A GENERAL LIMIT THEORY FOR NONLINEAR FUNCTIONALS OF NONSTATIONARY TIME SERIES

By

QIYING WANG AND PETER C. B. PHILLIPS

July 2022

COWLES FOUNDATION DISCUSSION PAPER NO. 2336

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

A GENERAL LIMIT THEORY FOR NONLINEAR FUNCTIONALS OF NONSTATIONARY TIME SERIES

BY QIYING WANG[1] AND PETER C. B. PHILLIPS[2]

[1] *The University of Sydney* qiying.wang@sydney.edu.au

[2] *Yale University, University of Auckland*
University of Southampton & Singapore Management University peter.phillips@yale.edu

Limit theory is provided for a wide class of covariance functionals of a nonstationary process and stationary time series. The results are relevant to estimation and inference in nonlinear nonstationary regressions that involve unit root, local unit root or fractional processes and they include both parametric and nonparametric regressions. Self normalized versions of these statistics are considered that are useful in inference. Numerical evidence reveals a strong bimodality in the finite sample distributions that persists for very large sample sizes although the limit theory is Gaussian. New self normalized versions are introduced that deliver improved approximations.

1. Introduction. Parametric and nonparametric regressions with nonstationary data have attracted considerable recent attention in statistical theory because of the prevalence of nonstationary time series in applied work across many different fields and the need for asymptotic theory to support methods of estimation and inference in the presence of nonstationarity. Much of this work has focussed on cointegrating regression where linkages between nonstationary processes and stationary innovations play an integral role in the notion of cointegration and its various extensions to fractional processes involving long memory time series. The literature in this area is now voluminous, as discussed in the recent work of Duffy and Kasparis (2021)[9] in this journal. In almost all of this literature a key role in the asymptotic development is played by sample covariance functionals that involve nonstationary processes and stationary time series. These functionals take similar but subtly different forms in parametric and nonparametric regressions that in both cases are critical in determining the limit theory needed for estimation, inference and specification testing in such regressions. The goal of the present paper is to accommodate these two forms in a general limit theory and analyze self normalized versions of the statistics that are useful in inference.

The formulation employed is as follows. Suppose an observable time series x_t is a scalar nonstationary process, either integrated $I(1)$, near $I(1)$, or a similar time series with fractional process innovations, as detailed in what follows, and $w_k = (w_{1k}, ..., w_{dk})$ is a sequence of stationary random vectors. The paper is concerned with sample quantities S_n of x_k and w_k defined by sample sums of nonlinear functions of x_k and w_k that take the general form

$$S_n = \sum_{k=1}^{n} f(x_k/h, w_k),$$

where $h \equiv h_n > 0$ is a sequence of positive constants indexed by the sample size n and $f(x, y)$ is a real function on R^{1+d}. The partial sum S_n is a scalar nonlinear functional of multivariate arguments that involve both stationary and nonstationary processes. Such functionals play a dominant role in the development of the theory of estimation and inference in

MSC2020 subject classifications: Primary 62F99, 62F99; secondary 62G99.
Keywords and phrases: Endogeneity, Limit theory, Local time, Nonlinear functional, Nonstationarity, Sample covariance, Zero energy.

nonlinear cointegrating regression, where the regressor is usually a nonstationary time series, including those with autoregressive unit roots and local unit root properties. In such regression contexts, a prominent example of S_n has the form of a sample covariance function that involves both the nonstationary regressor and the equation innovations. In this case, two covariance functions are most typical, one of the form $S_{1n} = \sum_{k=1}^n f(x_k, w_{2k}, ..., w_{dk})w_{1k}$ and the other of the form $S_{2n} = \sum_{k=1}^n f(x_k/h)w_{1k}$, where an auxiliary sequence $h = h_n$ may be present that depends on the sample size, as in nonparametric kernel regression.

As is now well known in the literature (see, for instance, Park and Phillips (2001)[18], Wang and Phillips (2009a[24], 2009b[25]), Chan and Wang (2015)[5], Dong and Linton (2018)[6], Duffy (2020)[8], Hu, et al. (2021)[13], and the many references therein), covariance expressions such as S_{1n} occur in nonlinear parametric cointegrating regression and expressions such as S_{2n} with the auxiliary sequence h arise naturally in Nadaraya-Watson estimation where $f(x)$ is a kernel function and $h \to 0$ is a bandwidth used in the nonparametric regression.

The limit behavior of S_n depends on the value of the integral $\int_{-\infty}^{\infty} g(s)\, ds$, where $g(x) = \mathbb{E}\, f(x, w_1)$. When $\int_{-\infty}^{\infty} g(s)\, ds \neq 0$, it was shown in Wang, Phillips and Kasparis (2021)[28] that upon suitable normalization S_n satisfies

$$\frac{d_n}{nh} S_n \to_D \int_{-\infty}^{\infty} g(x)dx\, L_{\mathcal{G}}(1, 0), \tag{1.1}$$

provided $d_n/nh \to 0$ and $d_n/h \to \infty$, with $d_n^2 = var(x_n)$ and where $L_{\mathcal{G}}(t, s)$ is the local time of a stochastic process $\mathcal{G}(t)$ at the spatial point s, as defined in the following section. Result (1.1) was established in quite general settings, generalizing and improving previous related work on convergence to local time given by Akonom (1993)[1], Borodin and Ibragimov (1995)[4], Phillips and Park (1998)[21], Jeganathan (2004)[14], Wang and Phillips (2009a[24], 2016[27]) and Duffy (2016)[7]. This fundamental limit result enabled the investigation of asymptotic theory for latent variable nonparametric cointegrating regression in which some variables were observed with measurement error.

The present work is concerned with developing a limit theory for the sample function S_n in the case where $\int_{-\infty}^{\infty} g(s)\, ds = 0$, commonly known as the zero-energy case. Towards this end in some specialized cases such as $f(x, y) = m(x)$ or $f(x, y) = m(x)\, y$, where $m(x)$ is bounded and integrable, the asymptotic behaviour of S_n is known and has been considered in Wang and Phillips (2009b[25], 2011[26]), with the attendant requirement that $h \to 0$, and in an unpublished manuscript by Jeganathan (2008)[15] (with $h = 1$). This paper provides a unified extension of these existing results that encompasses the two cases where $h = 1$ and $h \to 0$, together with the setting of general functionals $f(x, y)$ rather than the specialized forms $f(x, y) = m(x)\, y$ or $m(x)$.

It should be mentioned that the zero energy case where $\int_{-\infty}^{\infty} g(s)\, ds = 0$ [recall that $g(x) = \mathbb{E} f(x, w_1)$] arises naturally in regression applications. For instance, in nonparametric cointegrating regression, the development of a limit theory for normalized versions of functionals such as the sample covariance S_{2n} is vital for both estimation and inference. Thus, when x_k is an $I(1)$ regressor and w_{1k} is an error process, use of the natural centralizing condition $\mathbb{E}w_{11} = 0$ in turn implies that $\int_{-\infty}^{\infty} g(s)\, ds = \int_{-\infty}^{\infty} f(x)dx\, \mathbb{E}w_{11} = 0$. Such situations arise even in complex settings where endogeneity is present - see Wang and Phillips (2009b[25], 2011[26], 2016[27]) for details and econometric applications. Similarly, in regression with nonstationary nonlinear heteroskedasticity when nonstationary volatility is present in the errors [with $u_t = f(x_t, w_t)$, say], the zero energy condition $\int_{-\infty}^{\infty} g(s)\, ds = 0$ where $g(x) = \mathbb{E} f(x, w_1)$ is usually required for the development of an asymptotic theory. In this case, the use of general functionals such as $f(x, y)$ in the sample covariance limit theory enables a full representation of nonstationary nonlinear volatility in the regression errors.

The paper is organized as follows. Section 2 provides the main limit theory for nonlinear functionals of non-stationary time series and a series of remarks that analyze the findings and connect to later discussion. Section 3 provides numerical evidence which reveals an intriguing bimodality for self-normalized statistics that arises in finite samples and that can persist in extremely large samples even though the limit theory is Gaussian. Section 4 discusses these findings, explains the slow convergence, and shows how bimodal limit theory does arise in the presence nonstationary long memory innovations. Alternative self-normalized statistics are considered that substantially improve finite sample performance. Concluding remarks are in Section 5. Proofs of the main results are given in Section 6 and supporting propositions and lemmas that play key roles in proving the main results are in Section 7. Proofs of the lemmas are in the Appendix.

Throughout the paper $\Rightarrow$ denotes weak convergence of probability measures with respect to the uniform topology (see, for instance, Billingsley (1968)[2]) and $\to_D$ is distributional convergence in Euclidean space. For a vector $A = (A_1, .., A_d)$, we define $||A|| = |A_1| + ... + |A_d|$. Constants are represented by $C, C_1, C_2, ...$, which may differ in different locations.

2. Main Results.

2.1. *Assumptions and Preliminaries.* Let $\lambda_i = (\epsilon_i, e_i)'$, $i \in \mathbb{Z}$ be a sequence of iid random vectors with $\mathbb{E}||\lambda_0||^2 < \infty$. Let $\xi_k = \sum_{j=0}^{\infty} \phi_j \epsilon_{k-j}$ be a linear process where the coefficients $\phi_k, k \geq 0$, satisfy $\phi_0 \neq 0$ and one of the following conditions:

LM: $\phi_k \sim k^{-\mu}\rho(k), 1/2 < \mu < 1$ and $\rho(k)$ is a function slowly varying at ∞;
SM: $\sum_{k=0}^{\infty} |\phi_k| < \infty$ and $\phi \equiv \sum_{k=0}^{\infty} \phi_k \neq 0$.

The following assumptions are made about the components of $S_n = \sum_{k=1}^{n} f(x_k/h, w_k)$ for the development of the asymptotic theory in our main results.

A1 (i) $x_k = \rho_n x_{k-1} + \xi_k$, where $x_0 = 0$, $\rho_n = 1 - \gamma n^{-1}$ for some constant $\gamma \geq 0$;
(ii) $\mathbb{E}\epsilon_1 = 0$ and $\int_{-\infty}^{\infty} |\mathbb{E}e^{it\epsilon_1}|dt < \infty$.

A2 (a) $w_k = (w_{1k}, ..., w_{dk})$, where $w_{ik} = \Gamma_i(\lambda_k, ..., \lambda_{k-m_0})$ for some fixed $m_0 \geq 0$ and $\Gamma_i(.), i = 1, 2, ..., d$, are real measurable functions of their respective components;
(b) $\mathbb{E}\,||w_1||^{\max\{2,4\beta\}} < \infty$, where β is given in **A3**(I) below.

A3 (I) A bounded function $T(x)$ exists such that, for some $\beta > 0$,

$$|f(x,y)| \leq T(x)(1 + ||y||^{\beta}) \quad \text{and} \quad \int_{-\infty}^{\infty} (1 + |x|)T(x)dx < \infty;$$

(II) $\int_{-\infty}^{\infty} g(x)dx = 0$, where $g(x) = \mathbb{E}\,f(x, w_1)$;
(III) $\int_{-\infty}^{\infty} \mathbb{E}\,|\hat{f}(x, w_1)|dx < \infty$, where $\hat{f}(x, y) = \int_{-\infty}^{\infty} e^{itx} f(t, y)dt$.

Assumption **A1**(i) accommodates near integrated time series x_k that are derived from either short memory (under **SM**) or long memory (under **LM**) innovations, thereby covering a large class of nonstationary time series. The extra distributional assumption **A1**(ii) is a smoothness condition requiring integrability of the characteristic function $\mathbb{E}e^{it\epsilon_1}$ that is often useful in establishing convergence to a local time process. The condition can be relaxed to $\limsup_{|t|\to\infty} |t|^a \mathbb{E}e^{it\epsilon_1}| < \infty$ for some $a > 0$, but is generally difficult to eliminate completely in the development of limit theory for nonlinear cointegrating regression. The zero initialization $x_0 = 0$ is assumed for convenience to avoid notational clutter and can be considerably relaxed, as is well known from earlier research. In particular, all the main results still hold if instead $x_0 = o_P(d_n)$, where $d_n^2 = var(\sum_{k=1}^n \xi_k)$. It is also well-known (see Wang, Lin and Gulati (2003)[23], for instance) that

$$d_n^2 \sim \mathbb{E}\epsilon_0^2 \begin{cases} c_\mu \, n^{3-2\mu} \, \rho^2(n), & \text{under } \mathbf{LM}, \\ \phi^2 \, n, & \text{under } \mathbf{SM}, \end{cases}$$

and $x_{\lfloor nt \rfloor}/d_n \Rightarrow Z_t$ on $D[0,1]$, where $c_\mu = \frac{1}{(1-\mu)(3-2\mu)} \int_0^\infty x^{-\mu}(x+1)^{-\mu} dx$ and

$$Z_t = W(t) + \gamma \int_0^t e^{-\gamma(t-s)} W(s) ds, \quad t \geq 0$$

$$W(t) = \begin{cases} B_{3/2-\mu}(t), & \text{under } \mathbf{LM}, \\ B_{1/2}(t), & \text{under } \mathbf{SM}, \end{cases}$$

and $B_H(t)\}$ is fractional Brownian motion with Hurst exponent H and $B_{1/2}(t)$ is standard Brownian motion. In this event, Z_t is a fractional Ornstein-Uhlenbeck process, having a continuous local time process which we denote by $L_Z(t,x)$. As in Geman and Horowitz (1980)[12], the local time process $L_Z(t,x)$ is defined as

$$L_Z(t,x) = \lim_{\epsilon \to 0} \frac{1}{2\epsilon} \int_0^t I\big(|Z_r - x| \leq \epsilon\big) dr. \tag{2.1}$$

These notations will be used subsequently without further explanation.

Assumption **A2** ensures that $w_k, \ k \geq 1$, is a sequence of stationary random vectors. No restriction is imposed on the relationship between ϵ_k and e_k of $\lambda_k = (\epsilon_k, e_k)'$, which enables the results established here to be widely applicable in nonlinear cointegrating regression models with endogeneity. The extension of **A2** to include linear process formulations is possible if $f(x,y)$ has certain structure. We refer to Corollary 2.1 for further details on this extension.

Finally, Assumption **A3** provides conditions on the function $f(x,y)$. These, together with **A2**(b), ensure that,

$$\int_{-\infty}^{\infty} \big[\mathbb{E}f^2(x,w_1) + \mathbb{E}f^4(x,w_1)\big] dx \leq C\, \mathbb{E}\, ||w_1||^{\max\{2,4\beta\}} \int_{-\infty}^{\infty} T(x) dx < \infty, \tag{2.2}$$

the Fourier transform $\hat{f}(t,y) = \int_{-\infty}^{\infty} e^{itx} f(x,y) dx$ is well defined, $\sup_x g(x) < \infty$, $\int |g(x)| dx \leq \int \mathbb{E}\,|f(x,w_1)| dx < \infty$, and $\int_{-\infty}^{\infty} (1+|x|) \mathbb{E}\,|f(x,w_1)| dx < \infty$. Furthermore, it follows from $\mathbb{E}\hat{f}(0,w_1) = \int_{-\infty}^{\infty} \mathbb{E}\, f(x,w_1) dx = 0$ that

$$|\mathbb{E}\hat{f}(t,w_1)| \leq \int_{-\infty}^{\infty} \big| \big(e^{itx} - 1\big) \mathbb{E} f(x,w_1) \big| \, dx \leq C \min\{1, |t|\}. \tag{2.3}$$

On the other hand, using the inverse Fourier transformation, **A3**(III) ensures the representation of $f(x,w_k)$, almost surely,

$$f(x,w_k) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \hat{f}(t,w_k) dt. \tag{2.4}$$

These properties will be used in the main results that follow without further reference.

2.2. *Asymptotic theory.* Our main result is as follows.

THEOREM 2.1. *Suppose* **A1** – **A3** *hold. For any* $h \equiv h_n \to 0$ *satisfying* $nh/d_n \to \infty$, *we have*

$$\Big(\frac{d_n}{nh} \sum_{k=1}^{\lfloor nt \rfloor} f^2\big(x_k/h, w_k\big), \ \Big(\frac{d_n}{nh}\Big)^{1/2} \sum_{k=1}^{\lfloor nt \rfloor} f\big(x_k/h, w_k\big) \Big)$$

$$\Rightarrow \ \big(\tau^2 \, L_Z(t,0), \ \tau \, \mathbb{N} \, L_Z^{1/2}(t,0)\big), \tag{2.5}$$

on $D_{R^2}[0,1]$, where $\tau^2 = \int_{-\infty}^{\infty} \mathbb{E}\, f^2(s, w_1) ds$, and $\mathbb{N}$ is a standard normal variate independent of $L_Z(t,0)$ for $0 \le t \le 1$.

*If in addition $\gamma = 0$, where γ is used in **A1** (i), and $\int_{-\infty}^{\infty} \mathbb{E}\big\{|\hat{f}(t, w_0)\big(1 + ||w_r||^{\beta}\big)\big\} dt < \infty$ for any $r \ge 0$, then*

$$\Big(\frac{d_n}{n}\sum_{k=1}^{\lfloor nt\rfloor} f^2\big(x_k, w_k\big),\ \Big(\frac{d_n}{n}\Big)^{1/2}\sum_{k=1}^{\lfloor nt\rfloor} f(x_k, w_k)\Big)$$

$$\Rightarrow\ \big(\tau^2\, L_Z(t,0),\ \tau_1\, \mathbb{N}\, L_Z^{1/2}(t,0)\big), \tag{2.6}$$

on $D_{R^2}[0,1]$ (recall $Z_t = W(t)$ when $\gamma = 0$), where $\tau_1^2 = G_0 + 2\sum_{r=1}^{\infty} G_r$ with

$$G_r = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathbb{E}\big\{\hat{f}(s, w_0)\hat{f}(s, w_r) e^{-isx_r}\big\} ds$$

$$= \int_{-\infty}^{\infty} \mathbb{E}\big\{f(y, w_0) f(y + x_r, w_r)\big\} dy. \tag{2.7}$$

REMARK 2.1. *Different constants τ and τ_1 appear in the second components of results (2.5) and (2.6). In fact, as $h \to 0$, we have*

$$\frac{d_n}{nh}\sum_{k=1}^{n}\sum_{j=k+1}^{n} \mathbb{E}\big\{f(x_k/h, w_k) f(x_{k+j}/h, w_{k+j})\big\} = o(1),$$

(see the proof of (7.2) in Proposition 7.3); but when $h = 1$ and $\gamma = 0$

$$\frac{d_n}{n}\sum_{k=1}^{n} f(x_k, w_k) f(x_{k+j}, w_{k+j}) \to_D G_j L_Z(1,0), \tag{2.8}$$

for any $j \ge 1$ (see (7.5) of Proposition 7.4). These facts indicate that the influence of cross product terms such as $f(x_k/h, w_k) f(x_{k+j}/h, w_{k+j})$ on the variance of $\big(\frac{d_n}{nh}\big)^{1/2}\sum_{k=1}^{\lfloor nt\rfloor} f\big(x_k/h, w_k\big)$ is eliminated as $h \to 0$, but this is not the case when $h = 1$. In consequence, different constants appear in the two results (2.5) and (2.6). In addition to (2.6), the following joint convergence holds in which, for any $q > 0$,

$$\Big(\frac{d_n}{n}\sum_{k=1}^{\lfloor nt\rfloor} f^2\big(x_k, w_k\big),\ \frac{d_n}{n}\sum_{k=1}^{\lfloor nt\rfloor} f\big(x_k, w_k\big) f\big(x_{k+1}, w_{k+1}\big), ...,$$
$$\frac{d_n}{n}\sum_{k=1}^{\lfloor nt\rfloor} f\big(x_k, w_k\big) f\big(x_{k+q}, w_{k+q}\big),\ \Big(\frac{d_n}{n}\Big)^{1/2}\sum_{k=1}^{\lfloor nt\rfloor} f(x_k, w_k)\Big)$$

$$\Rightarrow\ \big(\tau^2\, L_Z(t,0),\ G_1\, L_Z(t,0), ...,\ G_q\, L_Z(t,0),\ \tau_1\, \mathbb{N}\, L_Z^{1/2}(t,0)\big), \tag{2.9}$$

on $D_{R^{q+1}}[0,1]$. The proof of (2.9) involves only minor additions to that of (2.6) and the details are omitted.

REMARK 2.2. *In special cases where $f(x,y) = K(x)\,y$ (with $K(x)$ bounded and integrable) and $f(x,y) = K(x)$ (with $\int K(x) dx = 0$ and $K(x)$ bounded and integrable), a similar result to (2.5) has been considered in Wang and Phillips (2009b)[25] and Wang and Phillips (2011)[26], respectively, and a similar result to (2.6) can be found in Jeganathan (2008)[15]. Theorem 2.1 provides a unified generalization of these existing results to functional limit theorems. Our proof makes use of the methodology initially developed in Wang and Phillips (2009b)[25], which seems simpler than that used in Jeganathan (2008)[15].*

REMARK 2.3. *The quantity m_0 given in* **A2** *(a) is set to be a fixed constant, but it can be chosen as large as required in applications. Further, careful examination the proof reveals that the result continues to hold when $m_0 = m_n \to \infty$ provided the expansion rate is slow enough. Moreover, when $f(x,y) = K(x)y$, the stationary component w_k in Theorem 2.1 can be extended to include linear processes and endogeneity, as the following corollary shows, thereby covering regression models with errors u_t and regressors x_t that allow for endogeneity.*

COROLLARY 2.1. *In addition to* **A1**, *suppose that*

(a) *$K(x)$ is a bounded continuous function satisfying $\int K(x)dx < \infty$ and $\int |\hat{K}(x)|dx < \infty$, where $\hat{K}(x) = \int e^{ixs}K(s)ds$;*
(b) *$u_k = \sum_{j=0}^{\infty} \psi_j \lambda_{k-j}$, where $\mathbb{E}\lambda_1 = 0$, $\mathbb{E}||\lambda_1||^4 < \infty$ and the coefficient vector $\psi_k = (\psi_{k1}, \psi_{k2})$ satisfies $\sum_{k=0}^{\infty} k(|\psi_{1k}| + |\psi_{2k}|) < \infty$ and $\sum_{k=0}^{\infty} \psi_k \neq 0$.*

For any $h \equiv h_n \to 0$ satisfying $nh/d_n \to \infty$, we have

$$\Big(\frac{d_n}{nh}\sum_{k=1}^{n} K^2\big(x_k/h\big)u_k^2,\ \Big(\frac{d_n}{nh}\Big)^{1/2}\sum_{k=1}^{n} K\big(x_k/h\big)\,u_k\Big)$$

$$\to_D \big(\tilde{\tau}^2 L_Z(1,0),\ \tilde{\tau}\,\mathbb{N}\,L_Z^{1/2}(1,0)\big), \tag{2.10}$$

where $\tilde{\tau}^2 = \int_{-\infty}^{\infty} K^2(s)ds\,\mathbb{E}\,u_1^2$ and $\mathbb{N}$ is a standard normal variate independent of $L_Z(1,0)$.
If $h = 1$ and in addition $\gamma = 0$, where γ is used in **A1** *(i), then*

$$\Big(\frac{d_n}{n}\sum_{k=1}^{n} K^2\big(x_k\big)u_k^2,\ \frac{d_n}{n}\,\mathcal{J}_n,\ \Big(\frac{d_n}{n}\Big)^{1/2}\sum_{k=1}^{n} K\big(x_k\big)u_k\Big)$$

$$\to_D \big(\tilde{\tau}^2 L_Z(1,0),\ \tilde{\tau}_1^2 L_Z(1,0),\ \tilde{\tau}_1\,\mathbb{N}\,L_Z^{1/2}(1,0)\big), \tag{2.11}$$

where, for some $M = M_n \to \infty$,

$$\mathcal{J}_n = \sum_{k=1}^{n} K^2\big(x_k\big)u_k^2 + 2\sum_{j=1}^{M} \ell\Big(\frac{j}{M}\Big)\sum_{k=1}^{n-j} K\big(x_k\big)K\big(x_{k+j}\big)\,u_k\,u_{k+j}, \tag{2.12}$$

takes the form of a heteroskedastic and autocorreltation consistent (HAC) estimator in which $\ell(\frac{j}{M})$ is a lag kernel weight function such as the Bartlett triangular kernel $\ell(\frac{j}{M}) = 1 - \frac{|j|}{M}$, and where $\tilde{\tau}_1^2 = \tilde{G}_0 + 2\sum_{r=1}^{\infty}\tilde{G}_r$ with

$$\tilde{G}_r = \frac{1}{2\pi}\int_{-\infty}^{\infty} |\hat{K}(s))|^2\mathbb{E}\big\{u_0\,u_r\,e^{-isx_r}\big\}ds = \int_{-\infty}^{\infty} K(y)\mathbb{E}\big\{u_0\,u_r\,K(y+x_r)\big\}dy.$$

REMARK 2.4. *Result (2.10) coincides with (7.4) of Proposition 7.1 in Wang and Phillips (2016)[27] but with less restrictions on h (the requirement $h\log n \to 0$ used there is removed here), indicating the following self-normalized result: as $h \to 0$ and $nh/d_n \to \infty$,*

$$J_n(h) := \frac{\sum_{k=1}^{n} K\big(x_k/h\big)\,u_k}{\sqrt{\sum_{k=1}^{n} K^2\big(x_k/h\big)u_k^2}} \to_D \mathcal{N}(0,1). \tag{2.13}$$

In view of the standard normal asymptotics this result is convenient and useful for purposes of estimation and inference in nonparametric regression models involving nonstationary time series and kernel estimation with a shrinking bandwidth parameter $h \to 0$. Result (2.11) with fixed $h = 1$ is similar to that of Theorem 5 in Jeganathan (2008)[15]. In this case, a *suitable self-normalized version of the sample covariance statistic can be constructed from the elements of (2.11) and (2.12) as*

$$J_n^*(1) := \mathcal{J}_n^{-1/2} \sum_{k=1}^{n} K(x_k)\, u_k \to_D \mathcal{N}(0,1), \tag{2.14}$$

which again has standard normal asymptotics making the formulation convenient in applications that involve nonlinear parametric regressions with nonstationary time series. We mention that, the result that $\mathcal{J}_n^2 \to_D \tilde{\tau}_1^2 L_Z(1,0)$ holds for any continuous function $\ell(x)$ satisfying $\ell(0)=1$, although we assume here that $\ell(\frac{j}{M})$is a lag kernel weight function to ensure the positivity of $\mathcal{J}_n$ in finite samples. Furthermore, we only prove (2.11) for some $M_n \to \infty$. The result should be true for all $M_n \le n$ and $M_n \to \infty$. In fact, when $l(x) \equiv 1$, the latter claim can be established by using (7.3) in Proposition 7.3. But the proof for a general continuous function $\ell(x)$ will involve complex calculations and the details are omitted.

