

ESTIMATION AND INFERENCE IN THE PRESENCE OF FRACTIONAL $d = 1/2$ AND WEAKLY NONSTATIONARY PROCESSES

BY JAMES A. DUFFY

AND

IOANNIS KASPARIS

We provide new limit theory for functionals of a general class of processes lying at the boundary between stationarity and nonstationarity – what we term weakly nonstationary processes (WNPs). This includes, as leading examples, fractional processes with $d = 1/2$, and arrays of autoregressive processes with roots drifting slowly towards unity. We first apply the theory to study inference in parametric and nonparametric regression models involving WNPs as covariates. We then use these results to develop a new specification test for parametric regression models. By construction, our specification test statistic has a χ^2 limiting distribution regardless of the form and extent of persistence of the regressor, implying that a practitioner can validly perform the test using a fixed critical value, while remaining agnostic about the mechanism generating the regressor. Simulation exercises confirm that the test controls size across a wide range of data generating processes, and outperforms a comparable test due to Wang and Phillips (2012, Ann. Stat.) against many alternatives.

1. Introduction. Inference in regression models when data is temporally dependent is a challenging problem, which has engendered a voluminous literature. Previous work has investigated the asymptotics of parametric and nonparametric regression estimators under a variety of assumptions on the form and extent of that dependence, including, for example: regressors generated by autoregressive fractionally integrated moving average (ARFIMA) models, general linear and nonlinear processes, and by partial sums and arrays formed from such processes. Both stationary and nonstationary processes have been considered, and quite distinct arguments – relying on stationary laws of large numbers (LLNs) and central limit theorems (CLTs) for the former, and the weak convergence of stochastic processes for the latter – have been utilised to handle these cases.

It might therefore seem as though little work remains to be done on this problem. However, in its treatment of nonstationary processes, previ-

MSC 2010 subject classifications: Primary, 62M10, 62F12, 62G08

Keywords and phrases: Fractional Process, Half Unit Root, Kernel Regression, Specification Testing, Mildly Integrated Process, Weakly Nonstationary Process

ous work has typically employed assumptions that prevent these from being wholly contiguous with their stationary counterparts, leaving some significant gaps in the domain of the existing theory. Thus, for example, when regressors are generated by arrays of ARIMA processes, both the cases of fixed stationary autoregressive roots and of an autoregressive root drifting towards unity at rate n^{-1} (a ‘nearly integrated’ process, henceforth ‘NI’; see Chan and Wei, 1987, 1988; Phillips, 1987a,b) have been closely studied, but the intermediate case of roots drifting towards unity at a strictly slower rate than n^{-1} (a ‘mildly integrated’ process, henceforth ‘MI’; see Giraitis and Phillips, 2006; Magdalinos and Phillips, 2007) has received less attention until very recently. Similarly, the asymptotics of regression estimators when applied to fractionally integrated regressors of order d (henceforth, $I(d)$) are well understood both when $d \in (-1/2, 1/2)$ or $d \in (1/2, 3/2)$, but not nearly so well when $d = 1/2$. In view of the importance of ARFIMA models to the statistical modelling of time series, it is striking that the limit theory for these processes still remains to be fully characterised.¹

Existing work on the asymptotics of regression estimators may thus be divided into two literatures: that dealing with stationary processes, and that with strongly dependent nonstationary processes, with a certain space left in between them. Though these literatures are too vast to be cited exhaustively here, some particularly notable contributions include the following. For stationary long memory processes, the asymptotics of nonparametric regression estimators were developed by Wu and Mielniczuk (2002) and Wu, Huang and Huang (2010). For nonstationary processes, parametric regression estimators have been studied by Chan and Wei (1987, 1988), Phillips (1987a,b), Phillips (1995), Park and Phillips (1999, 2001) and Chan and Wang (2015). Robinson and Hualde (2003), Christensen and Nielsen (2006), Hualde and Robinson (2010) and Johansen and Nielsen (2012a) consider fractional systems. Several papers consider the problem of inference in regressions with a NI covariate, including Mikusheva (2007), Phillips and Magdalinos (2009) and Kostakis, Magdalinos and Stamatogiannis (2015). Wang and Phillips (2009a,b, 2011, 2012) consider nonparametric methods for estimation and inference in regressions with a NI or nonstationary fractional covariate; some closely related work on nonparametric estimation in the setting of null recurrent Markov chains is the subject of the papers by Karlsen and Tjøstheim (2001) and Karlsen, Myklebust and Tjøstheim (2007).

The present work aims to fill the gap between these two literatures, by

¹Although recent work by Shimotsu and Phillips (2005) and Hualde and Robinson (2011) allow for cases where $d = 1/2$, the limit theory developed by these authors is specific to functionals in the frequency domain that arise in the context of memory estimation.

developing the asymptotics of regression estimators for a class of processes intermediate between the stationary and more strongly dependent nonstationary processes previously considered. We term these *weakly nonstationary processes* (WNPs), with leading examples being MI and $I(1/2)$ processes. To appreciate the significance of our results, consider the regression model

$$(1) \quad y_t = m(x_{t-1}) + u_t$$

where u_t is a martingale difference sequence; suppose for concreteness that x_t is $I(d)$ for an unknown $d \in (-1/2, 3/2)$. The object of interest is the regression function $m(\cdot)$; e.g. we would like to test such a null hypotheses as $\mathcal{H}_0 : m(x) = m_0(x)$ for a given x and m_0 . Since \mathcal{H}_0 places no restriction on the process followed by x_t , establishing the asymptotic validity of any test of \mathcal{H}_0 requires that its asymptotics be developed under all possible values of the nuisance parameter d (and indeed, under appropriate drifting sequences $\{d_n\}$: see e.g. Mikusheva, 2007; Andrews, Cheng and Guggenberger, 2020).² This requires results for the case where $d = 1/2$, no less than for $d \in (-1/2, 1/2)$ and $d \in (1/2, 3/2)$. Empirically, values of d in the vicinity of $1/2$ have been systematically found to provide a good description of the dynamics of inflation and realised volatility series (see e.g. Hassler & Wolter, 1995; Baillie, Chung and Tieslau, 1996; and Andersen, Bollerslev, Diebold and Labys, 2001). Indeed, recent work by Hassler and Pohle (2019) finds that when forecasting such series, an ARFIMA($p, d, 0$) model with d fixed at $1/2$ gives superior forecasts to those produced by the same model with an estimated value of d . As such, our results should be particularly relevant for inference in regressions involving such series as r.h.s. variables.

In the context of (1), we show that nonparametric kernel estimators of m are asymptotically (mixed) Gaussian when x_t is a WNP; consequently, the t statistic for testing \mathcal{H}_0 is asymptotically standard normal. This accords with previous results for both stationary and strongly dependent nonstationary processes (e.g. NI processes or $I(d)$ process with $d > 1/2$), which establish the asymptotic normality of the t statistic in these cases (see Wu and Mielniczuk, 2002; and Wang and Phillips, 2009a,b, 2011). It follows that conventional tests of \mathcal{H}_0 , involving the comparison of the t statistic to normal critical values, are asymptotically valid even when the regressor has an unknown, but possibly high, degree of persistence. This is of particular importance for practitioners, since it implies that tests of \mathcal{H}_0 can be conducted without having to in any way adjust for the persistence of x_t .

²Formally, the asymptotic size of a test $\phi_n \in \{0, 1\}$ would be defined as $\limsup_{n \rightarrow \infty} \sup_{d \in (-1/2, 3/2), m \in \mathcal{M}} \mathbb{P}_{d,m} \{\phi_n = 1\}$, for \mathcal{M} a class of functions respecting \mathcal{H}_0 .

We also consider the case where m is parametrised as $m(x) = \mu + \gamma g(x)$ for a known function g . When x_t is a WNP, least squares estimators of (μ, γ) are shown to exhibit the elevated rates of convergence familiar from when regressors are more strongly dependent, but with limiting distributions that are (mixed) Gaussian, similarly to the case of stationary regressors. This result is less directly useful to practitioners, since it breaks down when x_t is more strongly dependent, in which case the limiting distributions of these estimators are well known to be nonstandard (see e.g. Phillips, 1995; Marinucci and Robinson, 1998; Robinson and Hualde, 2003).

We build on these results to develop a test for parametric specifications of m the form $m(x) = \mu + \gamma g(x)$. The test is based on a comparison of the fit provided by parametric and nonparametric estimates of m , and is designed so as to inherit the asymptotic (mixed) normality of the nonparametric estimator. It therefore has the attractive property that the limiting distribution of the test statistic is invariant to the persistence of x_t , implying that a practitioner can perform the test using a fixed set of critical values, while remaining agnostic about the dependence properties of the regressor. Simulations confirm that the size of the test is successfully controlled in finite samples, as the process generating the regressor varies between weakly dependent and stationary, weakly nonstationary, and strongly dependent nonstationary. Relative to the specification test proposed by Wang and Phillips (2012), our test appears to have greater power against a broad range of alternatives, with these power improvements being especially pronounced for integrable and asymptotically vanishing alternatives.

Underpinning our limit theory for regression estimators and specification tests are a collection of new technical results concerning the asymptotics of additive functionals of WNPs, of the form

$$(2) \quad \frac{1}{n} \sum_{t=1}^n f(\beta_n^{-1} x_t) \quad \text{and} \quad \frac{\beta_n}{nh_n} \sum_{t=1}^n K\left(\frac{x_t - x}{h_n}\right)$$

where $\beta_n^2 = \text{Var}(x_n)$, f is locally integrable, K is integrable, and h_n denotes a bandwidth sequence. These new results are needed because WNPs are characterised by being: (a) sufficiently nonstationary to resist the application of existing LLNs; and (b) so weakly dependent that the finite-dimensional distributions of the process $r \mapsto \beta_n^{-1} x_{\lfloor nr \rfloor}$ converge to those of a nonseparable Gaussian process.³ Since this convergence cannot be strengthened to weak convergence with respect to the uniform or Skorokhod topologies, the

³By the *nonseparability* of a process G , we mean there does *not* exist a countable $T \subset [0, 1]$ such that for every open interval $I \subset [0, 1]$, $\inf_{r \in I \cap T} G(r) = \inf_{r \in I} G(r)$ and $\sup_{r \in I \cap T} G(r) = \sup_{r \in I} G(r)$ (see Loève, 1978, p. 171).

asymptotics of (2) are not amenable to an application of the continuous mapping theorem, or even to more general results on the convergence of integral functionals (see Gikhman and Skorokhod, 1969, p. 485, Thm 1). Despite this, we shall prove that for WNPs

$$(3) \quad \frac{1}{n} \sum_{t=1}^n f(\beta_n^{-1} x_t) \xrightarrow{d} \int_{\mathbb{R}} f(x + X^-) \varphi_{\sigma_+^2}(x) dx$$

where $X^- \sim N[0, \sigma_-^2]$ (with possibly $\sigma_-^2 = 0$) and $\varphi_{\sigma_+^2}$ denotes the $N[0, \sigma_+^2]$ density, and

$$(4) \quad \frac{\beta_n}{nh_n} \sum_{t=1}^n K\left(\frac{x_t - x}{h_n}\right) \xrightarrow{d} \varphi_{\sigma_+^2}(-X^-) \int_{\mathbb{R}} K(u) du.$$

The remainder of this paper is organised as follows. To help further motivate this work, two leading examples of WNPs, $I(1/2)$ and MI processes, are discussed in detail in Section 2. Our general results on the asymptotics of the functionals in (2) are presented in Section 3. These provide the basis for the asymptotics of both parametric and nonparametric estimators in regression models involving WNPs – or non-linear transformations thereof – developed in Section 4. These results are in turn used, in Section 5, to propose a specification test whose limiting distribution is invariant to the persistence of the regressor. Its finite-sample performance is evaluated through the simulation exercises presented in Section 6. Proofs of all results are given in the Appendices (in the Supplementary Material).

Notation. $\xrightarrow{a.s.}$, \xrightarrow{p} and \xrightarrow{d} respectively denote convergence almost surely, in probability, and in distribution. For deterministic sequences $\{a_n\}$ and $\{b_n\}$, $a_n \sim b_n$ denotes $\lim_{n \rightarrow \infty} a_n/b_n = 1$ and $a_n \asymp b_n$ denotes $\lim_{n \rightarrow \infty} |a_n/b_n| \in (0, \infty)$. For a random variable X , $X \sim F$ denotes that X has distribution F . For a positive real number x , $\lfloor x \rfloor$ denotes its integer part. $\mathbf{1}\{A\}$ denotes the indicator function for the set A . $\overline{\mathbb{R}}$, \mathbb{R}_+ and \mathbb{R}_+^* are the extended, the non-negative, and (strictly) positive real numbers respectively. $f^{(j)}(x)$ denotes the j th derivative of the function $f(x)$. All limits are taken as $n \rightarrow \infty$ unless otherwise indicated.

2. Leading examples of WNPs: $I(1/2)$ and MI processes. In the next section, we provide general limit theorems for the additive functionals in (2) under high-level conditions, which may be regarded as defining the class of weakly nonstationary processes (WNPs); these results will then be specialised to $I(1/2)$ and MI processes. Before doing so, we give a precise

definition of these two processes, which helps to motivate our high-level conditions. To bring these processes into a common framework, consider a linear process array of the form

$$(5) \quad x_t(n) = \sum_{j=0}^{t-1} \phi_j(n) v_{t-j}, \quad \text{where} \quad v_t = \sum_{i=0}^{\infty} c_i \xi_{t-i},$$

$t = 1, \dots, n \in \mathbb{N}$, and the coefficients $\phi_j(n)$ and c_i will be specified below. (Where there is no possibility of ambiguity, we shall generally denote $x_t(n)$ as simply x_t , for ease of notation.) $\{\xi_t\}_{t \in \mathbb{Z}}$ satisfies

ASSUMPTION INN.

- (i) ξ_t is i.i.d. with $\mathbf{E}\xi_1 = 0$ and $\text{Var}(\xi_1) = \sigma_\xi^2 < \infty$.
- (ii) ξ_1 has an absolutely continuous distribution, and a characteristic function $\psi_\xi(\lambda)$ that satisfies $\int_{\mathbb{R}} |\psi_\xi(\lambda)|^\theta d\lambda < \infty$, for some $\theta \in \mathbb{N}$.

2.1. *I(1/2) processes.* The definition of a ‘fractionally integrated process’ used in this paper closely follows that of Marinucci and Robinson (1999). These authors classify a non-stationary fractional process x_t as type I or type II according to the ‘type’ of the fractional Brownian motion (fBM) to which the finite dimensional distributions of $\beta_n^{-1} x_{[nr]}$ converge, where $\beta_n^2 := \text{Var}(x_n)$. Although these authors consider processes with a long memory component that is specified ‘parametrically’ – via the expansion of an autoregressive lag polynomial $(1 - L)^d$ (see also Remark 2.1(c) below) – their classification extends straightforwardly to the case where this is instead formulated ‘semi-parametrically’, in terms of the decay rate of the coefficients $\{\phi_j\}$ in (5). Thus we shall say that for $d \in (1/2, 1)$, x_t is an $I(d)$ process of

- type I: if $\phi_j = 1$, $c_s \sim \ell(s)s^{d-2}$, $\sum_{s=0}^{\infty} c_s = 0$; and
- type II: if $\phi_j \sim \ell(j)j^{d-1}$, $\sum_{s=0}^{\infty} |c_s| < \infty$, and $\sum_{s=0}^{\infty} c_s \neq 0$;

where $\ell : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is locally integrable on $[1, \infty)$, positive-valued and slowly varying at infinity (henceforth, ‘SV’) in the sense of Bingham, Goldie and Teugels (1987, p. 6). Fractional processes of this kind have been widely studied when $d > 1/2$: see e.g. Taqqu (1975), Kasahara and Maejima (1988) and Jeganathan (2004, 2008) for the type I case, and Robinson and Hualde (2003), Phillips and Shimotsu (2004), Shimotsu and Phillips (2005) and Hualde and Robinson (2011) for the type II case.

The preceding extend naturally to $d = 1/2$, and give the definitions of $I(1/2)$ processes (of each type) used throughout this paper, even though $\beta_n^{-1} x_{[nr]}$ will not converge weakly to an fBM of either type in the case. We shall accordingly develop our limit theory for these processes under

ASSUMPTION FR. $x_t(n)$ is generated by (5). ℓ is SV such that $L(n) := \int_1^n [\ell^2(x)/x]dx \rightarrow \infty$ as $n \rightarrow \infty$, and either:

- FR1** $\phi_j = 1 \ \forall j \geq 0$, $c_s \sim \ell(s)s^{-3/2}$, and $\sum_{s=0}^{\infty} c_s = 0$; or
FR2 $\phi_j \sim \ell(j)j^{-1/2}$, $\phi_0 \neq 0$, $\sum_{s=0}^{\infty} |c_s| < \infty$, $\sum_{s=0}^{\infty} c_s \neq 0$.

