subset C_n$ be a lattice of mesh $n^{-\delta}$. $p_n(a,b)$ denotes the projection of (a,b) onto a nearest neighbour in $\mathscr{C}_n$ (with some tie-breaking rule). We shall now prove that γ and δ may be chosen (sufficiently large) such that

$$\sup_{(a,b)\in C_n} |\mathcal{R}_n^f(a,b)| = \sup_{(a,b)\in\mathscr{C}_n} |\mathcal{R}_n^f(a,b)| + o_p(1), \tag{5.2}$$

$$\sup_{(a,b)\in[-n^\gamma, n^\gamma]^c \times \mathscr{B}_n} |\mathcal{R}_n^f(a,b)| = o_p(1), \tag{5.3}$$

with the aid of the following.

LEMMA 5.1. *For every $\gamma > 0$, there exists a $\delta > 0$ such that*

$$\sup_{(a,b)\in C_n} \frac{1}{e_n^{1/2}} \sum_{t=1}^{n} |f_{(a,b)}(x_t) - f_{p_n(a,b)}(x_t)||u_t^{(\leq)}| = o_p(1).$$

https://doi.org/10.1017/S0266466616000451 Published online by Cambridge University Press

LEMMA 5.2. *Suppose* $f \in \mathrm{BIL}$. *Then* $|f(x)| = o(|x|^{-1/2})$ *as* $x \to \pm\infty$.

Since $\|u_t^{(\leq)}\|_\infty \lesssim n^{1/q_0}$, Lemma 5.1 may be proved by an argument identical to that used in the proof of Lemma 6.1 in Duffy (2016), while Lemma 5.2 is a special case of Lemma B.1 in the Supplement to that paper. Observe that (5.2) follows immediately from Lemma 5.1. To establish (5.3), first note that w.p.a.1,

$$\inf_{t\leq n}\inf_{|a|\geq n^\gamma}|x_t - d_n a| \geq d_n n^\gamma\left(1 - n^{-\gamma}d_n^{-1}\max_{t\leq n}|x_t|\right) = d_n n^\gamma\,(1+o_p(1)),$$

provided that γ is chosen large enough that

$$\mathbb{E}\max_{t\leq n}|x_t| \leq n^2\mathbb{E}|v_0| = o(n^\gamma d_n).$$

For the proof that such a γ exists, see the arguments following (C.1) in the Supplement to Duffy (2016). Thence by Lemma 5.2

$$\begin{aligned}\max_{t\leq n}\sup_{(a,b)\in[-n^\gamma,n^\gamma]^c\times\mathscr{B}_n} b^{1/2}f[b(x_t - d_n a)] &\lesssim \max_{t\leq n}\sup_{|a|\geq n^\gamma}|x_t - d_n a|^{-1/2}\\ &\lesssim_p (d_n n^\gamma)^{-1/2},\end{aligned}$$

whence

$$\begin{aligned}\sup_{(a,b)\in[-n^\gamma,n^\gamma]^c\times\mathscr{B}_n}|\mathcal{R}_n^f(a,b)| &\leq \sup_{(a,b)\in[-n^\gamma,n^\gamma]^c\times\mathscr{B}_n}\frac{1}{e_n^{1/2}}\sum_{t=1}^n|f_{(a,b)}(x_t)||u_t^{(\leq)}|\\ &\lesssim_p\left(\frac{n^{2/q_0}}{e_n d_n n^\gamma}\right)^{1/2}\\ &= o(1)\end{aligned}$$

for $\gamma > 0$ sufficiently large.

5.3. Control Over the Finite Set

It remains to provide an estimate for

$$\sup_{(a,b)\in\mathscr{C}_n}|\mathcal{R}_n^f(a,b)| = \frac{1}{e_n^{1/2}}\sup_{(a,b)\in\mathscr{C}_n}\left|\sum_{t=1}^n f_{(a,b)}(x_t)u_t^{(\leq)}\right| = \frac{1}{e_n^{1/2}}\sup_{g\in\mathscr{G}_n}|\mathcal{S}_n g|,$$

where $\mathscr{G}_n := \{f_{(a,b)} \mid (a,b)\in\mathscr{C}_n\}$, and $\mathcal{S}_n g$ is defined as in (2.6) above. Since $f \in \mathrm{BI}$ and $\overline{b}_n \lesssim e_n n^{-2r_0}$, it is clear that

$$\delta_n(\mathscr{G}_n) \lesssim \|\eta\|_n\overline{b}_n^{1/2} + [\|\eta\|_n + e_n^{1/2}] = O[e_n^{1/2}(1+n^{1/q_0-r_0})]$$

and thus

$$\frac{1}{e_n^{1/2}}\sup_{g\in\mathscr{G}_n}|\mathcal{S}_n g| \lesssim_p (1+n^{1/q_0-r_0})\log n$$

by Proposition 2.2. This completes the proof of (2.4).