REMARK 2.5. *While these naturally constructed self-normalized statistics have elegant Gaussian limit results, numerical work (reported below in Section 3) reveals that neither (2.13) nor (2.14) perform well in finite sample simulations. In particular, when x_t is generated with long memory innovations in ξ_t and the memory parameter is large (μ close to 0.5), bimodality appears in the finite sample densities even when the sample size is as large as $n = 5,000$. Such bimodality is known to arise with self-normalized statistics and t ratios in other contexts, especially in the presence of heavy tailed data where individual large draws can dominate both the numerator and the denominator in the ratio (see Logan et al (1972)[16] and Fiorio et al (2010)[10]). The explanation of the phenomena in the present setting is unrelated to heavy tails but is instead related to strong dependence in the data. Heuristically, strong memory when μ is close to 0.5 ensures that the weight function $K(x_k)$ is generally so small that only a limited number of terms dominate the numerator and denominator summations $\sum_{k=1}^n K(x_k)u_k$ and $\sum_{k=1}^n K^2(x_k)u_k^2$ (see Fig. 4 for illustrative trajectories), thereby inducing bimodality in the finite sample densities of $\mathcal{J}_n^*(1)$ around the modes ± 1. To control this behavior, a modification of (2.14) such as the following*

$$\widehat{J_n^*}(1) := \widehat{\mathcal{J}_n}^{-1/2} \sum_{k=1}^{n} K(x_k)\, u_k \to_D \mathcal{N}(0,1), \tag{2.15}$$

might be considered where $\mathcal{J}_n$ in (2.12) is replaced by

$$\widehat{\mathcal{J}_n} = \widehat{\sigma}_n^2 \sum_{k=1}^{n} K^2(x_k) + 2\sum_{j=1}^{M} \ell\left(\frac{j}{M}\right) \sum_{k=1}^{n-j} K(x_k) K(x_{k+j})\, u_k\, u_{k+j}, \tag{2.16}$$

for some consistent estimator $\widehat{\sigma}_n^2$ of $\sigma^2 = \mathbb{E}u_1^2$ and with $M = M_n \to \infty$ as $n \to \infty$. The advantage of $\widehat{\mathcal{J}_n}$ is that the use of $\widehat{\sigma}_n^2 \sum_{k=1}^n K^2(x_k)$ in the first term, rather than $\sum_{k=1}^n K^2(x_k^2) u_k^2$, attenuates the bimodality induced by the numerator and denominator summations $\sum_{k=1}^n K(x_k)u_k$ and $\sum_{k=1}^n K^2(x_k)u_k^2$ discussed above and in the heuristic analysis of (3.4) below. However, the estimate $\widehat{\mathcal{J}_n}$ in (2.16) is not necessarily positive. For instance, in 40,000 replications when $n = 100$ around 15 cases of negative values occur with $d = 0.1$, rising to 60 cases with $d = 0.55$. To address this difficulty the following adjustment to (2.16) is employed

$$\widehat{\mathcal{J}}_{nM^*} = \widehat{\sigma}_n^2 \sum_{k=1}^{n} K^2(x_k) + 2\sum_{j=1}^{M^*} \ell\left(\frac{j}{M}\right) \sum_{k=1}^{n-j} K(x_k) K(x_{k+j})\, u_k\, u_{k+j}, \tag{2.17}$$

where

$$M^* := M \times \mathbb{1}\big(\widehat{\mathcal{J}_n} \geq 0\big) + M^* \times \mathbb{1}\big(\widehat{\mathcal{J}_n} < 0\big)\mathbb{1}\big(\widehat{\mathcal{J}}_{nM^*} > 0\big), \tag{2.18}$$

in which the truncation lag number M is reduced by one lag at a time when $\widehat{\mathcal{J}_n} < 0$ to the first value M^* for which $\widehat{\mathcal{J}}_{nM^*} > 0$. In 50,000 replications with n=100 and n=1,000 the modification (2.17), with the simple rule (2.18), was found to work well. Using $\widehat{\mathcal{J}}_{nM^*}$ in place of $\widehat{\mathcal{J}_n}$ leads to the same standard normal asymptotics as (2.15) for the statistic

$$\widetilde{J_n}(1) := \widehat{\mathcal{J}}_{nM^*}^{-1/2} \sum_{k=1}^{n} K\big(x_k\big)\, u_k \to_D \mathcal{N}(0,1), \tag{2.19}$$

provided $M^* \to \infty$ as $n \to \infty$. Simulation results for the statistic $\widetilde{J_n}(1)$ are shown in Fig. 3 in the following numerical section and confirm that the statistic removes bimodality in finite samples and has distributions considerably closer to the standard normal limit than the statistic $J_n^*(1)$ in (2.14) for various values of the long memory parameter d and samples as small as $n = 100$.

Similarly, we may use the following result instead of (2.13): as $h \to 0$ and $nh/d_n \to \infty$,

$$\widehat{J}_n(h) := \frac{\sum_{k=1}^{n} K\big(x_k/h\big)\, u_k}{\sqrt{\widehat{\sigma}_n^2 \sum_{k=1}^{n} K^2\big(x_k/h\big)}} \to_D \mathcal{N}(0,1). \tag{2.20}$$

The proofs of (2.15) and (2.20) follow easily from (2.14), (2.13) and the following fact by using (4.8) of Wang, et al (2021)[28] [see also (7.42) in the proof of Proposition 7.4 with $f(x,y) = K(x)y$]: for any $h > 0$,

$$\frac{d_n}{nh} \sum_{k=1}^{n} K^2\big(x_k/h\big)\big(\mathbb{E}u_k^2 - u_k^2\big) = o_P(1). \tag{2.21}$$

The details are omitted.

REMARK 2.6. *It is of some interest to investigate the asymptotics of $\sum_{k=1}^n f\big(x_k/h, w_k\big)$ when $h \to \infty$. In this case, it seems that new techniques may be required. As an illustration, suppose $f(x,y) = K(x)y$ with $K(x) = e^{-x^2}$ and the ϵ_i are iid with $\mathbb{E}\epsilon_1 = 0$, $|\epsilon_i| \leq 1$, and $x_k = \sum_{j=1}^k \epsilon_j$. Using Taylor expansion as $h \to \infty$ we can write*

$$f(x_k/h, \epsilon_k) = K(x_{k-1}/h)\epsilon_k + h^{-1}\, K'(x_{k-1}/h)\epsilon_k^2 + h^{-2}\, R(x_{k-1}/h, \epsilon_k),$$

where $|R(x,y)| \leq C\,(1+x^2)^{-1}|y|$. It is easy to show that, as $h \equiv h_n \to \infty$ and $h/n \to 0$,

$$\frac{h}{\sqrt{nh}} \sum_{k=1}^{n} f(x_k/h, \epsilon_k) \to_D \mathbb{E}\epsilon_1^2 \int K'(x)dx L_B(1,0), \tag{2.22}$$

where $B = \{B_t\}_{t\geq 0}$ is standard Brownian motion. Note that $\frac{h}{\sqrt{nh}} \neq (d_n/nh)^{1/2}$ (in fact, $(d_n/nh)^{1/2} = 1/(nh^2)^{1/4}$ for this example). In comparison with (2.5) and (2.6), result (2.22) has a different convergence rate, indicating that a different approach and technique must be used in investigating the asymptotics of $\sum_{k=1}^n f\big(x_k/h, w_k\big)$ when $h \to \infty$. A general analysis of such cases is an interesting topic for future research.

3. Numerical evidence. We explore the finite sample properties of the self-normalized statistics J_n and $J_n^*(1)$ defined as in (2.13) and (2.14). Since earlier research has considered models with shrinking bandwidths $h \to 0$, the model employed here focuses mainly on the case $h = 1$ for which the general limit theory is given in (2.9). As indicated above, the key difference in this case is that the cross product term (2.8) is not eliminated when $h \not\to 0$. The statistic $J_n^*(1)$ takes this into account by estimating the appropriate self-normalizing quantity. As is apparent from (2.9) and (2.11) the limiting form of the denominator of $J_n^*(1)$ has the form of a long run self-normalization, with the major difference that in the present case this quantity has a random limit since $\mathcal{J}_n \to \tilde{\tau}_1^2 L_Z(1,0)$ as $n \to \infty$ in place of the usual non-random quantity that arises in standard problems with stationary short memory time series.

In the simulations here, x_t is generated according to **A1** with autoregressive coefficient $\rho_n = 1$. The linear process $\xi_t = \sum_{j=0}^{\infty} \phi_j \epsilon_{t-j}$ in **LM** is generated using the fractional integration mechanism $\xi_t = (1-L)^{-d}\epsilon_t = \sum_{j=0}^{\infty} \frac{(d)_j}{j!}\epsilon_{t-j}$, where $(d)_j = \frac{\Gamma(d+j)}{\Gamma(1+j)}$, so that $\phi_j \sim \frac{1}{\Gamma(d) j^{1-d}}$, where $\Gamma(\cdot)$ is the gamma function and the memory parameter $d = 1-\mu \in (0, 0.5)$. Endogeneity in x_t is introduced by defining the innovations in the linear process ξ_t by $\epsilon_t = (1-\rho^2)^{1/2}\epsilon_{xt} + \rho u_t$ where u_t is the short memory autoregressive process $u_t = \theta u_{t-1} + e_{ut}$, $|\theta| < 1$, with $e_{ut} \sim_{iid} \mathcal{N}(0,1)$ and independent of $\epsilon_{xt} \sim_{iid} \mathcal{N}(0,1)$. With this specification of u_t we have

$$\begin{aligned}
\xi_t &= \sum_{j=0}^{\infty} \phi_j \epsilon_{t-j} = (1-\rho^2)^{1/2} \sum_{k=0}^{\infty} \phi_k \epsilon_{x\,t-k} + \rho \sum_{j=0}^{\infty} \phi_j \sum_{\ell=0}^{\infty} \theta^{\ell} \epsilon_{u\,t-j-\ell}\Bigg] \\
&= (1-\rho^2)^{1/2} \sum_{k=0}^{\infty} \phi_k \epsilon_{x\,t-k} + \rho \sum_{k=0}^{\infty} \left(\sum_{\ell=0}^{k} \phi_{k-\ell} \theta^{\ell} \right) \epsilon_{u\,t-k} \\
&= \sum_{k=0}^{\infty} \left[\bar{\psi}_{1k} \epsilon_{x\,t-k} + \bar{\psi}_{2k} \epsilon_{u\,t-k} \right] \qquad (3.1)
\end{aligned}$$

with $\bar{\psi}_{1k} = (1-\rho^2)^{1/2}\phi_k$ and $\bar{\psi}_{2k} = \rho \sum_{\ell=0}^{k} \phi_{k-\ell}$. The innovation ξ_t has long memory parameter d and endogeneity measured through the correlation coefficient ρ.

The self-normalized statistics $J_n(h)$, $J_n(1)$, and $J_n^*(1)$ defined in (2.13) and (2.14) are computed for $f(x_t/h, w_t) = K(x_t/h)u_t$ with $h = 2/n^{0.2}$ or $h = 1$. In the following computations we used $K(x) = (1/\sqrt{2\pi})e^{-x^2/2}$, $\theta = 0.5$, $\rho = 5.0$ and $d \in \{0.1, 0.25, 0.4, 0.55\}$, where $d = 0.55$ lies in the nonstationary long memory region and is included for comparison. Kernel estimates of the densities of $J_n(h)$ were computed using

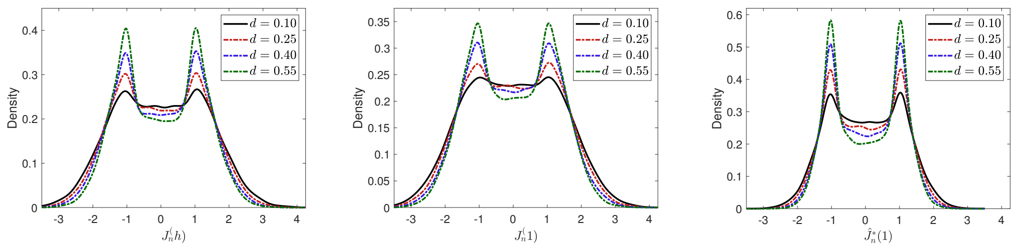
$$J_n(h) = \frac{\sum_{k=1}^{n} K(x_k/h)\, u_k}{\sqrt{\sum_{k=1}^{n} K^2(x_k/h) u_k^2}}, \qquad (3.2)$$

for $h = 2/n^{0.2}$ and $h = 1$ and are shown in Figs. 1(a) and 1(b). The self normalized statistic $J_n^*(1)$ was computed by the explicit formula

$$J_n^*(1) = \frac{\sum_{k=1}^{n} K(x_k)\, u_k}{\left[\sum_{k=1}^{n} K^2(x_k) u_k^2 + 2\sum_{j=1}^{M} \ell\left(\frac{j}{M}\right) \sum_{k=1}^{n-j} K(x_k) K(x_{k+j})\, u_k\, u_{k+j} \right]^{1/2}}. \qquad (3.3)$$

with lag truncation parameter $M = \lfloor 2n^{1/6} \rfloor$ and its densities are shown in Figs. 1(c) and 2(c). The number of replications employed was $40{,}000$, with sample size $n = 100$ in Fig. 1 and $n = 1{,}000$ in Fig. 2.

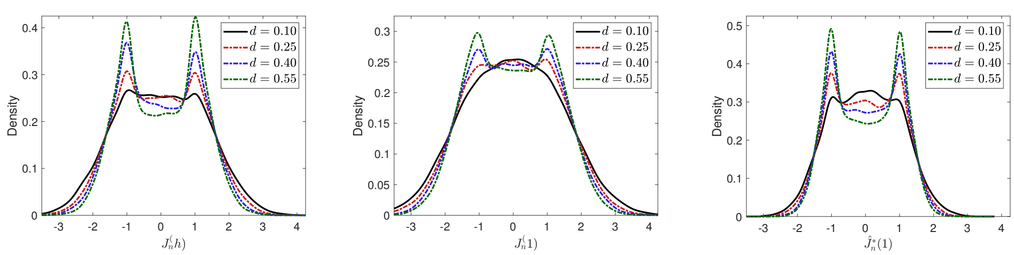
The densities in Fig. 1 where $n = 100$ are all non-normal. Bimodality with modes around ± 1 are clearly evident in all cases and all values of d. For $J_n(1)$ the dual modes are evident but somewhat less pronounced than for $J_n(h)$ with $h = 2/n^{0.2}$. The bimodality is clearly stronger in the presence of nonstationary long memory innovations ξ_t with $d = 0.55$ (shown by dashed green lines). Bimodality is most prominent and with greatest concentration for the statistic $J_n^*(1)$. Bimodality is evidently weaker for the lower memory parameters, particularly cases where $d = 0.10$ (shown by black unbroken lines).



(a) $J_n(h)$ densities (b) $J_n(1)$ densities (c) $J_n^*(1)$ densities

Fig 1: Empirical densities of $J_n(h)$ with $h = \frac{2}{n^{0.2}}$, $J_n(1)$, and $J_n^*(1)$ for sample size $n = 100$ and $d \in \{0.10, 0.25, 0.40, 0.55\}$.

In Fig. 2 the densities are computed for $n = 1,000$. In Fig. 2(a) bimodality is clearly evident for $J_n(h)$, applies for all values of d and is again stronger in the nonstationary case. The densities of $J_n(1)$ and $J_n^*(1)$ in Figs. 2(b) and 2(c), where $n = 1,000$, are closer to normal than when $n = 100$ except for the nonstationary innovation case ($d = 0.55$); and bimodality is still more pronounced for $J_n^*(1)$ than for $J_n(1)$. When $d = 0.1$, there are no apparent modes in the density of $J_n(1)$ and only minor modes in the density of $J_n^*(1)$. Nonetheless, convergence to normality when $0 < d < 0.5$ appears slow and shape differences in the densities persist between the stationary and nonstationary error cases. The tendency to bimodality continues to be more marked in the nonstationary case.



(a) $J_n(h)$ densities (b) $J_n(1)$ densities (c) $J_n^*(1)$ densities

Fig 2: Empirical densities of $J_n(h)$, $J_n(1)$, and $J_n^*(1)$ for sample size $n = 1,000$ and $d \in \{0.10, 0.25, 0.40, 0.55\}$.

As discussed in Remark 2.5, when the innovations ξ_k have strong dependence with memory parameter d close to the nonstationary boundary 0.5, the weight function $K(x_t)$ is negligible except for a very small number of terms in which $x_t = \sum_{k=1}^t \xi_k \approx 0$. Suppose x_t is closest to zero for $t = \tau$ then $K(x_\tau) \approx 1$ and so $J_n(1) \approx \pm 1$, thereby inducing a tendency to bimodality in the finite sample densities of $\mathcal{J}_n(1)$ around modes at ± 1. When $h \to 0$ this facet of the weight function is accentuated for $K(x_t/h)$ and we may therefore expect greater evidence of bimodality in finite samples for $J_n(h)$, which is corroborated by the results in Figs. 1(a) and 2(a).

Further, in Figs. 1 and 2 it is evident that $J_n^*(1)$ shows more evidence of bimodality than $J_n(1)$. This may be explained by the following heuristic. Suppose x_t is closest to zero in the sample at $t = \tau$ and next closest to zero at $t = \tau + 1$, so that $K(x_\tau) \approx K(0) \approx 1/\sqrt{2\pi}$ and then $K(x_{\tau+1}) \approx K(\xi_{\tau+1}) = e^{-\xi_{\tau+1}^2/2}/\sqrt{2\pi}$. (Fig. 4 below shows an illustrative case). With a Bartlett kernel $\ell(\cdot)$ we then have

$$J_n^*(1) \approx \frac{K(x_\tau)u_\tau + K(x_{\tau+1})u_{\tau+1}}{\left[K(x_\tau)^2 u_\tau^2 + K(x_{\tau+1})^2 u_{\tau+1}^2 + 2\left(1 - \frac{1}{M}\right) K(x_\tau)K(x_{\tau+1})u_\tau u_{\tau+1}\right]^{1/2}}$$

$$= \frac{K(x_\tau)u_\tau + K(x_{\tau+1})u_{\tau+1}}{|K(x_\tau)u_\tau + K(x_{\tau+1})u_{\tau+1}| + O_p\left(\frac{1}{M}\right)} = \pm 1 + O_p\left(\frac{1}{M}\right), \tag{3.4}$$

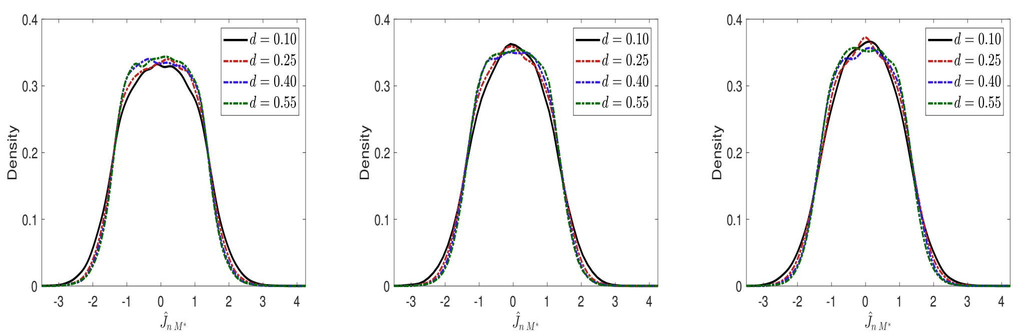
showing a clear tendency to bimodality.

Next note that $\xi_t = (1-L)^{-d}\epsilon_t$ has variance $\sigma_\xi^2 = \sigma_\epsilon^2 \frac{\Gamma(1-2d)}{\Gamma(1-d)^2} \sim_a \frac{\sigma_\epsilon^2/\pi}{1-2d} \to \infty$ as $d \to 0.5$. Let $\xi_t = \sigma_\xi \tilde{\xi}_t$ where $\tilde{\xi}_t$ has unit variance. Then $K(x_{\tau+1}) \approx K(\xi_{\tau+1}) = e^{-\sigma_\xi^2 \tilde{\xi}_{\tau+1}^2}/\sqrt{2\pi}$ and

$$J_n(1) \approx \frac{K(x_\tau)u_\tau + K(x_{\tau+1})u_{\tau+1}}{\left[K(x_\tau)^2 u_\tau^2 + K(x_{\tau+1})^2 u_{\tau+1}^2\right]^{1/2}} \approx \frac{u_\tau + e^{-\sigma_\xi^2 \tilde{\xi}_{\tau+1}^2} u_{\tau+1}}{\left[u_\tau^2 + e^{-\sigma_\xi^2 \tilde{\xi}_{\tau+1}^2}\right]^{1/2}} \approx \frac{u_\tau}{|u_\tau|} + O_p\left(e^{-\sigma_\xi^2}\right)$$

$$\approx \pm 1 + O_p\left(\frac{1}{1-2d}\right), \tag{3.5}$$

showing a tendency to bimodality as the memory parameter $d \to 0.5$. The same tendency to bimodality is also present in the approximation of $J_n^*(1)$ in addition to that given in (3.4), thereby implying that $J_n^*(1)$ is more likely to manifest bimodal behavior in finite samples than $J_n(1)$, corroborating the simulation findings.



(a) $\widetilde{\mathcal{J}}_n(1)$ densities, $n = 100$ (b) $\widetilde{\mathcal{J}}_n(1)$ densities, $n = 1{,}000$ (c) $\widetilde{\mathcal{J}}_n(1)$ densities, $n = 5{,}000$

Fig 3: Empirical densities of $\widetilde{\mathcal{J}}_n(1)$ for sample sizes $n = 100$ and for $n = 1{,}000$ and $d \in \{0.10, 0.25, 0.40, 0.55\}$.

Fig. 3 shows finite sample densities of the statistic $\widetilde{J}_n(1)$ in (2.17) using the same simulation design with the same set of long memory parameters, endogeneity correlation $\rho = 0.5$, and for sample sizes increasing from $n = 100$ to $n = 5{,}000$ based on $40{,}000$ replications. As evident in the graphics, the statistic removes bimodality in finite samples although there are extended shoulders on either side of the origin to around ± 1, particularly when $n = 100$. The distributions are far closer to the standard normal limit than those of the statistic $J_n^*(1)$

in (2.14) at every sample size with evident convergence in shape to normal for all values of the long memory parameter and clearest for $d = 0.1$, as would be expected. These findings support the heuristic analysis leading to (3.4) and (3.5). For when the variance estimate $\widehat{\mathcal{J}}_{nM^*}$ is employed, the scaling-out effect that leads to bimodality is removed, thereby explaining the finite sample distributions being closer to the standard normal.