REMARK 2.1. (a) **FR1** and **FR2** respectively imply that $x_t(n)$ is an $I(1/2)$ process of types I and II.

(b) $\text{Var}(x_n) \asymp L(n)$ under **FR**, where $L(n)$ is itself SV. The divergence or convergence of $L(n)$ effectively demarcates the boundary between WNP and stationary long memory processes. Either is possible, depending on ℓ : e.g. $\ell(n) = 1$ gives $L(n) \sim \ln n$ and $\ell(n) = (\ln n)^{-1/2}$ gives $L(n) \asymp \ln \ln n$, whereas $\ell(n) = (\ln n)^{-1}$ gives a bounded $L(n)$. When $L(n)$ diverges, $L(n)^{-1/2}x_n$ will obey a CLT. But when $L(n)$ is bounded, no CLT applies: and indeed in this case we have under **FR2** that $\sum_{j=0}^{\infty} \phi_j^2 < \infty$, so that x_t is stationary. Such processes fall within the purview of existing results, and so have been excluded by our assumption that $L(n) \rightarrow \infty$.

(c) **FR** encompasses both types of parametric ARFIMA models with $d = 1/2$. For example, consider the ARFIMA(1/2) type II model

$$(6) \quad (1 - L)^{1/2}x_t = v_t 1\{t > 0\}, \quad \text{where } a(L)v_t = b(L)\xi_t,$$

where a and b denote finite-order polynomials in the lag operator L . In this case, it is possible to write $x_t = \sum_{j=0}^{t-1} \phi_j v_{t-j}$, where $\{\phi_j\}_{j \geq 0}$ are the coefficients in the power series expansion of $(1 - L)^{-1/2}$; and so $\phi_0 = 1$ and $\phi_j \asymp j^{-1/2}$ (see e.g. p. 673 in Johansen and Nielsen, 2012b). Further, if all the roots of a lie outside the unit circle, $v_t = a(L)^{-1}b(L)\xi_t$ is a linear process with geometrically decaying coefficients. Thus **FR2** is satisfied. Since $\ell(x) = 1$, we have trivially that $L(n) \rightarrow \infty$. Similar arguments show that **FR1** is consistent with an ARFIMA(1/2) type I model, under which x_t is the partial sum of a ARFIMA(-1/2) type II process.

2.2. *MI processes.* Mildly integrated (MI) processes are closely related to the nearly integrated (NI) processes studied by Chan and Wei (1987) and Phillips (1987), and more recently extended by Buchmann and Chan (2007). Both MI and NI processes may be defined in terms of an array as

$$(7) \quad x_t(n) = (1 - \kappa_n^{-1})x_{t-1}(n) + v_t,$$

where $x_0(n) = 0$ and v_t is a stationary process and $\kappa_n > 1$ with $\kappa_n \rightarrow \infty$, so that the autoregressive coefficient approaches unity as n grows. They can

thus be encompassed within the framework of (5) if we allow ϕ_j to depend on n as per

$$(8) \quad \phi_j = \phi_j(n) = (1 - \kappa_n^{-1})^j.$$

Both NI and MI processes thus describe highly persistent autoregressive processes, which have a root in the vicinity of unity. They have accordingly been used to investigate the behaviour of various inferential procedures under local departures from unit roots (e.g. Mikusheva, 2007, and Duffy, 2020), and in the construction of robust inferential procedures (e.g. Magdalinos and Phillips, 2011; Kostakis, Magdalinos and Stamatogiannis, 2015; Demetrescu et al 2019; Yang, Long, Peng and Cai, 2019). The crucial difference between NI and MI processes concerns the assumed growth rate of the sequence κ_n . NI processes are defined by $\kappa_n/n \rightarrow c \neq 0$, with the consequence that $n^{-1/2}x_{[nr]}$ converges weakly to an Ornstein-Uhlenbeck process. MI processes have $\kappa_n/n \rightarrow 0$, which tilts $x_t(n)$ closer to stationarity: and as a consequence, FCLTs cannot be used to derive the asymptotics of functionals of these processes. Their variance also grows at a slower rate, since $Var(x_n(n)) \asymp \kappa_n = o(n)$.

Formally, we define an MI process as follows, specifying regularity conditions that are helpful in unifying notation and simplifying some derivations.⁴

ASSUMPTION MI. $x_t(n)$ and $\phi_j(n)$ are as in (5) and (8). $\{c_s\}_{s \in \mathbb{Z}}$ is such that $\sum_{s=0}^{\infty} |c_s| < \infty$ and $\sum_{s=0}^{\infty} c_s \neq 0$. $\{\kappa_n\}_{n \in \mathbb{N}}$ has $\kappa_n > 1$, $\kappa_n = n^{\alpha_\kappa} \ell_\kappa(n)$ for ℓ_κ SV and $\alpha_\kappa \in [0, 1)$, $\kappa_n \rightarrow \infty$ and $\sup_{n \geq 1} \sup_{1 \leq t \leq n} \kappa_n^{-1} \kappa_t < \infty$.

3. Limit theory for functionals of WNPs.

3.1. *Additive functionals of standardised processes.* Consider

$$(9) \quad \frac{1}{n} \sum_{t=1}^n f(\beta_n^{-1} x_t)$$

where f is locally integrable, and $\beta_n^2 := Var(x_n(n))$. Here we provide high-level conditions (**Assumption HL**) under which the asymptotics of (9) may be derived. These conditions, particularly **HL0–2** and **HL4** below, may be

⁴Previous work on these processes has assumed that v_t is short memory in the sense that $\sum_{s=0}^{\infty} |c_s| < \infty$. In a previous working paper version of the present work (available as arXiv:1812.07944v1), we allowed v_t to have long memory in the sense that $c_s \sim s^{-m}$ for $m \in (1/2, 1)$, thereby extending this previous work much in the manner of Buchmann and Chan's (2007) extension of earlier work on NI processes. To keep the length of the paper manageable, this generalisation is not reported here.

taken as providing an abstract definition of a WNP – with **HL3**, **HL5** and **HL6** being merely regularity conditions that may be dispensed with, if f satisfies certain assumptions. These high-level conditions are stated in terms of a general random array denoted $\{X_t(n)\}$, to distinguish it from the linear process array $\{x_t(n)\}$ introduced in (5) above.

ASSUMPTION HL (high-level conditions).

HL0 Let $\{X_t(n)\}_{t=1}^n$, $n \in \mathbb{N}$ be a random array and $\{\mathcal{F}_t\}_{t=-\infty}^\infty$ a filtration such that $X_t(n)$ is \mathcal{F}_t -measurable for all t and n . Let $\{\beta_n\}$ denote a positive sequence with $\beta_n \rightarrow \infty$.

HL1 $X_t(n) = X_t^+(n) + X_t^-(n) + R_t(n)$, where $X_t(n)^-$ is \mathcal{F}_0 -measurable, and $\sup_{1 \leq t \leq n} \mathbf{P}\{\beta_n^{-1}|R_t(n)| > \epsilon\} \rightarrow 0$ for every $\epsilon > 0$.

HL2 There are random variables X^+ and X^- , where X^+ has bounded continuous density Φ_{X^+} such that: for every $\delta \in (0, 1)$ and $\{t_n\}$ with $\lfloor n\delta \rfloor \leq t_n \leq n$

(a) $\beta_n^{-1}X_{t_n}^+(n) \xrightarrow{d} X^+$, conditionally on \mathcal{F}_0 in the sense that for all bounded and continuous $h : \mathbb{R} \rightarrow \mathbb{R}$

$$\mathbf{E}[h(\beta_n^{-1}X_{t_n}^+(n)) \mid \mathcal{F}_0] \xrightarrow{p} \int_{\mathbb{R}} h(x)\Phi_{X^+}(x)dx; \text{ and}$$

(b) $\beta_n^{-1}X_{t_n}^-(n) \xrightarrow{d} X^-$, and $\beta_n^{-1}[X_n^-(n) - X_{t_n}^-(n)] \xrightarrow{p} 0$.

HL3 $\beta_t^{-1}X_t(n)$ has density $\mathcal{D}_{n,t}(x)$ such that for some $n_0 \geq t_0 \geq 1$,

$$\sup_{n \geq n_0, t_0 \leq t \leq n} \sup_x \mathcal{D}_{n,t}(x) < \infty$$

HL4 For every bounded and Lipschitz continuous $g : \mathbb{R} \rightarrow \mathbb{R}$

$$\frac{1}{n} \sum_{t=1}^n g(\beta_n^{-1}X_t(n)) = \frac{1}{n} \sum_{t=1}^n \mathbf{E}[g(\beta_n^{-1}X_t(n)) \mid \mathcal{F}_0] + o_p(1).$$

HL5 For some $\lambda \in (0, \infty)$ and $n_0 \geq 1$

(a) $\sup_{n \geq n_0, 1 \leq t \leq n} \mathbf{E}|\beta_n^{-1}X_t(n)|^\lambda < \infty$; or

(b) $\sup_{n \geq n_0, 1 \leq t \leq n} \mathbf{E} \exp(\lambda |\beta_n^{-1}X_t(n)|) < \infty$.

HL6 $\sup_{n \in \mathbb{N}} \frac{\beta_n}{n} \sum_{t=t_0}^n \beta_t^{-1} < \infty$, for t_0 as in **HL3**.

REMARK 3.1. (a) When $X_t(n)$ is a linear process array formed from an underlying i.i.d. sequence $\{\xi_t\}$ as in (5), **HL1** is trivially satisfied by splitting it into terms depending on $\{\xi_s\}_{s \leq 0}$ and $\{\xi_s\}_{s=1}^t$.

(b) **HL2** expresses one of the key properties of a WNP: that its finite dimensional distributions should converge (upon standardisation), albeit not to those of a separable process. Its requirements may be illustrated by an $I(1/2)$ type I process with $\ell(x) = 1$, denoted $\{x_t\}$. Per the previous remark, write $x_t = x_t^+ + x_t^-$, where x_t^+ and x_t^- are respectively weighted sums of $\{\xi_s\}_{s=1}^t$ and $\{\xi_s\}_{s \leq 0}$. Then a CLT for weighted sums of linear processes (see Abadir, Distaso, Giraitis, and Koul, 2014) yields that for every $r, s \in (0, 1]$

$$(10) \quad \beta_n^{-1}(x_{\lfloor nr \rfloor}^+, x_{\lfloor ns \rfloor}^+) \xrightarrow{d} (\eta_r, \eta_s)$$

where $\beta_n^2 \asymp \ln n$, and η_r and η_s are independent $N[0, 1/2]$ random variables. We thus have the marginal convergence of each coordinate of $\beta_n^{-1}x_{\lfloor nr \rfloor}^+$ to identical distributional limits, as per **HL2(a)** – and if ξ_t is i.i.d., the required conditional convergence holds trivially. In that case, (10) holds jointly with (and independently of)

$$(11) \quad \beta_n^{-1}(x_{\lfloor nr \rfloor}^-, x_{\lfloor ns \rfloor}^-) \xrightarrow{d} (\eta^-, \eta^-)$$

where $\eta^- \sim N[0, 1/2]$. Note the degeneracy in the joint distribution of the limit in (11), consistent with the second part of **HL2(b)**.

(c) **HL3** is useful for establishing L_1 -approximations to functionals of WNPs, which permit convergence results proved under the requirement that f in (9) be bounded and continuous to be extended to a much broader class of integrable functions. High-level conditions similar to **HL3** have been employed for similar purposes in many previous works, e.g. Jeganathan (2004, 2008), Pötscher (2004), Gao, King, Lu and Tjøstheim (2009), Wang and Phillips (2009a,b; 2012) among others.

(d) **HL4** expresses the requirement that a WNP should not be too strongly dependent. It would fail both for $I(d)$ processes with $d > 1/2$, and for NI processes – and indeed for any process for which $\beta_n^{-1}X_{\lfloor nr \rfloor}(n)$ converges weakly to a process with continuous sample paths.

THEOREM 3.1. *Suppose $f : \mathbb{R} \rightarrow \mathbb{R}$ is locally integrable, and **HL0–4** and **HL6** hold. Further suppose f is bounded or that:*

- (i) *There is a $\mathcal{Y} \subseteq \mathbb{R}$ such that $\int_{\mathbb{R}} |f(x+y)| \Phi_{X^+}(x) dx < \infty$ for all $y \in \mathcal{Y}$, and $\mathbf{P}(X^- \in \mathcal{Y}) = 1$;*
- (ii) *$n^{-1} \sum_{t=1}^{t_0-1} f(\beta_n^{-1}X_t(n)) = o_p(1)$, for t_0 as in **HL3**;*
- (iii) *For some $\lambda' \in (0, \lambda)$, where λ is as in **HL5**, either:*
 - (a) *$|f(x)| = O(|x|^{\lambda'})$, as $|x| \rightarrow \infty$ and **HL5(a)** holds; or*
 - (b) *$|f(x)| = O(\exp(\lambda'|x|))$, as $|x| \rightarrow \infty$ and **HL5(b)** holds.*

Then as $n \rightarrow \infty$,

$$(12) \quad \frac{1}{n} \sum_{t=1}^n f(\beta_n^{-1} X_t(n)) \xrightarrow{d} \int_{\mathbb{R}} f(x + X^-) \Phi_{X^+}(x) dx.$$

REMARK 3.2. (a) If f is bounded, conditions (i)–(iii) hold trivially, and **HL5** is unnecessary for (12). If f is additionally Lipschitz, then **HL3** and **HL6** may also be dispensed with.

(b) Condition (ii) of Theorem 3.1 is a technical requirement that has also been employed in other studies that develop limit theory for functionals of nonstationary processes, e.g. Jeganathan (2004) and Pötscher (2004). This condition is redundant if **HL3** holds with $t_0 = 1$.

(c) Since $X_t(n)$ is continuously distributed under **Assumption INN**, (12) continues to hold if f is modified on a set of Lebesgue measure zero. Thus e.g. if f has an integrable pole at some $x_0 \in \mathbb{R}$, and otherwise satisfies the requirements of Theorem 3.1, then (12) holds regardless of how f is defined at x_0 .

For $I(1/2)$ and MI processes, it may be shown that **Assumption INN** and each of **FR** and **MI** are sufficient for **Assumption HL**. Theorem 3.1 therefore specialises as follows. Recall that $\varphi_{\sigma^2}(x)$ denotes the $N[0, \sigma^2]$ density, and take $\beta_n^2 = \text{Var}(x_n(n))$. For $X^- \sim N[0, 1/2]$, let

$$(13) \quad \varrho(x) := \begin{cases} \varphi_{1/2}(x - X^-) & \text{under } \mathbf{FR1} \\ \varphi_1(x) & \text{under } \mathbf{FR2}, \mathbf{MI}, \end{cases}$$

noting that this is a density, albeit a random one under **FR1**.

THEOREM 3.2. Suppose $f : \mathbb{R} \rightarrow \mathbb{R}$ is locally Lebesgue integrable, and **Assumption INN** and either **FR** or **MI** hold. Further suppose:

1. Either: (a) for each $t' \in \mathbb{N}$ fixed, $n^{-1} \sum_{t=1}^{t'} f(\beta_n^{-1} x_t(n)) = o_p(1)$; or (b) **Assumption INN** holds with $\theta = 1$.
2. For some $\lambda' \in (0, \infty)$, as $|x| \rightarrow \infty$ either
 - (a) $|f(x)| = O(|x|^{\lambda'})$, and $\mathbf{E} |\xi_1|^{2\vee\lambda} < \infty$ for some $\lambda > \lambda'$; or
 - (b) $|f(x)| = O(\exp(\lambda' |x|))$, and ξ_1 has a finite m.g.f. in a neighbourhood of zero.

Then

$$(14) \quad \frac{1}{n} \sum_{t=1}^n f(\beta_n^{-1} x_t(n)) \xrightarrow{d} \int_{\mathbb{R}} f(x) \varrho(x) dx.$$

REMARK 3.3. (a) Theorem 3.2 bridges existing asymptotic results for $I(d)$ processes of order $d \in (-1/2, 1/2)$ and $d \in (1/2, 3/2)$. There are similarities and differences between (14) and the limit theory that applies in these two cases. Firstly, there is some analogy with the LLN results that hold when $d \in (-1/2, 1/2)$. As in that case, the limit in (14) is determined by an expectation, but with the difference that the expectation in (14) is with respect to a limiting distribution, rather than the invariant distribution of a strictly stationary process. Secondly, the limiting density (φ_1 or $\varphi_{1/2}$) is obtained by an application of a CLT, and in this respect the limit theory is analogous to that for $d \in (1/2, 3/2)$, which involves FCLTs. In that case, the weak limits of additive functionals are stochastic, being a functional of a limiting fBM. This nondegeneracy of the limit carries over to the $I(1/2)$ type I process (FR1), as evinced by the dependence of ϱ on $X^- \sim N[0, 1/2]$ in this case. (The type II process can be regarded as a truncated type I process, and the additional variability of the latter appears to make its behaviour closer to that of an $I(d)$ processes with $d > 1/2$).