https://doi.org/10.1017/S0266466616000451 Published online by Cambridge University Press

5.4. Modifications Required for the Proof of (2.5)

The proof of (2.5) is almost identical to the preceding, albeit somewhat simpler. The truncation performed in Section 5.1 is not necessary, while the same argument as given in Section 5.2 may be used to reduce the problem to that of providing a suitable bound for

$$\frac{1}{e_n^{1/2}} \sup_{g \in \mathscr{G}_n} |\mathcal{S}_n^* g|,$$

where $\mathcal{S}_n^* g := e_n^{-1/2} \sum_{t=1}^n g(x_t)$. Define

$$\delta_n^*(\mathscr{F}) := \|\mathscr{F}\|_\infty + e_n^{1/2}(\|\mathscr{F}\|_1 + \|\mathscr{F}\|_2) + \left[\sum_{k=1}^n d_k^{-2} + e_n^{1/2} \sum_{k=1}^{n-1} k^{-1/2} d_k^{-3/2}\right] \|\mathscr{F}\|_{[1]}$$
$$\lesssim \|\mathscr{F}\|_\infty + e_n^{1/2}(\|\mathscr{F}\|_1 + \|\mathscr{F}\|_2 + \|\mathscr{F}\|_{[1]}),$$

where $\|f\|_{[1]} := \inf\{c \in \mathbb{R}_+ \mid |\hat{f}(\lambda)| \leq c|\lambda|\}$, and the final bound follows from parts (i) and (ii) of Lemma 3.4. With the aid of Lemma 9.1(ii) in Duffy (2016), it is easily verified that

$$\|\mathscr{G}_n\|_1 \vee \|\mathscr{G}_n\|_2 \vee \|\mathscr{G}_n\|_{[1]} \lesssim 1 \qquad \|\mathscr{G}_n\|_\infty \lesssim \overline{b}_n^{1/2} = o(e_n^{1/2}),$$

whence by Remark 4.1 in Duffy (2016),

$$\frac{1}{e_n^{1/2}} \sup_{g \in \mathscr{G}_n} |\mathcal{S}_n^* g| \lesssim_p e_n^{-1/2} \delta_n^*(\mathscr{G}_n) \log n \lesssim \log n$$

as required.

6. PROOFS OF THE CONVERGENCE RATES

Recall that

$$\mathcal{L}_n(a) := \frac{1}{e_n} \sum_{t=1}^n K_{h_n}(x_t - d_n a),$$

where K satisfies Assumption 3.

Proof of Proposition 2.3. Define

$$\mathfrak{m}_n := d_n^{-1} x_{(1)} \qquad \mathfrak{m} := \inf_{r \in [0,1]} |X(r)|$$
$$\mathfrak{M}_n := d_n^{-1} x_{(n)} \qquad \mathfrak{M} := \sup_{r \in [0,1]} |X(r)|.$$

Since $(\alpha, H) = (2, \frac{1}{2})$, X is a Brownian motion, and so by Ray's (1963) theorem,

$$\mathbb{P}\left\{\inf_{a \in [\mathfrak{m}, \mathfrak{M}]_\varepsilon} \mathcal{L}(a) > 0\right\} = 1. \tag{6.1}$$

For a more detailed argument as to why (6.1) follows from Ray's theorem, see (6.4.36)–(6.4.38) and the surrounding discussion in Karatzas and Shreve (1991).

https://doi.org/10.1017/S0266466616000451 Published online by Cambridge University Press