4. Further analysis of the finite sample bimodality. As noted in Remark 2.5, natural self-normalization of sample covariance statistics does not perform well in finite samples relative to the asymptotic theory when strong effects of long memory are present in the data. This result seems new to the literature. But the observed finite sample bimodality has a subtle connection in its origins with earlier findings on bimodal t ratios where behavior is dominated by a few observations when there is heavy tailed data. In the present case, behavior is dominated by the few neighboring observations whose impact is not diminished by the kernel weights under strong dependence. Fig. 4 illustrates with a single shot of the data trajectories generated for x_t and u_t with $d = 0.1$ and $n = 1,000$.

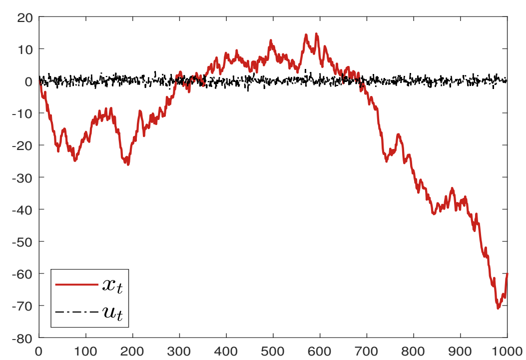


Fig 4: Single shot trajectories of x_t and u_t generated with $d = 0.10$ and $n = 1,000$ according to the simulation design given below.

Some additional analysis and computations are now provided to shed light on the finite sample properties of self-normalized sample covariance statistics in which nonstationarity originates in partial sums of long memory processes. The following simple framework with no endogeneity is used for the following discussion and data generation.

Simulation design

- both ϵ_k and u_k are iid $\mathcal{N}(0,1)$ and the ϵ_k are independent of the u_k;
- $x_k = \sum_{j=1}^{k} \xi_j$, where $(1-L)^d \xi_j = \epsilon_j$, with $0 < d < 1/2$ and $1/2 < \mu = 1 - d < 1$, so that $\xi_j = (1-L)^{-d}\epsilon_j = \sum_{i=0}^{\infty} \phi_i \epsilon_{j-i}$ with $\phi_i \sim \frac{1}{\Gamma(d)} i^{-(1-d)}$;
- $K(x) = e^{-x^2/2}/\sqrt{2\pi}$.

For $j = 1$ and 2, define

$$S_{jn} = \mathcal{J}_{jn}{}^{-1/2} \sum_{k=1}^{n} K(x_k)\, u_k,$$

where $\mathcal{J}_{1n} = \sum_{k=1}^{n} K^2(x_k)$ and $\mathcal{J}_{2n} = \sum_{k=1}^{n} K^2(x_k) u_k^2$. Under these conditions ξ_k is a long memory process with memory parameter $0 < d = 1 - \mu < 1/2$ and x_k is nonstationary with memory parameter $1+d$. S_{2n} is a natural self-normalized sample covariance statistic, matching $J_n^*(1)$ in (2.14). [1]

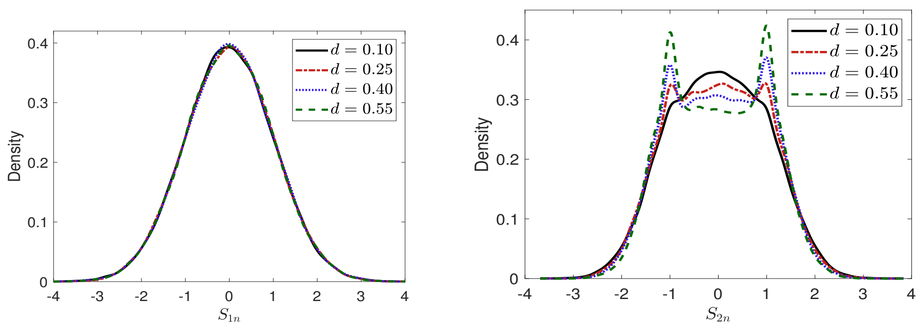
Recall that $d_n^2 = var(x_n) \sim A_d\, n^{1+2d}$, where A_d is a positive constant depending only on d. It is readily seen from (2.11) and (2.21) that

$$\frac{1}{n^{1/2-d}}\mathcal{J}_{1n}, \quad \frac{1}{n^{1/2-d}}\mathcal{J}_{2n} \to_D \left(\frac{A_d}{2}\right)^{1/2} L_{B_{(1+2d)/2}}(1,0),$$
$$\frac{\mathcal{J}_{2n}-\mathcal{J}_{1n}}{\mathcal{J}_{1n}} \to_P 0, \tag{4.1}$$

where $B_H(t)\}$ is fractional Brownian motion with Hurst exponent H and $L_{B_H}(t,s)$ is the local time process of $\{B_H(t)\}_{t\geq 0}$. In view of the independence of x_k and u_k and since $u_k \sim_{iid} \mathcal{N}(0,1)$, we have $S_{1n} \sim_d \mathcal{N}(0,1)$ for all $n \geq 1$ and

$$S_{2n} = \left(\frac{\mathcal{J}_{1n}}{\mathcal{J}_{2n}}\right)^{1/2} S_{1n} \to_D \mathcal{N}(0,1), \tag{4.2}$$

so that S_{2n} has a standard normal limit distribution. Now consider the finite sample performance of the statistics S_{1n} and S_{2n}.



(a) S_{1n} densities, $n = 100$

(b) S_{2n} densities, $n = 100$

Fig 5: Empirical densities of S_{1n} and S_{2n} for $n = 100$, $d \in \{0.1, 0.25, 0.4, 0.55\}$.

A. Simulation results for $\mathbf{S_{1n}}$: Kernel density estimates of the finite sample distributions of S_{1n} are shown in Fig. 5(a) for sample size $n = 100$ with $d \in \{0.1, 0.25, 0.4, 0.55\}$ from $40,000$ replications. The graphs confirm the exact finite sample $\mathcal{N}(0,1)$ distribution for all values of the memory parameter d, including the nonstationary case $d = 0.55$.

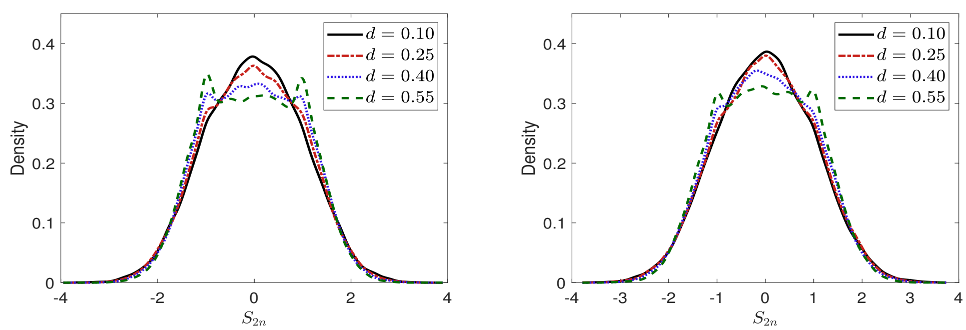
B. Simulation results for $\mathbf{S_{2n}}$: Fig. 5(b) shows the finite sample densities of S_{2n} for $n = 100$ and same memory parameter values $d \in \{0.1, 0.25, 0.4, 0.55\}$ again from $40,000$ replications. Bimodality in these distributions around the points ± 1 is clearly evident for all $d > 0.10$ and strong in the nonstationary case $d = 0.55$; for $d = 0.10$ the density has shoulders at the same points ± 1. Figs. 6(a) and 6(b) show the corresponding densities for $n = 1,000$ and $n = 5,000$. The slow convergence of these distributions to normality in the presence of stationary long memory is evident, especially for $d = 0.4$ where shoulders in the density around ± 1 are evident even when $n = 5,000$. In the nonstationary $d = 0.55$ case bimodality remains evident, although it is not as strong as it is for smaller sample sizes.

[1] When ϵ_k and x_k are independent of u_k the term $2\sum_{j=1}^{M}\ell(\frac{j}{M})\sum_{k=1}^{n-j}K(x_k)K(x_{k+j})\,u_k\,u_{k+j}$ that is included in $\mathcal{J}_n$ is unnecesssary since the terms $\tilde{G}_r$ appearing in Corollary 2.1 are zero for all $r \geq 1$.

Although S_{2n} has a normal limit distribution for all memory parameters $d \in (0, 0.5)$ the finite sample performance of S_{2n} depends on the value of d, in contrast to S_{1n}. Bimodality is strongest for stationary values of d closest to the boundary $d = 0.5$ and remains present even for very large sample sizes. This anomalous behavior can be explained in terms of relative convergence rates as follows. Recalling (4.1), when $d = 0.4$ we have

$$\Big(\frac{\mathcal{J}_{1n}}{\mathcal{J}_{2n}}\Big)^{1/2} - 1 = \frac{\mathcal{J}_{1n} - \mathcal{J}_{2n}}{\mathcal{J}_{2n}^{1/2}(\mathcal{J}_{1n}^{1/2} + \mathcal{J}_{2n}^{1/2})} = O_P(n^{-0.05}),$$

whence $\mathcal{J}_{2n}/\mathcal{J}_{1n} \to_P 1$ as $n \to \infty$; but the convergence rate is seen to be very slow. With such a slow convergence rate, even for $n = 5,000$ (where $n^{-0.05} \approx 0.65$) and with $S_{1n} \sim_d \mathcal{N}(0,1)$ for all $n \geq 1$, the value of $S_{2n} = \big(\frac{\mathcal{J}_{1n}}{\mathcal{J}_{2n}}\big)^{1/2} S_{1n}$ can be substantially impacted by the factor $\big(\frac{\mathcal{J}_{1n}}{\mathcal{J}_{2n}}\big)^{1/2}$, leading to departures from the normality of S_{1n} and the presence of bimodality in the distribution.



(a) S_{2n} densities, $n = 1,000$ (b) S_{2n} densities, $n = 5000$

Fig 6: Empirical densities of S_{2n} for sample sizes $n = 1,000$ and $n = 5,000$ and $d \in \{0.1, 0.25, 0.4, 0.55\}$.

When $x_k = \sum_{j=1}^k \xi_j$ with $(1-L)^d \xi_j = \epsilon_j$ and $d > 1/2$, the input ξ_j is a nonstationary long memory process) and the limit distribution S_{2n} is not normal, i.e., bimodality must appear. Indeed, in this case, we have

$$\mathcal{J}_{1n} \to_P A := \sum_{k=1}^{\infty} K^2(x_k), \quad \mathcal{J}_{2n} \to_P B := \sum_{k=1}^{\infty} K^2(x_k) u_k^2, \tag{4.3}$$

where A and B ($A \neq B$) are well defined positive random variables. Hence, as $n \to \infty$,

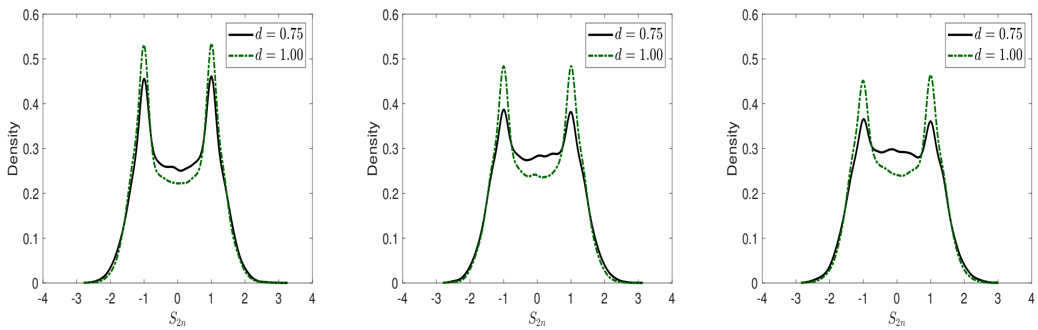
$$S_{2n} = \Big(\frac{\mathcal{J}_{1n}}{\mathcal{J}_{2n}}\Big)^{1/2} S_{1n} \to_D \Big(\frac{A}{B}\Big)^{1/2} \mathcal{N}(0,1),$$

since $S_{1n} \sim N(0,1)$ for all $n \geq 1$. The proof of (4.3) is straightforward. Let $A_{m,n} = \sum_{k=m}^n K^2(x_k)$ and recall that $x_n \sim_d \mathcal{N}(0, d_n)$ where $d_n^2 = var(x_n) \sim_a A_d\, n^{1+2d}$ as $n \to \infty$, it is readily seen that, whenever $d > 1/2$ and $m, n \to \infty$,

$$\begin{aligned} \mathbb{E} A_{m,n} &= \sum_{k=m}^n \mathbb{E} K^2(x_k) = \sum_{k=m}^n \int K^2(d_k y) e^{-y^2/2} dy \\ &\leq C \sum_{k=m}^n d_k^{-1} = C_1 \sum_{k=m}^n k^{-(1+2d)/2} \to 0. \end{aligned}$$

Hence, $A := \sum_{k=1}^{\infty} K^2(x_k)$ is a well defined random variable and $\mathcal{J}_{1n} \to_P A$. Similarly, we have $\mathbb{E}B_{m,n} \to 0$ where $B_{m,n} = \sum_{k=m}^{n} K^2(x_k)\eta_k^2$, and hence $\mathcal{J}_{2n} \to_P B$.

Fig. 7 gives simulation results for S_{2n} in the nonstationary innovation cases $d = 0.75$ and $d = 1$ for $n = 100, 1,000$, and $5,000$ based on $25,000$ replications. Bimodality appears a prominent feature of the densities of S_{2n} for both $d = 0.75$ and $d = 1$, showing little tendency to diminish even in very large sample sizes, corroborating the non-Gaussian limit theory in the nonstationary case. The bimodality is stronger when $d = 1$ than when $d = 0.75$ for all sample sizes.



(a) S_{2n} densities, $n = 100$ (b) S_{2n} densities, $n = 1000$ (c) S_{2n} densities, $n = 5000$

Fig 7: Empirical densities of S_{2n} for sample sizes $n = 100, 1000$ and $n = 5,000$ and $d \in \{0.75, 1.00\}$.

5. Conclusion. Sample covariance functionals of regressors and innovations play a key role in nonlinear nonstationary regression models and self normalized versions of these statistics are a foundation for inference. The limit theory given here covers a wide class of such functionals and reveals important differences between stationary and nonstationary long memory innovations. Numerical work shows strong bimodality in the finite sample distributions, slow convergence to the Gaussian limit theory under stationary long memory innovations and non-Gaussian limit theory when the innovations have nonstationary long memory. New forms of self normalization are shown to provide improved finite sample performance suitable for practical work.

6. Proofs of the main results. *Proof of Theorem 2.1.* First note that, for any bounded $h > 0$ and $nh/d_n \to \infty$,

$$\Big(\frac{d_n}{nh}\Big)^{1/2} \max_{1\le k\le n} |f(x_k/h, w_k)| = o_{\mathrm{P}}(1), \tag{6.1}$$

by a similar augument as in Proposition 7.4[2]. Due to (6.1), without loss of generality, we assume

$$f(x_k/h, w_k) = 0 \quad \text{for} \;\; k = 1, ..., A_0, \tag{6.3}$$

[2] Indeed, as in (7.4) of Proposition 7.4, it follows from $nh/d_n \to \infty$ that, for any $A > 0$,

$$\Big(\frac{d_n}{nh}\Big)^{1/2} \max_{1\le k\le n} |f(x_k/h, w_k)|$$
$$\le \Big[\frac{d_n}{nh}\sum_{k=1}^{n} f^2(x_k/h, w_k) I(|f(x_k/h, w_k)| \ge A)\Big]^{1/2} + A\Big(\frac{d_n}{nh}\Big)^{1/2}$$

where A_0 is a fixed constant that can be chosen large enough. This convention will reduce notational complexity in the proofs of propositions that are given in next section and the lemmas in the Appendix.

We adopt the methodology employed in Wang and Phillips (2009b)[25], starting with an outline of the proof of (2.6), where some useful propositions will be given in the next section. Define, for $0 \le t \le 1$,

$$S_n(t) = \big(\frac{d_n}{n}\big)^{1/2} \sum_{k=1}^{\lfloor nt \rfloor} f(x_k, w_k),$$

$$Y_{nq}(t) = \psi_{n0}(t) + 2\sum_{j=1}^{q} \psi_{nj}(t),$$

where for $j = 0, 1, ..., q$,

$$\psi_{nj}(t) = \frac{d_n}{n} \sum_{k=1}^{\lfloor nt \rfloor} f(x_k, w_k)\, f(x_{k+j}, w_{k+j}),$$

and for all $\alpha_i, \beta_j \in \mathrm{R}$, $0 \le s_0 < s_1 < ... < s_m < \infty$ and $0 \le t_0 < t_1 < ... < t_l < \infty$,

$$Z_{n2} = \sum_{i=1}^{l} \alpha_i \big[\zeta_{n1}(t_i) - \zeta_{n1}(t_{i-1})\big] + \sum_{i=1}^{m} \beta_i \big[\zeta_{n2}(s_i) - \zeta_{n2}(s_{i-1})\big],$$

where $\zeta_{n1}(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor nt \rfloor} \epsilon_j$ and $\zeta_{n2}(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{\lfloor nt \rfloor} \epsilon_{-j}$. An application of Proposition 7.4 implies that, for any $q \ge 1$,

(6.4) $$\Big(\psi_{n0}, \psi_{n1}, ..., \psi_{nq}, Y_{nq}(t)\Big) \Rightarrow \Big(G_0, G_1, ..., G_q, \Lambda_q\Big)\, L_Z(t, 0),$$

on $D_{R^{q+2}}0, 1]$, where $\Lambda_q = G_0 + 2\sum_{r=1}^{q} G_r$. This, together with the tightness of $\{S_n(t)\}_{n\ge1}$ (see Proposition 7.2 with $h = 1$), yields

(6.5) $$\{S_n(t),\ Y_{nq}(t),\ Z_{n2}\}_{n\ge1} \text{ is tight on } D_{\mathrm{R}^3}[0, 1].$$

Hence, for each $\{n'\} \subseteq \{n\}$, there exists a subsequence $\{n''\} \subseteq \{n'\}$ such that

(6.6) $$\big\{S_{n''}(t),\ Y_{n''q}(t),\ Z_{n''2}\big\} \Rightarrow \big\{\eta(t),\ \Lambda_q\, L_Z(t, 0),\ Z_2\big\},$$

on $D_{\mathrm{R}^3}[0, 1]$, where

$$Z_2 = \sum_{i=1}^{l} \alpha_i \big(B_{1t_i} - B_{1,t_{i-1}}\big) + \sum_{i=1}^{m} \beta_i \big(B_{2s_i} - B_{2,s_{i-1}}\big),$$

and $\eta(t)$ is a process continuous with probability one due to (6.1).

Let $Z_{n3} = \sum_{i=1}^{v} \gamma_i \big[S_n(t_i) - S_n(t_{i-1})\big]$ and $Z_3 = \sum_{i=1}^{v} \gamma_i \big[\eta(t_i) - \eta(t_{i-1})\big]$, where $\gamma_j \in \mathrm{R}$ and $0 \le t_0 < t_1 < ... < t_v \le s$. Since, for each $0 \le t \le 1$, $S_n(t)$ is uniformly integrable (see Proposition 7.1 with $h = 1$), it follows from Proposition 7.3 (i) with $h = 1$ that, for any $s < t$,

$$\mathbb{E}\, e^{i(Z_3+Z_2)} \big[\eta(t) - \eta(s)\big]$$

(6.7) $$= \lim_{n''\to\infty} \mathbb{E}\, e^{i(Z_{n''3}+Z_{n''2})} [S_{n''}(t) - S_{n''}(s)] = 0.$$

(6.2) $$\to_D \Big[\int_{-\infty}^{\infty} \mathbb{E} f^2(x, w_1) I(|f(x, w_1)| \ge A) dx\, L_Z(1, 0)\Big]^{1/2}, \quad \text{as } n \to \infty.$$

This implies (6.1) since $\int_{-\infty}^{\infty} \mathbb{E} f^2(x, w_1) I(|f(x, w_1)| \ge A) dx \le A^{-2} \int_{-\infty}^{\infty} \mathbb{E} f^4(x, w_1) dx \to 0$ by (2.2), as $A \to \infty$.

See, e.g., Theorem 5.4 of Billingsley (1968)[2]. Similarly, by Propositions 7.1 with $h=1$ and 7.3 (iii) with $h=1$, we have

$$\mathbb{E}\, e^{i(Z_3+Z_2)}\big\{[\eta(t)-\eta(s)]^2-[Y(t)-Y(s)]\big\}=0, \tag{6.8}$$

where $Y(t)=\tau_1^2 L_Z(t,0)$. Indeed, by letting $Y_q(t)=\Lambda_q L_Z(t,0)$ and noting

$$\sup_{0\le t\le 1}\mathbb{E}\,|Y_q(t)-Y(t)|\le 2\,|\Lambda_q-\tau_1^2|E\sup_{0\le t\le 1}L_Z(t,0)\le C\sum_{r=q+1}^{\infty}|G_r|\to 0,$$

due to Proposition 7.5, it follows from Propositions 7.1 with $h=1$ and 7.3 (iii) with $h=1$ that, for any $\epsilon>0$,

$$\begin{aligned}
&\big|\mathbb{E}\, e^{i(Z_3+Z_2)}\big\{[\eta(t)-\eta(s)]^2-\big[Y(t)-Y(s)\big]\big\}\big|\\
&\le \big|\mathbb{E}\, e^{i(Z_3+Z_2)}\big\{[\eta(t)-\eta(s)]^2-\big[Y_q(t)-Y_q(s)\big]\big\}\big|\\
&\quad +\mathbb{E}\,\big|\big[Y_q(t)-Y(t)\big]\big|+\mathbb{E}\,\big|\big[Y_q(s)-Y(s)\big]\big|\\
&\le \lim_{n''\to\infty}\big|\mathbb{E}\, e^{i(Z_{n''3}+Z_{n''2})}\big\{[S_{n''}(t)-S_{n''}(s)]^2-[Y_{n''q}(t)-Y_{n''q}(s)]\big\}\big|+2\epsilon\\
&\le 3\epsilon, \qquad (6.9)
\end{aligned}$$

by letting $q\to\infty$. This yields (6.8) as the left side of (6.9) does not depend on ϵ.

Let $\mathcal{F}_s=\sigma\{B_{1t},0\le t\le 1; B_{2t},0\le t<\infty,\eta(t),0\le t\le s\}$. Results (6.7) and (6.8) imply that, for any $0\le s<t\le 1$,

$$\mathbb{E}\Big(\big[\eta(t)-\eta(s)\big]\mid \mathcal{F}_s\Big)=0,\quad a.s.,$$

$$\mathbb{E}\Big(\big\{[\eta(t)-\eta(s)]^2-\big[Y(t)-Y(s)\big]\big\}\mid \mathcal{F}_s\Big)=0,\quad a.s.$$

Note that $\mathcal{F}_s\uparrow$, $\eta(s)$ is $\mathcal{F}_s$-measurable for each $0\le s\le 1$ and $Y(t)=\tau_1^2L_Z(t,0)$ (for any fixed $t\in[0,1]$) is $\mathcal{F}_s$-measurable for each $0\le s\le 1$. It follows from Lemma 3.4 of Wang (2015)[22] that the finite-dimensional distributions of $(\eta(t),Y(t))$ coincide with those of $\{\mathbb{N}Y^{1/2}(t),\,Y(t)\}$, where $\mathbb{N}$ is a normal variate independent of $Y(t)$. Since $\eta(t)$ does not depend on the choice of the subsequence $\{n''\}$, it follows from (6.5) and (6.6) that

$$\big\{S_n(t),\,Y_{nq}(t)\big\}\Rightarrow\big\{[\tau_1L_Z(t,0)]^{1/2}\,\mathbb{N},\,\Lambda_qL_Z(t,0)\big\}, \tag{6.10}$$

on $D_{R^2}[0,1]$, where $\mathbb{N}$ is normal variate independent of $L_Z(t,0)$. This, together with (6.4) and the continuous mapping theorem, yields (2.6).