(b) Theorem 3.2 generalises the limit theory of Giraitis and Phillips (2006) and Phillips and Magdalinos (2007) for MI processes to general nonlinear functionals. Those two papers consider quadratic functions (i.e. $f(x) = x^2$) of MI processes driven by short memory linear processes errors. Using a direct approach they show that $(n\beta_n^2)^{-1} \sum_{t=1}^n x_t(n)^2 \xrightarrow{P} 1$. This result can be understood as a special case of Theorem 3.2; indeed in this case the r.h.s. of (14) is

$$\int_{\mathbb{R}} f(x) \varphi_1(x) dx = \int_{\mathbb{R}} x^2 \varphi_1(x) dx = 1.$$

Theorem 3.2 also generalises a result due to Tanaka (1999, p. 555), who shows that $\frac{1}{n \ln n} \sum_{t=1}^n x_t^2 = O_p(1)$ for an ARFIMA(1/2) processes, in the course of deriving the asymptotics of the maximum likelihood estimator for d over a parameter space that includes $d = 1/2$.

(c) **Assumption INN** entails that ξ_t has finite second moment, and so $\beta_n^{-1} x_{\lfloor nr \rfloor}(n)$ satisfies a CLT; the limiting distributions that appear in (14) are therefore Gaussian.⁵ Note that the process $r \mapsto \beta_n^{-1} x_{\lfloor nr \rfloor}(n)$ does not converge weakly (with respect to the uniform or Skorokhod topologies): as the example of an $I(1/2)$ type I process given in Remark 3.1(b) illustrates, this is impossible because the finite-dimensional distributions of $\beta_n^{-1} x_{\lfloor nr \rfloor}(n)$

⁵Some preliminary work of the authors' shows that Theorem 3.2 can be extended to the case where ξ_t is in the domain of attraction of an α -stable law with parameter $\alpha \in (0, 2)$, in which case other stable distributions will appear in the limit. We leave extensions of this kind for future work.

converge to those of a nonseparable process. (Similar calculations show that if x_t is $I(1/2)$ type II or MI, then the fidis converge to those of a Gaussian ‘white noise’ process G for which $G(r) \sim N[0, 1]$ is independent of $G(s)$ for all $r, s \in [0, 1]$.) This accords with the results of Johansen and Nielsen (2012b), who show that for an $I(d)$ process with $d \in (1/2, 3/2)$ of either type, $\{\xi_t\}$ must have moments of order greater than $(d - 1/2)^{-1}$ if $\beta_n^{-1}x_{\lfloor nr \rfloor}(n)$ is to converge weakly to a fBM, a requirement that becomes progressively more demanding as d approaches $1/2$.

3.2. *Kernel functionals.* We next consider kernel functionals of the form

$$(15) \quad \frac{\beta_n}{h_n n} \sum_{t=1}^n K \left(\frac{x_t(n) - x}{h_n} \right),$$

where $x \in \mathbb{R}$, h_n is a bandwidth sequence, and K is an integrable kernel function satisfying

ASSUMPTION K (kernel). $K : \mathbb{R} \rightarrow \overline{\mathbb{R}}$ is such that K and K^2 are Lebesgue integrable.

Whereas in (9) the nonlinear transformation f is applied to the standardised process $\beta_n^{-1}x_t(n)$, in (15) K is applied to the unstandardised process $x_t(n)$. This leads to a different limit theory, which is partly reflected in the different normalisations of the sums in (9) and (15). A notable difference between (15) and the usual expression for a kernel density estimator is the appearance of β_n . The more variable that x_t is, the less frequent are its visits to the support of $K[(\cdot - x)/h_n]$ – and since the variance β_n of a WNP grows with n , these visits accumulate only at rate nh_n/β_n , a fact reflected exactly in the normalisation in (15).

Nonetheless, the results of the preceding section turn out to be highly relevant for the asymptotics of (15). To explain why this is the case, we return to the setting of **Assumption HL** above, which we now augment by the following additional smoothness conditions on the density of the increments of $X_t(n)$. To state these, let

$$\Omega_n(\eta) := \{ \{s, t\} \in \mathbb{N} : \lfloor \eta n \rfloor \leq s \leq \lfloor (1 - \eta)n \rfloor, \lfloor \eta n \rfloor + s \leq t \leq n \},$$

for $\eta \in (0, 1)$.

ASSUMPTION HL (continued).

HL7 Let $X_0(n) := 0$ and $t > s \geq 0$. Conditionally on \mathcal{F}_s , $\beta_{t-s}^{-1}(X_t(n) - X_s(n))$ has density $\mathcal{D}_{t,s,n}(x)$ such that for some $n_0, t_0 \geq 1$

$$\sup_{n \geq n_0, 0 \leq s < t \leq n, t-s \geq t_0} \sup_x \mathcal{D}_{t,s,n}(x) < \infty.$$

HL8 For all $q_0, q_1 > 0$

$$\lim_{\eta \downarrow 0} \limsup_{n \rightarrow \infty} \sup_{(s,t) \in \Omega_n(\eta)} \sup_{|x| \leq q_0 \eta^{q_1}} |\mathcal{D}_{t,s,n}(x) - \mathcal{D}_{t,s,n}(0)| = 0$$

HL9 For t_0 as in **HL7**:

- (a) $\lim_{\eta \rightarrow 0} \lim_{n \rightarrow \infty} \frac{\beta_n}{n} \sum_{t=\lfloor (1-\eta)n \rfloor}^n \beta_t^{-1} = 0$;
- (b) $\lim_{\eta \rightarrow 0} \lim_{n \rightarrow \infty} \frac{\beta_n}{n} \sum_{t=t_0}^{\lfloor \eta n \rfloor} \beta_t^{-1} = 0$;
- (c) $\limsup_{n \rightarrow \infty} \frac{\beta_n}{n} \sum_{t=t_0}^n \beta_t^{-1} < \infty$;
- (d) there exist $l_0, l_1 > 0$ such that $\liminf_{n \rightarrow \infty} \beta_n^{-1} \inf_{(s,t) \in \Omega_n(\eta)} \beta_{t-s} \geq \eta^{l_1}/l_0$ for all $\eta \in (0, 1)$;
- (e) $\sup_{n \geq 1} \sup_{1 \leq t \leq n} \beta_n^{-1} \beta_t < \infty$.

HL7 and **HL9** may be regarded as strengthened versions of **HL3** and **HL6**, and are closely related to Assumption 2.3 in Wang and Phillips (2009a). Under these conditions, an L_1 -approximation argument developed by those authors and Jeganathan (2004) yields that, for t_0 as in **HL7**,

$$\frac{\beta_n}{h_n n} \sum_{t=t_0}^n K\left(\frac{X_t(n) - x}{h_n}\right) = \frac{1}{n} \sum_{t=t_0}^n \varphi_{\varepsilon^2}(\beta_n^{-1} X_t(n)) \int_{\mathbb{R}} K(u) du + o_p(1),$$

as $n \rightarrow \infty$ and then $\varepsilon \rightarrow 0$. The leading order term on the r.h.s. clearly has the same form as the l.h.s. of (12) and is thus amenable to a direct application of Theorem 3.1, which entails

$$(16) \quad \begin{aligned} \frac{1}{n} \sum_{t=1}^n \varphi_{\varepsilon^2}(\beta_n^{-1} X_t(n)) &\xrightarrow{d} \int \varphi_{\varepsilon^2}(x + X^-) \Phi_{X^+}(x) dx, \quad \text{as } n \rightarrow \infty \\ &\xrightarrow{a.s.} \Phi_{X^+}(-X^-), \quad \text{as } \varepsilon \rightarrow 0. \end{aligned}$$

We thus have the following counterpart of Theorem 3.1 for kernel functionals.

THEOREM 3.3. Suppose that, in addition to **Assumptions K**, **HL0–2**, **HL4** and **HL6–9**, the following hold:

- (i) $\{h_n\}$ is a positive sequence with $\beta_n^{-1} h_n + \beta_n(nh_n)^{-1} \rightarrow 0$; and

- (ii) for each $x \in \mathbb{R}$ and t_0 as in **HL7**, $\frac{\beta_n}{nh_n} \sum_{t=1}^{t_0-1} K\left(\frac{X_t(n)-x}{h_n}\right) = o_p(1)$.

Then

$$\frac{\beta_n}{h_n n} \sum_{t=1}^n K\left(\frac{X_t(n)-x}{h_n}\right) \xrightarrow{d} \Phi_{X^+}(-X^-) \int_{\mathbb{R}} K(u) du.$$

For $I(1/2)$ and MI processes, the preceding specialises as follows.

THEOREM 3.4. Suppose that, in addition to **Assumption K**:

- (i) $x_t(n)$ satisfies **Assumption INN**, and **FR** or **MI**;
- (ii) $\{h_n\}$ satisfies condition (i) of Theorem 3.3; and
- (iii) for each $x \in \mathbb{R}$ and $t' \in \mathbb{N}$, $\frac{\beta_n}{nh_n} \sum_{t=1}^{t'} K\left(\frac{x_t(n)-x}{h_n}\right) = o_p(1)$.

Then

$$\frac{\beta_n}{h_n n} \sum_{t=1}^n K\left(\frac{x_t(n)-x}{h_n}\right) \xrightarrow{d} \varrho(0) \int_{\mathbb{R}} K(u) du.$$

REMARK 3.4. (a) Theorem 3.4 fills a gap in existing asymptotic theory for kernel functionals of linear processes. A general theory for stationary linear processes, including $I(d)$ processes with $|d| < 1/2$, is given in Wu and Mielniczuk (2002). Supposing that $\int_{\mathbb{R}} K = 1$, under their conditions kernel functionals converge in probability to the invariant density of the stationary process. Jeganathan (2004, 2008) provides limit theorems for kernel functionals of $I(d)$ processes with $1/2 < d < 3/2$. In that case, kernel functionals converge to the local time of a fractional Brownian motion (or fractional stable motion if innovations are in the domain of attraction of a stable law) – so their limit is an occupation density rather than the invariant density of some stationary process. The limiting behaviour of kernel functionals of $I(1/2)$ processes is intermediate between these two cases. These converge to the density of a random variable, rather than to an occupation density, but the density corresponds to a limiting random variate, rather than the invariant density of a stationary process.

(b) Theorem 3.4 nests a similar result provided by Duffy (2020) for bounded kernel functionals of MI processes, which unlike **Assumption K** requires K to be bounded and Lipschitz continuous.

4. Estimation and inference in regressions with WNPs. The preceding results are fundamental to the asymptotics of parametric and non-parametric least squares estimators, in models involving WNPs as regressors. In this section, we show that these estimators have either Gaussian or mixed Gaussian limit distributions, and in consequence their associated t statistics

are asymptotically standard Gaussian. These results are in turn used, in Section 5, to derive the asymptotic distribution of a proposed regression specification test statistic, and in particular to show that it is asymptotically pivotal, being unaffected by the persistence of the regressor process.

4.1. *Parametric regression.* Consider the ordinary least squares (OLS) estimator of (μ, γ) in the model

$$(17) \quad y_t = \mu + \gamma g(x_{t-1}) + u_t$$

given by $(\hat{\mu}, \hat{\gamma}) := \operatorname{argmin}_{(a,b)} \sum_{t=1}^n [y_t - a - bg(x_{t-1})]^2$, where g is a known nonlinear transformation. Since the regressor is predetermined (i.e. \mathcal{F}_{t-1} -measurable) relative to the error u_t , (17) is an instance of a so-called ‘predictive’ or ‘reduced form’ regression model. If x_t is stationary, the OLS estimator will be asymptotically normal; whereas if x_t is strongly dependent, the OLS estimator has a non-standard limiting distribution, unless either g is itself integrable, or x_t and u_t satisfy a very restrictive ‘long-run orthogonality’ condition (see e.g. Park and Phillips, 1999, 2001).

When x_t is a WNP, the OLS estimator is either asymptotically normal or mixed normal, depending on the type of process. In either case, the t statistic is asymptotically $N[0, 1]$, due to self-normalisation. In this respect, the asymptotics are similar to those when x_t is stationary; but since the variance of a WNP grows without bound, the analysis requires arguments more appropriate to nonstationary processes. In particular, the following property of g , first introduced by Park and Phillips (1999, 2001), plays a key role.

DEFINITION AHF (asymptotically homogeneous function). Let $\{x_t(n)\}$ denote a random array and $\beta_n^2 = \operatorname{Var}(x_n(n))$. $g : \mathbb{R} \rightarrow \overline{\mathbb{R}}$ is *asymptotically homogeneous* for $\{x_t(n)\}$, if for each $\lambda > 0$ it admits the decomposition

$$g(x) = \kappa_g(\lambda) H_g(x/\lambda) + R_g(x, \lambda),$$

where $\kappa_g : \mathbb{R}_+^* \rightarrow \mathbb{R}_+^*$, $H_g : \mathbb{R} \rightarrow \overline{\mathbb{R}}$, $R_g : \mathbb{R} \times \mathbb{R}_+^* \rightarrow \overline{\mathbb{R}}$ and for $j = \{1, 2\}$,

$$(18) \quad R_{g,n}^j := \frac{1}{\kappa_g^j(\beta_n) n} \sum_{t=1}^n \mathbf{E} |R_g(x_t(n), \beta_n)|^j = o(1).$$

AHFs encompass a wide range of commonly used regression functions, such as polynomial functions, cumulative distribution functions (with $H_g(u) = \mathbf{1}\{u > 0\}$), and logarithmic functions (with $H_g(u) = 1$); see Park and Phillips (1999, 2001) for some further examples. Such a condition as

$$\lim_{\lambda \rightarrow \infty} \kappa_g(\lambda)^{-1} \sup_x |R_g(x, \lambda)| = 0$$

is sufficient, but not necessary, for (18) to hold. The relevance of AHFs for the OLS estimator can be seen most easily in the case where $\mu = 0$ is known and imposed, so that the OLS estimator for γ satisfies

$$\hat{\gamma}_n - \gamma = \frac{\sum_{t=2}^n g(x_{t-1})u_t}{\sum_{t=2}^n g^2(x_{t-1})} = (1 + o_p(1)) \frac{\sum_{t=2}^n H_g(\beta_n^{-1}x_{t-1})u_t}{\sum_{t=2}^n H_g^2(\beta_n^{-1}x_{t-1})}.$$

Upon standardisation, the denominator on the r.h.s. is directly amenable to an application of Theorem 3.2. Since the numerator is a sum of martingale differences, it can be handled via an appropriate martingale CLT (either Hall and Heyde, 1980; or Wang, 2014); here Theorem 3.2 is used to verify the stability condition pertaining to its conditional variance.

Reasoning along these lines yields our main result on parametric OLS estimators. To state it, let $MN[0, \varsigma^2]$ denote a mixed normal distribution with mixing variate ς^2 (i.e. which has characteristic function $u \mapsto \mathbf{E}e^{-\varsigma^2 u^2/2}$), and $\sigma(\{\xi_s, u_s\}_{s \leq t})$ denote the σ -field generated by $\{\xi_s, u_s\}_{s \leq t}$.

THEOREM 4.1. *Let $\{y_t\}_{t=1}^n$ be generated by (17), $\mathcal{F}_t := \sigma(\{\xi_s, u_s\}_{s \leq t})$, and suppose that:*

- (i) $x_t(n)$ satisfies (5), **Assumption INN** and either **FR** or **MI**;
- (ii) $\{u_t, \mathcal{F}_t\}_{t \geq 1}$ is a martingale difference sequence such that $\mathbf{E}[u_t^2 | \mathcal{F}_{t-1}] = \sigma_u^2$ a.s. for some constant $\sigma_u^2 < \infty$;
- (iii) $\sup_{1 \leq t \leq n} \mathbf{E}[u_t^2 \mathbf{1}\{|u_t| \geq A_n\} | \mathcal{F}_{t-1}] = o_p(1)$ for non-random $A_n \rightarrow \infty$;
- (iv) $g(x)$ is AHF for $\{x_t(n)\}$, with limit homogeneous component H_g that is not a.e. constant, and is such that H_g^2 satisfies the conditions of Theorem 3.2.

Then

$$(19) \quad n^{1/2} \begin{bmatrix} \hat{\mu} - \mu \\ \kappa_g(\beta_n)(\hat{\gamma} - \gamma) \end{bmatrix} \xrightarrow{d} MN \left[0, \sigma_u^2 \left(\int \begin{bmatrix} 1 & H_g \\ H_g & H_g^2 \end{bmatrix} \varrho \right)^{-1} \right].$$

REMARK 4.1. (a) For $I(1/2)$ type II and MI processes $\hat{\gamma}_n$ is asymptotically normal; for $I(1/2)$ type I processes it is mixed normal, because then $\varrho(x) = \varphi(x - X^-)$ is random. In either case, the t statistics for testing hypotheses about μ or γ will be asymptotically standard normal, so that inferences may be drawn in the usual manner. This contrasts with the case where regressors are $I(d)$ for $d > 1/2$, e.g. see Phillips (1995), Park and Phillips (1999, 2001), Robinson and Hualde (2003).