(Note that it is necessary that $\varepsilon > 0$ here, since $\mathcal{L}(\mathfrak{m}) = \mathcal{L}(\mathfrak{M}) = 0$ by the continuity of $\mathcal{L}$.) Under the assumption that $\mathbb{E}\epsilon_0^2$, $X_n \rightsquigarrow X$ on $\ell_\infty[0,1]$ by Hannan (1979), whence $(\mathfrak{m}_n, \mathfrak{M}_n) \rightsquigarrow (\mathfrak{m}, \mathfrak{M})$ by the continuous mapping theorem (CMT). By Theorem 3.1 in Duffy (2016), $\mathcal{L}_n \rightsquigarrow \mathcal{L}$ on $\ell_\infty(\mathbb{R})$, and thus a further application of the CMT yields

$$\inf_{x \in R_n^\varepsilon} \frac{1}{e_n} \sum_{t=1}^n K_{h_n}(x_t - x) = \inf_{a \in [\mathfrak{m}_n, \mathfrak{M}_n]_\varepsilon} \mathcal{L}_n(a) \rightsquigarrow \inf_{a \in [\mathfrak{m}, \mathfrak{M}]_\varepsilon} \mathcal{L}(a). \tag{6.2}$$

Together, (6.1) and (6.2) yield (2.8).

To obtain (2.9), we note that

$$\frac{1}{n} \sum_{t=1}^n \mathbf{1}\{x_t > (1-\varepsilon)x_{(n)}\} \le \int_0^1 \mathbf{1}\{X_n(r) \ge (1-\varepsilon)\mathfrak{M}_n\}\,\mathrm{d}r + o_p(1). \tag{6.3}$$

Let $\{g_k\}$ denote a uniformly bounded sequence of functions that converges pointwise to $x \mapsto \mathbf{1}\{x \ge 0\}$, from above. Then

$$\begin{aligned} \int_0^1 \mathbf{1}\{X_n(r) \ge (1-\varepsilon)\mathfrak{M}_n\}\,\mathrm{d}r &\le \int_0^1 g_k[X_n(r) - (1-\varepsilon)\mathfrak{M}_n]\,\mathrm{d}r \\ &\rightsquigarrow \int_0^1 g_k[X(r) - (1-\varepsilon)\mathfrak{M}]\,\mathrm{d}r \\ &\xrightarrow{\text{a.s.}} \int_0^1 \mathbf{1}\{X(r) \ge (1-\varepsilon)\mathfrak{M}\}\,\mathrm{d}r, \end{aligned} \tag{6.4}$$

as $n \to \infty$ and then $k \to \infty$, by the CMT and the dominated convergence theorem. Further,

$$\int_0^1 \mathbf{1}\{X(r) \ge (1-\varepsilon)\mathfrak{M}\}\,\mathrm{d}r \xrightarrow{\text{a.s.}} \int_0^1 \mathbf{1}\{X(r) = \mathfrak{M}\}\,\mathrm{d}r = \int_{\mathbb{R}} \mathbf{1}\{x = \mathfrak{M}\}\mathcal{L}(x)\,\mathrm{d}x = 0 \tag{6.5}$$

as $\varepsilon \to 0$, by dominated convergence, (2.2), and the fact that $\mathcal{L}(\mathfrak{M}) = 0$. It follows from (6.3)–(6.5) that $\varepsilon > 0$ may be chosen such that

$$\limsup_{n \to \infty} \mathbb{P}\left\{ \frac{1}{n} \sum_{t=1}^n \mathbf{1}\{x_t > (1-\varepsilon)x_{(n)}\} \ge \delta \right\} \le \frac{\delta}{2}.$$

By an analogous argument, this holds also when $\mathbf{1}\{x_t > (1-\varepsilon)x_{(n)}\}$ is replaced by $\mathbf{1}\{x_t < (1-\varepsilon)x_{(1)}\}$. ■

The proof of Theorem 2.2 requires the following two results. For a matrix A, let $\|A\|_T := \sup_{\|x\|=1} \|Ax\|_2$.

LEMMA 6.1. *For every* $g \in \mathrm{BIL}$ *with* $\int |g(x)x|\,\mathrm{d}x < \infty$,

$$\sup_{a \in \mathbb{R}} \left| \frac{1}{e_n h_n} \sum_{t=1}^n g\left(\frac{x_t - d_n a}{h_n} \right) - \mathcal{L}_n(a) \int g \right| = o_p(1). \tag{6.6}$$

https://doi.org/10.1017/S0266466616000451 Published online by Cambridge University Press