The proof of (2.5) is similar. Set, for $0\le t\le 1$ and $h>0$,

$$S_{n,h}(t)=\big(\frac{d_n}{nh}\big)^{1/2}\sum_{k=1}^{\lfloor nt\rfloor}f(x_k/h,w_k),\quad Z_{n,h}(t)=\frac{d_n}{nh}\sum_{k=1}^{\lfloor nt\rfloor}f^2(x_k/h,w_k).$$

As $h\to 0$ and $nh/d_n\to\infty$, $Z_{n,h}(t)\Rightarrow Z(t):=\tau^2L_Z(t,0)$ by (7.4) in Proposition 7.4. The same arguments as those leading to (2.6) can be used to establish (2.5) except that $S_n(t),Y_{nq}(t)$ and $Y(t)$ are replaced by $S_{n,h}(t)$, $Z_{n,h}(t)$ and $Z(t)$, respectively. The corresponding propositions with $h\to 0$ are given in next section. □

Proof of Corollary 2.1. We only prove (2.11). The proof of the other result is similar. Let $u_{1k}=\sum_{j=0}^{m_0}\psi_j\,\lambda'_{k-j}$, $u_{2k}=u_k-u_{1k}=\sum_{j=m_0+1}^{\infty}\psi_j\,\lambda'_{k-j}$ and, for $r=0,1,2,...$

$$\tilde G_{r,m_0}=\int_{-\infty}^{\infty}K(y)\,\mathbb{E}\,\big\{u_{10}\,u_{1r}\,K(y+x_r)\big\}dy.$$

Using (2.9), for any $m_0 > 0$ and $q \geq 0$, we have

$$\Big(\frac{d_n}{n}\sum_{k=1}^n K^2(x_k)u_{1k}^2,\ \frac{d_n}{n}\sum_{k=1}^n K(x_k)u_{1k}K(x_{k+1})u_{1,k+1},...,\ \frac{d_n}{n}\sum_{k=1}^n K(x_k)u_{1k}K(x_{k+q})u_{1,k+q},\ \Big(\frac{d_n}{n}\Big)^{1/2}\sum_{k=1}^n K(x_k)u_{1k}\Big)$$
$$\Rightarrow\ \big(\tilde G_{0,m_0}\,L_Z(1,0),\ \tilde G_{1,m_0}\,L_Z(1,0),...,\ \tilde G_{q,m_0}\,L_Z(1,0),\ \tilde\tau_{1,m_0}\,\mathbb{N}\,L_Z^{1/2}(1,0)\big),$$

where $\tilde\tau_{1,m_0} = \tilde G_{0,m_0} + 2\sum_{r=1}^\infty \tilde G_{r,m_0}$. This implies that, for any $m_0 > 0, q \geq 0$ and any continuous function with $l(0) = 1$,

$$\Big(\frac{d_n}{n}\sum_{k=1}^n K^2(x_k)u_{1k}^2,\ \widetilde{\mathcal{J}_{n,q}},\ \Big(\frac{d_n}{n}\Big)^{1/2}\sum_{k=1}^n K(x_k)u_{1k}\Big)$$
$$\to_D\ \big(\tilde\tau^2\,L_Z(1,0),\ \tilde\tau_{1,q}^2\,L_Z(1,0),\ \tilde\tau_{1,q}\,\mathbb{N}\,L_Z^{1/2}(1,0)\big),$$

where $\tilde\tau_{1,q}^2 = \tilde G_{0,m_0} + 2\sum_{r=1}^q \tilde G_{r,m_0}$ and

$$\widetilde{\mathcal{J}_{n,q}} = \frac{d_n}{n}\sum_{k=1}^n K^2(x_k)u_{1k}^2 + \frac{2d_n}{n}\sum_{j=1}^q \ell\Big(\frac{j}{M}\Big)\sum_{k=1}^{n-j} K(x_k)K(x_{k+j})\,u_{1k}\,u_{1,k+j}.$$

Consequently, to prove Corollary 2.1, it suffices to show the following:

(a) as $m_0 \to \infty$,

$$|\tilde G_0 - \tilde G_{0,m_0}| + \sum_{r=1}^\infty |\tilde G_r - \tilde G_{r,m_0}| \to 0; \tag{6.11}$$

(b) for any $m_0 \geq 1$,

$$\mathbb{E}\,\Big|\sum_{k=1}^n u_{2k}\,K(x_k)\Big|^2 \leq C\,(n/d_n)\Big[\sum_{j=m_0}^\infty j^{1/4}(|\psi_{1j}| + |\psi_{2j}|)\Big]^2; \tag{6.12}$$

(c) for any $r \geq 0$, as $n \to \infty$ first and then $m_0 \to \infty$,

$$\frac{d_n}{n}\sum_{k=1}^{n-r} K(x_k)K(x_{k+r})\,\big(u_{1k}\,u_{1,k+r} - u_k u_{k+r}\big) = o_P(1); \tag{6.13}$$

Further, if $m_0 = m_0(n) \to \infty$, i.e., m_0 depends on n, it also follows that there exists $M_1 \equiv M_{1n}$ depending on m_0 such that, as $n \to \infty$,

$$R_n := \frac{d_n}{n}\sum_{r=1}^{M_1}\Big|\sum_{k=1}^{n-r} K(x_k)K(x_{k+r})\,\big(u_{1k}\,u_{1,k+r} - u_k u_{k+r}\big)\Big| = o_P(1). \tag{6.14}$$

(d) there exists $M \equiv M_n \to \infty$ so that, as $n \to \infty$ first and then $q \to \infty$,

$$\frac{d_n}{n}\sum_{r=q+1}^{M} \ell\Big(\frac{r}{M}\Big)\sum_{k=1}^{n-r} K(x_k)K(x_{k+r})\,u_k\,u_{k+r} = o_P(1). \tag{6.15}$$

For the proofs of (6.11), (6.12) - (6.14) and (6.15), we refer to Propositions 7.5, 7.6 and 7.7, respectively. □

7. Subsidiary propositions. This section proves the following propositions which are required in the proofs of Theorem 2.1 and Corollary 2.1. The notation is the same as in the previous section except where explicitly mentioned.

PROPOSITION 7.1. *For any fixed* $0 \le t \le 1$, $r \ge 0$ *and any bounded* $h > 0$ *satisfying* $nh/d_n \to \infty$, $\psi_{nr}(t)$, $Z_{n,h}(t)$ *and* $S^2_{n,h}(t)$, $n \ge 1$, *are uniformly integrable.*

PROPOSITION 7.2. *For any bounded* $h > 0$ *satisfying* $nh/d_n \to \infty$, $\{Z_{n,h}(t)\}_{n\ge}$ *and* $\{S_{n,h}(t)\}_{n\ge1}$ *are tight on* $D[0,1]$.

PROPOSITION 7.3. *For any* $0 \le s < t \le 1$, *we have that*

(i) *if* $h > 0$ *is bounded satisfying* $nh/d_n \to \infty$, *then*

$$\lim_{n\to\infty} \mathbb{E}\, e^{i(Z_{n3}+Z_{n2})}[S_{n,h}(t) - S_{n,h}(s)] = 0; \tag{7.1}$$

(ii) *if* $h \to 0$ *satisfying* $nh/d_n \to \infty$, *then*

$$\lim_{n\to\infty} \mathbb{E}\, e^{i(Z_{n3}+Z_{n2})}\big\{[S_{n,h}(t) - S_{n,h}(s)]^2 - [Z_{n,h}(t) - Z_{n,h}(s)]\big\} = 0; \tag{7.2}$$

(iii) *for any* $\epsilon > 0$, *there exists a* $q_0 > 0$ *such that*

$$\lim_{n\to\infty} \big|\mathbb{E}\, e^{i(Z_{n3}+Z_{n2})}\big\{[S_n(t) - S_n(s)]^2 - [Y_{nq}(t) - Y_{nq}(s)]\big\}\big| \le \epsilon, \tag{7.3}$$

for all $q \ge q_0$.

PROPOSITION 7.4. *For any bounded* $h > 0$ *satisfying* $nh/d_n \to \infty$, *we have*

$$Z_{n,h}(t) \Rightarrow \tau^2 L_Z(t,0), \tag{7.4}$$

on $D_R[0,1]$. *If, in addition,* $\gamma = 0$ *and* $\int \mathbb{E}\big\{|\hat f(t,w_0)\big(1 + ||w_r||^\beta\big)\big\}dt < \infty, 0 \le r \le m$, *then*

$$\big\{\psi_{n0}(t), \psi_{n1}(t), ..., \psi_{nm}(t)\big\} \Rightarrow \big\{G_0\,, G_1\,, ..., G_m\,\big\}\, L_Z(t,0), \tag{7.5}$$

on $D_{R^{m+1}}[0,1]$.

PROPOSITION 7.5. *If* $\gamma = 0$, *we have* $\sum_{r=1}^\infty |G_r| < \infty$ *and* $\sum_{r=1}^\infty |\tilde G_r| < \infty$, *and (6.11) also holds.*

PROPOSITION 7.6. *Results (6.13) and (6.14) hold and, for any bounded* $h > 0$ *satisfying* $nh/d_n \to \infty$, *we have*

$$\mathbb{E}\,\Big|\sum_{k=1}^n u_{2k}\, K(x_k/h)\Big|^2 \le C\,(nh/d_n)\,\Big[\sum_{j=m_0}^\infty j^{1/4}(|\psi_{1j}| + |\psi_{2j}|)\Big]^2. \tag{7.6}$$

PROPOSITION 7.7. *Result (6.15) holds.*

7.1. *Preliminary lemmas.* Except where explicitly mentioned, the proofs of all lemmas are given in the Appendix. Throughout this section, we let $\mathcal{F}_k = \sigma(\lambda_k, \lambda_{k-1}, ...)$.

LEMMA 7.1. *Let* $p(s, s_1, ..., s_m)$ *be a real function of its components and* $t_1, ..., t_m \in \mathbb{Z}$, *where* $m \ge 0$. *There exists an* $A_0 > 0$ *such that the following results hold.*

(i) *For any* $h > 0$ *and* $k \geq 2m + A_0$, *we have*

$$\mathbb{E}|\, p(x_k/h, \lambda_{t_1}, ..., \lambda_{t_m})| \leq \frac{C\, h}{d_k} \int_{-\infty}^{\infty} \mathbb{E}|p(t, \lambda_1, ..., \lambda_m)|dt. \tag{7.7}$$

(ii) *For any* $h > 0$, $k - j \geq 2m + A_0$ *and* $j + 1 \leq t_1, ..., t_m \leq k$, *we have*

$$\mathbb{E}\big[\big|p(x_k/h, \lambda_{t_1}, ..., \lambda_{t_m})\big| \mid \mathcal{F}_j\big] \leq \frac{C\, h}{d_{k-j}} \int_{-\infty}^{\infty} \mathbb{E}|p(t, \lambda_1, ..., \lambda_m)|dt. \tag{7.8}$$

(iii) *For any* $h > 0$ *and* $k - j \geq 1$, *we have*

$$\mathbb{E}\left[|p(x_k/h)|\,\big|\mathcal{F}_j\right] \leq \frac{C\, h}{d_{k-j}} \int_{-\infty}^{\infty} |p(x)|dx, \tag{7.9}$$

Proof. For the proofs of (7.7) and (7.8), we refer to Lemma A.1 of Wang et al (2021)[28]. As $\phi_0 \neq 0$, the proof of (7.9) is simple. See, for instance, Lemma 2.1 (iii) of Wang (2015)[22]. □

Recalling (6.3), $f(x, y) \leq T(x)(1 + ||y||^{\beta})$ and $\mathbb{E}\,||w_1||^{\max\{2,4\beta\}} < \infty$, where $T(x)$ is bounded and integrable, a simple application of Lemma 7.1 (i) and (ii) yields that, for any $h > 0$,

$$\sum_{k=1}^{n} \mathbb{E}\, f^2(x_k/h, w_k) \leq Cnh/d_n, \quad \mathbb{E}\Big[\sum_{k=1}^{n} f^2(x_k/h, w_k)\Big]^2 \leq C\left(nh/d_n\right)^2. \tag{7.10}$$

and (7.10) still holds if $f^2(x_k/h, w_k)$ is replaced by Y_{kj}^2 defined by

$$Y_{kj} = \mathbb{E}\left[f(x_k/h, w_k)|\mathcal{F}_{k-j}\right] - \mathbb{E}\left[f(x_k/h, w_k)|\mathcal{F}_{k-j-1}\right],$$

where $j \geq 0$ is a fixed integer. Furthermore, it follows from Lemma 7.1 (iii) that, for any $r \geq 1$,

$$\begin{aligned}\mathbb{E}\left[|f(x_{k+r}/h, w_{k+r})||\mathcal{F}_k\right] &\leq \left\{\mathbb{E}\left[T^2(x_{k+r}/h) \mid \mathcal{F}_k\right]\right\}^{1/2} \left\{\mathbb{E}\left[(1 + ||w_{k+r}||^{2\beta})|\mathcal{F}_k\right]\right\}^{1/2} \\ &\leq Ch^{1/2}\, R_k,\end{aligned}$$

where $R_k = \left\{\mathbb{E}\left[(1 + ||w_{k+r}||^{2\beta})|\mathcal{F}_k\right]\right\}^{1/2}$ depending only on $\lambda_k, ..., \lambda_{k-m_0}$. Hence, for any $r \geq 1$, $h > 0$ and $0 \leq s < t \leq 1$, we also have

$$\begin{aligned}&\sum_{k=[ns]+B_0}^{\lfloor nt \rfloor} \mathbb{E}\Big[|f(x_k/h, w_k)|\,|f(x_{k+r}/h, w_{k+r})| \mid \mathcal{F}_{[ns]}\Big] \\ &\leq \sum_{k=[ns]+B_0}^{\lfloor nt \rfloor} \mathbb{E}\Big[|f(x_k/h, w_k)|\,\mathbb{E}\left\{|f(x_{k+r}/h, w_{k+r})| \mid \mathcal{F}_k\right\} \mid \mathcal{F}_{[ns]}\Big] \\ &\leq Ch^{1/2} \sum_{k=[ns]+B_0}^{\lfloor nt \rfloor} \mathbb{E}\left\{|f(x_k/h, w_k)|\, R_k \mid \mathcal{F}_{[ns]}\right\} \\ &\leq C\, nh^{3/2}(t-s)^{\alpha}/d_n, \qquad (7.11)\end{aligned}$$

for some $\alpha > 0$, whenever B_0 is sufficiently large so that (7.8) is applicable. We remark that (7.11) holds for $r = 0$ if $h^{3/2}$ is replaced by h. These results will be used later.

In the next lemma, Ω_1 is set to be a subset of $\Omega = \{1, 2, ..., k\}$, $\Omega_2 = \Omega - \Omega_1$ and

$$z_k(t) = \sum_{v=1}^{k} \epsilon_v \big(t\alpha_v + \beta_v\big).$$

LEMMA 7.2. *Suppose that $\sum_{v=1}^k \alpha_v^2 \le C\tau_k^2$ and, for any Ω_1 satisfying $\#\Omega_1 \le \sqrt{k}$,*

$$B_{1k} := \sum_{v\in\Omega_2} \alpha_v^2 \ge \tau_k^2, \tag{7.12}$$

for some constants sequence τ_k. Then, for any $\delta \ge 0$ and $s_1, s_2 \in R^+$, we have

$$\int \min\{1, s_1\,|t|^\delta + s_2\}\big|\mathbb{E}\, e^{iz_k(t)}\big|dt$$

$$\le\; C\Big(k^{-3} + s_1\,\tau_k^{-1-\delta}\,\big[1 + \big(\sum_{v=1}^{k} \beta_v^2\big)^{\delta/2}\big] + s_2\,\tau_k^{-1}\Big); \tag{7.13}$$

$$\int \min\{1, s_1\,|t|\}\min\{1, |t|\}\big|\mathbb{E}\, e^{iz_k(t)}\big|dt$$

$$\le\; C\Big(k^{-3} + s_1\,\tau_k^{-3}\,\big[1 + \sum_{v=1}^{k} \beta_v^2\big]\Big). \tag{7.14}$$

If in addition $\sum_{v=1}^k \beta_l^2 \le a < \infty$, then

$$\int_{|t|\ge B/\tau_k} \big|\mathbb{E}\, e^{iz_k(t)}\big|dt \le C\big(k^{-3} + \tau_k^{-1}B^{-1}\big), \tag{7.15}$$

for any $B \ge 2a^{1/2}$.

Proof. The proof of Lemma 7.2 is similar to that of Wang and Phillips (2011, pages 246-247)[26] and is therefore omitted. But an outline of the proof is given in Appendix A.1 for completeness. □

Since Lemma 7.2 still holds when $z_k(t)$ is replaced by $z_{k-m_0}(t)$ when $k \ge m_0^2$ and since w_k depends only on $\lambda_k, ..., \lambda_{k-m_0}$, the following lemma is a direct consequence of Lemma 7.2.

LEMMA 7.3. *Let $g(x, y)$ be a real function satisfying*

- $|\mathbb{E}\, g(t, w_1)| \le C \min\{1, |t|\}$ *and* $\sup_t \mathbb{E}\,\big\{(1 + |\epsilon_0|)|g(t, w_1)|\big\} < \infty$.

For any bounded $h > 0$ and $\tau_k \le C\,k^2$, we have

$$\int_{-\infty}^{\infty} \big|\mathbb{E}\, e^{iz_k(t/h)}\, g(t, w_k)\big|dt \le Ch\,\tau_k^{-1}, \tag{7.16}$$

for all $k \ge m_0^2$. Instead of (7.16), we also have

$$\int_{-\infty}^{\infty} \big|\mathbb{E}\, e^{iz_k(t/h)} g(t, w_k)\big|dt$$

$$\le Ch\,\big\{(1 + \alpha_{k0})\tau_k^{-2}\,\big[1 + \big(\sum_{v=1}^{k} \beta_v^2\big)^{1/2}\big] + \beta_{k0}\,\tau_k^{-1}\big\}, \tag{7.17}$$

where $\alpha_{k0} = \max_{0\le i\le m_0\vee(k-1)} |\alpha_{k-i}|$ *and* $\beta_{k0} = \max_{0\le i\le m_0\vee(k-1)} |\beta_{k-i}|$. *Similarly, when* $\sup_k \alpha_{k0} = O(1)$, *we have*

$$\int_{-\infty}^{\infty} \min\{1, |t|/h\} \left|\mathbb{E}\, e^{iz_k(t/h)} g(t, w_k)\right| dt$$

$$\le C\, h \left\{k^{-3} + \left[\beta_{k0}(\tau_k^{-2} + k^{-3}) + \tau_k^{-3}\right]\left(1 + \sum_{v=1}^{k} \beta_v^2\right)\right\}. \tag{7.18}$$

Proof. See Appendix A.2. □

Let $\mathrm{I}_k(m) = \int \mathbb{E}\left(e^{isx_k/h + i\sum_{j=m+1}^{l}\gamma_j\epsilon_j} g(s, w_k) \mid \mathcal{F}_m\right) ds$ and

$$\mathrm{I}_{k,l}(m) = \int\int \mathbb{E}\left(e^{isx_k/h + itx_l/h + i\sum_{j=m+1}^{l}\gamma_j\epsilon_j} g(s, w_k) g(t, w_l) \mid \mathcal{F}_m\right) ds dt,$$

where $g(x, y)$ is a real function given in Lemma 7.3, and let

$$\mathrm{II}_{k,l}(B) = \int_{|s|\ge B/d_k} \int_{|t|\ge B/d_l} g_1(t) g_2(t)\, \mathbb{E}\left(e^{isx_k/h + itx_l/h + i\sum_{j=1}^{l}\gamma_j\epsilon_j} \mid \mathcal{F}_0\right) ds dt,$$

where $g_1(t)$ and $g_2(t)$ are bounded real functions. The next lemma is an application of Lemma 7.3.

LEMMA 7.4. *Let* $m \ge 0$, $l - k \ge A_0^2 + 1$ *and* $k - m \ge A_0^2 + 1$, *where* $A_0 \ge m_0$ *and* m_0 *is given as in Lemma 7.3. Suppose* $a := \sum_{j=1}^{l} \gamma_j^2 < \infty$.

(i) *For any* $h > 0$, *we have*

$$|\mathrm{I}_k(m)| \le C\, h \left[d_{k-m}^{-2}(1 + a^{1/2}) + \beta_{l0}\, d_{k-m}^{-1}\right], \tag{7.19}$$

$$|\mathrm{I}_{k,l}(m)| \le C\, h^2\, d_{k-m}^{-1} \left[d_{l-k}^{-2}(1 + a^{1/2}) + \beta_{l0}\, d_{l-k}^{-1}\right]. \tag{7.20}$$

where $\beta_{l0} = \max_{0\le j\le m_0} |\gamma_{l-j}|$.

(ii) *Under* **SM**, *if* $|\gamma_j| \le C/\sqrt{n}$ *where* $m \le j \le l$, *for any* $h > 0$, *we have*

$$|\mathrm{I}_k(m)| \le C\, h \left((k - m)^{-1} + \sqrt{k - m}/\sqrt{n}\right), \tag{7.21}$$

$$|\mathrm{I}_{k,l}(m)| \le C\, h^2 \left[(l - k)^{-1}(k - m)^{-1} + (l - k)^{-3/2}(k - m)^{-1/2}\right]. \tag{7.22}$$

(iii) *For any* $h > 0$ *and* $B \ge 2a^{1/2}$, *we have*

$$|\mathrm{II}_{k,l}(B)| \le C\, h^2 \left[(l - k)^{-2} + B^{-1} d_{l-k}^{-1}\right] d_k^{-1}. \tag{7.23}$$

Proof. See Appendix A.3. □

Let $I_k(h) = f(x_k/h, w_k) \exp\left\{i \sum_{j=m+1}^{n} \mu_j\epsilon_j/\sqrt{n}\right\}$ and

$$II_{lk}(h) = f(x_k/h, w_k)\, f(x_l/h, w_l)\, \exp\left\{i \sum_{j=m+1}^{n} \mu_j\epsilon_j/\sqrt{n}\right\},$$

where μ_l are constants satisfying $|\mu_l| \le C$. Using Lemma 7.4, we have the following results.

LEMMA 7.5. *There exists a* $B_0 \ge m_0$ *such that, for all* $m \ge 0$, $l - k \ge B_0$, $k - m \ge B_0$ *and bounded* $h > 0$,

(i) *under* **LM**,

$$\left|\mathbb{E}\left[I_k(h)\mid\mathcal{F}_m\right]\right|\le C\,h\left(d_{k-m}^{-2}+d_{k-m}/\sqrt{n}\right), \tag{7.24}$$

$$\left|\mathbb{E}\left[II_{lk}(h)\mid\mathcal{F}_m\right]\right|\le C\,h^2\,d_{k-m}^{-1}\left(d_{l-k}^{-2}+d_{l-k}/\sqrt{n}\right), \tag{7.25}$$

(ii) *under* **SM**,

$$\left|\mathbb{E}\left[I_k(h)\mid\mathcal{F}_m\right]\right|\le C\,h\left((k-m)^{-1}+\sqrt{k-m}/\sqrt{n}\right), \tag{7.26}$$

$$\begin{aligned}\left|\mathbb{E}\left[II_{lk}(h)\mid\mathcal{F}_m\right]\right|\le C\,h^2&\left[(l-k)^{-1}(k-m)^{-1}\right.\\&\left.+(l-k)^{-3/2}(k-m)^{-1/2}\right].\end{aligned} \tag{7.27}$$

LEMMA 7.6. *There exists a $B_0\ge m_0$ such that, for all $m\ge 0$, $l-k\ge B_0$, $k-m\ge B_0$ and bounded $h>0$,*

(i) *under* **LM**,

$$\left|\mathbb{E}\left\{f(x_l/h,w_l)\mathbb{E}\left[f(x_k/h,w_k)\mid\mathcal{F}_{k-m}\right]\right\}\right|\le C\,h^2\,d_k^{-1}\,d_{l-k}^{-2}, \tag{7.28}$$

(ii) *under* **SM**,

$$\begin{aligned}&\left|\mathbb{E}\left\{f(x_l/h,w_l)\mathbb{E}\left[f(x_k/h,w_k)\mid\mathcal{F}_{k-m}\right]\right\}\right|\\&\quad\le\ C\,h^2\left[(l-k)^{-1}\,k^{-1}+(l-k)^{-3/2}\,k^{-1/2}\right].\end{aligned} \tag{7.29}$$

The proofs of Lemmas 7.5 and 7.6 will be given in Appendix A.4 and A.5, respectively.