(b) In a linear regression model, i.e. $g(x) = x$, we have $H_g(u) = u$ and $\kappa_g(\lambda) = \lambda$, and it follows that the OLS estimator for γ has convergence rate

$\beta_n n^{1/2}$, which is faster than the $n^{1/2}$ -convergence rate that obtains when the regressor is stationary. For $I(1/2)$ processes the gain in convergence rate is given by the slowly varying factor $L^{1/2}(n)$ (see Remark 2.1(b) above).

(c) Suppose instead that $y_t = \mu + \gamma g(x_t) + u_t$, so that the regressor is no longer predetermined. In this case, the asymptotics of the OLS estimator are different from (19): for example, if it is known that $\mu = 0$ and $g(x) = x$, (so that $\kappa_g(\beta_n) = \beta_n$) we have

$$\beta_n^2 (\hat{\gamma} - \gamma) \xrightarrow{d} \left[\int x^2 \varrho(x) dx \right]^{-1} \lim_{n \rightarrow \infty} \mathbf{E}(x_t(n) - x_{t-1}(n)) u_t.$$

In this case there is a severe reduction in the convergence rate, by a factor of $n^{1/2}/\beta_n$, due to the endogeneity of x_t . This result is comparable to Theorem 5.2 of Marinucci and Robinson (1998) which gives the asymptotics of the OLS estimator when x_t is $I(d)$ with $d \in (1/2, 1)$. We expect that in this setting such methods as narrowband LS (see e.g. Marinucci and Robinson, 1998; Robinson and Hualde 2003; Christensen and Nielsen 2006) or those that use lagged regressors as instruments will be more efficient. We leave the exploration of alternative estimation procedures for future work.

4.2. *Nonparametric regression.* We next consider the nonparametric estimation of m in the predictive regression

$$(20) \quad y_t = m(x_{t-1}) + u_t.$$

In particular, we consider the kernel regression (Nadaraya–Watson; NW) estimator

$$\hat{m}(x) := \sum_{t=2}^n K_{th}(x) y_t / \sum_{t=2}^n K_{th}(x),$$

and the local linear (LL) estimator

$$\begin{bmatrix} \tilde{m}(x) \\ \tilde{m}^{(1)}(x) \end{bmatrix} := \operatorname{argmin}_{(a,b) \in \mathbb{R}^2} \sum_{t=2}^n [y_t - a - b(x_{t-1} - x)]^2 K_{th}(x),$$

where $K_{th}(x) := K[(x_{t-1} - x)/h_n]$. The following theorem is a direct consequence of Theorem 3.4 and certain martingale central limit theorems, and is complementary to the recent work of Wang and Phillips (2009a,b; 2012) who develop estimation and testing procedures in the context of nonparametric regression with NI and $I(d)$ processes with $d \in (1/2, 3/2)$. Let

$$Q := \left\{ \int \begin{bmatrix} 1 & x \\ x & x^2 \end{bmatrix} K \right\}^{-1} \left\{ \int \begin{bmatrix} 1 & x \\ x & x^2 \end{bmatrix} K^2 \right\} \left\{ \int \begin{bmatrix} 1 & x \\ x & x^2 \end{bmatrix} K \right\}^{-1},$$

and $\nu_{K,i} := \int x^i K(x) dx$ for $i \in \mathbb{N}$.

THEOREM 4.2. *Let $\{y_t\}_{t=1}^n$ be generated by (20) and suppose that:*

- (i) *conditions (i)–(iii) of Theorem 4.1 hold, $\nu_{K,0} = 1$ and $\nu_{K,2} \neq \nu_{K,1}^2$.*
- (ii) *$x^j[K(x) + K^2(x)]$ are bounded and integrable for $j \in [0, 3]$;*
- (iii) *$h_n + \beta_n/nh_n \rightarrow 0$;*

If m has a bounded first derivative and $nh_n^3/\beta_n \rightarrow 0$, then

$$(21) \quad \left(\frac{nh_n}{\beta_n} \right)^{1/2} (\hat{m}(x) - m(x)) \xrightarrow{d} MN \left[0, \sigma_u^2 \varrho(0)^{-1} \int K^2 \right].$$

Alternatively, if m has a bounded second derivative and $nh_n^5/\beta_n \rightarrow 0$, then

$$(22) \quad \left(\frac{nh_n}{\beta_n} \right)^{1/2} \left[\frac{\tilde{m}(x) - m(x)}{h_n (\tilde{m}^{(1)}(x) - m^{(1)}(x))} \right] \xrightarrow{d} MN [0, \sigma_u^2 \varrho(0)^{-1} Q].$$

REMARK 4.2. **(a)** Since $\beta_n \rightarrow \infty$, the convergence rate of both \hat{m} and \tilde{m} when x_t is a WNP is slower than when x_t is stationary. For $I(1/2)$ processes this convergence rate is reduced by the slowly varying factor $L(n)^{1/2}$.

(b) Let $\tilde{\sigma}_u^2$ denote a consistent estimator of σ_u^2 , and consider the nonparametric t statistic for the hypothesis $\mathcal{H}_0 : m(x) = m_0(x)$ based on the local linear estimator, as given by

$$\tilde{t}(x; m_0) := \left(\frac{\sum_{t=2}^n K_{th}(x)}{\tilde{\sigma}_u^2 Q_{11}} \right)^{1/2} [\tilde{m}(x) - m_0(x)].$$

It follows directly from Theorem 4.2 that $\tilde{t}(x; m_0) \xrightarrow{d} N[0, 1]$, and similarly when (\tilde{m}, Q_{11}) is replaced by $(\hat{m}, \int K^2)$.

Thus in conjunction with the existing literature, Theorem 4.2 implies that kernel nonparametric t statistics are asymptotically standard Gaussian across a wide range of regressor processes, including: stationary fractional (with $-1/2 < d < 1/2$; Wu and Mielniczuk, 2002), weakly nonstationary (fractional with $d = 1/2$ or mildly integrated), nonstationary fractional ($1/2 < d < 3/2$) and (nearly) integrated processes (Wang and Phillips, 2009a,b, 2011, 2012). This is in marked contrast to parametric t statistics, which when regressors are nonstationary have limiting distributions that are typically nonstandard and dependent on nuisance parameters relating to the persistence of the regressor, which cannot be consistently estimated – a fact that greatly complicates parametric inference in these models (for an overview of this problem and the relevant literature, see Phillips and Lee, 2013, pp. 251–254).

(c) Suppose that x_{t-1} on the r.h.s. of (20) is replaced by x_t , so that the regressor is no longer predetermined. If x_t is stationary, then the correlation between it and u_t prevents m from being consistently estimated. However, for the case where x_t is NI, Wang and Phillips (2009b) show that the nonparametric regression estimator is consistent for m and asymptotically mixed Gaussian, even when x_t is correlated with u_t , and u_t is serially dependent. In other words, for NI covariates the asymptotics of the nonparametric regression estimator are unaffected by whether x_t or x_{t-1} appears in (20). We conjecture that a similar results also holds for WNPs, but leave an examination of this for future work.

(d) Our smoothness assumptions on m could be relaxed along the lines of Wang and Phillips (2009a,b) and Wang and Phillips (2011), for the NW and LL estimators respectively; we have refrained from doing so here to permit Theorem 4.2 to be more concisely stated.

5. Specification testing when a regressor has an unknown degree of persistence. In this section, we exploit the asymptotic normality of the nonparametric t statistic to develop a specification test statistic for parametric regression models that has the same asymptotic distribution regardless of the extent of the persistence of the regressor, and indeed regardless of whether that persistence is modelled in terms of long memory (i.e. as $I(d)$ for some $d \in (-1/2, 3/2)$) or in terms of an autoregressive root localised to unity (as in an MI or NI process). The proposed test can thus be validly conducted, in a straightforward manner, without requiring practitioners either to make an assumption on the persistence of the regressor, or to somehow estimate this and take account of it when carrying out the test.

The hypothesis to be tested is that the true regression function m in (20) belongs to a certain parametric family, as e.g. postulated in (17). Formally, the null is

$$(23) \quad \mathcal{H}_0 : m(x) = \mu + \gamma g(x), \text{ for some } (\mu, \gamma) \in \mathbb{R}^2 \text{ and all } x \in \mathbb{R};$$

where g is a known function; the alternative is that no such μ and γ exist. Tests of \mathcal{H}_0 , in a setting with (possibly) nonstationary regressors, have also been considered by Gao, King, Lu and Tjøstheim (2009), Wang and Phillips (2012; hereafter ‘WP’), and Dong, Gao, Tjøstheim and Yin (2017). WP test a parametric fit in the presence of a NI regressor, while Dong et al (2017) test for a parametric fit in regressions with a $d = 0$ and a $d = 1$ covariate. The test statistic of WP closely resembles that of Gao et al (2009), who propose a studentised U-statistic formed of kernel-weighted OLS regression

residuals.⁶

We propose to test \mathcal{H}_0 by comparing parametric OLS and kernel non-parametric estimates of m . The model specified in (23) can be estimated parametrically by OLS regression, and also nonparametrically at each x as

$$\tilde{m}_g(x) := \operatorname{argmin}_{a \in \mathbb{R}} \min_{b \in \mathbb{R}} \sum_{t=1}^n \{y_t - a - b[g(x_{t-1}) - g(x)]\}^2 K_{th}(x)$$

which under \mathcal{H}_0 has no asymptotic bias, even if h remains fixed as $n \rightarrow \infty$. If $g(x) = x$, so that the null of linearity is being tested, $\tilde{m}_g(x)$ specialises to the local linear regression estimator; but in general \tilde{m}_g should be chosen consistent with the model under test, so that it has no bias under the null. Provided that $g^{(1)}(x) \neq 0$, a slight modification of the proof of Theorem 4.2 shows that $\tilde{m}_g(x)$ has the same limiting distribution as displayed in (22), under \mathcal{H}_0 . Letting $(\hat{\mu}, \hat{\gamma})$ denote the OLS estimates of (μ, γ) , we can therefore compare the fit provided by the parametric and local nonparametric estimates of the model via an ensemble of t statistics of the form

$$\tilde{t}(x; \hat{\mu}, \hat{\gamma}) := \left[\frac{\sum_{t=2}^n K_{th}(x)}{\tilde{\sigma}_u^2(x) Q_{11}} \right]^{1/2} [\tilde{m}_g(x) - \hat{\mu} - \hat{\gamma}g(x)],$$

where $\tilde{\sigma}_u^2(x) := [\sum_{t=2}^n K_{th}(x)]^{-1} \sum_{t=2}^n [y_t - \hat{\mu} - \hat{\gamma}g(x_{t-1})]^2 K_{th}(x)$. Under \mathcal{H}_0 , both the parametric and nonparametric estimators converge to identical limits and so for each $x \in \mathbb{R}$,

$$\tilde{t}(x; \hat{\mu}, \hat{\gamma}) = \tilde{t}(x; \mu, \gamma) + o_p(1) \xrightarrow{d} N[0, 1],$$

where the equality is due to the relatively faster convergence rate of the parametric estimator, and the distributional limit follows as per Remark 4.2(b). Under the alternative, only the nonparametric estimator is consistent for m , and thus $|\tilde{t}| \xrightarrow{p} \infty$.

Our proposed specification test statistic is based on these t statistics evaluated at a set of p points $\mathcal{X} \subset \mathbb{R}$, constructed as per

$$(24) \quad \tilde{F} := \sum_{x \in \mathcal{X}} \tilde{t}(x; \hat{\mu}, \hat{\gamma})^2.$$

⁶Gao et al (2009) apply their statistic to the problem of testing the null of a random walk (of the form $x_t = x_{t-1} + \xi_t$), against a (possibly nonlinear) stationary alternative. The underlying idea is to test for a neglected nonlinear component in an autoregression, whose presence would make the process stationary. The specification test proposed in this paper could also be potentially used for this purpose, but we leave explorations in this direction for future work.

\tilde{F} is related to the ‘non-predictability sum test’ developed by Kasparis, Andreou and Phillips (2015), who were concerned with testing the null that x_{t-1} cannot predict y_t , which in the present framework can be expressed as $m(x) = \mu$ for all $x \in \mathbb{R}$. Relative to other specification tests available in the literature, the principal advantage of a test based on \tilde{F} is that the limiting distribution of this statistic is invariant to the extent of persistence in the regressor, making valid inference in the presence of data with an unknown degree of persistence straightforward.⁷

5.1. *Asymptotics for WNP.s.* Our final result gives the limiting distribution of \tilde{F} under \mathcal{H}_0 , and under a sequence of local alternatives of the form

$$(25) \quad \mathcal{H}_1 : m(x) = \mu + \gamma g(x) + r_n g_1(x)$$

for some g_1 , where $r_n \rightarrow 0$. This formulation of the alternative is similar to that of Horowitz and Spokoiny (2001) and WP. We only provide explicit results for the boundary case where $\{x_t\}$ is a WNP, which is the main focus of the present work. However, as discussed in Section 5.2 below, analogous results may be derived for the stationary fractional ($-1/2 < d < 1/2$), nonstationary fractional ($1/2 < d < 3/2$) and nearly integrated cases, on the basis of the limit theory presented in Wu and Mielniczuk (2002) and Wang and Phillips (2009a,b).

For the purposes of the next result, assume $g_1(x)$ is either integrable or an asymptotically homogeneous function (AHF) of asymptotic order κ_{g_1} ; this helps to characterise the limiting behaviour of the test statistic under \mathcal{H}_1 . Define $\mu_* := -[\int H_{g^2} \varrho \int H_{g_1} \varrho - \int H_g \varrho \int H_g H_{g_1} \varrho] / [\int H_{g^2} \varrho - (\int H_g \varrho)^2]$, which is the distributional limit of $-\kappa_{g_1}(\beta_n) r_n^{-1}(\hat{\mu} - \mu)$ when g_1 is AHF.

THEOREM 5.1. *Suppose that \mathcal{X} has p elements and*

- (i) *conditions (i)-(iii) of Theorem 4.2 hold, $x^j[K(x) + K^2(x)]$ are bounded and integrable for $j \in [0, 4]$, $\sup_t \mathbf{E}[u_t^4 | \mathcal{F}_{t-1}] < \infty$ a.s., and $nh_n^5/\beta_n \rightarrow 0$; and*
- (ii) *g is AHF for the array $\{x_t(n)\}$, with limit homogeneous component H_g such that $\kappa_g(\beta_n) \rightarrow \infty$ and H_g^2 satisfies the requirements of Theorem 3.2. Further, g has bounded second derivative, and $g^{(1)}(x) \neq 0$ for each $x \in \mathcal{X}$.*

⁷The use of kernel methods in specification testing does not in and of itself lead to conventional inference: for example, a test recently proposed by Dong et al (2017) is also based on nonparametric methods, but the limiting distribution of their test statistic, and therefore the critical values for their test, depends on precise assumptions as to the form and extent of the persistence of the regressor.

Then under \mathcal{H}_0 ,

$$(26) \quad \tilde{F} \xrightarrow{d} \chi_p^2$$

Suppose that in addition $p \geq 2$ and:

(iii) $r_n \rightarrow 0$ and either

- (a) $g_1, g \cdot g_1$ are bounded and integrable, $r_n^{-1}n^{-1/2}\beta_n \rightarrow 0$; or
- (b) g_1 is AHF for the array $\{x_t(n)\}$, with limit homogeneous function H_{g_1} such that $H_{g_1}H_g$ satisfies the requirements of Theorem 3.2, and $r_n^{-1}n^{-1/2}\kappa_{g_1}^{-1}(\beta_n) \rightarrow 0$. Further, each of the following limits exist (allowing ‘convergence’ to ∞):

$$\kappa_* := \lim_{n \rightarrow \infty} \kappa_{g_1}(\beta_n) \quad \kappa_{**} := \lim_{n \rightarrow \infty} \kappa_{g_1}(\beta_n)r_n.$$

(iv) $l_n n h_n / \beta_n \rightarrow \infty$ where

$$l_n = \begin{cases} r_n^2 & \text{under (iii.a) or (iii.b) with } \kappa_* \in [0, \infty) \\ r_n^2 \kappa_{g_1}^2(\beta_n) & \text{under (iii.b), } \kappa_{**} = 0 \text{ and } \kappa_* = \infty \\ 1 & \text{under (iii.b) and } \kappa_{**} \in (0, \infty]. \end{cases}$$

(v) g_1 has bounded second derivative.

(vi) Either: (iii.a) holds and $g_1(x) \neq 0$ for some $x \in \mathcal{X}$; (iii.b) holds with $\kappa_* \in (0, \infty]$ and $g_1(x) \neq g_1(x')$ for some $x, x' \in \mathcal{X}$; or (iii.b) holds with $\kappa_* = \infty$ and $\mu_* \neq 0$ a.s.