LEMMA 6.2. *Suppose that $Y_n(a)$ is a $(k \times k)$ matrix-valued process, such that $Y_n(a)$ is positive semi-definite for every $a \in \mathbb{R}$ and $n \in \mathbb{N}$, and let Γ be a positive definite $(k \times k)$ matrix, for which $\sup_{a\in\mathbb{R}} \|Y_n(a) - \Gamma \mathcal{L}_n(a)\|_T = o_p(1)$. Then*

$$\sup_{\{a|\mathcal{L}_n(a)\geq\varepsilon\}} \|Y_n(a)^{-1}\|_T \lesssim_p 1.$$

Proof of Lemma 6.1. Setting $f(x) := g(x) - K(x)\int g$, the left side of (6.6) may be written as

$$\sup_{a\in\mathbb{R}} \left| \frac{1}{e_n h_n} \sum_{t=1}^{n} f\left(\frac{x_t - d_n a}{h_n}\right) \right| \lesssim_p \frac{\log n}{(e_n h_n)^{1/2}} = o_p(1)$$

by Theorem 2.1. ■

Proof of Lemma 6.2. By multiplying both Y_n and $A\mathcal{L}_n$ by A^{-1}, we may reduce the problem to one in which $A = I_k$. We may also replace $(Y_n, \mathcal{L}_n)$ by a distributionally equivalent sequence for which $(Y_n, \mathcal{L}_n) \xrightarrow{\text{a.s.}} (\mathcal{L}, \mathcal{L})$ in $\ell_{\text{ucc}}(\mathbb{R}^2)$: see Theorem 1.10.3 in van der Vaart and Wellner (1996). Define $r_n^\omega(a) := \|Y_n^\omega(a) - I_k \mathcal{L}_n^\omega(a)\|_T$, and let $\Omega_0 \subset \Omega$ denote a set, with $\mathbb{P}\Omega_0 = 1$, on which $\mathcal{L}_n^\omega \to \mathcal{L}^\omega$ in $\ell_{\text{ucc}}(\mathbb{R})$, and $r_n^\omega \to 0$ in $\ell_\infty(\mathbb{R})$.

It is easily verified that, for B a real symmetric matrix, and $z > 0$,

$$\lambda_{\min}(B) = z + \lambda_{\min}(B - zI) \geq z - \|B - zI\|_T,$$

where $\lambda_{\min}(B)$ denotes the smallest eigenvalue of B. Thus, fixing an $\omega \in \Omega_0$,

$$\begin{aligned}\inf_{\{a|\mathcal{L}_n^\omega(a)\geq\varepsilon\}} \lambda_{\min}[Y_n^\omega(a)] &\geq \inf_{a\in\mathbb{R}} \left[\mathcal{L}_n^\omega(a) - r_n^\omega(a)\right] \mathbf{1}\{\mathcal{L}_n^\omega(a) \geq \varepsilon\} \\ &\geq \varepsilon - \sup_{a\in\mathbb{R}} r_n^\omega(a) \\ &\to \varepsilon,\end{aligned}$$

whence

$$\sup_{\{a|\mathcal{L}_n^\omega(a)\geq\varepsilon\}} \|Y_n^\omega(a)^{-1}\|_T = \left(\inf_{\{a|\mathcal{L}_n^\omega(a)\geq\varepsilon\}} \lambda_{\min}[Y_n^\omega(a)] \right)^{-1} \to \varepsilon^{-1},$$

from which the result follows. ■

Proof of Theorem 2.2. Since (2.12) follows from arguments given in the course of Section 2.4, we provide the proof of only (2.13) here. In the notation of Fan and Gijbels (1996, pp. 58f.), $\hat{m}_L(x)$ is given by the first element of

$$\hat{\beta}(x) := (X'WX)^{-1} X'Wy,$$

which admits the decomposition

$$\hat{\beta}(x) - \beta(x) = (X'WX)^{-1} X'W[\vec{m}_0 - X'\beta] + (X'WX)^{-1} X'Wu,$$

where $y = (y_1, \ldots, y_n)'$, $u = (u_1, \ldots, u_n)'$,

https://doi.org/10.1017/S0266466616000451 Published online by Cambridge University Press

$$X := \begin{bmatrix} 1\ x_1 - x \\ 1\ x_2 - x \\ \vdots\ \ \vdots \\ 1\ x_n - x \end{bmatrix} \qquad W := \operatorname{diag} \begin{bmatrix} K_{h_n}(x_1 - x) \\ K_{h_n}(x_2 - x) \\ \vdots \\ K_{h_n}(x_n - x) \end{bmatrix} \qquad \vec{m}_0 := \begin{bmatrix} m_0(x_1) \\ m_0(x_2) \\ \vdots \\ m_0(x_n) \end{bmatrix}$$

and $\beta(x) = (m_0(x), m_0'(x))'$.