LEMMA 7.7. *Let $\Gamma(.)$ be a measurable function with $\Gamma(\lambda_1)=0$ and $\mathbb{E}\Gamma^2(\lambda_1)<\infty$. There exists a A_0 so that*

(a) *for all $k\ge A_0$ and $|l-k|\le A_0$,*

$$\left|\mathbb{E}\left\{\Gamma(\lambda_{k-j})\,\Gamma(\lambda_{l-j})\,K(x_k/h)\,K(x_l/h)\right\}\right|\le C\,h\,d_k^{-1} \tag{7.30}$$

(b) *for all $k\ge A_0$, $l-k\ge A_0$ and $l-j\le k$,*

$$\left|\mathbb{E}\left\{\Gamma(\lambda_{k-j})\,\Gamma(\lambda_{l-j})\,K(x_k/h)\,K(x_l/h)\right\}\right|\le C\,h^2\,d_k^{-1}\,d_{l-k}^{-1}. \tag{7.31}$$

(c) *for all $k\ge A_0$, $l-k\ge A_0$ and $l-j>k$,*

$$\begin{aligned}&\left|\mathbb{E}\left\{\Gamma(\lambda_{k-j})\,\Gamma(\lambda_{l-j})\,K(x_k/h)\,K(x_l/h)\right\}\right|\\&\quad\le C\,h^2\begin{cases}\sum_{k=0}^{j}|\phi_k|\,d_k^{-1}\,d_{l-k}^{-2} & \textit{under } \mathbf{LM},\\ k^{-1}(l-k)^{-1}+k^{-1/2}\,(l-k)^{-3/2}. & \textit{under } \mathbf{SM}\end{cases}\end{aligned} \tag{7.32}$$

Similarly, uniformly for $y\in R$, we have

$$\begin{aligned}&\left|\mathbb{E}\left\{K(y+x_l/h)\Gamma(\lambda_{l-j})\,\Gamma(\lambda_{-k})\,\right\}\right|\\&\quad\le C\,h\begin{cases}d_l^{-1} & \textit{if } |l-j+k|\le A_0,\\ \sum_{s=0}^{j}|\phi_s|\,\sum_{s=k}^{l+k}|\phi_s|\,|(d_l^{-3}+l^{-3}), & \textit{if } |l-j+k|>A_0,\end{cases}\end{aligned} \tag{7.33}$$

for any $A_0\ge 1$ and $j,k\ge 0$.

Proof. See Appendix A.6.

Our final lemma gives a tightness criterion for a class of stochastic processes on $D[0,1]$.

LEMMA 7.8. *Let X_{nk} be a sequence of random variables and $X_n(t) = \sum_{k=1}^{\lfloor nt \rfloor} X_{nk}$. The sequence $\{X_n(t)\}$ is tight in $D[0,1]$ if $\max_{1\le k\le n}|X_{nk}| = o_P(1)$ and there exist an integer $A_0 \ge 0$ and a number $\alpha_n(\epsilon,\delta)$ such that*

$$P\Big(\big|\sum_{k=[nt_m]+A_0}^{[ns]} X_{nk}\big| \ge \epsilon \,\big|\, X_n(t_1),...,X_n(t_m)\Big) \le \alpha_n(\epsilon,\delta),$$

and

$$\lim_{\delta\to 0}\lim\sup_{n\to\infty} \alpha_n(\epsilon,\delta) = 0,$$

for each positive $\epsilon > 0$, where $0 \le t_1 \le t_2 \le ... \le t_m \le s \le 1$ and $s - t_m \le \delta$.

Proof. If $A_0 = 0$, Lemma 7.8 is a special case of Theorem 4 in Billingsley (1974)[3]. Extension to integer $A_0 \ge 1$ is trivial under the condition that $\max_{1\le k\le n}|X_{nk}| = o_P(1)$. The details are omitted. □

7.2. *Proofs of propositions.* Propositions 7.4 and 7.7 are treated separately due to their complexity and their proofs are given later in Sections 7.3 and 7.4, respectively.

Proof of Proposition 7.1. We only prove uniformity of $S_{n,h}^2(1)$ for bounded $h > 0$ satisfying $nh/d_n \to \infty$. The other results are similar and simpler. Let $m \ge m_0$ be a constant that will be specified later. Let

$$S_{1n} = \Big(\frac{d_n}{nh}\Big)^{1/2} \sum_{k=1}^n \mathbb{E}\left[f(x_k/h, w_k)|\mathcal{F}_{k-m}\right],$$

$$S_{2n} = \Big(\frac{d_n}{nh}\Big)^{1/2} \sum_{k=1}^n \big\{f(x_k/h, w_k) - \mathbb{E}\left[f(x_k/h, w_k)|\mathcal{F}_{k-m}\right]\big\}.$$

Note that, for any $A \ge 2$,

$$\begin{aligned} &\mathbb{E}S_{n,h}^2(1)\,\mathrm{I}(S_{n,h}^2(1) \ge A) \le 2\,\mathbb{E}S_{1n}^2 + 2\mathbb{E}S_{2n}^2\mathrm{I}(S_{1n}^2 + S_{2n}^2 \ge A/2) \\ &\le 2\,\mathbb{E}S_{1n}^2 + 8A^{-1}\mathbb{E}S_{2n}^4 + 2\mathbb{E}S_{2n}^2\mathrm{I}(S_{1n}^2 \ge A/4) \\ &\le 4\,\mathbb{E}S_{1n}^2 + 16A^{-1}\mathbb{E}S_{2n}^4. \end{aligned}$$

It suffices to show that, for some $c_0 > 0$,

(a) $\mathbb{E}S_{2n}^4 \le c_0 m^4$;
(b) under **LM**, $\mathbb{E}S_{1n}^2 \le c_0\, d_m^{1/2-\mu}$;
(c) under **SM**, $\mathbb{E}S_{1n}^2 \le c_0\,\big(d_m^{-1/2} + \log^2 n/\sqrt{n}\big)$.

Indeed, for any $\epsilon > 0$, by taking A, n sufficiently large and $m = A^{1/8}$, it follows from (a)-(c) that

$$\mathbb{E}S_{n,h}^2(1)\,\mathrm{I}(S_{n,h}^2(1) \ge A) \le 4c_0(d_m^{-1/2} + d_m^{1/2-\mu}) + 16c_1A^{-1/2} + c_0\log^2 n/\sqrt{n} \le \epsilon,$$

under both **LM** and **SM**, due to $d_m \to 0$ and $\mu > 1/2$.

To prove (a), let $Y_{kj} = \mathbb{E}\left[f(x_k/h, w_k)|\mathcal{F}_{k-j}\right] - \mathbb{E}\left[f(x_k/h, w_k)|\mathcal{F}_{k-j-1}\right], 0 \le j \le m-1$. We may write

$$S_{2n} = \big(\frac{d_n}{nh}\big)^{1/2}\sum_{j=0}^{m-1}\sum_{k=1}^n Y_{kj}.$$

Note that Y_{kj} forms a martingale difference. Hölder's and Burkholder's inequalities imply that

$$\mathbb{E}S_{2n}^4 \le m^3\big(\frac{d_n}{nh}\big)^2 \sum_{j=0}^{m-1} \mathbb{E}\big(\sum_{k=1}^n Y_{kj}\big)^4$$
$$\le C_2 m^3\big(\frac{d_n}{nh}\big)^2 \sum_{j=0}^{m-1} \mathbb{E}\big(\sum_{k=1}^n Y_{kj}^2\big)^2 \le c_o\, m^4,$$

for some $c_0 > 0$, which yields (a), where we have used the result (7.10) with $f^2(.)$ replaced by Y_{kj}^2.

We next prove (b) and (c). Let $g_k = \mathbb{E}\big[f(x_k/h, w_k)|\mathcal{F}_{k-m}\big]$. For some $q \ge 1$, we may write

$$\mathbb{E}S_{1n}^2 = \frac{d_n}{nh}\Big[\sum_{k=1}^n \mathbb{E}\, g_k^2 + 2\sum_{k=1}^n \sum_{j=k+1}^{k+q} \mathbb{E}\, g_k g_j + 2\sum_{k=1}^n \sum_{j=k+q}^{n} \mathbb{E}\,(g_k g_j)\Big]$$
$$= R_{n1} + R_{n2} + R_{n3}. \tag{7.34}$$

Recall (6.3). It follows from (7.8) in Lemma 7.1 that $|g_k| \le Ch/d_m$. On the other hand, $\mathbb{E}\,|g_k| \le \mathbb{E}\,|f(x_k/h, w_k)| \le Ch/d_k$. As a consequence, we have

$$|R_{n1}| + |R_{n2}| \le Cqh\,/d_m\,\frac{d_n}{nh}\sum_{k=l_n}^n \mathbb{E}\,|q_k| \le Cqhd_m^{-1}.$$

As for R_{n3}, by taking $m \ge B_0$ where B_0 is given in Lemma 7.6,

(i) under **LM**, it follows from (7.28) that, for any $q \ge B_0$,

$$|R_{n3}| \le \frac{2d_n}{nh}\sum_{k=1}^n \sum_{j=k+q}^n |\mathbb{E}\,(g_k g_j)| \le C\,\frac{hd_n}{n}\sum_{k=1}^n \sum_{j=k+q}^n d_k^{-1} d_{j-k}^{-2}$$
$$\le Ch\int_q^\infty x^{2\mu-3}\rho^{-2}(x)dx.$$

(ii) under **SM**, it follows from (7.29) that, for any $q \ge B_0$,

$$|R_{n3}| \le \frac{2}{\sqrt{nh}}\sum_{k=1}^n \sum_{j=k+q}^n |\mathbb{E}\,(g_k g_j)|$$
$$\le \frac{C\,h}{\sqrt{n}}\sum_{k=1}^n \sum_{j=k+q}^n \big[(j-k)^{-1}\,k^{-1} + (j-k)^{-3/2}\,k^{-1/2}\big]$$
$$\le Ch\big(\log^2 n/\sqrt{n} + \int_q^\infty x^{-3/2}dx\big).$$

Taking these estimates into (7.34), we obtain (b) and (c) by letting $q = \sqrt{d_m}$, as h is bounded. This completes the proof. □

Proof of Proposition 7.2. We prove tightness of $S_{n,h}(t)$. Tightness of $Z_{n,h}(t)$ is shown in a similar way to Theorem 2.20 in Wang (2015)[22] and the details are omitted.

Recalling (6.1) and Lemma 7.8, it suffices to prove the folllowing: for any fixed $s \in [0,1]$, for each $\epsilon > 0$ and any bounded $h > 0$ satisfying $nh/d_n \to \infty$, there exists a sequence of $\alpha_n(\epsilon, \delta)$ satisfying $\lim_{\delta \to 0} \limsup_{n\to\infty} \alpha_n(\epsilon,\delta) = 0$ such that

$$I_n := \sup_{|t-s|\le\delta} \mathrm{P}\Big(\Big| \sum_{k=[ns]+B_0}^{\lfloor nt \rfloor} f(x_k/h, w_k) \Big| \ge \epsilon\, (nh/d_n)^{1/2} \mid \mathcal{F}_{[ns]} \Big) \le \alpha_n(\epsilon,\delta), \tag{7.35}$$

where B_0 is chosen as in Lemma 7.5. In fact, by noting

$$\begin{aligned} J_n(s,t) := \mathbb{E}\Big[\Big| \sum_{k=[ns]+B_0}^{\lfloor nt \rfloor} f(x_k/h, w_k) \Big|^2 \mid \mathcal{F}_{[ns]} \Big] &\le 2 \sum_{k=[ns]+B_0}^{\lfloor nt \rfloor} \sum_{k\le l \le 2B_0} \mathbb{E}\big(|f(x_k/h, w_k)|\, |f(x_l/h, w_l)| \mid \mathcal{F}_{[ns]} \big) \\ &\quad +2 \sum_{k=[ns]+B_0}^{\lfloor nt \rfloor} \sum_{l=k+2B_0}^{n} \Big| \mathbb{E}\big\{ f(x_k/h, w_k)\, f(x_l/h, w_l) \mid \mathcal{F}_{[ns]} \big\} \Big|, \end{aligned}$$

it follows from (7.11) and Lemma 7.5 that, for some $\alpha > 0$:

(a) under **LM** [using (7.25)],

$$\begin{aligned} J_n(s,t) &\le C\, nh(t-s)^\alpha / d_n + C h^2 \sum_{k=[ns]+1}^{\lfloor nt \rfloor} \sum_{l=k+1}^{n} d_{k-[ns]}^{-1}\, d_{l-k}^{-2} \\ &\le 2C\, nh(t-s)^\alpha / d_n; \end{aligned}$$

(b) under **SM** [using (7.27)],

$$\begin{aligned} J_n(s,t) \le\ & C\sqrt{n} h(t-s)^\alpha + \\ & C h^2 \sum_{k=[ns]+1}^{\lfloor nt \rfloor} \sum_{l=k+1}^{n} \big[(l-k)^{-1} (k-[ns])^{-1} + (l-k)^{-3/2} (k-[ns])^{-1/2} \big] \\ &\le 2C\sqrt{n}\, h(t-s)^\alpha. \end{aligned}$$

Now (7.35) follows by choosing $\alpha_n(\epsilon, \delta) = 2C\epsilon^{-2}\delta^\alpha$ and the fact that

$$I_n \le \epsilon^{-2} d_n/(nh) \sup_{|t-s|\le\delta} J_n(s,t) \le \alpha_n(\epsilon,\delta).$$

□

Proof of Proposition 7.3. We start with (7.2). Due to the iid properties of λ_k, there exist constants μ_j with $|\mu_j| \le C$,

$$\begin{aligned} & \Big| \mathbb{E}\, e^{i(Z_{n3}+Z_{n2})} \big\{ [S_{n,h}(t) - S_{n,h}(s)]^2 - [Z_{n,h}(t) - Z_{n,h}(s)] \big\} \Big| \\ \le\ & \mathbb{E} \Big| \mathbb{E}\Big[e^{i\sum_{j=[ns]+1}^{\lfloor nt \rfloor} \mu_j \epsilon_j} \big\{ [S_{n,h}(t) - S_{n,h}(s)]^2 - [Z_{n,h}(t) - Z_{n,h}(s)] \big\} \mid \mathcal{F}_{[ns]} \Big] \Big| \\ \le\ & \frac{d_n}{nh} \sum_{k=[ns]+1}^{n} \sum_{l=k+1}^{n} E \big| \mathbb{E}\big[II_{lk}(h) \mid \mathcal{F}_{[ns]} \big] \big| \end{aligned}$$

$$\leq \frac{d_n}{nh} \sum_{k=[ns]+1}^{n} \Big(\sum_{l=k+1}^{k+B_0} + \sum_{l=k+B_0}^{n} \Big) E\big|\mathbb{E}\left[II_{lk}(h) \mid \mathcal{F}_{[ns]}\right]\big|$$

(7.36) $=: R_{n4} + R_{n5},$

where B_0 and $II_{lk}(h)$ are defined as in Lemma 7.5. Similar to (7.11) with minor modifications, under both **LM** and **SM**, we have $R_{n4} \leq C\, h^{1/2}$. To estimate R_{n5}, under **LM**, it follows from (7.25) that

$$R_{n5} \leq \frac{Cd_n}{nh} h^2 \sum_{k=1}^{n} \sum_{l=k+B_0}^{n} d_k^{-1} \left(d_{l-k}^{-2} + d_{l-k}/\sqrt{n}\right) \leq Ch.$$

Similarly, under **SM**, we have $R_{n5} \leq Ch$ by (7.27). Taking these estimates into (7.36), we have (7.2) as $h \to 0$.

In a similar way for any $q \geq B_0$, we have

$$\begin{aligned}
&\big|\mathbb{E}\, e^{i(Z_{n3}+Z_{n2})}\{[S_n(t) - S_n(s)]^2 - [Y_{nq}(t) - Y_{nq}(s)]\}\big| \\
&\leq \frac{d_n}{n} \sum_{k=[ns]+1}^{n} \sum_{l=k+q}^{n} E\big|\mathbb{E}\left[II_{lk}(1) \mid \mathcal{F}_{[ns]}\right]\big| \\
&\leq \begin{cases} \frac{d_n}{n} \sum_{k=[ns]+1}^{n} \sum_{l=k+q}^{n} d_{k-[ns]}^{-1}\, d_{l-k}^{-2}, \quad \text{under } \textbf{LM}, \\ \frac{1}{\sqrt{n}} \sum_{k=[ns]+1}^{n} \sum_{l=k+q}^{n} \left[(l-k)^{-1}(k-[ns])^{-1} + (l-k)^{-3/2}(k-[ns])^{-1/2}\right], \\ \qquad\qquad \text{under } \textbf{SM}, \end{cases} \\
&\leq C \begin{cases} \int_q^\infty x^{2\mu-3} dx, \qquad \text{under } \textbf{LM}, \\ \int_q^\infty x^{-3/2} dx + \log^2 n/\sqrt{n}, \text{ under } \textbf{SM}, \end{cases} \\
&\leq \epsilon + C \log^2 n/\sqrt{n},
\end{aligned}$$

by choosing q sufficiently large. This proves (7.3). The proof of (7.1) is similar and simpler, so the details are omitted. □

Proof of Proposition 7.5. With $\gamma = 0$ where γ is used in **A1** (i), we may write

(7.37) $$x_r = \sum_{i=1}^{r} \sum_{j=0}^{\infty} \phi_j \epsilon_{i-j} = \sum_{j=1}^{r} a_{r-j}\epsilon_j + \sum_{j=0}^{\infty} [a_{r+j} - a_j]\epsilon_{-j},$$

where $a_l = \sum_{s=0}^{l} \phi_s$ and $a_l = 0$ if $l < 0$. Let $z_r = \sum_{k=1}^{r} \epsilon_k a_{r-k}$ and $z_{1r} = \sum_{j=0}^{m_0}[a_{r+j} - a_j]\epsilon_{-j}$. We have $var(z_r) \sim d_r^2$ for $r \geq 2m_0$ and, when m_0 is fixed,

$$\begin{aligned}
|\mathbb{E}\hat{f}(s, w_0)e^{-isz_{1r}}| &\leq \mathbb{E}\,|\hat{f}(s, w_0)(e^{-isz_{1r}} - 1)| + |\mathbb{E}\hat{f}(s, w_0)| \\
&\leq C\,(1 + |a_r|) \min\{1, |s|\}.
\end{aligned}$$

Now it is readily seen from the iid properties of ϵ_k and (7.18) in Lemma 7.3 that

$$\begin{aligned}
|G_r| &\leq \frac{1}{2\pi} \int_{-\infty}^{\infty} |\mathbb{E}\left\{\hat{f}(s, w_0)e^{-isz_{1r}}\right\}|\,|\mathbb{E}\left\{\hat{f}(s, w_r)e^{-isz_r}\right\}|ds \\
&\leq C\,(1 + |a_r|) \int_{-\infty}^{\infty} \min\{1, |s|\}\,|\mathbb{E}\left\{\hat{f}(s, w_r)e^{-isz_r}\right\}|ds \\
&\leq C\,(1 + |a_r|)(d_r^{-3} + r^{-3}).
\end{aligned}$$

Hence $\sum_{r=2m_0}^{\infty}|G_r|<\infty$ due to $|a_r|\le C$ under **SM** and $|a_r|\le d_r$ under **LM**.

To prove (6.11) and $\sum_{r=1}^{\infty}|\tilde G_r|<\infty$, we make use of (7.33) in Lemma 7.7. In fact, for any $r\ge 1$ and $y\in R$, it follows from (7.33) that

$$\begin{aligned}
&|\mathbb{E}\left\{(u_{10}u_{1r}-u_0u_r)K(y+x_r)\right\}| \\
&\le \Big(\sum_{k=m_0+1}^{\infty}\sum_{j=0}^{\infty}+\sum_{k=0}^{\infty}\sum_{j=m_0+1}^{\infty}\Big)\big|\mathbb{E}\left\{\psi_k\lambda'_{-k}\,\psi_j\lambda'_{r-j}K(y+x_r)\right\}\big| \\
&\le 2\sum_{k=m_0+1}^{\infty}\sum_{j=r+k-1}^{r+k+1} d_r^{-1}\,||\psi_k||\,||\psi_j|| \\
&\qquad +2\sum_{k=m_0+1}^{\infty}\sum_{j=0}^{\infty}||\psi_k||\,||\psi_j||\sum_{s=0}^{j}|\phi_s|\sum_{s=k}^{r+k}|\phi_s|\,|(d_r^{-3}+r^{-3}) \\
&\le 2d_r^{-1}\sum_{k=m_0+1}^{\infty}||\psi_k||\sum_{j=-1}^{1}||\psi_{j+r+k}|| \\
&\qquad +2C\sum_{k=m_0+1}^{\infty}\sum_{j=0}^{\infty}||\psi_k||\,||\psi_j||\sum_{s=0}^{j}|\phi_s|\sum_{s=k}^{r+k}|\phi_s|\,|(d_r^{-3}+r^{-3}).
\end{aligned}$$

Note that $\sum_{s=0}^{j}|\phi_s|\sum_{s=k}^{r+k}|\phi_s|(d_r^{-3}+r^{-3})\le Cj^{1/2}\,k^{1/2}r^{-3/2}$ under both **SM** and **LM**. It is readily seen from $\sum_{k=0}^{\infty}k^{1/2}\,||\psi_k||<\infty$ that

$$\begin{aligned}
\sum_{r=1}^{\infty}|\tilde G_r-\tilde G_{r,m_0}| &\le \int_{-\infty}^{\infty}K(y)\sum_{r=1}^{\infty}|\mathbb{E}\left\{(u_{10}u_{1r}-u_0u_r)K(y+x_r)\right\}|dy \\
&\le C\sum_{k=m_0+1}^{\infty}k^{1/2}\,||\psi_k||\int K(y)dy\to 0, \qquad (7.38)
\end{aligned}$$

as $m_0\to\infty$. Similarly, we have $|\tilde G_0-\tilde G_{0,m_0}|\to 0$, as $m_0\to\infty$, and $\sum_{r=1}^{\infty}|\tilde G_r|<\infty$. The proof of Proposition 7.5 is then complete. □

Proof of Proposition 7.6. The proofs of (6.13) and (6.14) are simply established using Lemma 7.1. Indeed, by noting that

$$\begin{aligned}
&\Big|\sum_{k=1}^{n-r}K(x_k)K(x_{k+r})\,(u_{1k}\,u_{1,k+r}-u_ku_{k+r})\Big| \\
&\le C\Big(\sum_{l=m_0+1}^{\infty}\sum_{l_1=0}^{\infty}+\sum_{l=0}^{\infty}\sum_{l_1=m_0+1}^{\infty}\Big)\sum_{k=1}^{n-r}K(x_k)\,\big|\psi_l\lambda'_{k-l}\,\psi_{l_1}\lambda'_{k+r-l_1}\big|,
\end{aligned}$$

it follows from Lemma 7.1 (i) and $\sum_{l=0}^{\infty}l||\psi_l|<\infty$ that, for some constant $A_0>0$,

$$\begin{aligned}
\mathbb{E}|R_n| &\le C\,M_1\sum_{l=m_0+1}^{\infty}\sum_{l_1=0}^{\infty}||\psi_l||\,||\psi_{l_1}||\frac{d_n}{n}\,\Big[(A_0+2)+\sum_{k=1}^{n}d_k^{-1}\Big] \\
&\le C_1\,M_1\sum_{l=m_0+1}^{\infty}||\psi_l|\le CM_1m_0^{-1}.
\end{aligned}$$

Hence (6.14) follows if we take $M_1 = \sqrt{m_0}$. The proof of (6.13) is similar.