Then under \mathcal{H}_1 ,

$$(27) \quad \tilde{F} \xrightarrow{p} \infty.$$

REMARK 5.1. (a) The requirement $\kappa_g(\beta_n) \rightarrow \infty$ is a technical condition that is satisfied in most specifications employed in empirical work, e.g. linear models. It can be relaxed at the cost of a more involved exposition.

(b) If (iii.b) holds with $\kappa_* = \infty$, then $\mu_* \neq 0$ a.s. is sufficient for the test to be consistent, in the sense that (27) holds. Consistency may still obtain when $\mu_* = 0$, but that requires a more detailed analysis than we are able to provide here.

(c) The sequence $l_n n h_n / \beta_n$ gives the divergence rate of the test statistic under \mathcal{H}_1 , which is closely related the power of the test. The maximal divergence rate (i.e. $n h_n / \beta_n$) is attained when g_1 is AHF of diverging asymptotic order ($\kappa_{g_1}(\beta_n) \rightarrow \infty$) and $\kappa_{g_1}(\beta_n)r_n \rightarrow \infty$. In such cases, the divergence rate is otherwise unaffected by r_n . When $\kappa_{g_1}(\beta_n) \rightarrow \infty$ but $\kappa_{g_1}(\beta_n)r_n \rightarrow 0$, the

divergence rate reduces to $\kappa_{g_1}^2(\beta_n)r_n^2nh_n/\beta_n$. The divergence rate is smallest (i.e. $r_n^2nh_n/\beta_n$) in cases where g_1 is integrable or AHF of vanishing asymptotic order (i.e. $\kappa_{g_1}(\beta_n) \rightarrow 0$).

(d) We have assumed that the set \mathcal{X} comprises a fixed number (p) of points. We can get some idea of the large-sample distribution of \tilde{F} if p is allowed to grow with the sample size from the fact that $(2p)^{-1/2}(\tilde{F} - p) \xrightarrow{d} (2p)^{-1/2}(\chi_p^2 - p) \xrightarrow{d} N(0, 1)$ as $n \rightarrow \infty$ and then $p \rightarrow \infty$.

5.2. *Asymptotics in other cases.* \tilde{F} will have the same limiting distribution as given by (26) of Theorem 5.1, even when x_t is a stationary, nonstationary fractional, or near-integrated process. This can be established using results available in the existing literature, by arguments outlined in the remainder of this section, and is confirmed by the simulation exercises presented in Section 6 below. (These also permit our results on the consistency of the test against local alternatives to be extended beyond weakly nonstationary processes.)

From the proof of Theorem 5.1, it is clear that (26) holds if

- (i) the OLS estimator for (μ, γ) converges faster than the nonparametric estimator for m (e.g. NW/LL);
- (ii) the nonparametric estimator is asymptotically (mixed) Gaussian; and
- (iii) for every $x, x' \in \mathcal{X}$, such that $x \neq x'$,

$$(28) \quad \frac{\beta_n}{nh_n} \sum_{t=1}^n K_j[(x_t - x)/h_n] K_{j'}[(x_t - x')/h_n] = o_p(1),$$

for all $j, j' \in \{0, 1, 2\}$, where $K_j(x) := x^j K(x)$.

This last requirement ensures that the component t statistics in (24) are asymptotically independent of each other.

When x_t is nonstationary fractional ($1/2 < d < 3/2$) or nearly integrated, the arguments of Park and Phillips (2001) (see also Christopeit, 2009 and the references therein) show that the convergence rates of $(\hat{\mu}, \hat{\gamma})$ are as in (19); that the convergence rates of the kernel nonparametric regression estimators are slower follows from Wang and Phillips (2009a,b). Conditions (ii) and (iii) also follow from Wang and Phillips (2009a, and particularly pp. 1910–11 of 2009b for (28)).

When x_t is stationary, it is well known that the OLS estimator is $n^{-1/2}$ -consistent while the $(nh_n)^{-1/2}$ -consistency and asymptotic normality of the nonparametric estimators follows e.g. from the results of Wu and Mielniczuk (2002). Thus conditions (i)–(ii) hold. Finally, if x_t has a bounded density

TABLE 1
Size: maximum rejection frequency over $\rho \in \{-0.5, 0, 0.5\}$; $\alpha = 0.1$

d	n	WP (2012)			$p = 17$			$p = 25$		
$h = n^b, b =$		-0.2	-0.1	-0.05	-0.2	-0.1	-0.05	-0.2	-0.1	-0.05
0.25	100	0.04	0.02	0.01	0.07	0.05	0.04	0.09	0.06	0.04
	200	0.05	0.02	0.01	0.07	0.06	0.04	0.09	0.07	0.05
	500	0.06	0.02	0.01	0.07	0.06	0.04	0.09	0.07	0.05
0.50	100	0.05	0.02	0.01	0.07	0.06	0.04	0.09	0.07	0.05
	200	0.06	0.03	0.02	0.08	0.07	0.05	0.11	0.09	0.07
	500	0.07	0.04	0.03	0.07	0.07	0.05	0.10	0.09	0.07
0.75	100	0.06	0.04	0.03	0.08	0.07	0.06	0.11	0.09	0.07
	200	0.08	0.06	0.04	0.08	0.08	0.07	0.10	0.10	0.09
	500	0.08	0.07	0.05	0.07	0.08	0.07	0.09	0.09	0.09
1.00	100	0.08	0.06	0.05	0.09	0.09	0.08	0.13	0.11	0.11
	200	0.09	0.07	0.06	0.08	0.10	0.10	0.11	0.12	0.12
	500	0.09	0.08	0.08	0.09	0.09	0.09	0.10	0.11	0.11

$\mathcal{D}_x(u)$ (see Wu and Mielniczuk, 2002, Lemma 1) and under standard regularity conditions on K , we have

$$\begin{aligned} \frac{1}{nh_n} \sum_{t=1}^n \mathbf{E} |K_j[(x_t - x)/h_n] K_{j'}[(x_t - x')/h_n]| \\ \leq \sup_u \mathcal{D}_x(u) \int_{\mathbb{R}} |K_j[z] K_{j'}[z + (x - x')/h_n]| dz \rightarrow 0, \end{aligned}$$

so that condition (iii) holds, since $\beta_n \asymp 1$ in this case.

6. Simulations. We conducted simulations to evaluate the finite-sample performance of the proposed specification test, in terms of size and power against a range of alternatives. For this exercise, a natural comparison is with the specification test of WP, which is known to have a standard Gaussian limiting distribution when x_t is NI. (In our simulation exercises, we assume that this also holds when x_t is fractionally integrated and/or stationary, and so compare their statistic to normal critical values in these cases.)

TABLE 2
Size-adjusted power when $d \in \{0.5, 1.0\}$; $\alpha = 0.1$

n		WP (2012)			$p = 17$			$p = 25$		
$h = n^b, b =$		-0.2	-0.1	-0.05	-0.2	-0.1	-0.05	-0.2	-0.1	-0.05
$d = 0.5$		Size adj. power			Size adjusted, relative to WP					
$\varphi_1(x)$	100	0.08	0.06	0.04	0.08	0.09	0.09	0.11	0.11	0.11
	200	0.14	0.13	0.11	0.11	0.14	0.14	0.15	0.17	0.17
	500	0.37	0.42	0.41	0.15	0.18	0.19	0.20	0.22	0.22
$\varphi_1(2x)$	100	0.07	0.04	0.03	0.07	0.08	0.06	0.10	0.09	0.08
	200	0.12	0.09	0.07	0.09	0.10	0.09	0.13	0.13	0.12
	500	0.29	0.27	0.23	0.13	0.15	0.14	0.19	0.19	0.18
$ x ^{-2} \wedge 1$ ($\times 0.5$)	100	0.06	0.03	0.02	0.11	0.12	0.12	0.15	0.15	0.14
	200	0.10	0.06	0.04	0.16	0.19	0.18	0.21	0.23	0.22
	500	0.19	0.16	0.12	0.12	0.14	0.14	0.16	0.16	0.16
$ x ^{-1} \wedge 1$ ($\times 0.5$)	100	0.06	0.03	0.02	0.07	0.09	0.08	0.10	0.11	0.10
	200	0.07	0.05	0.04	0.10	0.14	0.14	0.15	0.17	0.17
	500	0.14	0.13	0.11	0.14	0.18	0.19	0.19	0.21	0.22
$ x ^{1.5}$ ($\times 0.02$)	100	0.11	0.09	0.07	0.03	0.04	0.04	0.05	0.06	0.05
	200	0.22	0.23	0.22	0.04	0.05	0.06	0.06	0.07	0.07
	500	0.61	0.66	0.66	0.04	0.09	0.11	0.08	0.12	0.13
x^2 ($\times 0.02$)	100	0.07	0.05	0.04	0.06	0.07	0.07	0.09	0.09	0.09
	200	0.13	0.12	0.11	0.08	0.12	0.13	0.12	0.15	0.16
	500	0.37	0.42	0.43	0.07	0.12	0.13	0.12	0.15	0.15
$d = 1.0$		Size adj. power			Size adjusted, relative to WP					
$\varphi_1(x)$	100	0.08	0.07	0.06	0.04	0.06	0.06	0.08	0.09	0.09
	200	0.10	0.09	0.08	0.02	0.05	0.07	0.06	0.09	0.10
	500	0.12	0.12	0.12	0.02	0.05	0.05	0.05	0.07	0.09
$\varphi_1(2x)$	100	0.08	0.06	0.05	0.02	0.04	0.05	0.07	0.08	0.07
	200	0.09	0.08	0.07	0.01	0.03	0.04	0.05	0.06	0.07
	500	0.10	0.09	0.09	0.02	0.02	0.03	0.03	0.04	0.05
$ x ^{-2} \wedge 1$ ($\times 0.5$)	100	0.08	0.06	0.05	0.05	0.08	0.08	0.10	0.12	0.12
	200	0.09	0.07	0.07	0.04	0.08	0.10	0.09	0.12	0.14
	500	0.09	0.09	0.08	0.02	0.06	0.08	0.05	0.09	0.11
$ x ^{-1} \wedge 1$ ($\times 0.5$)	100	0.08	0.06	0.05	0.04	0.07	0.07	0.09	0.10	0.10
	200	0.08	0.07	0.07	0.03	0.07	0.08	0.08	0.11	0.12
	500	0.11	0.11	0.11	0.02	0.06	0.08	0.06	0.09	0.12
$ x ^{1.5}$ ($\times 0.02$)	100	0.10	0.09	0.08	0.04	0.07	0.08	0.09	0.10	0.11
	200	0.13	0.13	0.13	-0.02	0.04	0.07	0.05	0.08	0.11
	500	0.19	0.21	0.22	-0.10	-0.01	0.00	-0.04	0.01	0.02
x^2 ($\times 0.02$)	100	0.09	0.07	0.06	0.04	0.06	0.08	0.07	0.09	0.09
	200	0.11	0.11	0.11	0.00	0.01	0.01	0.01	0.01	0.01
	500	0.16	0.19	0.19	0.00	0.00	0.00	0.00	0.00	0.00

For all simulation exercises, the null hypothesis is $\mathcal{H}_0 : m(x) = \mu + \beta x$, so that the proposed specification test can be implemented by comparing the fit of a local linear regression with an OLS regression. The data generating process is

$$y_t = x_{t-1} + g_1(x_{t-1}) + u_t \quad (1 - L)^d x_t = \xi_t + 0.5\xi_{t-1}$$

with $x_0 = 0$, where (u_t, ξ_t) are i.i.d. bivariate Gaussian with unit variances and correlation ρ . For each value of $d \in \{0.25, 0.50, 0.75, 1.00\}$, we evaluate the size of the test by computing the maximum rejection frequency under the null (i.e. when $g_1(x) = 0$) for $\rho \in \{-0.5, 0.0, +0.5\}$ (with 5000 replications). We consider sample sizes $n \in \{100, 200, 500\}$, bandwidths of the form $h = n^b$ for $b \in \{-0.2, -0.1, -0.05\}$, and use the Gaussian kernel in all cases. For our test, which requires the choice of points at which to compare the nonparametric and parametric estimates of the regression function, we consider two choices: $p = 17$ or 25 points, evaluated at the quantiles of $\{x_t\}_{t=1}^n$ equally spaced between the 0.1 and 0.9 quantiles.

The results are displayed in Table 1, for a test having 10 per cent nominal significance level. They clearly illustrate that our test has good size control across the range of bandwidths considered, both when the data is in the stationary and nonstationary regions, and on the boundary between these, suggesting that the asymptotics developed in Sections 3–5 provide a good approximation to the finite-sample distribution of the statistic.

We also computed the size-adjusted power of our test, and of WP's, against alternatives that are either integrable ($g_1(x) = \varphi_1(x)$, $\varphi_1(2x)$, or $x^{-2} \wedge 1$), non-integrable but vanishing at infinity ($g_1(x) = x^{-1} \wedge 1$), or polynomials ($g_1(x) = |x|^v$ with $v \in \{1.5, 2.0\}$).⁸ The simulation designs are the same as for the size calculations, except that we here only report results for $\rho = 0$ and $d \in \{0.5, 1.0\}$ (results for $d \in \{0.25, 0.75\}$ and MI processes are broadly similar, and are provided in Appendix F in the Supplementary Material). The alternatives are scaled by the factors indicated in Table 2 so as to ensure non-trivial power for these designs. To facilitate the comparison between the power of our test and that of WP, we report the size-adjusted power of their procedure in the first three columns of Table 2, and the *relative* size-adjusted power of our test alongside, i.e. the difference between the power of our test and of theirs.

It is noticeable that our test generally outperforms WP's: indeed, the

⁸By size-adjusted power, we mean that if the test is found to reject at rate $\hat{\alpha} > 0.1$ under the null (as reported in the top panel), then the power of the test is adjusted downwards by subtracting $\hat{\alpha} - 0.1$ from the rejection rate under each alternative (so if $\hat{\alpha} \leq 0.1$, no adjustment is made).

relevant entries of the table are almost uniformly positive, with the exception of a few cases where $g_1(x) = |x|^{1.5}$ and $d = 1$. The most pronounced power improvements are for those cases where d is smaller, and the alternatives are either integrable or asymptotically vanishing (i.e. transformations that exhibit weaker signal and are therefore harder to detect); whereas the performance of the two tests is less easily distinguishable for polynomially growing alternatives, when $d = 1$.

References.

- [1] Abadir, K.M., Distaso, W., Giraitis, L., and Koul, H.L. (2014). Asymptotic normality for weighted sums of linear processes. *Econometric Theory*, **30**(1), 252–284.
- [2] Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2001) The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, **96**, 42–55.
- [3] Andrews, D. W. K., Cheng, X. and Guggenberger, P. (2020) Generic results for establishing the asymptotic size of confidence sets and tests. *Journal of Econometrics* in press, DOI: 10.1016/j.jeconom.2020.04.027.
- [4] Baillie, R. T., Chung, C.-F. and Tieslau, M. T. (1996) Analysing inflation by the fractionally integrated ARFIMA–GARCH model. *Journal of Applied Econometrics*, **11**, 23–40.
- [5] Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987). *Regular Variation*. Cambridge University Press.
- [6] Buchmann, B. and Chan, N.H. (2007). Asymptotic theory of least squares estimators for nearly unstable processes under strong dependence. *Annals of Statistics*, **35**(5), 2001–2017.
- [7] Chan, N. and Wang, Q. (2015). Nonlinear regressions with nonstationary time series. *Journal of Econometrics*, **185**, 182–195.
- [8] Chan, N.H. and Wei, C.Z. (1987). Asymptotic inference for nearly nonstationary AR(1) processes. *Annals of Statistics*, **15**(3), 1050–1063.
- [9] Chan, N.H. and Wei, C.Z. (1988). Limiting distributions of least squares estimates of unstable autoregressive processes. *Annals of Statistics*, **16**(1), 367–401.
- [10] Christensen, B.J. and Nielsen M.Ø. (2006). Asymptotic normality of narrow-band least squares in the stationary fractional cointegration model and volatility forecasting. *Journal of Econometrics*, **133**(1), 343–371.
- [11] Christopeit, N. (2009). Weak convergence of nonlinear transformations of integrated processes: The multivariate case. *Econometric Theory*, **25**(5), 1180–1207.
- [12] Demetrescu, M., Georgiev, I., Rodrigues, P. and Taylor, R. (2019). Testing for Episodic Predictability in Stock Returns. *Journal of Econometrics*, in press.
- [13] Dong, C., Gao, J., Tjøstheim, D. and Yin, J. (2017). Specification testing for nonlinear multivariate cointegrating regressions. *Journal of Econometrics*, **200**(1), 104–117.
- [14] Duffy, J.A. (2020). Asymptotic theory for kernel estimators under moderate deviations from a unit root'. *Econometric Theory*, **36**, pp. 559–582.
- [15] Gao, J., King, M., Lu, Z. and Tjøstheim, D. (2009). Specification testing in nonlinear and nonstationary time series autoregression. *Annals of Statistics*, **37**(6B), 3893–3928.
- [16] Gikhman, I.I. and Skorokhod, A.V. (1969). *Introduction to the theory of random processes*. Saunders, Philadelphia.
- [17] Giraitis, L. and Phillips, P.C.B. (2006). Uniform limit theory for stationary autoregression. *Journal of Time Series Analysis*, **27**(1), 51–60.