Observe $X'WX$ is a (2×2) matrix with (i, j)th element

$$(X'WX)_{ij}(x) = h_n^{(i+j-2)} \sum_{t=1}^{n} K_{h_n}^{[i+j-2]}(x_t - x).$$

Note that by Lemma 6.1,

$$\frac{1}{e_n} \sum_{t=1}^{n} K_{h_n}^{[i+j-2]}(x_t - d_n a) - \mathcal{L}_n(a) \int K^{[i+j-2]} \xrightarrow{p} 0$$

in $\ell_\infty(\mathbb{R})$. Hence, for $D := \operatorname{diag}[1, h_n]$,

$$e_n^{-1}[D^{-1}(X'WX)D^{-1}](d_n a) - \mathcal{K}\mathcal{L}(a) \xrightarrow{p} 0$$

in $\ell_\infty(\mathbb{R})$, where $\mathcal{K} := [\int K^{[i+j-2]}]$ is positive definite. Thus by Lemma 6.2,

$$\sup_{x \in A_n^\varepsilon} \|[D(X'WX)^{-1}D](x)\|_T = \sup_{\{a | \mathcal{L}_n(a) \geq \varepsilon\}} \|[D(X'WX)^{-1}D](d_n a)\|_T \lesssim_p e_n^{-1}. \tag{6.7}$$

To handle the bias term, note that by a Taylor series expansion

$$|m_0(x_t) - \beta_0(x) - \beta_1(x)(x_t - x)| \leq m''(\tilde{x}_t)|x_t - x|^2$$

for all $x \in A_n^\varepsilon$, where $\tilde{x}_t \in [x, x_t]$. Hence for $i \in \{1, 2\}$,

$$|\{D^{-1}X'W[\vec{m}_0 - X'\beta]\}_i(x)| \leq h_n^2 \overline{m}_2(\tilde{A}_n^\varepsilon) \sum_{t=1}^{n} |K_{h_n}^{[1+i]}(x_t - x)|,$$

whence by Proposition 2.1,

$$\frac{1}{e_n} \sup_{x \in A_n^\varepsilon} |\{D^{-1}X'W[\vec{m}_0 - X'\beta]\}_i(x)| \leq \frac{h_n^2 \overline{m}_2(\tilde{A}_n^\varepsilon)}{e_n} \sup_{x \in \mathbb{R}} \sum_{t=1}^{n} |K_{h_n}^{[1+i]}(x_t - x)| \lesssim_p h_n^2 \overline{m}_2(\tilde{A}_n^\varepsilon). \tag{6.8}$$

For the variance term, note that for $i \in \{1, 2\}$,

$$\frac{1}{e_n} \sup_{x \in A_n^\varepsilon} |[D^{-1}X'u]_i(x)| = \frac{1}{e_n} \sup_{x \in A_n^\varepsilon} \left| \sum_{t=1}^{n} K_{h_n}^{[i-1]}(x_t - x) u_t \right| \lesssim_p \frac{\log n}{(e_n h_n)^{1/2}} \tag{6.9}$$

by Theorem 2.1. (6.7)–(6.9) now yield the stated result. ■

https://doi.org/10.1017/S0266466616000451 Published online by Cambridge University Press

NOTES

1. For a complete index of the notation used in this paper, please consult the preprint (arXiv:1505.01787).

2. Supplementary material to this article is provided in the "Online Supplementary Material for 'Uniform Convergence Rates over Maximal Domains in Structural Nonparametric Cointegrating Regression'", which is available at Cambridge Journals Online (journals.cambridge.org/ect).

REFERENCES

Bercu, B. & A. Touati (2008) Exponential inequalities for self-normalized martingales with applications. *Annals of Applied Probability* 18(5), 1848–1869.

Bingham, N.H., C.M. Goldie, & J.L. Teugels (1987) *Regular Variation*. Cambridge University Press.

Borodin, A.N. & I.A. Ibragimov (1995) Limit theorems for functionals of random walks. *Proceedings of the Steklov Institute of Mathematics* 195(2), 1–259.

Chan, N. & Q. Wang (2014) Uniform convergence for nonparametric estimators with nonstationary data. *Econometric Theory* 30(5), 1110–1133.