We next prove (7.6). Let $\sum_{j=k}^{l} = 0$ for $k > l$ and $\Gamma(.)$ be a measurable function with $\Gamma(\lambda_1) = 0$ and $\mathbb{E}\Gamma^2(\lambda_1) < \infty$. Since $K(x)$ is bounded, for A_0 being chosen as in Lemma 7.7, we have

$$\begin{aligned}
\Delta_n &\equiv \Big|\sum_{k=1}^{n} \Gamma(\lambda_{k-j})\, K(x_k/h)\Big|^2 \\
&\le 2\Big|\sum_{k=A_0}^{n} \Gamma(\lambda_{k-j})\, K(x_k/h)\Big|^2 + C\Big(\sum_{k=1}^{A_0} |\Gamma(\lambda_{k-j})|\Big)^2 \\
&= 2\Big(\sum_{k=A_0}^{n} \sum_{|k-l|<A_0}^{n} + 2\sum_{k=A_0}^{n-1} \sum_{l=k+A_0}^{n}\Big) \Gamma(\lambda_{k-j})\, \Gamma(\lambda_{l-j})\, K(x_k/h)\, K(x_l/h) \\
&\qquad + C\Big(\sum_{k=1}^{A_0} |\Gamma(\lambda_{k-j})|\Big)^2 \\
&=: \Delta_{1n} + \Delta_{2n} + \Delta_{3n}, \quad say. \qquad (7.39)
\end{aligned}$$

It follows from Lemma 7.7 that

$$\mathbb{E}|\Delta_{1n}| \le C\, h \sum_{k=1}^{n} \sum_{|k-l|<A_0}^{n} 1/d_k \le C_1\, nh/d_n,$$

$$\begin{aligned}
\mathbb{E}|\Delta_{2n}| &\le C\, h^2 \begin{cases} \sum_{k=A_0}^{n-1} d_k^{-1} \Big(\sum_{l=k+A_0}^{n\wedge(k+j)} d_{l-k}^{-1} + \sum_{k=0}^{j} |\phi_k| \sum_{l=k+j}^{n} d_{l-k}^{-2}\Big), & \text{under } \mathbf{LM}, \\ \sum_{k=A_0}^{n-1} k^{-1/2} \sum_{l=k+A_0}^{n\wedge(k+j)} (l-k)^{-1/2} + \\ \quad \sum_{k=A_0}^{n-1} \sum_{l=k+j}^{n} \big[k^{-1}(l-k)^{-1} + k^{-1/2}\,(l-k)^{-3/2}\big], & \text{under } \mathbf{SM} \end{cases} \\
&\le C\,(nh^2/d_n) \begin{cases} j/d_j + \sum_{k=0}^{j} |\phi_k|, & \text{under } \mathbf{LM}, \\ j^{1/2} + \log^2 n/\sqrt{n} + 1, & \text{under } \mathbf{SM}, \end{cases} \\
&\le C\, j^{1/2}\, nh^2/d_n,
\end{aligned}$$

where we have used the fact $\sum_{k=0}^{j} |\phi_k| \le C\, j/d_j \le C\, j^{1/2}$ under **LM**. On the other hand, it is readily seen that $\mathbb{E}|\Delta_{3n}| \le C\, A_0^2$.

Taking these estimates into (7.39), for any bounded h, we have

$$\mathbb{E}\Big|\sum_{k=1}^{n} \Gamma(\lambda_{k-j})\, K(x_k/h)\Big|^2 \le C\, j^{1/2}\, nh/d_n. \qquad (7.40)$$

The result (6.12) now follows from

$$\begin{aligned}
&\mathbb{E}\Big|\sum_{k=1}^{n} u_{k,m_0}\, K(x_k/h)\Big|^2 = \mathbb{E}\Big|\sum_{j=m_0}^{\infty} \sum_{k=1}^{n} \psi_j\, \lambda'_{k-j}\, K(x_k/h)\Big|^2 \\
&\le \sum_{j=m_0}^{\infty} j^{1/4}(|\psi_{1j}| + |\psi_{2j}|) \sum_{j=m_0}^{\infty} j^{-1/4}(|\psi_{1j}| + |\psi_{2j}|)^{-1}\mathbb{E}\Big|\sum_{k=1}^{n} \psi_j\, \lambda'_{k-j}\, K(x_k/h)\Big|^2 \\
&\le 2 \sum_{j=m_0}^{\infty} j^{1/4}(|\psi_{1j}| + |\psi_{2j}|) \sum_{j=m_0}^{\infty} j^{-1/4}(|\psi_{1j}| + |\psi_{2j}|)
\end{aligned}$$

$$\Big(\mathbb{E}\big|\sum_{k=1}^{n}\epsilon_{k-j}\,K(x_k/h)\,\big|^2+\mathbb{E}\big|\sum_{k=1}^{n}e_{k-j}\,K(x_k/h)\,\big|^2\Big)$$

$$\le C\,(nh/d_n)\,\big[\sum_{j=m_0}^{\infty}j^{1/4}(|\psi_{1j}|+|\psi_{2j}|)\big]^2,$$

where we employ Hölder's inequality and (7.40) with $\Lambda(\lambda_k)=\epsilon_k$ and e_k, respectively. The proof of Proposition 7.6 is complete. □

7.3. *Proof of Proposition 7.4.* We start with (7.4). The tightness of $Z_{n,h}(t)$ has been established in Proposition 7.2. It suffices to show that the finite-dimensional distributions of $Z_{n,h}(t)$ converge to those of $\tau^2 L_Z(t,0)$. To this end, let $g(x)=\mathbb{E}\,f^2(x,w_1)$. Under **A2**(b) and **A3**(I), $g(x)$ is bounded and integrable. Furthermore, by using Theorem 2.20 of Wang (2015), we have

$$\frac{d_n}{nh}\sum_{k=1}^{\lfloor nt\rfloor}g(x_k/h)\Rightarrow\tau^2L_Z(t,0), \tag{7.41}$$

whenever $d_n/h\to\infty$ and $d_n/nh\to 0$. In terms of (7.41), the finite-dimensional distribution of $Z_{n,h}(t)$ will converge to those of $\tau^2 L_Z(t,0)$ if we show that, for any fixed $0<t\le 1$,

$$\frac{d_n}{nh}\sum_{k=1}^{\lfloor nt\rfloor}\big[g(x_k/h)-f^2(x_k/h,w_k)\big]=o_P(1). \tag{7.42}$$

This is essentially the same as in the proof of (A.20) for $i=2$ in Wang, et al. (2021)[28] (also see (4.8) in the paper) and hence the details are omitted. (7.4) is now proved.

We next prove (7.5). It suffices to show the following:

(a) for each $0\le r\le m$, $\{\psi_{nr}(t)\}_{n\ge1}$ is tight on $D[0,1]$; and
(b) the finite-dimensional distributions of $\big\{\psi_{n0}(t),\psi_{n1}(t),...,\psi_{nm}(t)\big\}$ converge to those of $\big\{G_0\,,G_1\,,...,G_m\,\big\}\,L_Z(t,0)$.

The proof of part (a) is simple. Indeed, by noting

$$|\psi_{nr}(t)-\psi_{nr}(s)|\le\frac{d_n}{n}\sum_{k=[ns]+1}^{\lfloor nt\rfloor}|f(x_k,w_k)f(x_{k+r},w_{k+r})|$$
$$\le\frac{d_n}{n}\sum_{k=[ns]+1}^{\lfloor nt\rfloor+r}f^2(x_k,w_k)\le|Z_{n,1}(t)-Z_{n,1}(s)|+o_P(1),$$

uniformly for $s<t$, the tightness of $\psi_{nr}(t)$ is implied by that of $Z_{n,1}(t)$.

To prove part (b), let $h_r(y)=\mathbb{E}\big\{f(y,w_0)f(y+x_r,w_r)\big\}$. We have $h_r(y)$ is bounded and integrable due to **A2**(b) and **A3**(I). Hence, as in (7.41),

$$\frac{d_n}{n}\sum_{k=1}^{\lfloor nt\rfloor}\big[\alpha_0h_0(x_k)+...+\alpha_mh_m(x_k)\big]\Rightarrow\sum_{r=0}^{m}\alpha_rG_rL_Z(t,0),$$

on $D[0,1]$, for any $(\alpha_0,...,\alpha_m)\in R^{m+1}$. The Cramér-Wold theorem now implies that part (b) will follow if we prove

$$\big|\psi_{nr}(t)-\frac{d_n}{n}\sum_{k=1}^{\lfloor nt\rfloor}h_r(x_k)\big|=o_P(1), \tag{7.43}$$

for any $r \geq 0$ and any fixed $0 \leq t \leq 1$[3].

The proof of (7.43) is quite technical, starting with some preliminaries. Let $a_l = \sum_{s=0}^{l} \phi_s$ and $a_l = 0$ if $l < 0$. With $\gamma = 0$, we may write

$$x_k = \sum_{j=-\infty}^{0} [a_{k-j} - a_{-j}]\epsilon_j + \sum_{j=1}^{k} a_{k-j}\epsilon_j, \tag{7.44}$$

and

$$\begin{aligned} x_{k+r} - x_k &= \sum_{j=-\infty}^{k} [a_{k+r-j} - a_{k-j}]\epsilon_j + \sum_{j=k+1}^{k+r} a_{k+r-j}\epsilon_j \\ &= \sum_{j=-\infty}^{0} [a_{r-j} - a_{-j}]\epsilon_{j+k} + \sum_{j=1}^{r} a_{r-j}\epsilon_{j+k} \\ &= x_{1k,r} + x_{2k,r}, \end{aligned} \tag{7.45}$$

where

$$x_{1k,r} = \sum_{j=-\infty}^{-A_0} [a_{r-j} - a_{-j}]\epsilon_{j+k},$$

$$x_{2k,r} = \sum_{j=-A_0+1}^{0} [a_{r-j} - a_{-j}]\epsilon_{j+k} + \sum_{j=1}^{r} a_{r-j}\epsilon_{j+k}.$$

It is readily seen that, for any $A_0 > 0$, $x_{1k,r}$ is independent of $x_{2k,r}$ and $x_{1k,r}$ is independent of w_k and w_{k+r} when $A_0 \geq m_0 + 1$. By letting $\gamma_j = a_{r+j} - a_j$, we further have $\sum_{j=1}^{\infty} \gamma_j^2 < \infty$ and

$$x_{1k,r} = \sum_{j=-\infty}^{-A_0} [a_{r-j} - a_{-j}]\epsilon_{j+k} = \sum_{q=1}^{k-A_0} \gamma_{k-q}\epsilon_q + \sum_{q=-\infty}^{0} \gamma_{k-q}\epsilon_q. \tag{7.46}$$

We next let $\hat{f}(t,s) = \int_{-\infty}^{\infty} e^{itx} f(x,s)dx$,

$$V_k(t,s) = \hat{f}(-t, w_k)\hat{f}(s, w_{k+r})e^{-isx_{2k,r}},$$
$$A_r(t,s) = \mathbb{E}\left\{\hat{f}(-t, w_0)\hat{f}(s, w_r)e^{-isx_r}\right\}.$$

Using the Fourier transformations, under **A3** (III), it is readily seen that

$$h_{1r}(y,s) := \frac{1}{2\pi}\int e^{i(t-s)y}\mathbb{E}\,V_0(t,s)dt = e^{-isy}\,\mathbb{E}\left\{f(y,w_0)\hat{f}(s,w_r)e^{-isx_{20,r}}\right\},$$

$$h_{2r}(y,s) := \frac{1}{2\pi}\int e^{i(t-s)y}A_r(t,s)dt = e^{-isy}\,\mathbb{E}\left\{f(y,w_0)\hat{f}(s,w_r)e^{-isx_r}\right\},$$

$$h_r(y) = \mathbb{E}\left\{f(y,w_0)f(y+x_r,w_r)\right\} = \frac{1}{2\pi}\int h_{2r}(y,s)ds.$$

[3]We remark that the r in (7.43) is allowed to depend on n and we have in fact established the convergence in (7.43) in L_1 rather than in probability. These enhanced properties will be useful in the proof of Proposition 7.7.

We are now ready to consider (7.43). Without loss of generality, assume $t=1$. We have

$$\begin{aligned}
\psi_{nr}(1) &= \frac{d_n}{2\pi n}\sum_{k=1}^n f(x_k,w_k)\int \hat f(s,w_{k+r})e^{-isx_{k+r}}ds \\
&= \frac{d_n}{(2\pi)^2 n}\sum_{k=1}^n \int\int_{|s|\le A}\hat f(-t,w_k)\hat f(s,w_{k+r})e^{i(t-s)x_k-is(x_{k+r}-x_k)}dsdt+R_{0A}, \\
&= \frac{d_n}{(2\pi)^2 n}\sum_{k=1}^n \int\int_{|s|\le A} e^{i(t-s)x_k-isx_{1k,r}}\mathbb{E}V_k(t,s)dsdt+R_{1A}+R_{0A} \\
&= \frac{d_n}{2\pi n}\sum_{k=1}^n \int_{|s|\le A} e^{-isx_{1k,r}}h_{1r}(x_k,s)ds+R_{1A}+R_{0A} \\
&= \frac{d_n}{2\pi n}\sum_{k=1}^n \int_{|s|\le A} e^{-isx_{1k,r}}h_{2r}(x_k,s)ds+R_{2A}+R_{1A}+R_{0A} \\
&= \frac{d_n}{2\pi n}\sum_{k=1}^n \int_{|s|\le A} h_{2r}(x_k,s)ds+R_{3A}+R_{2A}+R_{1A}+R_{0A} \\
&=: \frac{d_n}{n}\sum_{k=1}^n h_r(x_k)-R_{4A}+R_{3A}+R_{2A}+R_{1A}+R_{0A}, \qquad (7.47)
\end{aligned}$$

where

$$\begin{aligned}
R_{0A} &= \frac{d_n}{2\pi n}\sum_{k=1}^n f(x_k,w_k)\int_{|s|>A}\hat f(s,w_{k+r})e^{-isx_{k+r}}ds, \\
R_{1A} &= \frac{d_n}{(2\pi)^2 n}\sum_{k=1}^n \int_{|s|\le A}\int e^{i(t-s)x_k-isx_{1k,r}}\big[V_k(t,s)-\mathbb{E}V_k(t,s)\big]dt\,ds, \\
R_{2A} &= \frac{d_n}{2\pi n}\int_{|s|\le A}\sum_{k=1}^n e^{-isx_{1k,r}}\big[h_{1r}(x_k,s)-h_{2r}(x_k,s)\big]ds, \\
R_{3A} &= \frac{d_n}{2\pi n}\int_{|s|\le A}\sum_{k=1}^n \big(e^{-isx_{1k,r}}-1\big)\,h_{2r}(x_k,s)dtds \\
&= \frac{d_n}{(2\pi)^2 n}\sum_{k=1}^n \int_{|s|\le A}\int e^{i(t-s)x_k}\big(e^{-isx_{1k,r}}-1\big)A_r(t,s)ds, \\
R_{4A} &= \frac{d_n}{2\pi n}\sum_{k=1}^n \int_{|s|>A} h_{2r}(x_k,s)ds\,.
\end{aligned}$$

Recalling w_k depends only on $\lambda_k,...,\lambda_{k-m_0}$, where m_0 is a fixed integer, it follows from Lemma 7.1 (i) and $|f(y,w_0)|\le T(y)(1+||w_0||^\beta)$ that

$$\begin{aligned}
\mathbb{E}\,|R_{0A}| &\le C\,\frac{d_n}{n}\sum_{k=1}^n \int_{|s|>A}\mathbb{E}\big\{\,|f(x_k,w_k)|\,|\hat f\big(s,w_{k+r}\big)|\big\}ds \\
&\le C\,\frac{d_n}{n}\sum_{k=1}^n d_k^{-1}\int_{|s|>A}\int \mathbb{E}\big\{|f(y,w_0|\,|\hat f(s,w_r)|\big\}dyds
\end{aligned}$$

$$\leq C \int T(y)dy \int_{|s|>A} \mathbb{E}\{|\hat{f}(s,w_r)|(1+||w_0||^{\beta})\}ds \to 0,$$

as $A \to \infty$. Similarly,

$$\begin{aligned}\mathbb{E}\,|R_{4A}| &\leq C\,\frac{d_n}{n}\sum_{k=1}^{n}\int_{|s|>A}\mathbb{E}\,|h_{2r}(x_k,s)|ds \\ &\leq C\,\frac{d_n}{n}\sum_{k=1}^{n} d_k^{-1}\int_{|s|>A}\int |h_{2r}(y,s)|dyds \\ &\leq C\int_{|s|>A}\int \mathbb{E}\{|f(y,w_0|\,|\hat{f}(s,w_r)|\}dyds \to 0,\end{aligned}$$

as $A \to \infty$. Hence, $|R_{0A}| + |R_{4A}| = o_P(1)$, as $n \to \infty$ first and then $A \to \infty$. This, together with (7.47), implies that (7.43) will follow if we prove: for any fixed $A > 0$,

$$R_{jA} = o_P(1), \qquad j = 1,\, 2,\, 3, \tag{7.48}$$

as $n \to \infty$ frist and then $A_0 \to \infty$.

The proof of (7.48) for $j = 2$ is simple. Indeed, due to the independence between $x_{10,r}$ and w_1, w_r, we have

$$\begin{aligned}&\int_{|s|\leq A}\int \big|h_{1r}(y,s) - h_{2r}(y,s)\big|\,\big|dyds \\ &\leq \int_{|s|\leq A}\int \mathbb{E}\{|f(y,w_0)|\,|\hat{f}(s,w_r)|\,|e^{-isx_{10,r}} - 1|\}dyds \\ &\leq A\int\int \mathbb{E}\{|f(y,w_0)|\,|\hat{f}(s,w_r)|dyds\,\mathbb{E}\,|x_{10,r}| \\ &\leq CA\Big[\sum_{j=A_0}^{\infty}(a_{r+j} - a_j)^2\Big]^{1/2},\end{aligned}$$

for any fixed $A > 0$. This yields that

$$\begin{aligned}\mathbb{E}\,|R_{2A}| &\leq \frac{d_n}{2\pi n}\int_{|s|\leq A}\sum_{k=1}^{n}\mathbb{E}\,\big|h_{1r}(x_k,s) - h_{2r}(x_k,s)\big|ds \\ &\leq \frac{d_n}{n}\sum_{k=1}^{n} d_k^{-1}\int_{|s|\leq A}\int |h_{1r}(y,s) - h_{2r}(y,s)|dyds \\ &\leq C\,A\Big[\sum_{j=A_0}^{\infty}(a_{r+j} - a_j)^2\Big]^{1/2} \to 0,\end{aligned}$$

as $n \to \infty$ first and then $A_0 \to \infty$, as required.

It is readily seen that (7.48) for $j = 1$ and 3 will follow if we prove: for any fixed $A > 0$,

$$\frac{d_n}{n}\sup_{|s|\leq A}\mathbb{E}\Big|\sum_{k=1}^{n}\int e^{i(u-s)x_k - isx_{1k,r}}\big[V_k(u,s) - \mathbb{E}V_k(u,s)\big]du\Big| = o(1), \tag{7.49}$$

$$\frac{d_n}{n}\sup_{|s|\leq A}\mathbb{E}\Big|\sum_{k=1}^{n}\int e^{i(u-s)x_k}\big(e^{-isx_{1k,r}} - 1\big)A_r(u,s)du\Big| = o(1), \tag{7.50}$$

as $n\to\infty$ first and then $A_0\to\infty$.

We first prove (7.50). We may write, for any $B\geq 1$ and $|s|\leq A$,

$$\begin{aligned}
&\sum_{k=1}^{n}\int e^{iux_k}\big(e^{-isx_{1k,r}}-1\big)A_r(u+s,s)du\\
&=\sum_{k=1}^{n}\Big(\int_{|u|\geq B/d_k}+\int_{|u|<B/d_k}\Big)e^{iux_k}\big(e^{-isx_{1k,r}}-1\big)A_r(u+s,s)du\\
&=\Delta_{1n}(s)+\Delta_{2n}(s),\quad \text{say.} \qquad (7.51)
\end{aligned}$$

Recalling $|f(x,y)|\leq T(x)(1+||y||^{\beta})$, where $T(x)$ is a bounded and integrable function, we have

$$\sup_{u,s}|A_r(u,s)|\leq\int\int\mathbb{E}\big\{|f(x,w_0)|f(y,w_r)|\big\}dxdy<\infty, \qquad (7.52)$$

$$\begin{aligned}
\sup_s\int|A_r(u,s)|du&\leq\int\int\mathbb{E}\big\{|\hat f(t,w_0)|f(x,w_r)|\big\}dtdx\\
&\leq\int T(x)dx\int\mathbb{E}\big\{|\hat f(t,w_0)\big(1+||w_r||^{\beta}\big)\big\}dt<\infty, \qquad (7.53)
\end{aligned}$$

$$\begin{aligned}
\frac{1}{2\pi}\sup_s\int\Big|\int A_r(t+s,s)e^{ity}dt\Big|dy&=\sup_s\int|h_{2r}(y,s)|dy\\
&\leq C\int\int\mathbb{E}\big\{|f(t,w_0)|\hat f(x,w_r)|\big\}dtdx<\infty. \qquad (7.54)
\end{aligned}$$

Due to (7.52), it is readily seen that, uniformly for $|s|\leq A$ and any $B>0$,

$$\begin{aligned}
\mathbb{E}|\Delta_{2n}(s)|&\leq C\sup_{|u|,|s|\leq A}|A_r(u+s,s)|\,B\sum_{k=1}^{n}d_k^{-1}\,\mathbb{E}\,|x_{1k,r}|\\
&\leq C\,Bn/d_n\,\Big[\sum_{k=A_0}^{\infty}(a_{r+k}-a_k)^2\Big]^{1/2}. \qquad (7.55)
\end{aligned}$$

To consider $\Delta_{1n}(s)$, writing $\Delta_{1n}(s)=\Delta_{1n,1}(s)+\Delta_{1n,2}(s)$, where

$$\Delta_{1n,1}(s)=\sum_{k=1}^{n}\int_{|u|\geq B/d_k}e^{iux_k-isx_{1k,r}}A_r(u+s,s)du,$$

$$\Delta_{1n,2}(s)=\sum_{k=1}^{n}\int_{|u|\geq B/d_k}e^{iux_k}A_r(u+s,s)du,$$

then (7.50) will follow if we prove

$$\frac{d_n}{n}\sup_{|s|\leq A}\mathbb{E}\,|\Delta_{1n,i}(s)|\leq C\,(n/d_n)\sqrt{B^{-1}+BA_0^2d_n/n},\qquad i=1,2. \qquad (7.56)$$

Indeed, due to (7.51) - (7.56) and $\tau_{A_0}:=\sum_{k=A_0}^{\infty}(a_{r+k}-a_k)^2\to 0$ as $A_0\to\infty$, (7.50) follows by taking $B=\tau_{A_0}^{-1/3}$.

We only prove (7.56) for $i=1$ as the result for $i=2$ is similar. We have

$$\mathbb{E}|\Delta_{1n,1}(s)|^2\leq\sum_{k=1}^{n}\sum_{j=1}^{n}\Big|\int_{|t|\geq B/d_k}\int_{|u|\geq B/d_j}A_r(t+s,s)\,A_r(u+s,s)\,\mathbb{E}\,T_{kj}\,dtdu\Big|$$

$$= \Big(\sum_{|k-j|\ge A_0^2+1} + \sum_{|k-j|\le A_0^2} \Big) \Big| \int_{|t|\ge B/d_k} \int_{|u|\ge B/d_j} A_r(t+s,s)\, A_r(u+s,s)\, \mathbb{E} T_{kj} dt du \Big|$$

$$=: \Omega_{1n} + \Omega_{2n}, \quad \text{say}, \tag{7.57}$$

where $T_{k,j} = e^{itx_k + iux_j}\, e^{-is(x_{1k,r}+x_{1j,r})}$. Recalling (7.46), it follows that

$$\begin{aligned} &\big|\mathbb{E}(T_{kj} \mid \mathcal{F}_0)\big| \\ &\le \big|\mathbb{E}(e^{itx_k+iux_j}\, e^{-is\sum_{q=1}^{k-A_0}\gamma_{k-q}\epsilon_q} e^{-is\sum_{q=1}^{j-A_0}\gamma_{j-q}\epsilon_q} \mid \mathcal{F}_0)\big| \\ &= \big|\mathbb{E}(e^{itx_k+iux_j}\, e^{-i\sum_{q=1}^{k\vee j} s\gamma_q'\epsilon_q} \mid \mathcal{F}_0)\big|, \end{aligned} \tag{7.58}$$

where

$$\gamma_q' = \begin{cases} \gamma_{k-q}+\gamma_{j-q}, & \text{if } 1\le q< k\wedge j, \\ \gamma_{k\vee j-q}, & \text{if } k\wedge j \le q < k\vee j - A_0, \\ 0, & \text{if } q\ge k\vee j - A_0, \end{cases}$$

satisfying $\sum_{q=1}^{\infty} \gamma_q'^2 < \infty$. Now, by noting (7.52) and using (7.23), we have that, uniformly for $|s|\le A$,

$$\begin{aligned} \Omega_{1n} &\le 2\,\mathbb{E} \sum_{k-j\ge A_0^2+1} \int_{|t|\ge B/d_k}\int_{|u|\ge B/d_j} |A_r(t+s,s)A_r(u+s,s)|\,|\mathbb{E}(T_{kj}\mid\mathcal{F}_0)| dt du \\ &\le C \sum_{l-k\ge A_0^2+1} \big[(l-k)^{-2} + B^{-1}\, d_{l-k}^{-1}\big] d_k^{-1} \\ &\le C\, B^{-1}\, (n/d_n)^2. \end{aligned}$$

Turning to consider Ω_{2n}, note that

$$\begin{aligned} \mathbb{E}\Big|\int_{|t|\ge B/d_k} A_r(t+s,s)e^{itx_k}dt\Big| &\le B/d_k \sup_{t,s}|A_r(t+s,s)| \\ &\quad + \mathbb{E}\Big|\int A_r(t+s,s)e^{itx_k}dt\Big| \\ &\le CB/d_k + Cd_k^{-1}\int\Big|\int A_r(t+s,s)e^{ity}dt\Big|dy \le CB/d_k, \end{aligned}$$

due to (7.52) and (7.53). Uniformly for $|s|\le A$, we have

$$\begin{aligned} |\Omega_{2n}| &\le \sum_{|k-j|\le A_0^2}\int_{|u|\ge B/d_j}|A_r(u+s,s)|du\,\mathbb{E}\Big|\int_{|t|\ge B/d_k}A_r(t+s,s)\,T_{kj}dt\Big| \\ &\le \sum_{|k-j|\le A_0^2}\int_{|u|\ge B/d_j}|A_r(u+s,s)|\,\mathbb{E}\Big|\int_{|t|\ge B/d_k}A_r(t+s,s)e^{itx_k}dt\Big|du \\ &\le CBA_0^2\, n/d_n. \end{aligned}$$

Taking this estimate into (7.57), for any fixed $A>0$, we have

$$\sup_{|s|\le A}\mathbb{E}|\Delta_{1n,1}(s)|^2 \le C\big(B^{-1} + B\,A_0^2 d_n/n\big)\,(n/d_n)^2, \tag{7.59}$$

yielding (7.56). Then (7.50) is established.