- [18] Hall, P. and Heyde, C.C. (1980). *Martingale limit theory and its application*. Academic Press, New York.
- [19] Hassler, U. and Pohle, J.-M. (2019). Forecasting under long memory and nonstationarity. Working paper, Goethe University Frankfurt.
- [20] Hassler, U. and Wolters, J. (1995). Long memory in inflation rates: international evidence. *Journal of Business and Economic Statistics*, **13**(1), 37–45.
- [21] Horowitz, J.L. and Spokoiny, V.G. (2001). An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica*, **69**(3), 599–631.
- [22] Hualde, J. and Robinson, P.M. (2010). Semiparametric inference in multivariate fractionally cointegrated systems. *Journal of Econometrics*, **157**(2), 492–511.
- [23] Hualde, J. and Robinson, P.M. (2011). Gaussian pseudo-maximum likelihood estimation of fractional time series models. *Annals of Statistics*, **39**(6), 3152–3181.
- [24] Jeganathan, P. (2004). Convergence of functionals of sums of r.v.s to local times of fractional stable motions. *Annals of Probability*, **32**(3), 1771–1795.
- [25] Jeganathan, P. (2008). Limit theorems for functionals of sums that converge to fractional Brownian and stable motions. *Cowles Foundation Discussion Paper 1949*.
- [26] Johansen, S. and Nielsen, M.Ø. (2012a). Likelihood inference for a fractionally cointegrated vector autoregressive model. *Econometrica*, **80**(6), 2667–2732.
- [27] Johansen, S. and Nielsen, M.Ø. (2012b). A necessary moment condition for the fractional functional central limit theorem. *Econometric Theory*, **28**(3), 671–679.
- [28] Kasahara, Y. and Maejima, M. (1988). Weighted sums of i.i.d. random variables attracted to integrals of stable processes. *Probability Theory Related Fields*, **78**, 75–96.
- [29] Karlsen, A. and Tjøstheim, D. (2001). Nonparametric estimation in null recurrent time series. *Annals of Statistics*, **29**(2), 372–416.
- [30] Karlsen, A. Myklebust, T. and Tjøstheim, D. (2007). Nonparametric estimation in a nonlinear cointegration type model. *Annals of Statistics*, **35**(1), 252–299.
- [31] Kasparsis, I., Andreou, E., and Phillips, P.C.B. (2015) Nonparametric predictive regression. *Journal of Econometrics*, **185**(2), 468–494.
- [32] Kostakis, A., Stamatogiannis M.P. and Magdalinos, T. (2015). Robust econometric inference for stock return predictability. *Review of Financial Studies*, **28**(5), 1506–1553.
- [33] Loève, M. (1978) *Probability Theory*, Vol. II, 4th ed. Springer-Verlag, New York.
- [34] Magdalinos, T. and Phillips, P.C.B. (2011). Econometric inference in the vicinity of unity. *Mimeo*, Singapore Management University.
- [35] Marinucci, D. and Robinson, P.M. (1998). Semiparametric frequency domain analysis of fractional cointegration. *Mimeo*, London School of Economics.
- [36] Marinucci, D. and Robinson, P.M. (1999). Alternative forms of fractional Brownian motion. *Journal of Statistical Planning and Inference*, **80**(1–2), 111–122.
- [37] Mikusheva, A. (2007). Uniform inference in autoregressive models. *Econometrica*, **75**(5), 1411–1452.
- [38] Park, J.Y. and Phillips P.C.B. (1999). Asymptotics for nonlinear transformations of integrated time series. *Econometric Theory*, **15**(3), 269–298.
- [39] Park, J.Y. and Phillips P.C.B. (2001). Nonlinear Regressions with Integrated Time Series, *Econometrica*, **69**(1), 117–161.
- [40] Phillips, P.C.B. (1987). Time series regression with a unit root. *Econometrica*, **55**(2), 277–301.
- [41] Phillips, P.C.B. (1987). Towards a unified asymptotic theory for autoregression. *Biometrika*, **74**(3), 535–547.

- [42] Phillips, P.C.B. (1995). Fully modified least squares and vector autoregression. *Econometrica*, **63**(5) 1023-1078.
- [43] Phillips P.C.B. and Lee, J.H. (2013). Predictive regression under various degrees of persistence and robust long-horizon regression. *Journal of Econometrics*, **177**, 250-264.
- [44] Phillips, P.C.B. and Magdalinos, T. (2007). Limit theory for moderate deviations from a unit root. *Journal of Econometrics*, **136**(1), 115-130.
- [45] Phillips, P.C.B. and Shimotsu, K. (2004). Local Whittle estimation in nonstationary and unit root cases. *Annals of Statistics*, **32**(2), 656-692.
- [46] Pötscher, B.M. (2004). Nonlinear functions and convergence to Brownian motion: Beyond the continuous mapping theorem. *Econometric Theory*, **20**(1), 1-22.
- [47] Robinson, P.M. and Hualde, J. (2003). Cointegration in fractional systems with unknown integration orders. *Econometrica*, **71**(6), 1727-1766.
- [48] Shimotsu, K. and Phillips, P.C.B. (2005). Exact local Whittle estimation of fractional integration. *Annals of Statistics*, **33**(4), 1890-1933.
- [49] Taqqu, M.S. (1975). Weak convergence to fractional Brownian motion and to the Rosenblatt process, *Probability Theory and Related Fields*, **31**(4), 287-302.
- [50] Wang, Q. (2014). Martingale limit theorem revisited and nonlinear cointegrating regression. *Econometric Theory*, **30**(3), 509-535.
- [51] Wang, Q. and Phillips P.C.B. (2009a). Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory*, **25**(3), 710-738.
- [52] Wang, Q. and Phillips P.C.B. (2009b). Structural nonparametric cointegrating regression. *Econometrica*, **77**(6), 1901-1948.
- [53] Wang, Q. and Phillips P.C.B. (2011). Asymptotic theory for zero energy functionals with nonparametric regression applications. *Econometric Theory*, **27**(2), 235-259.
- [54] Wang, Q. and Phillips P.C.B. (2012). A specification test for nonlinear nonstationary models. *Annals of Statistics*, **40**(2), 727-758.
- [55] Wu, W. B., Huang, Y. and Huang Y. (2010) Kernel estimation for time series: an asymptotic theory. *Stochastic Processes and their Applications*, **120**, 2412-2431.
- [56] Wu, W. B. and Mielniczuk, J. (2002). Kernel density estimation for linear processes. *Annals of Statistics*, **30**(5), 1441-1459.
- [57] Yang, B., Long, W., Peng, L. and Cai, Z. (2019). Testing the Predictability of U.S. Housing Price Index Returns Based on an IVX-AR Model, *Journal of the American Statistical Association*, in press.

**SUPPLEMENT TO: ESTIMATION AND INFERENCE IN
THE PRESENCE OF FRACTIONAL $d = 1/2$ AND
WEAKLY NONSTATIONARY PROCESSES**

CONTENTS

A	Proofs under high-level conditions	32
A.1	Technical lemmas	32
A.2	Proofs of Theorems 3.1 and 3.3	32
A.3	Proofs of Lemmas A.1–A.3	35
B	Proofs under low-level conditions	37
B.1	Sufficient conditions for Assumption HL	37
B.2	Lemmas for $I(1/2)$ and MI processes	43
B.3	Proofs of Theorems 3.2 and 3.4	44
C	Proofs auxiliary to Appendix B	51
C.1	Proofs of Lemmas COEF, LVAR and CLT	51
C.2	Proofs of Lemmas B.1 and B.2	58
D	Proofs of Theorems 4.1 and 4.2	62
E	Proof of Theorem 5.1	67
F	Additional simulations for Section 6	77

Throughout, $C, C_1, C_2, \dots \in (0, \infty)$ denote generic constants which may take different values at each appearance, even within the same proof. $\xrightarrow{a.s.}$, \xrightarrow{p} and \xrightarrow{d} respectively denote convergence almost surely, in probability, and in distribution. For deterministic sequences $\{a_n\}$ and $\{b_n\}$, $a_n \sim b_n$ denotes $\lim_{n \rightarrow \infty} a_n/b_n = 1$ and $a_n \asymp b_n$ denotes $\lim_{n \rightarrow \infty} |a_n/b_n| \in (0, \infty)$. For random variables X and Y , $X \sim F$ denotes that X has distribution F , and $X \stackrel{d}{=} Y$ that X has the same distribution as Y . For a positive real number x , $[x]$ denotes its integer part. $\mathbf{1}\{A\}$ denotes the indicator function for the set A . $\overline{\mathbb{R}}$, \mathbb{R}_+ and \mathbb{R}_+^* are the extended, the nonnegative, and (strictly) positive real numbers respectively. $f^{(j)}(x)$ denotes the j th derivative of the function $f(x)$. All limits are taken as $n \rightarrow \infty$ unless otherwise indicated.

References to Bingham, Goldie and Teugels (1987) are henceforth abbreviated to ‘BGT’.

APPENDIX A: PROOFS UNDER HIGH-LEVEL CONDITIONS

This appendix provides proofs of our main results under high level conditions (Assumption **HL**), i.e. Theorems 3.1 and 3.3. We begin by stating some auxiliary technical lemmas, whose proofs appear in Appendix A.3.

A.1. Technical lemmas.

LEMMA A.1. *Suppose $f : \mathbb{R} \rightarrow \mathbb{R}$ is locally integrable, and let $\varepsilon, \eta > 0$. Then there is a Lipschitz continuous $f_{\varepsilon, \eta}$ such that $\int_{|x| \leq \eta} |f(x) - f_{\varepsilon, \eta}(x)| dx < \varepsilon$ and $f_{\varepsilon, \eta}(x) = 0$ for $|x| > \eta$.*

LEMMA A.2. *Let $\{X_n\}$ and $\{Y_n\}$ be real valued random sequences on some probability space $(\Omega, \mathcal{A}, \mathbf{P})$ and $\mathcal{F} \subset \mathcal{A}$ a σ -field, for which*

- (i) $X_n \xrightarrow{d} X \sim F_X$, conditionally on \mathcal{F} , in the sense that $\mathbf{E}(h(X_n) | \mathcal{F}) \xrightarrow{P} \int_{\mathbb{R}} h(x) dF_X(x)$ for all $h : \mathbb{R} \rightarrow \mathbb{R}$ bounded and continuous; and
- (ii) $Y_n \xrightarrow{d} Y$, where Y_n is \mathcal{F} -measurable for each n .

Then for all $g : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ bounded and Lipschitz continuous,

$$\mathbf{E}(g(X_n, Y_n) | \mathcal{F}) \xrightarrow{d} \int_{\mathbb{R}} g(x, Y) dF_X(x).$$

LEMMA A.3. *Suppose that:*

- (i) **HLO** and **HL7–9** hold;
- (ii) K satisfies Assumption **K**; and
- (iii) $\{h_n\}$ is such that $\beta_n^{-1} h_n + \beta_n (nh_n)^{-1} \rightarrow 0$.

*Then for all $x \in \mathbb{R}$, and t_0 as in **HL7**,*

$$\lim_{\varepsilon \downarrow 0} \limsup_{n \rightarrow \infty} \mathbf{E} \left| \frac{\beta_n}{h_n n} \sum_{t=t_0}^n K \left(\frac{X_t(n) - x}{h_n} \right) - \frac{1}{n} \sum_{t=t_0}^n \varphi_{\varepsilon^2}(\beta_n^{-1} X_t(n)) \right| = 0.$$

A.2. Proofs of Theorems 3.1 and 3.3.

PROOF OF THEOREM 3.1. By Lemma A.1, for each $\varepsilon > 0$ there is a Lipschitz continuous $f_{\varepsilon}(x)$ such that $\int_{|x| \leq \varepsilon^{-1}} |f(x) - f_{\varepsilon}(x)| dx < \varepsilon$ and $f_{\varepsilon}(x) = 0$ for $|x| > \varepsilon^{-1}$. We shall prove that:

$$(S.1) \quad \frac{1}{n} \sum_{t=1}^n f(\beta_n^{-1} X_t(n)) = \frac{1}{n} \sum_{t=1}^n f_{\varepsilon}(\beta_n^{-1} X_t(n)) + o_p(1),$$

as $n \rightarrow \infty$ then $\varepsilon \rightarrow 0$; that for each $\varepsilon > 0$

$$(S.2) \quad \frac{1}{n} \sum_{t=1}^n f_\varepsilon(\beta_n^{-1} X_t(n)) \xrightarrow{d} \int_{\mathbb{R}} f_\varepsilon(x + X^-) \Phi_{X^+}(x) dx,$$

as $n \rightarrow \infty$; and that

$$(S.3) \quad \int_{\mathbb{R}} f_\varepsilon(x + X^-) \Phi_{X^+}(x) dx \xrightarrow{a.s.} \int_{\mathbb{R}} f(x + X^-) \Phi_{X^+}(x) dx.$$

as $\varepsilon \rightarrow 0$, where the integral on the r.h.s. exists a.s. by condition (i) of Theorem 3.1. In view of (S.1)-(S.3), (12) then follows from Theorem 4.2 in Billingsley (1968).

Proof of (S.1). We prove (S.1) under condition (iii.a) of Theorem 3.1. The proof under (iii.b) is identical and therefore omitted. By condition (iii.a) we may choose $\varepsilon > 0$ sufficiently small that $|f(x)| \leq C|x|^{\lambda'}$ for all $|x| \geq \varepsilon^{-1}$. Decompose

$$f(x) = f(x) 1\{|x| \leq \varepsilon^{-1}\} + f(x) 1\{|x| > \varepsilon^{-1}\} =: f_{1,\varepsilon}(x) + f_{2,\varepsilon}(x),$$

where $\int_{|x| \leq \varepsilon^{-1}} |f_{1,\varepsilon}(x) - f_\varepsilon(x)| dx < \varepsilon$, and $|f_{2,\varepsilon}(x)| \leq C|x|^{\lambda'} 1\{|x| > \varepsilon^{-1}\}$. Letting $\tilde{X}_t(n) := \beta_n^{-1} X_t(n)$, in view of condition (ii) of Theorem 3.1, (S.1) will follow once we have shown that

$$(S.4) \quad \frac{1}{n} \sum_{t=t_0}^n \mathbf{E}|f_{1,\varepsilon}(\tilde{X}_t(n)) - f_\varepsilon(\tilde{X}_t(n))| + \frac{1}{n} \sum_{t=1}^n \mathbf{E}|f_{2,\varepsilon}(\tilde{X}_t(n))| \rightarrow 0,$$

as $n \rightarrow \infty$ and then $\varepsilon \rightarrow 0$, for t_0 as in condition (ii).

To that end, note that by **HL5(a)** and condition (iii.a) of Theorem 3.1, there is an $n_0 \geq 1$ and a $\lambda > \lambda'$ such that $\sup_{n \geq n_0, 1 \leq t \leq n} \mathbf{E}|\tilde{X}_t(n)|^\lambda < \infty$. Hence $|\tilde{X}_t(n)|^{\lambda'}$ is uniformly integrable, and so for $n \geq n_0$

$$\frac{1}{n} \sum_{t=1}^n \mathbf{E}|f_{2,\varepsilon}(\tilde{X}_t(n))| \leq \sup_{n \geq n_0, 1 \leq t \leq n} \mathbf{E}|\tilde{X}_t(n)|^{\lambda'} 1\{|\tilde{X}_t(n)| > \varepsilon^{-1}\} \rightarrow 0,$$

as $\varepsilon \rightarrow 0$. This gives the required negligibility of the second l.h.s. term in (S.4). For the first l.h.s. term, we note **HL3** implies that for n sufficiently large

$$\frac{1}{n} \sum_{t=t_0}^n \mathbf{E}|f_{1,\varepsilon}(\tilde{X}_t(n)) - f_\varepsilon(\tilde{X}_t(n))|$$

$$\begin{aligned}
&= \frac{1}{n} \sum_{t=t_0}^n \int_{\mathbb{R}} \left| f_{1,\varepsilon} \left(\frac{\beta_t}{\beta_n} x \right) - f_{\varepsilon} \left(\frac{\beta_t}{\beta_n} x \right) \right| \mathcal{D}_{n,t}(x) dx \\
&\leq \sup_{n \geq n_0, t_0 \leq t \leq n} \sup_u \mathcal{D}_{n,t}(u) \int_{\mathbb{R}} |f_{1,\varepsilon}(x) - f_{\varepsilon}(x)| dx \cdot \frac{\beta_n}{n} \sum_{t=t_0}^n \beta_t^{-1} \\
&\leq C \int_{|x| \leq \varepsilon^{-1}} |f_{1,\varepsilon}(x) - f_{\varepsilon}(x)| dx \\
&\leq C\varepsilon \rightarrow 0
\end{aligned}$$

as $\varepsilon \rightarrow 0$, where the second inequality holds by **HL6**.