Duffy, J.A. (2016) A uniform law for convergence to the local times of linear fractional stable motions. *Annals of Applied Probability* 25(1), 45–72.

Einmahl, U. & D.M. Mason (2005) Uniform in bandwidth consistency of kernel-type function estimators. *Annals of Statistics* 33(3), 1380–1403.

Fan, J. & I. Gijbels (1996) *Local Polynomial Modelling and its Application*. Chapman & Hall.

Freedman, D.A. (1975) On tail probabilities for martingales. *Annals of Probability* 3(1), 100–118.

Gao, J., S. Kanaya, D. Li, & D. Tjøstheim (2015) Uniform consistency for nonparametric estimators in null recurrent time series. *Econometric Theory* 31(5), 911–952.

Hannan, E.J. (1979) The central limit theorem for time series regression. *Stochastic Processes and Their Applications* 9(3), 281–289.

Hansen, B.E. (2008) Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory* 24(3), 726–748.

Hansen, B.E. (2012) *Econometrics*. Unpublished. Available at www.ssc.wisc.edu/∼bhansen/econometrics/Econometrics2012.pdf.

Ibragimov, I.A. & Y.V. Linnik (1971) *Independent and Stationary Sequences of Random Variables*. Wolters–Noordhoff.

Jeganathan, P. (2004) Convergence of functionals of sums of r.v.s to local times of fractional stable motions. *Annals of Probability* 32, 1771–1795.

Jeganathan, P. (2008) Limit Theorems for Functionals of Sums that Converge to Fractional Brownian and Stable Motions. Cowles Foundation Discussion Paper No. 1649, Yale University.

Karatzas, I. & S.E. Shreve (1991) *Brownian Motion and Stochastic Calculus*, 2nd ed. Springer.

Karlsen, H.A., T. Myklebust, & D. Tjøstheim (2007) Nonparametric estimation in a nonlinear cointegration type model. *Annals of Statistics* 35(1), 252–299.

Kristensen, D. (2009) Uniform convergence rates of kernel estimators with heterogeneous dependent data. *Econometric Theory* 25(5), 1433–1445.

Li, D., Z. Lu, & O. Linton (2012) Local linear fitting under near epoch dependence: uniform consistency with convergence rates. *Econometric Theory* 28(5), 935–958.

Liu, W., N. Chan, & Q. Wang (2014) Uniform approximation to local time with applications in nonlinear cointegrating regression. Unpublished, University of Sydney.

Park, J.Y. & P.C.B. Phillips (2001) Nonlinear regressions with integrated time series. *Econometrica* 69(1), 117–161.

Phillips, P.C.B. & L. Su (2011) Non-parametric regression under location shifts. *Econometrics Journal* 14(3), 457–486.

Ray, D. (1963) Sojourn times of diffusion processes. *Illinois Journal of Mathematics* 7(4), 615–630.

Revuz, D. & M. Yor (1999) *Continuous Martingales and Brownian Motion*, 3rd ed. Springer.

https://doi.org/10.1017/S0266466616000451 Published online by Cambridge University Press

Samorodnitsky, G. & M.S. Taqqu (1994) *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. CRC Press.

Stone, C.J. (1982) Optimal global rates of convergence for nonparametric regression. *Annals of Statistics* 10(4), 1040–1053.

van der Vaart, A.W. & J.A. Wellner (1996) *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer.

Wang, Q. & N. Chan (2014) Uniform convergence rates for a class of martingales with application in non-linear cointegrating regression. *Bernoulli* 20(1), 207–230.

Wang, Q. & P.C.B. Phillips (2009a) Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory* 25(3), 710–738.

Wang, Q. & P.C.B. Phillips (2009b) Structural nonparametric cointegrating regression. *Econometrica* 77(6), 1901–1948.

Wang, Q. & P.C.B. Phillips (2011) Asymptotic theory for zero energy functionals with nonparametric regression applications. *Econometric Theory* 27(2), 235–259.

Wang, Q. & P.C.B. Phillips (2016) Nonlinear cointegrating regression with endogeneity and long memory. *Econometric Theory* 32(2), 359–401.

Wang, Q. & Y.X.R. Wang (2013) Nonparametric cointegrating regression with NNH errors. *Econometric Theory* 29(1), 1–27.

https://doi.org/10.1017/S0266466616000451 Published online by Cambridge University Press