Finally, we prove (7.49). Let $\sigma_k(t,s) = V_k(t,s) - \mathbb{E}V_k(t,s)$. Uniformly for $|s| \le A$ where A is fixed, we have

$$\mathbb{E}\Big|\sum_{k=1}^n \int e^{itx_k - isx_{1k,r}} \sigma_k(t+s,s)dt\Big|^2$$
$$= \sum_{k=1}^n \sum_{j=1}^n \mathbb{E} \int\int e^{-is(x_{1k,r}+x_{1j,r})} e^{itx_k + iux_j} \sigma_k(t+s,s)\sigma_j(u+s,s)dtdu$$
$$= \Big(\sum_{|j-k|\ge A_0^2+1} + \sum_{|j-k|\le A_0^2}\Big)\mathbb{E}\int\int e^{-is(x_{1k,r}+x_{1j,r})} e^{itx_k+iux_j}\sigma_k(t+s,s)\sigma_j(u+s,s)dtdu$$
$$=: R_{n6} + R_{n7}, \quad \text{say.} \tag{7.60}$$

Note that $\sigma_k(t+s,s)$ depends only on $\epsilon_{k+r}, ..., \epsilon_{k-A_0}$, $\mathbb{E}\sigma_k(u+s,s) = 0$ and

$$\sup_{t,s} |\sigma_k(t+s,s)| \le C + \sup_t |\hat f(t,w_k)| \sup_t |\hat f(t, w_{k+r})|$$
$$\le C(1 + ||w_k||^{2\beta} + ||w_{k+r}||^{2\beta}).$$

As in the proof of (7.50), it follows from (7.20) in Lemma 7.4 that

$$|R_{n6}| \le \sum_{|j-k|\ge A_0^2+1} \Big|\mathbb{E}\int\int e^{-is(x_{1k,r}+x_{1j,r})} e^{itx_k+iux_j}\sigma_k(t,s)\sigma_j(u,s)dtdu\Big|$$
$$\le \sum_{|j-k|\ge A_0^2+1} \mathbb{E}\int\int \Big|\mathbb{E}\big[e^{itx_k+iux_j - is\sum_{q=1}^{k\vee j}\gamma_q'\epsilon_q}\sigma_k(t+s,s)\sigma_j(u+s,s) \mid \mathcal{F}_0\big]\Big|dtdu$$

(where γ_q' is given as in (7.58))

$$\le C \sum_{|j-k|\ge A_0^2+1} d_k^{-1} d_{|j-k|}^{-2}$$
$$\le C \begin{cases} n/d_n, & \text{under } \mathbf{LM}, \\ n \log n/d_n, & \text{under } \mathbf{SM}. \end{cases} \tag{7.61}$$

To consider R_{n7}, let $l_k(y) = \int e^{ity}\sigma_k(t+s,s)dt$. It is readily seen that

$$|l_k(y)| \le |f(y,w_k)||\hat f(s,w_{k+r})| + \mathbb{E}\big\{|f(y,w_k)||\hat f(s,w_{k+r})|\big\}$$
$$\le C\,|f(y,w_k)|(1+||w_{k+r}||^\beta) + C\,\mathbb{E}\big\{|f(y,w_k)|(1+||w_{k+r}||^\beta)\big\}$$

and by Lemma 7.1

$$\mathbb{E}\,|l_k(x_k)|^2 \le C d_k^{-1}\mathbb{E}\,(1+||w_1||^{4\beta}) \le C_1 d_k^{-1}.$$

This yields that

$$|R_{n7}| \le \sum_{|j-k|\le A_0^2+1} \mathbb{E}\big\{|l_k(x_k)|\,|l_j(x_j)|\big\} \le C_1 \sum_{|j-k|\le A_0^2+1} d_k^{-1} \le C A_0^2 n/d_n. \tag{7.62}$$

It follows from (7.60)-(7.62) that

$$\frac{d_n}{n}\mathbb{E}\Big|\sum_{k=1}^n\int e^{itx_k - isx_{1k,r}}\sigma_k(t+s,s)dt\Big|$$
$$\le C\,(A_0^2 + \log n)\,(\frac{d_n}{n})^{1/2} \to 0,$$

as $n \to \infty$ first and then $A_0 \to \infty$. This proves (7.49) and also completes the proof of Proposition 7.4. □

7.4. *Proof of Proposition 7.7.* Recall (6.14) and that $l(x)$ is continuous with $l(0) = 1$. It suffices to show that there exists $M \equiv M_n \to \infty$ so that, as $n \to \infty$ first and then $q \to \infty$,

$$\frac{d_n}{n} \sum_{r=q+1}^{M} \ell\Big(\frac{r}{M}\Big) \sum_{k=1}^{n-r} K\big(x_k\big) K\big(x_{k+r}\big)\, u_{1k}\, u_{1,k+r} = o_P(1), \tag{7.63}$$

where $u_{1j} (= u_{1,j}) = \sum_{i=0}^{m_0} \psi_i \lambda'_{j-i}$ for some $m_0 = m_0(n) \to \infty$ and $m_0 = o\big(\sqrt{n/d_n}\big)$.

To this end, as in (7.45) and (7.46), for $A_0 = m_0 + 1$, we write

$$x_{k+r} - x_k = x_{1k,r} + x_{2k,r},$$

where, by using the notations $a_l = \sum_{s=0}^{l} \phi_s$ with $a_l = 0$ if $l < 0$ and $\gamma_l = a_{r+l} - a_l$,

$$x_{1k,r} = \sum_{j=-\infty}^{-A_0} [a_{r-j} - a_{-j}]\epsilon_{j+k} = \sum_{j=1}^{k-A_0} \gamma_{k-j}\epsilon_j + \sum_{j=-\infty}^{0} \gamma_{k-j}\epsilon_j,$$

$$x_{2k,r} = \sum_{j=-A_0+1}^{0} [a_{r-j} - a_{-j}]\epsilon_{j+k} + \sum_{j=1}^{r} a_{r-j}\epsilon_{j+k}.$$

Recall that $K(x) = \frac{1}{2\pi}\int e^{itx}\hat{K}(t)dt$ under the condition (a). For any $r \ge 0$ and $l_n \ge 0$, we have

$$\begin{aligned}
&\sum_{k=1}^{n-r} K\big(x_k\big) K\big(x_{k+r}\big)\, u_{1k}\, u_{1,k+r} \\
&= \frac{1}{2\pi}\sum_{k=1}^{n-r} K\big(x_k\big)\, u_{1k}\, u_{1,k+r} \int_{|s|\le l_n} \hat{K}(s) e^{-isx_{k+r}} ds + L_{1n} \\
&= L_{1n}(r) + L_{2n}(r) + L_{3n}(r),
\end{aligned} \tag{7.64}$$

where, with $V_k(s) = e^{-isx_{2k,r}}\, u_{1k}\, u_{1,k+r}$,

$$L_{1n}(r) = \frac{1}{2\pi}\sum_{k=1}^{n-r} K\big(x_k\big)\, u_{1k}\, u_{1,k+r} \int_{|s|>l_n} \hat{K}(s) e^{-isx_{k+r}} ds,$$

$$L_{2n}(r) = \frac{1}{2\pi}\sum_{k=1}^{n-r} K\big(x_k\big) \int_{|s|\le l_n} \hat{K}(s)\, e^{-is(x_k + x_{1k,r})}\, \mathbb{E}V_k(s)\, ds,$$

$$L_{3n}(r) = \frac{1}{2\pi}\sum_{k=1}^{n-r} K\big(x_k\big) \int_{|s|\le l_n} \hat{K}(s)\, e^{-is(x_k + x_{1k,r})}\, \big[V_k(s) - \mathbb{E}V_k(s)\big]\, ds.$$

Using Lemma 7.1(i) and $\int |\hat{K}(s)|ds < \infty$, for any $m_0 \to \infty$ satisfying $m_0 = O(n/d_n)$, there exists $M_1 = M_{1n} \to \infty$ so that, whenever $l_n \to \infty$,

$$\begin{aligned}
&\frac{d_n}{n} \sum_{r=q+1}^{M_1} \mathbb{E}|L_{1n}(r)| \\
&\le C \sum_{r=q+1}^{M_1} \frac{d_n}{n} \Big[\sum_{k=1}^{3m_0} \mathbb{E}|u_{1k}\, u_{1,k+r}| + \sum_{3m_0+1}^{n} d_k^{-1}\Big] \int_{|s|>l_n} |\hat{K}(s)| ds
\end{aligned}$$

$$\leq C\, M_1 \int_{|s|>l_n} |\hat{K}(s)|ds \to 0. \tag{7.65}$$

To estimate $L_{2n}(r)$, let $\widetilde{h}_r(y) = \mathbb{E}\big[K(y + x_{20,r})u_{10}u_{1r}\big]$. It is readily seen that $\widetilde{h}_r(y)$ is bounded and integrable. Furthermore, using (7.33) in Lemma 7.7 with minor modifications0, we have

$$\begin{aligned}
\big|\widetilde{h}_r(y)\big| &\leq \sum_{l=0}^{m_0}\sum_{v=0}^{m_0} \big|\mathbb{E}\left[K(y + x_{20,r})\psi_l\lambda'_{-l}\,\psi_v\lambda'_{r-v}\right]\big| \\
&\leq C\sum_{l=0}^{m_0}\sum_{v=0}^{m_0} ||\psi_l||\,||\psi_v|| \begin{cases} d_r^{-1} & \text{if } |r-v+l|\leq 1, \\ \sum_{s=0}^{v}|\phi_s|\,\sum_{s=l}^{r+l}|\phi_s|\,|(d_r^{-3}+r^{-3}), & \text{if } |r-v+l|\geq 2, \end{cases} \\
&\leq C\sum_{l=0}^{m_0}||\psi_l|| \sum_{v=r+l-1}^{r+l+1}||\psi_v||\,d_r^{-1} + C\sum_{l=0}^{m_0}\sum_{v=0}^{m_0} l^{1/2}||\psi_l||\,v^{1/2}||\psi_v||\,d_r^{-3/2} \\
&\leq Cr^{-1}d_r^{-1} + C\,r^{-3/2} \leq Cr^{-3/2},
\end{aligned}$$

uniformly in $y \in R$, where we have used the facts that $d_r^{-1} \leq Cr^{-1/2}$ and $\sum_{s=0}^{v}|\phi_s|\,\sum_{s=l}^{r+l}|\phi_s|(d_r^{-3} + r^{-3}) \leq Cv^{1/2}\,l^{1/2}r^{-3/2}$ under both **SM** and **LM** and $\sum_{v=0}^{\infty} v||\psi_v|| < \infty$. Now, by noting that $\mathbb{E}V_k(s) = \mathbb{E}V_0(s)$, $\sup_s \mathbb{E}|V_0(s)| \leq \mathbb{E}|u_{10}\,u_{1r}| \leq C < \infty$ and

$$\widetilde{h}_r(y) = \frac{1}{2\pi}\int \hat{K}(s)\,e^{-isy}\,\mathbb{E}V_0(s)\,ds,$$

standard calculations, together with the Hölder inquality, show[4] that

$$\begin{aligned}
\frac{d_n}{n}\,\mathbb{E}|L_{2n}(r)| &\leq \frac{d_n}{n}\sum_{k=1}^{n}\mathbb{E}\big[K\big(x_k\big)\,\big|\widetilde{h}_r(x_k + x_{1k,r})\,\big|\big] + C\,\frac{d_n}{n}\sum_{k=1}^{n}\mathbb{E}K\big(x_k\big)\int_{|s|>l_n}|\hat{K}(s)|ds \\
&\leq \Big[\frac{d_n}{n}\sum_{k=1}^{n}\mathbb{E}K^{4/3}\big(x_k\big)\Big]^{3/4}\Big[\frac{d_n}{n}\sum_{k=1}^{n}\mathbb{E}\big|\widetilde{h}_r(x_k + x_{1k,r})\,\big|^4\Big]^{1/4} + C\int_{|s|>l_n}|\hat{K}(s)|ds \\
&\leq C\Big[\int K^{4/3}(y)dy\Big]^{3/4}\Big[\int\big|\widetilde{h}_r(y)\big|^4dy\Big]^{1/4} + C\int_{|s|>l_n}|\hat{K}(s)|ds \\
&\leq Cr^{-9/8} + C\int_{|s|>l_n}|\hat{K}(s)|ds.
\end{aligned}$$

As a consequence, for any $l_n \to \infty$ and $M_1 \to \infty$ as given in (7.65), we have

$$\begin{aligned}
&\frac{d_n}{n}\sum_{r=q+1}^{M_1}\mathbb{E}|L_{2n}(r)| \\
&\leq C\sum_{r=q+1}^{M_1} r^{-9/8} + CM_1\int_{|s|>l_n}|\hat{K}(s)|ds \\
&\leq Cq^{-1/8} + CM_1\int_{|s|>l_n}|\hat{K}(s)|ds \to 0,
\end{aligned} \tag{7.66}$$

[4]Note that $x_k + x_{1k,r} = \sum_{j=-\infty}^{k}\tilde{a}_{k-j}\epsilon_j$ where $\tilde{a}_{k-j} = a_{k-j} + \gamma_{k-j}I(j \leq k - A_0)$ if $j \geq 1$ and $\tilde{a}_{k-j} = a_{k-j} - a_{-j} + \gamma_{k-j}$ if $j \leq 0$, and $\sum_{j=-\infty}^{n}\tilde{a}_j^2 \asymp d_n^2$. Lemma 7.1 still holds when the x_k is replaced by $x_k + x_{1k,r}$.

as $n \to \infty$ first and then $q \to \infty$.

We finally estimate $L_{3n}(r)$. It follows from the Fourier transformation that

$$L_{3n}(r) = \frac{1}{(2\pi)^2} \sum_{k=1}^{n-r} \int \int_{|s| \le l_n} \hat{K}(-t)\hat{K}(s)\, e^{i(t-s)x_k}\, e^{-isx_{1k,r}} \big[V_k(s) - \mathbb{E}V_k(s)\big]\, dsdt$$

$$= \frac{1}{2\pi} \int_{|s| \le l_n} \hat{K}(s)\mathcal{L}_n(s,r)ds, \tag{7.67}$$

where $\mathcal{L}_n(s,r) = \sum_{k=1}^{n-r} \int \hat{K}(-t)\, e^{i(t-s)x_k}\, e^{-isx_{1k,r}} \big[V_k(s) - \mathbb{E}V_k(s)\big]\, dt$. Let $\sigma_k(s) = V_k(s) - \mathbb{E}V_k(s)$. Uniformly for $|s| \le l_n$, we have

$$\mathbb{E}\mathcal{L}_n^2(s,r) = \mathbb{E}\Big|\sum_{k=1}^{n} \int \hat{K}(t+s)\, e^{itx_k - isx_{1k,r}} \sigma_k(s)dt\Big|^2$$

$$= \sum_{k=1}^{n}\sum_{j=1}^{n} \mathbb{E} \int\int \hat{K}(t+s)\, \hat{K}(u+s)\, e^{-is(x_{1k,r}+x_{1j,r})} e^{itx_k + iux_j} \sigma_k(s)\sigma_j(s)dtdu$$

$$= \Big(\sum_{|j-k| \ge A_0^2+1} + \sum_{|j-k| \le A_0^2}\Big)\mathbb{E} \int\int \hat{K}(t+s)\, \hat{K}(u+s)\, e^{-is(x_{1k,r}+x_{1j,r})} e^{itx_k + iux_j} \sigma_k(s)\sigma_j(s)dtdu$$

$$=: R_{n1}(s) + R_{n2}(s), \tag{7.68}$$

Note that $\sigma_k(s)$ depends only on $\epsilon_{k+r}, ..., \epsilon_{k-A_0}$, $\mathbb{E}\sigma_k(s) = 0$ and

$$\sup_s |\sigma_k(s)| \le C\big(1 + |u_{1k}|\, |u_{1,k+r}|\big).$$

As in the proof of (7.50), it follows from (7.20) in Lemma 7.4 that

$$|R_{n1}(s)| \le \sum_{|j-k| \ge A_0^2+1} \Big|\mathbb{E} \int\int e^{-is(x_{1k,r}+x_{1j,r})} e^{itx_k + iux_j} \sigma_k(s)\sigma_j(s)dtdu\Big|$$

$$\le \sum_{|j-k| \ge A_0^2+1} \mathbb{E} \int\int \Big|\mathbb{E}\big[e^{itx_k + iux_j - is\sum_{q=1}^{k\vee j} \gamma'_q \epsilon_q} \sigma_k(s)\sigma_j(s) \mid \mathcal{F}_0\big]\Big| dtdu$$

(where γ'_q is given as in (7.58))

$$\le C \sum_{|j-k| \ge A_0^2+1} d_k^{-1} d_{|j-k|}^{-2} (1 + |s|)$$

$$\le C\,(1 + |s|) \begin{cases} n/d_n, & \text{under } \mathbf{LM}, \\ n\, \log n/d_n, & \text{under } \mathbf{SM}. \end{cases} \tag{7.69}$$

As for $R_{n2}(s)$, by recalling $K(x) = \frac{1}{2\pi}\int \hat{K}(t)e^{itx}dx$ and $A_0 = m_0 + 1$, we have

$$|R_{n2}(s)| \le \sum_{|j-k| \le A_0^2+1} \mathbb{E}\left[K(x_k)\, K(x_j) \sup_s |\sigma_k(s)| \sup_s |\sigma_j(s)|\right]$$

$$\le C_1 \sum_{|j-k| \le A_0^2+1} d_k^{-1} \le C m_0^2 n/d_n. \tag{7.70}$$

It follows from (7.67)-(7.70) that, for any $l_n \to \infty$ satisfying $l_n = o(\sqrt{n/d_n})$ and $m_0 = o(\sqrt{n/d_n})$, there exists $M_2 \equiv M_{2n} \to \infty$,

$$\frac{d_n}{n} \sum_{r=q+1}^{M_2} \mathbb{E}|L_{3n}(r)|$$

$$\leq C\, M_2 \sup_{|s|\leq l_n} \mathbb{E}\big|\mathcal{L}_n(s,r)\big| \int_{|s|\leq l_n} |\hat{K}(s)|ds \leq C M_2 \sup_{|s|\leq l_n} \big[\mathbb{E}\mathcal{L}_n^2(s,r)\big]^{1/2}$$

$$\leq C\, M_2 \big[l_n(1+\log n) + m_0^2\big]^{1/2} \big(\frac{d_n}{n}\big)^{1/2} \to 0. \tag{7.71}$$

By virtue of (7.64), (7.65), (7.66) and (7.71), for any $M \equiv M_n \to \infty$ and $M_n \leq \min\{M_{1n}, M_{2n}\}$, we have

$$\frac{d_n}{n} \sum_{r=q+1}^{M} \ell\Big(\frac{r}{M}\Big)\, \mathbb{E}\Big| \sum_{k=1}^{n-r} K\big(x_k\big) K\big(x_{k+r}\big)\, u_{1k}\, u_{1,k+r}\Big|$$

$$\leq C\, \frac{d_n}{n} \sum_{r=q+1}^{M_1} \big(\mathbb{E}|L_{1n}(r)| + \mathbb{E}|L_{2n}(r)|\big) + \frac{C\, d_n}{n} \sum_{r=q+1}^{M_2} \mathbb{E}|L_{3n}(r)| \to 0,$$

as $n \to \infty$ first and then $q \to \infty$. This proves (7.63) and completes the proof of Proposition 7.7. □

APPENDIX: PROOFS OF LEMMAS

A.1. Proof of Lemma 7.2. The idea of the proof is similar to that of Wang and Phillips (2011, pages 246-247)[26] and the following outline is provided here for completeness. We first prove (7.13). Write $\Omega_1 \equiv \Omega_1(t)$ (Ω_2, respectively) for the set of $1 \le v \le k$ such that $|t\,\alpha_v + \beta_v| \ge 1$ ($|t\,\alpha_v + \beta_v| < 1$, respectively), and

$$B_2 = \sum_{v\in\Omega_2} \alpha_v\beta_v \quad \text{and} \quad B_3 = \sum_{v\in\Omega_2} \beta_v^2.$$

Since $B_2^2 \le B_{1k}\, B_3$ by Hölder's inequality, we have

$$\begin{aligned}\sum_{q\in\Omega_2} (t\,\alpha_v + \beta_v)^2 &= t^2\, B_{1k} + 2tB_2 + B_3 \\ &= B_{1k}(t + B_2/B_{1k})^2 + (B_3 - B_2^2/B_{1k}) \\ &\ge B_{1k}(t + B_2/B_{1k})^2.\end{aligned}$$

On the other hand, there exist constants $\gamma_1 > 0$ and $\gamma_2 > 0$ such that

$$\big|\mathbb{E}\, e^{i\,\epsilon_1\, l}\big| \le \begin{cases} e^{-\gamma_1} & \text{if } |l| \ge 1, \\ e^{-\gamma_2 l^2} & \text{if } |l| \le 1, \end{cases} \tag{A.1}$$

since $\mathbb{E}\epsilon_1 = 0$, $\mathbb{E}\epsilon_1^2 < \infty$ and ϵ_1 satisfies the Cramér's condition due to $\int |\mathbb{E}\, e^{it\epsilon_0}|dt < \infty$. See, e.g., Chapter 1 of Petrov (1995)[20]. Without loss of generality, assume $\alpha_1 \ne 0$ and let $g(t) = \mathbb{E}\, e^{it\alpha_1\epsilon_0}$. From these facts and the independence of ϵ_t it follows that, for k sufficiently large and all t,

$$\begin{aligned}\big|\mathbb{E}\, e^{iz_k(t)}\big| &\le |g(t)| \prod_{q=2}^{k} |\mathbb{E}\, e^{i\epsilon_1(t\alpha_q+\beta_q)}| \\ &\le |g(t)| \exp\big\{ -\gamma_1\#(\Omega_1) - \gamma_2 \sum_{v\in\Omega_2} (t\,\alpha_v + \beta_v)^2\big\} \\ &\le |g(t)| \exp\big\{ -\gamma_1\#(\Omega_1) - \gamma_2\, B_{1k}\,(t + B_2/B_{1k})^2\big\}. \end{aligned}\tag{A.2}$$

Hence, by recalling (7.12) and $|B_2| \le \sum_{v=1}^k |\alpha_v\beta_v|$, simple calculations show that

$$\begin{aligned}&\int \min\{1, s_1\,|t|^\delta + s_2\}\,\big|\mathbb{E}\, e^{iz_k(t)}\big|dt \\ &\le \int_{\#(\Omega_1)\ge\sqrt{k}} |g(t)|\, e^{-\sqrt{k}}dt + C\int_{\#(\Omega_1)\le\sqrt{k}} (s_1\,|t|^\delta + s_2)e^{-\gamma_2\, B_{1k}\,(t+B_2/B_{1k})^2}dt \\ &\le Ce^{-\sqrt{k}} + C\, s_1 \int \big(|t| + |B_2|/B_{1k}\big)^\delta e^{-\gamma_2\, B_{1k}\, t^2} I(B_{1k} \ge m_k^2)dt \\ &\qquad + C\, s_2 \int e^{-\gamma_2\, B_{1k}\, t^2} I(B_{1k} \ge m_k^2)dt \\ &\le C\,\big(k^{-3} + s_1\big[m_k^{-1-\delta} + m_k^{-1-2\delta}\big(\sum_{v=1}^k |\alpha_v\,\beta_v|\big)^\delta\big] + s_2\, m_k^{-1}\big).\end{aligned}$$