Proof of (S.2). Let $\mathbf{E}_0(\cdot) := \mathbf{E}(\cdot \mid \mathcal{F}_0)$, and $\delta \in (0, 1)$. By **HL4** and the boundedness of f_{ε} ,

$$(S.5) \quad \frac{1}{n} \sum_{t=1}^n f_{\varepsilon}(\tilde{X}_t(n)) = \frac{1}{n} \sum_{t=\lfloor n\delta \rfloor + 1}^n \mathbf{E}_0[f_{\varepsilon}(\tilde{X}_t(n))] + o_p(1)$$

as $n \rightarrow \infty$ and then $\delta \rightarrow 0$. Now let $\tilde{X}_t^+(n) := \beta_n^{-1} X_t^+(n)$ and $\tilde{X}_t^-(n) := \beta_n^{-1} X_t^-(n)$. Since f_{ε} is bounded and Lipschitz, it follows from **HL1** that

$$\begin{aligned}
(S.6) \quad \frac{1}{n} \sum_{t=\lfloor n\delta \rfloor + 1}^n \mathbf{E} \left| \mathbf{E}_0[f_{\varepsilon}(\tilde{X}_t(n))] - \mathbf{E}_0[f_{\varepsilon}(\tilde{X}_t^+(n) + \tilde{X}_t^-(n))] \right| \\
\leq C \sup_{1 \leq t \leq n} \mathbf{E}(\beta_n^{-1} |R_t(n)| \wedge 1) \rightarrow 0
\end{aligned}$$

as $n \rightarrow \infty$ for each $\delta \in (0, 1)$; and from **HL2(b)** that

$$\begin{aligned}
&\frac{1}{n} \sum_{t=\lfloor n\delta \rfloor + 1}^n \mathbf{E} \left| \mathbf{E}_0[f_{\varepsilon}(\tilde{X}_t^+(n) + \tilde{X}_t^-(n))] - \mathbf{E}_0[f_{\varepsilon}(\tilde{X}_t^+(n) + \tilde{X}_n^-(n))] \right| \\
&\leq C \sup_{\lfloor n\delta \rfloor + 1 \leq t \leq n} \mathbf{E}(|\tilde{X}_t^-(n) - \tilde{X}_n^-(n)| \wedge 1) \\
(S.7) \quad &= C \mathbf{E}(|\tilde{X}_{l_n}^-(n) - \tilde{X}_n^-(n)| \wedge 1) \rightarrow 0
\end{aligned}$$

as $n \rightarrow \infty$ for each $\delta \in (0, 1)$, where $l_n \in \{\lfloor n\delta \rfloor + 1, \dots, n\}$ may always be chosen such that the final equality holds. Finally, by **HL2(a)**, Theorem 2.1 in Billingsley (1968) and Lemma A.2 we have as

$$(S.8) \quad \frac{1}{n} \sum_{t=\lfloor n\delta \rfloor + 1}^n \mathbf{E}_0[f_{\varepsilon}(\tilde{X}_t^+(n) + \tilde{X}_n^-(n))]$$

$$\xrightarrow{d} (1 - \delta) \int_{\mathbb{R}} f_{\varepsilon}(x + X^-) \Phi_{X^+}(x) dx$$

as $n \rightarrow \infty$, for each $\delta \in (0, 1)$. Hence (S.2) follows from (S.5)-(S.8) and Theorem 4.2 in Billingsley (1968).

Proof of (S.3). Let $y \in \mathcal{Y}$ for \mathcal{Y} as in condition (i) of Theorem 3.1. Noting $f_{\varepsilon}(x) = 0$ for $|x| > \varepsilon$, and that Φ_{X^+} is bounded under **HL2**, we have

$$\begin{aligned} & \int |f_{\varepsilon}(x + y) - f(x + y)| \Phi_{X^+}(x) dx \\ & \leq \sup_u \Phi_{X^+}(u) \int_{|x| \leq \varepsilon^{-1}} |f_{\varepsilon}(x) - f(x)| dx + \int_{|x| > \varepsilon^{-1}} |f(x)| \Phi_{X^+}(x - y) dx \\ & = C\varepsilon + o(1) \end{aligned}$$

as $\varepsilon \rightarrow 0$, where the negligibility of the second r.h.s. term follows by condition (i) of Theorem 3.1 and the dominated convergence theorem. Noting that $\mathbf{P}\{X^- \in \mathcal{Y}\} = 1$ completes the proof. \square

PROOF OF THEOREM 3.3. By Lemma A.3 and condition (ii) of the theorem,

$$\frac{\beta_n}{h_n n} \sum_{t=1}^n K\left(\frac{X_t(n) - x}{h_n}\right) = \frac{1}{n} \sum_{t=1}^n \varphi_{\varepsilon^2}(\beta_n^{-1} X_t(n)) + o_p(1),$$

as $n \rightarrow \infty$ and then $\varepsilon \rightarrow 0$. By Theorem 3.1 we have

$$\frac{1}{n} \sum_{t=1}^n \varphi_{\varepsilon^2}(\beta_n^{-1} X_t(n)) \xrightarrow{d} \int_{\mathbb{R}} \varphi_{\varepsilon^2}(x + X^-) \Phi_{X^+}(x) dx,$$

as $n \rightarrow \infty$. Finally, as noted in (16),

$$\int_{\mathbb{R}} \varphi_{\varepsilon^2}(x + X^-) \Phi_{X^+}(x) dx \rightarrow \Phi_{X^+}(-X^-),$$

as $\varepsilon \rightarrow 0$ by the continuity of Φ_{X^+} under **HL2**. The result then follows by Theorem 4.2 in Billingsley (1968). \square

A.3. Proofs of Lemmas A.1–A.3.

PROOF OF LEMMA A.1. See Theorem 2.26 in Folland (1999). \square

PROOF OF LEMMA A.2. By Theorems 6.3 and 6.4 in Kallenberg (2001), there is a probability kernel ν_n from (Ω, \mathcal{F}) to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that for each \mathcal{F} -measurable random variable η ,

$$\mathbf{E}(g(X_n, \eta) \mid \mathcal{F}) \stackrel{a.s.}{=} \int_{\mathbb{R}} g(x, \eta) d\nu_n(x).$$

For each $y \in \mathbb{R}$, define $h_n(y) := \int_{\mathbb{R}} g(x, y) d\nu_n(x)$: so by the preceding and condition (i) of the lemma, we have

$$h_n(y) \stackrel{a.s.}{=} \mathbf{E}(g(X_n, y) \mid \mathcal{F}) \xrightarrow{p} \int_{\mathbb{R}} g(x, y) dF_X(x) =: h(y).$$

Moreover, by the Lipschitz continuity of g

$$|h_n(y) - h_n(y')| \leq \int_{\mathbb{R}} |g(x, y) - g(x, y')| d\nu_n(x) \leq C |y - y'|.$$

Hence, $\{h_n(y)\}$ is stochastically equicontinuous on \mathbb{R} , whence $h_n(y) \xrightarrow{p} h(y)$ uniformly on every compact subset of \mathbb{R} (see for example Theorem 1 and Lemma 1 in Andrews, 1992). Finally, fix $\varepsilon > 0$ and choose M_ε such that $\limsup_{n \rightarrow \infty} \mathbf{P}(|Y_n| > M_\varepsilon) < \varepsilon$, which is possible since $Y_n \xrightarrow{d} Y$. Then

$$\begin{aligned} & \mathbf{P}(|h_n(Y_n) - h(Y_n)| > \varepsilon) \\ & \leq \mathbf{P}(\{|h_n(Y_n) - h(Y_n)| > \varepsilon\} \cap \{|Y_n| \leq M_\varepsilon\}) + \mathbf{P}(|Y_n| > M_\varepsilon) \\ & \leq \mathbf{P}\left(\sup_{|y| \leq M_\varepsilon} |h_n(y) - h(y)| > \varepsilon\right) + \varepsilon \\ & \rightarrow \varepsilon, \end{aligned}$$

as $n \rightarrow \infty$, by the uniform convergence in probability of h_n on compacta. In view of the preceding,

$$\mathbf{E}(g(X_n, Y_n) \mid \mathcal{F}) \stackrel{a.s.}{=} h_n(Y_n) = h(Y_n) + o_p(1) \xrightarrow{d}_{(1)} \int_{\mathbb{R}} g(x, Y) dF_X(x),$$

where $\xrightarrow{d}_{(1)}$ is due to condition (ii) and the continuity of $h(y)$. \square

PROOF OF LEMMA A.3. The result follows from Lemma 7 in Jeganathan (2004) and arguments similar to those used in Wang and Phillips (2009, pp. 725-728). Note, in particular, that if we define $x_{k,n} := \beta_n^{-1} X_k(n)$ and $d_{l,k,n} := \beta_n^{-1} \beta_{l-k}$, then **HL7–9** ensure that their Assumption 2.3 is satisfied, with the exception that the density of $x_{k,n}$ is bounded only for $k \geq t_0$. If $t_0 = 1$, the result follows immediately. Otherwise, it follows via very minor modifications of the arguments leading to (5.2) in Wang and Phillips (2009), making use in particular of the additional condition imposed on $\{\beta_n\}$ by **HL9(e)**. \square

APPENDIX B: PROOFS UNDER LOW-LEVEL CONDITIONS

This appendix provides proofs for the remaining results of Section 3, i.e. Theorems 3.2 and 3.4. Technical lemmas stated in this appendix are proved in Appendix C.

B.1. Sufficient conditions for Assumption HL. Preliminary to the proof of Theorem 3.2, we first present a set of ‘intermediate-level conditions’ for general linear processes (Assumption **LP** below); their sufficiency for Assumptions **HL0–HL6** is established by Proposition B.1 below. These conditions are of interest in their own right, insofar as they allow the conclusions of Theorem 3.1 to be extended to a broad class of linear processes. They also allow the proof of Theorem 3.2 to be reduced essentially to verifying that $I(1/2)$ and MI processes satisfy Assumption **LP** (see Appendix B.3).

Our conditions on linear processes shall be stated in terms of arrays of the form

$$(S.9) \quad x_t(n) = \sum_{k=0}^{\infty} a_{k,t}(n) \xi_{t-k}.$$

where $\{\xi_t\}$ is the i.i.d. sequence appearing in Assumption **INN**. We showed in Section 2 that it was possible to write both $I(1/2)$ and MI processes in the form

$$(S.10) \quad x_t(n) = \sum_{j=0}^{t-1} \phi_j(n) v_{t-j}, \quad \text{where} \quad v_t = \sum_{i=0}^{\infty} c_i \xi_{t-i}$$

for appropriate coefficients $\{c_i\}$ and $\{\phi_j(n)\}$. Suppressing the dependence of these quantities on n for the sake a readability, as we shall do freely below, (S.10) implies that

$$(S.11) \quad a_{k,t} = \sum_{j=0}^{(t-1) \wedge k} \phi_j c_{k-j} = \begin{cases} \sum_{j=0}^k \phi_j c_{k-j} =: a_k & \text{if } 0 \leq k \leq t-1, \\ \sum_{j=0}^{t-1} \phi_j c_{k-j} =: a_{k,t}^- & \text{if } k \geq t. \end{cases}$$

Note, in particular, that $a_{k,t}$ does not depend on t for $1 \leq k \leq t-1$, and we accordingly denote these coefficients by simply a_k . For $k \geq t$, the notation $a_{k,t}^-$ reminds us that these coefficients refer to innovations dated $t \leq 0$.

The following conditions on linear processes do *not* require these to have been generated according to a specific time series model (e.g. Assumption **FR/MI**), or indeed as in (S.10). We shall, however, impose one restriction

consistent with that model: that the coefficients $a_{k,t}$ should not depend on t for $1 \leq k \leq t-1$. Our conditions thus envisage an array of the form

$$(S.12) \quad x_t(n) = \sum_{k=0}^{t-1} a_k(n) \xi_{t-k} + \sum_{k=t}^{\infty} a_{k,t}^-(n) \xi_{t-k} =: x_t^+(n) + x_t^-(n),$$

associated to which, define

$$\beta_{n,t}^2 := \text{Var}(x_t(n)) = \sigma_{\xi}^2 \sum_{k=0}^{t-1} a_k(n)^2 + \sigma_{\xi}^2 \sum_{k=t}^{\infty} a_{k,t}^-(n)^2 =: (\beta_{n,t}^+)^2 + (\beta_{n,t}^-)^2$$

and set $\beta_n := \beta_{n,n}$, for $n \in \mathbb{N}$ and $t \in \{1, \dots, n\}$. The following will always be applied in conjunction with Assumption **INN**, and are stated in terms of the θ and the i.i.d. sequence $\{\xi_t\}$ appearing in that assumption.

ASSUMPTION LP (linear process).

LP1 $x_t(n)$ is as in (S.12), with $\beta_{n,t} \in (0, \infty)$ for all $n, t \in \mathbb{N}$.

LP2 Either:

- (a) $\mathbf{E}|\xi_1|^\lambda < \infty$ for some $\lambda \in [2, \infty)$; or
- (b) ξ_1 has a finite moment generating function (m.g.f.) in a neighbourhood of zero.

LP3 There exists $t_0 \in \mathbb{N}$ such that

- (a) $\liminf_{n \rightarrow \infty} \inf_{t_0 \leq t \leq n} \beta_t^{-1} \beta_{n,t}^+ > 0$;
- (b) $\limsup_{t \rightarrow \infty} \sup_{n \geq t} (\beta_{n,t}^+)^{-1} \max_{0 \leq k \leq t-1} |a_k(n)| = 0$

LP4 There exist $\delta > 0$ and $n_0 \in \mathbb{N}$ such that: for each $1 \leq t \leq t_0$ (as in **LP3**) and $n \geq n_0$, there exist $\{k_1, \dots, k_\theta\} \subset \mathbb{N}$ with

$$\min_{1 \leq l \leq \theta} |a_{k_l, t}(n)| > \delta.$$

LP5 $\lim_{n \rightarrow \infty} n^{-1/2} \beta_n^{-1} \sum_{k=0}^{n-1} |a_k(n)| = 0$.

LP6 For any $\delta \in (0, 1)$ and $\{t_n\}$ such that $\lfloor n\delta \rfloor \leq t_n \leq n$:

- (a) $\beta_n^{-1} [\beta_{n,t_n}^+, \beta_{n,t_n}^-] \rightarrow [\sigma_+, \sigma_-]$ with $\sigma_+ > 0$ and $\sigma_- \geq 0$;
- (b) $\beta_n^{-1} (\max_{0 \leq k \leq n-1} |a_k(n)| + \sup_{1 \leq t \leq n, k \geq t} |a_{k,t}^-(n)|) \rightarrow 0$;
- (c) $\beta_n^{-2} \sum_{l=-\infty}^0 [a_{t_n-l, t_n}^-(n) - a_{n-l, n}^-(n)]^2 \rightarrow 0$.

LP7 $\sup_{n \geq n_0, 1 \leq t \leq n} \beta_n^{-1} \beta_{n,t} < \infty$ for some $n_0 \in \mathbb{N}$.

REMARK B.1. The principal relationships between the preceding conditions, and the high-level conditions (Assumption **HL**) as they would be applied to $X_t(n) = x_t(n)$ may be summarised as follows; these are formally established by Proposition B.1 below.

(a) **LP2** and **LP7** imply that $\beta_n^{-1}x_t(n)$ has uniformly bounded moments of a sufficient order (as per **HL5**). For $\beta_n^{-1}x_t(n)$ to have finite λ -moments, it is sufficient that ξ_t also have finite λ -moments; $\beta_n^{-1}x_t(n)$ will have finite exponential moments if ξ_t has a finite m.g.f. in a neighbourhood of zero.

(b) **LP3** and **LP4** ensure that $\beta_t^{-1}x_t(n)$ has a uniformly bounded density, as required by **HL3**. Under **INN**, a weighted sum involving at least θ of the innovations ξ_t will have an integrable characteristic function (c.f.). $\beta_t^{-1}x_t^+(n) = \sum_{k=0}^{t-1} \beta_t^{-1}a_k(n)\xi_{t-k}$ will thus have a density bounded uniformly over n and t , provided that the L_1 norm of its c.f. can be uniformly bounded. This in turn requires that: (i) the variance of $\beta_t^{-1}x_t^+(n)$ can be bounded away from zero; and (ii) it is never dominated by less than θ of the innovations that contribute to it. Both are ensured by **LP3**, at least for n and t sufficiently large. **LP4** entails that for *every* t , a sufficient number of coefficients $\{a_{k,t}(n)\}$ are bounded away from zero; together with **LP3** it is sufficient for **HL3** to hold with $t_0 = 1$.

(c) **LP5** can be understood as a kind of weak dependence condition, which is used solely to verify **HL4**.