Result (7.13) now follows from the fact that

$$\sum_{v=1}^k |\alpha_v\,\beta_v| \le \big(\sum_{v=1}^k |\alpha_v|^2\big)^{1/2}\big(\sum_{v=1}^k |\beta_v|^2\big)^{1/2} \le Cm_k\,\big(\sum_{v=1}^k |\beta_v|^2\big)^{1/2}.$$

The proof of (7.14) is similar and hence the details are omitted. We finally prove (7.15). In fact, by recalling $B_2^2/B_{1k} \le B_3 \le a$, i.e, $B_2/B_{1k} \le a^{1/2}/m_k$ due to (7.12), it follows from (A.2) that

$$\begin{aligned}
&\int_{|t|\ge B/m_k} \big|\mathbb{E}\, e^{iz_k(t)}\big| dt \\
&\le \int_{\#(\Omega_1)\ge\sqrt{k}} |g(t)|\, e^{-\sqrt{k}} dt + C \int_{\#(\Omega_1)\le\sqrt{k}, |t|\ge B/m_k} e^{-\gamma_2 B_{1k}(t+B_2/B_{1k})^2} dt \\
&\le Ck^{-3} + \int_{|t|\ge 2^{-1}B/m_k} e^{-\gamma_2 B_{1k} t^2} I(B_{1k}\ge m_k^2) dt \\
&\le C(k^{-3} + m_k^{-1} B^{-1}),
\end{aligned}$$

as required. □

A.2. Proof of Lemma 7.3. Let $V_k(t) = \sum_{v=k-m_0+1}^{k}(t\alpha_v + \beta_v)\epsilon_v$. Note that

$$\begin{aligned}
\big|\mathbb{E}\, e^{iz_k(t/h)} g(t, w_k)\big| &\le \big|\mathbb{E}\, e^{iz_{k-m_0}(t/h)}\big|\, \big|\mathbb{E}\, e^{iV_k(t/h)} g(t, w_k)\big| \\
&\le \mathbb{E}\, |g(t, w_1)|\, \big|\mathbb{E}\, e^{iz_{k-m_0}(t/h)}\big|.
\end{aligned}$$

It follows from (7.13) with $s_1 = 0$ and $s_2 = 1$ that

$$\int \big|\mathbb{E}\, e^{iz_k(t/h)}\, g(t, w_k)\big| dt \le Ch \int \big|\mathbb{E}\, e^{iz_{k-m_0}(t)}\big| dt \le Ch\,\big(k^{-3} + \tau_k^{-1}\big),$$

yielding (7.16). Similarly, by noting that

$$\begin{aligned}
\big|\mathbb{E}\, e^{iV_k(t/h)} g(t, w_k)\big| &\le \big|\mathbb{E}\,(e^{iV_k(t/h)} - 1) g(t, w_k)\big| + \big|\mathbb{E}\, g(t, w_k)\big| \\
&\le 2\min\{1, \alpha_{k0}\,|t|/h + \beta_{k0}\}\mathbb{E}\left\{|\epsilon_0||g(t, w_1)|\right\} + C\min\{1, |t|\} \\
&\le C\,\beta_{k0} + C\min\{1, \alpha_{k0}\,|t|/h\} + C\min\{1, |t|\}, \qquad \text{(A.3)}
\end{aligned}$$

we have

$$\begin{aligned}
&\int \big|\mathbb{E}\left\{e^{iz_k(t/h)} g(t, w_k)\right\}\big| dt \\
&\le C \int \min\{1, \alpha_{k0}\,|t|/h\}\,\big|\mathbb{E}e^{iz_{k-m_0}(t/h)}\big| dt + C\beta_{k0}\int \big|\mathbb{E}e^{iz_{k-m_0}(t/h)}\big| dt \\
&\qquad + C\int \min\{1, |t|\}\,\big|\mathbb{E}e^{iz_{k-m_0}(t/h)}\big| dt \\
&\le Ch\left\{(1+\alpha_{k0})\tau_k^{-2}\left[1 + \Big(\sum_{v=1}^{k}\beta_v^2\Big)^{1/2}\right] + \beta_{k0}\,\tau_k^{-1}\right\},
\end{aligned}$$

as required in (7.17). As for (7.18), by noting that

$$\big|\mathbb{E}\, e^{iV_k(t/h)} g(t, w_k)\big| \le C\,\beta_{k0} + C\min\{1, |t|\} + C\min\{1, |t|/h\},$$

due to (A.3) and $\sup_k \alpha_{k0} = O(1)$, it follows from (7.13) and (7.14) that

$$\begin{aligned}
&\int \min\{1, |t|/h\}\,\big|\mathbb{E}\left\{e^{iz_k(t/h)} g(t, w_k)\right\}\big| dt \\
&\le C\beta_{k0}\int \min\{1, |t|/h\}\big|\mathbb{E}e^{iz_{k-m_0}(t/h)}\big| dt + C\int \min\{1, (|t|/h)^2\}\big|\mathbb{E}e^{iz_{k-m_0}(t/h)}\big| dt
\end{aligned}$$

$$+C\int \min\{1,|t|\}\,\min\{1,|t|/h\}\big|\,\mathbb{E}e^{iz_{k-m_0}(t/h)}\,\big|dt$$

$$\le C\,h\,\{k^{-3}+\big[\beta_{k0}(\tau_k^{-2}+k^{-3})+\tau_k^{-3}\big]\big(1+\sum_{v=1}^{k}\beta_v^2\big)\}.$$

This proves (7.18). □

A.3. Proof of Lemma 7.4. We only prove (7.20) and (7.22). The other proofs are similar and simpler. Note that

$$x_k=\sum_{j=1}^{k}\rho_n^{k-j}\,\xi_j=\sum_{j=1}^{k}\rho_n^{k-j}\,\big(\sum_{u=1}^{j}+\sum_{u=-\infty}^{0}\big)\epsilon_u\phi_{j-u}$$

$$=\sum_{u=1}^{k}\epsilon_u\,a_{k-u}+\sum_{u=0}^{\infty}\epsilon_{-u}b_{u,k}, \tag{A.4}$$

where $a_{k-u}=\sum_{s=0}^{k-u}\rho_n^{k-u-s}\phi_s$ and $b_{u,k}=\sum_{s=1}^{k}\rho_n^{k-s}\,\phi_{s+u}$. It follows from the independence of the ϵ_j that

$$|\mathrm{I}_{k,l}(m)|$$

$$\le\int\int\Big|\mathbb{E}\Big\{e^{is\sum_{v=m+1}^{k}a_{k-v}\epsilon_v/h+it\sum_{v=m+1}^{l}a_{l-v}\epsilon_v/h+i\sum_{j=m+1}^{l}\gamma_j\epsilon_j}g(s,w_k)\,g(t,w_l)\Big\}\Big|dsdt$$

$$\le C\int\Big|\mathbb{E}\Big\{e^{i\sum_{v=k+1}^{l}(ta_{l-v}/h+\gamma_v)\epsilon_v}\,g(t,w_l)\Big\}\Big|\,\Lambda(t,k)dt, \tag{A.5}$$

where

$$\Lambda(t,k)=\int\Big|\mathbb{E}\Big\{e^{i\sum_{v=m+1}^{k}(sa_{k-v}/h+ta_{l-v}/h+\gamma_v)\epsilon_v}\,g(s,w_k)\Big\}\Big|ds.$$

As in Lemma 7.2, denote by Ω_1 a subset of $\Omega=\{m+1,2,...,k\}$ and $\Omega_2=\Omega-\Omega_1$. Note that, for any $k-m\ge 1$, $\sum_{v\in\Omega_2}a_{k-v}^2\asymp d_{k-m}^2$ whenever $\#\Omega_1\le\sqrt{k-m}$. It is readily seen from (7.16) with $\alpha_v=a_{k-v}$ and $\beta_v=ta_{l-v}/h+\gamma_v$ that

$$\Lambda(t,k)\le Chd_{k-m}^{-1}, \tag{A.6}$$

By similar arguments it follows from (7.17) with $\alpha_v=a_{l-v}$ and $\beta_v=\gamma_v$ that

$$\int\Big|\mathbb{E}\Big\{e^{i\sum_{v=k+1}^{l}(ta_{l-v}/h+\gamma_v)\epsilon_v}\,g(t,w_l)\Big\}\Big|\,dt$$

$$\le Ch\big\{(l-k)^{-3}+\alpha_{l0}\,d_{l-k}^{-2}\,\big[1+\big(\sum_{v=k+1}^{l}\gamma_v^2\big)^{1/2}\big]+\beta_{l0}\,d_{l-k}^{-1}\big\}$$

$$\le Ch\big[d_{l-k}^{-2}(1+a^{1/2})+\beta_{l0}d_{l-k}^{-1}\big], \tag{A.7}$$

where $a=\sum_{v=1}^{l}\gamma_v^2$, $\beta_{l0}=\max_{0\le j\le m_0}|\gamma_{l-j}|$ and we have used the fact:

$$\alpha_{l0}=\max_{0\le i\le m_0}|\alpha_{l-i}|=\max_{0\le i\le m_0}|a_i|=O(1).$$

It follows from (A.5)-(A.7) that

$$|\mathrm{I}_{k,l}(m)| \le Ch\, d_{k-m}^{-1} \int \big|\mathbb{E}\, e^{i\sum_{v=k+1}^{l}(ta_{l-v}/h+\gamma_v)\epsilon_v}\, g(t,w_l)\big|\, dt$$
$$\le C\, h^2\, d_{k-m}^{-1}\, \big[d_{l-k}^{-2}(1+a^{1/2}) + \beta_{l0} d_{l-k}^{-1}\big],$$

implying (7.20).

The proof of (7.22) requires some modifications. First notice that, under **SM**, we have

$$\Lambda(t,k) \le Ch\big[(k-m)^{-1} + \min\{1, |t|/h\}\,(k-m)^{-1/2}\big], \tag{A.8}$$

rather than (A.6). Indeed, under **SM**, it follows that

(a) $\Lambda(t,k) \le Ch(k-m)^{-1/2}$ by (7.16) and, for any $t \in R$,
(b) $\Lambda(t,k) \le Ch\big[(k-m)^{-1} + |t|/h\,(k-m)^{-1/2}\big]$ by (7.17) with $\alpha_v = a_{k-v}$ and $\beta_v = ta_{l-v}/h + \mu_v/\sqrt{n}$,

implying (A.8). Now, by using (A.5) first and then (7.17) and (7.18), we have

$$|\mathrm{I}_{k,l}(m)|$$
$$\le Ch(k-m)^{-1}\int \big|\mathbb{E}\,\big\{e^{i\sum_{v=k+1}^{l}(ta_{l-v}/h+\gamma_v)\epsilon_v}\, g(t,w_l)\big\}\big|dt$$
$$+Ch(k-m)^{-1/2}\int \min\{1,|t|/h\}\big|\mathbb{E}\,\big\{e^{i\sum_{v=k+1}^{l}(ta_{l-v}/h+\gamma_v)\epsilon_v}\, g(t,w_l)\big\}\big|dt$$
$$\le Ch^2\big[(l-k)^{-1}(k-m)^{-1} + (l-k)^{-3/2}(k-m)^{-1/2}\big],$$

which yields (7.22). □

A.4. Proof of Lemma 7.5. We only prove (7.25). The other proofs are similar and use the corresponding results in Lemma 7.4. Recalling (2.4), we may write

$$II_{lk}(h) = \frac{1}{(2\pi)^2}\int\int \hat f(t,w_k)\hat f(s,w_l) e^{itx_k/h + isx_l/h} e^{i\sum_{j=m+1}^{n}\mu_j\epsilon_j/\sqrt{n}} dtds. \tag{A.9}$$

It follows from (A.4), the independence of ϵ_j and (7.20) with $\gamma_j = \mu_j/\sqrt{n}$ and $g(s,w_k) = \hat f(s,w_k)$ that

$$\big|\mathbb{E}\,\big[II_{lk}(h) \mid \mathcal{F}_m\big]\big|$$
$$\le \frac{1}{(2\pi)^2}\int\int \mathbb{E}\,\big(e^{isx_k/h+itx_l/h++i\sum_{j=m+1}^{l}\mu_j\epsilon_j/\sqrt{n}}\hat f(s,w_k)\hat f(t,w_l) \mid \mathcal{F}_m\big)dsdt$$
$$\le C\, h^2\, d_{k-m}^{-1}\,\big(d_{l-k}^{-2} + d_{l-k}/\sqrt{n}\big),$$

as required. □

A.5. Proof of Lemma 7.6. Recalling (A.4), as in (A.9) we have

$$\big|\mathbb{E}\,\Big\{f(x_l/h,w_l)\mathbb{E}\,\big[f(x_k/h,w_k) \mid \mathcal{F}_{k-m}\big]\Big\}\big|$$
$$= \int\int \big|\mathbb{E}\,\Big\{e^{itx_l/h}\hat f(-t,w_l)\,\mathbb{E}\big[e^{isx_k/h}\hat f(s,w_k)|\mathcal{F}_{k-m}\big]\Big\}\big|dsdt$$
$$\le \int\int \big|\mathbb{E}\,\Big\{e^{ith^{-1}\sum_{v=k}^{l}a_{l-v}\epsilon_v}\,\hat f(-t,w_l)\Big\}\big|$$
$$\mathbb{E}\big[e^{(ish^{-1}\sum_{v=1}^{k}a_{k-v}\epsilon_v + ith^{-1}\sum_{v=1}^{k-m}a_{l-v}\epsilon_v)}\hat f(s,w_k)\big]dsdt$$
$$\le C\int \big|\mathbb{E}\,\Big\{e^{ith^{-1}\sum_{v=k}^{l}a_{l-v}\epsilon_v}\,\hat f(-t,w_l)\Big\}\big|\,\Lambda(t,k)dt,$$

where, by letting $a^*_{l-v} = 0$ if $k-m+1 \le v \le k$ and $a^*_{l-v} = a_{l-v}$ if $1 \le v \le k-m$, we have

$$\Lambda(t,k) = \int \left|\mathbb{E}\left\{e^{i\sum_{v=1}^{k}(sa_{k-v}/h+ta^*_{l-v}/h)\epsilon_v}\,\hat{f}(s,w_k)\right\}\right|ds.$$

The remainder of the proof is the same as that of Lemma 7.4 and is omitted. □

A.6. Proof of Lemma 7.7. Take A_0 as required in Lemma 7.1. Recalling $K(x)$ is bounded, (7.30) follows immediately from Lemma 7.1 (i). If $k \ge A_0$, $l-k \ge A_0$ and $l-j \le k$, it follows from Lemma 7.1 (ii) and the conditional arguments that

$$\begin{aligned} I &:= \left|\mathbb{E}\left\{\Gamma(\lambda_{k-j})\,\Gamma(\lambda_{l-j})\,K(x_k/h)\,K(x_l/h)\right\}\right| \\ &\le \mathbb{E}\left\{\left|\Gamma(\lambda_{k-j})\,\Gamma(\lambda_{l-j})\,K(x_k/h)\right|\left|\mathbb{E}\left[K(x_l/h)|\mathcal{F}_{||}\right]\right|\right\} \\ &\le C\,\mathbb{E}\Gamma^2(\lambda_1)\,h^2\,d_k^{-1}\,d_{l-k}^{-1}, \end{aligned}$$

indicating (7.31).

We next assume that $k \ge A_0$, $l-k \ge A_0$ and $l-j > k$. Recalling (A.4), as in (A.9), we have

$$\begin{aligned} I &= \int\int \left|\mathbb{E}\left\{e^{itx_l/h}\,e^{isx_k/h}\Gamma(\lambda_{k-j})\,\Gamma(\lambda_{l-j})\right\}\right||\hat{K}(-s))|\,|\hat{K}(-t)|dsdt \\ &\le C\int \left|\mathbb{E}\left\{e^{ith^{-1}\sum_{v=k}^{l}a_{l-v}\epsilon_v}\,\Gamma(\lambda_{l-j})\right\}\right|\Lambda(t,k)dt \end{aligned}$$

where

$$\Lambda(t,k) = \begin{cases} \int\left|\mathbb{E}\left\{e^{i\sum_{v=1}^{k}(sa_{k-v}/h+ta_{l-v}/h)\epsilon_v}\,e^{-i(s\epsilon_{k-j}b_{j-k,k}/h+t\epsilon_{l-k}b_{l-k,k}/h)}\,\Gamma(\lambda_{k-j})\right\}\right|ds, & \\ & \text{if } k-j \le 0, \\ \int\left|\mathbb{E}\left\{e^{i\sum_{v=1}^{k}(sa_{k-v}/h+ta_{l-v}/h)\epsilon_v}\,\Gamma(\lambda_{k-j})\right\}\right|ds, & \text{if } k-j \ge 1. \end{cases}$$

It follows from arguments similar to those given in the proof of Lemma 7.4 with some minor modifications[5] that:

(a) under **LM**, $\Lambda(t,k) \le C\,h\,d_k^{-1}$ and

$$\begin{aligned} I &\le C\,h\,d_k^{-1}\int\left|\mathbb{E}\left\{e^{ith^{-1}\sum_{v=k}^{l}a_{l-v}\epsilon_v}\,\Gamma(\lambda_{l-j})\right\}\right|dt \\ &\le C\sum_{s=0}^{j}|\phi_s|\,h^2\,d_k^{-1}d_{l-k}^{-2}; \end{aligned}$$

(b) under **SM** (noting $|b_{j-m,m}| \le \sum_{i=j-m}^{j}|\phi_i| \le C < \infty$ for any $m \ge 0$ and $\max_{1\le v\le k}|a_v| \le C < \infty$),

$$\begin{aligned} \Lambda(t,k) &\le \int \left|\mathbb{E}\left\{e^{i\sum_{v=1,v\ne k-j}^{k}(sa_{k-v}/h+ta_{l-v}/h)\epsilon_v}\right\}\right|\left(\min\{1,|s|/h\}+\min\{1,|t|/h\}\right)ds \\ &\le Ch\left(k^{-1}+\min\{1,|t|/h\}k^{-1/2}\right) \end{aligned}$$

[5]Replace m_0 by j, set $\gamma_v = 0$ and take $m=0$. In this case, α_{l0} used in (A.7) satisfies

$$\alpha_{l0} = \max_{0\le i\le j}|\alpha_{l-i}| = \max_{0\le i\le j}|a_i| \le \sum_{s=0}^{j}|\phi_s|,$$

which can not be eliminated.

and

$$I \le Ch\,k^{-1} \int \left|\mathbb{E}\left\{e^{ith^{-1}\sum_{v=k}^{l} a_{l-v}\epsilon_v}\,\Gamma(\lambda_{l-j})\right\}\right| dt$$
$$+Ch\,k^{-1/2} \int \min\{1,|t|/h\}\left|\mathbb{E}\left\{e^{ith^{-1}\sum_{v=k}^{l} a_{l-v}\epsilon_v}\,\Gamma(\lambda_{l-j})\right\}\right| dt$$
$$\le C\,h^2\,k^{-1}(l-k)^{-1} + Ch^2\,k^{-1/2}\,(l-k)^{-3/2}.$$

This proves (7.32).

Similarly, by letting $z_{2r} = \sum_{k=1,k\neq r-j}^{r} \epsilon_k a_{r-k}$, we have

$$\left|\mathbb{E}\left\{\Gamma(\lambda_{r-j})\Gamma(\lambda_{-k})e^{isx_r/h}\right\}\right|$$
$$\le C\left|\mathbb{E}\,e^{isz_{2r}/h}\right| \begin{cases} 1, & \text{if } |r-j+k| \le A_0, \\ |a_j|\,|a_{r+k}-a_k|\min\{1,|s|^2\}, & \text{if } |r-j+k| > A_0, \end{cases}$$

implying that, uniformly for $y \in R$,

$$\left|\mathbb{E}\left\{K(y+x_l/h)\Gamma(\lambda_{l-j})\,\Gamma(\lambda_{-k})\right\}\right|$$
$$\le \int |\hat{K}(s))|\left|\mathbb{E}\left\{e^{isx_l/h}\,\Gamma(\lambda_{l-j})\Gamma(\lambda_{-k})\right\}\right| ds$$
$$\le C\,h \begin{cases} d_l^{-1} & \text{if } |l-j+k| \le A_0, \\ \sum_{s=0}^{j}|\phi_s|\,\sum_{s=k}^{l+k}|\phi_s|\,|(d_l^{-3}+l^{-3}), & \text{if } |l-j+k| > A_0, \end{cases}$$

as required in (7.33). □.

Acknowledgements

Phillips acknowledges research support from the NSF (Grant No. SES 18-50860) and a Kelly Fellowship at the University of Auckland.

REFERENCES

[1] Akonom, J. (1993). Comportement asymptotique du temps d'occupation du processus des sommes partielles *Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques*, **29**, 57-81.

[2] Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley.

[3] Billingsley, P. (1974). Conditional distributions and tightness. *Annals of Probability,* **2**, 480-485.

[4] Borodin, A. N. and Ibragimov, I. A. (1995). *Limit theorems for functionals of random walks*. Proceedings of the Steklov Institute of Mathematics, American Mathematical Society

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

[6] Dong, C. and Linton, O. (2018). Additive nonparametric models with time variables and both stationary and nonstationary regressors. *Journal of Econometrics*, **207**, 212-236.

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

[8] Duffy, J. A. (2020). Asymptotic theory for kernel estimators under moderate deviations from a unit room, with an application to the asymptotic size of nonparametric tests. *Econometric Theory,* **36,** 559-582.

[9] Duffy, J. A. and Kasparis, I. (2021). Estimation and inference in the presence of fractional $d = 1/2$ and weakly nonstationary processes. *Annals of Statistics*, **49**, 1195-1217.

[10] Fiorio C. V., Hajivassiliou, V. A., Phillips, P. C. B. (2010). Bimodal t-ratios: the impact of thick tails on inference. *Econometrics Journal,* **13**, 271-289.

[11] Gao, J. and Phillips, P.C.B. (2013). Semiparametric estimation in triangular system equations with nonstationarity, *Journal of Econometrics,* **176**, 59–79.

[12] Geman, D. and Horowitz, J. (1980). Occupation densities. *Ann. Probab.* **8**, 1-67.

[13] Hu, Z., Phillips P. C. B. and Wang Q. (2021). Nonlinear cointegrating power function regression with endogeneity. *Econometric Theory*, **37**, 1173-1213.

[14] 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.

[15] Jeganathan, P. (2008). Limit theorems for functional sums that converge to fractional Brownian and stable motions. Cowles Foundation Discussion Paper No. 1649, Cowles Foundation for Research in Economics, Yale University.

[16] Logan, B., Mallows, C., Rice S., and Shepp L. (1972). Limit distributions of self-normalized sums. *Annals of Probability* **1**, 788–809.

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

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

[19] Peng, J. and Wang, Q. (2018). Weak convergence to stochastic integrals under primitive conditions in nonlinear econometric models. *Econometric Theory*, **34**, 1132-1157.

[20] Petrov, V. V. (1995). *Limit Theorems of Probability Theory: Sequences of Independent Random Variables.* Oxford Studies in Probability.

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

[22] Wang, Q. (2015). *Limit Theorems for Nonlinear Cointegrating Regression.* World Scientific.

[23] Wang, Q., Lin, Y. X., Gulati, C. M., (2003). Strong approximation for long memory processes with applications. *Journal of Theoretical Probability*, **16**, 377–389.

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

[25] Wang, Q. and Phillips, P. C. B., (2009b). Structural nonparametric cointegrating regression. *Econometrica*, **77**, 1901-1948.

[26] Wang, Q. and Phillips, P. C. B. (2011). Asymptotic Theory for Zero Energy Functionals with Nonparametric Regression Applications. *Econometric Theory*, **27**, 235-259.

[27] Wang, Q. and Phillips, P. C. B. (2016). Nonparametric cointegrating regression with endogeneity and long memory. *Econometric Theory,* **32**, 359-401.

[28] Wang, Q., Phillips, P. C. B. and Kasparis, I. (2021). Latent variable nonparametric cointegrating regression. *Econometric Theory,* **37**, 130-168.