(d) **LP6** permits a central limit theorem for weighted sums to be applied to each of $\beta_n^{-1}x_t^+(n)$ and $\beta_n^{-1}x_t^-(n)$, as required by **HL2**. **LP6(a)** determines the limiting variance of each of these two terms, while **LP6(b)** is a negligibility requirement on the linear process coefficients, akin to a Lindeberg condition. Finally, **LP6(c)** implies the second part of **HL2(b)**.

PROPOSITION B.1 (LP \Rightarrow HL). *Suppose Assumptions **LP1** and **INN** hold. Then **HL0–1** hold for $\mathcal{F}_t := \sigma(\{\xi_r\}_{r \leq t})$, $X_t^+(n) := x_t^+(n)$, $X_t^-(n) := x_t^-(n)$, and $\beta_n^2 = \text{Var}[x_n(n)]$. Moreover:*

- (i) **LP6** \Rightarrow **HL2** with $(X^+, X^-) \sim N[0, \text{diag}\{\sigma_+^2, \sigma_-^2\}]$.
- (ii) (a) **LP3** \Rightarrow **HL3**;
- (b) **LP3** and **LP4** \Rightarrow **HL3** with $t_0 = 1$.
- (iii) **LP5** \Rightarrow **HL4**.
- (iv) (a) **LP2(a)** and **LP7** \Rightarrow **HL5(a)** with $\lambda \geq 2$;
- (b) **LP2(b)**, **LP6(b)** and **LP7** \Rightarrow **HL5(b)**.

The proof of Proposition B.1 (and subsequently, Theorem 3.4) requires the following technical lemma, whose proof is deferred to Appendix C.2. Let $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$.

LEMMA B.1. *Let $\{\xi_t\}$ be as in Assumption **INN** with $\sigma_\xi^2 = 1$, $\{\vartheta_k\}_{k \in \mathbb{N}_0}$ be a real sequence, and $\delta > 0$.*

- (i) *Suppose $\sigma_\vartheta^2 := \sum_{k=0}^{\infty} \vartheta_k^2 > 0$ and $\max_{k \in \mathbb{N}_0} \vartheta_k^2 \leq \sigma_\vartheta^2/2\theta$. Then there exists a function $G(A; \sigma^2, \psi_\xi)$, not otherwise depending on $\{\vartheta_k\}$, such that $\sigma^2 \mapsto G(A; \sigma^2, \psi_\xi)$ is weakly decreasing in σ^2 ,*

$$\int_{\{|\lambda| \geq A\}} \left| \mathbf{E} \left(i\lambda \sum_{k=0}^{\infty} \vartheta_k \xi_k \right) \right| d\lambda \leq G(A; \sigma_\vartheta^2, \psi_\xi) \leq C\sigma_\vartheta^{-1}$$

where $C < \infty$ depends only on ψ_ξ , and $\lim_{A \rightarrow \infty} G(A; \sigma_\vartheta^2, \psi_\xi) = 0$.

- (ii) *Suppose there exist $\{k_1, \dots, k_\theta\} \subset \mathbb{N}_0$ such that $|\vartheta_{k_i}| > \delta$ for all $i \in \{1, \dots, \theta\}$. Then*

$$\int_{\mathbb{R}} \left| \mathbf{E} \left(i\lambda \sum_{k=0}^{\infty} \vartheta_k \xi_k \right) \right| d\lambda < C\delta^{-1}$$

where $C < \infty$ depends only on ψ_ξ .

PROOF OF PROPOSITION B.1. That **HL0–1** are satisfied (with $R_t(n) = 0$) is immediate from (S.12), and the fact that $\beta_n = \beta_{n,n} \in (0, \infty)$ by **LP1**.

(i). Since $\{\xi_t\}$ is i.i.d., in view of (S.12) we have

$$\beta_n^{-1} [x_{t_n}^+(n), x_{t_n}^-(n)] \stackrel{d}{=} \beta_n^{-1} \left[\sum_{k=0}^{t_n-1} a_k(n) \xi_k, \sum_{k=t_n}^{\infty} a_{k,t_n}^-(n) \xi_k^* \right].$$

where $\{\xi_k^*\} \stackrel{d}{=} \{\xi_k\}$ and $\{\xi_k^*\} \perp \{\xi_k\}$. Using **LP6(a)–(b)**, the weak convergence of these quantities to a $N[0, \text{diag}\{\sigma_+^2, \sigma_-^2\}]$ distribution, as required by **HL2**, follows immediately from Lemma 2.1 of Abadir, Distaso, Giraitis and Koul (2014). Finally, note that due to **LP6(c)**

$$\begin{aligned} \beta_n^{-2} \mathbf{E} [x_{t_n}^-(n) - x_n^-(n)]^2 &= \beta_n^{-2} \mathbf{E} \left[\sum_{k=t_n}^{\infty} a_{k,t_n}^-(n) \xi_{t_n-k} - \sum_{k=n}^{\infty} a_{k,n}^-(n) \xi_{n-k} \right]^2 \\ &= \beta_n^{-2} \mathbf{E} \left[\sum_{l=-\infty}^0 a_{t_n-l,t_n}^-(n) \xi_l - \sum_{l=-\infty}^0 a_{n-l,n}^-(n) \xi_l \right]^2 \\ (S.13) \quad &= \beta_n^{-2} \sigma_\xi^2 \sum_{l=-\infty}^0 [a_{t_n-l,t_n}^-(n) - a_{n-l,n}^-(n)]^2 = o(1), \end{aligned}$$

which establishes the final part of **HL2(b)**.

(ii). By (S.12) and the Fourier inversion theorem (e.g. Feller, 1971, Theorem XV.3), it suffices for **HL3** to show that the c.f. of $\beta_t^{-1}x_t^+(n)$ has an L^1 norm that is bounded uniformly for all n and $t \leq n$ sufficiently large. Since $\beta_t^{-1}x_t^+(n)$ is a linear process with coefficients $\vartheta_{k,n,t} := \beta_t^{-1}a_k(n)\mathbf{1}\{0 \leq k \leq t-1\}$ and variance $\sigma_{\vartheta,n,t}^2 := \beta_t^{-2}(\beta_{n,t}^+)^2$, we can do this with the aid of Lemma B.1. **LP3(b)** implies that there exists $n_0, t_0 \in \mathbb{N}$ such that

$$\vartheta_{k,n,t}^2 = \beta_t^{-2}a_k^2(n)\mathbf{1}\{0 \leq k \leq t-1\} \leq \beta_t^{-2}(\beta_{n,t}^+)^2/2\theta = \sigma_{\vartheta,n,t}^2/2\theta$$

for all $k \in \mathbb{N}_0$, $t_0 \leq t \leq n$ and $n \geq n_0$, while **LP3(a)** implies that n_0 and t_0 may be additionally chosen such that

$$\sigma_{\vartheta,n,t} = \beta_t^{-1}\beta_{n,t}^+ > \epsilon > 0$$

for some $\epsilon > 0$, for all $t_0 \leq t \leq n$ and $n \geq n_0$. By Lemma B.1(i), the L^1 norm of $\mathbf{E}[i\lambda\beta_t^{-1}x_t^+(n)]$ is bounded by $C\epsilon^{-1} < \infty$ for such n and t , and thus **HL3** holds for some $t_0 \in \mathbb{N}$.

HL3 will hold with $t_0 = 1$ if we can additionally bound the L^1 norm of $\mathbf{E}[i\lambda\beta_t^{-1}x_t(n)]$ (note the deliberate omission of the ‘+’ superscript) for $1 \leq t \leq t_0$ and all $n \geq n_0$. To this end, note that $\beta_t^{-1}x_t(n)$ is a linear process with coefficients $\vartheta_{k,n,t} := \beta_t^{-1}a_{k,t}(n)\mathbf{1}\{k \geq 0\}$, and that for $\delta > 0$ as in **LP4**, for each $1 \leq t \leq t_0$ and $n \in \mathbb{N}$ there exist $\{k_1, \dots, k_\theta\} \in \mathbb{N}_0$ such that

$$|\vartheta_{k_i,n,t}| = \beta_t^{-1}|a_{k_i,t}(n)| \geq \left(\min_{1 \leq t \leq t_0} \beta_t^{-1}\right) \delta$$

where the r.h.s. is nonzero and independent of n and t . Applying Lemma B.1(ii) completes the proof.

(iii). The argument is similar to that of Wu and Mielniczuk (2002, p. 1452). Let $\mathbf{E}_t[\cdot] := \mathbf{E}[\cdot | \mathcal{F}_t]$, and decompose

$$g(x_t) - \mathbf{E}_0 g(x_t) = \sum_{s=0}^{t-1} \{\mathbf{E}_{t-s} g(x_t) - \mathbf{E}_{(t-1)-s} g(x_t)\}$$

so that

$$\begin{aligned} \sum_{t=1}^n [g(x_t) - \mathbf{E}_0 g(x_t)] &= \sum_{t=1}^n \sum_{s=0}^{t-1} [\mathbf{E}_{t-s} g(x_t) - \mathbf{E}_{(t-1)-s} g(x_t)] \\ (S.14) \quad &= \sum_{s=0}^{n-1} \sum_{t=s+1}^n [\mathbf{E}_{t-s} g(x_t) - \mathbf{E}_{(t-1)-s} g(x_t)] =: \sum_{s=0}^{n-1} M_{n,s}, \end{aligned}$$

where each $M_{n,s}$ is a sum of martingale differences.

By (S.9) we can write

$$\begin{aligned} x_t &= \sum_{k=0}^{s-1} a_{k,t} \xi_{t-s} + a_{s,t} \xi_{t-s} + \sum_{k=s+1}^{\infty} a_{k,t} \xi_{t-k} \\ &\stackrel{d}{=} \sum_{k=0}^{s-1} a_{k,t} \xi_{t-s} + a_{s,t} \xi^* + \sum_{k=s+1}^{\infty} a_{k,t} \xi_{t-k} =: x_t^* \end{aligned}$$

where $\xi^* \stackrel{d}{=} \xi_0$ is independent of $\{\xi_t\}$. Since g is Lipschitz,

$$\begin{aligned} |\mathbf{E}_{t-s} g(x_t) - \mathbf{E}_{(t-1)-s} g(x_t)| &= |\mathbf{E}_{t-s} [g(x_t) - g(x_t^*)]| \\ &\leq C |a_{s,t}| |\mathbf{E}_{t-s} [\xi_{t-s} - \xi^*]| \end{aligned}$$

whence, by the orthogonality of martingale differences

$$\mathbf{E} M_{n,s}^2 = \sum_{t=s+1}^n \mathbf{E} [\mathbf{E}_{t-s} g(x_t) - \mathbf{E}_{(t-1)-s} g(x_t)]^2 \leq C \sum_{t=s+1}^n a_{s,t}^2 \leq C n a_s^2,$$

where the final inequality follows since $a_{s,t} = a_s$ for $0 \leq s \leq t-1$. Deduce from (S.14) and the preceding that

$$\mathbf{E} \left| \frac{1}{n} \sum_{t=1}^n [g(x_t) - \mathbf{E}_0 g(x_t)] \right| \leq \frac{1}{n} \sum_{s=0}^{n-1} (\mathbf{E} M_{n,s}^2)^{1/2} \leq \frac{C}{n^{1/2}} \sum_{s=0}^{n-1} |a_s|.$$

Finally, note that if each x_t were divided by a positive constant, the preceding would hold with the r.h.s. divided by the same constant. Hence by **LP5**

$$\mathbf{E} \left| \frac{1}{n} \sum_{t=1}^n [g(\beta_n^{-1} x_t(n)) - \mathbf{E}_0 g(\beta_n^{-1} x_t(n))] \right| \leq \frac{C}{n^{1/2} \beta_n} \sum_{s=0}^{n-1} |a_s| \rightarrow 0,$$

which yields **HL4**.

(iv). Suppose **LP2(a)** holds for some $\lambda \geq 2$. By Theorem 2 of Whittle (1960), there exists a C (depending only λ and $\mathbf{E}|\xi_1|^\lambda$) such that

$$\mathbf{E} \left| \frac{x_t(n)}{\beta_n} \right|^\lambda \leq C \left(\beta_n^{-2} \sum_{k=0}^{\infty} a_{k,t}^2(n) \right)^{\lambda/2} = \frac{C}{(\mathbf{E} \xi_1^2)^{\lambda/2}} (\beta_n^{-2} \beta_{n,t}^2)^{\lambda/2}$$

and thus **HL5(a)** follows by **LP7**.

Next, suppose **LP2(b)** holds. For any $\lambda \in (0, \infty)$, we have

$$\mathbf{E} e^{\lambda \beta_n^{-1} |x_t(n)|} \leq \mathbf{E} e^{\lambda \beta_n^{-1} x_t(n)} + \mathbf{E} e^{-\lambda \beta_n^{-1} x_t(n)}.$$

We shall show that the first r.h.s. term is finite for sufficiently small λ ; the same argument delivers the bound for the second term and thence **HL5(b)**. Since $\mathbf{E} \exp(\mu \xi_1) < \infty$ for sufficiently small μ , there exist $\varsigma_\xi, b_\xi \in (0, \infty)$ such that $\mathbf{E} \exp(\mu \xi_t) \leq \exp(\mu^2 \varsigma_\xi^2)$ for all $|\mu| < b_\xi$ (see e.g. Theorem 2.13 in Wainwright, 2019). In view of **LP6(b)**, we may choose $\lambda \in (0, \infty)$ such that

$$\sup_{1 \leq t \leq n} \sup_{k \geq 0} \lambda |\beta_n^{-1} a_{k,t}(n)| < b_\xi$$

for all n sufficiently large. Hence by Fatou's lemma, for such n ,

$$\begin{aligned} \mathbf{E} e^{\beta_n^{-1} \lambda x_t(n)} &= \mathbf{E} \exp \left(\beta_n^{-1} \lambda \sum_{k=0}^{\infty} a_{k,t}(n) \xi_{t-k} \right) \\ &\leq \liminf_{M \rightarrow \infty} \prod_{k=0}^M \mathbf{E} \exp \left(\lambda \beta_n^{-1} a_{k,t}(n) \xi_{t-k} \right) \\ &\leq \liminf_{M \rightarrow \infty} \prod_{k=0}^M \exp \left(\lambda^2 \beta_n^{-2} a_{k,t}^2(n) \varsigma_\xi^2 \right) \\ &= \exp \left(\lambda^2 \beta_n^{-2} \sum_{k=0}^{\infty} a_{k,t}^2(n) \varsigma_\xi^2 \right) = \exp(\lambda^2 \beta_n^{-2} \beta_{n,t}^2 \varsigma_\xi^2), \end{aligned}$$

where the final term is bounded uniformly over $1 \leq t \leq n$ and $n \geq n_0$ by **LP7**. \square

B.2. Lemmas for $I(1/2)$ and MI processes. The proof of Theorem 3.2 is now mostly a matter of verifying that each of **FR** and **MI** (in conjunction with **INN**) imply **LP**, whereupon the result will follow by appeals to Proposition B.1 and Theorem 3.1. The following lemmas establish some key properties of $I(1/2)$ and MI processes under these assumptions, which will be used in the proofs of each of Theorems 3.2 and 3.4. Their proofs appear in Appendix C.1.

LEMMA COEF. *Suppose that **FR** or **MI** holds. Then there exist $D_1, D_2 \in (0, \infty)$ and $n_0, k_0 \in \mathbb{N}$ such that for all $n \geq n_0$*

$$(i) \sup_{1 \leq t \leq n, k \in \mathbb{N}_0} |a_{k,t}(n)| \leq D_2;$$

*and under **MI**,*

(ii) $\min_{k_0 \leq k \leq \kappa_n \wedge (n-1)} |a_k(n)| \geq D_1$.

For the purposes of the next two lemmas, define

$$(S.15) \quad [\gamma_n^2, \mathcal{V}_\infty] := \begin{cases} [L(n), 8\sigma_\xi^2] & \text{under } \mathbf{FR1}, \\ [L(n), \sigma_\xi^2(\sum_{s=0}^\infty c_s)^2] & \text{under } \mathbf{FR2}, \\ [\kappa_n, \sigma_\xi^2(\sum_{s=0}^\infty c_s)^2/2] & \text{under } \mathbf{MI}. \end{cases}$$

By Proposition 1.5.9a in BGT, $L(n) = \int_1^n x^{-1} \ell^2(x) dx$ is a slowly varying function, a fact that shall be used freely throughout the following. Recall also that $L(n) \rightarrow \infty$ under **FR**. **MI** entails that κ_n is regularly varying with index $\alpha \in [0, 1)$, and thus in all cases, γ_n is regularly varying with index $\alpha \in [0, 1/2)$.

LEMMA LVAR. *Suppose **INN** and either **FR** or **MI** holds. Let $t_n \in [\lfloor rn \rfloor, n] \cap \mathbb{N}$, where $0 < r \leq 1$. Then*

(i) *Under **FR1**:*

$$[\text{Var}[\gamma_n^{-1} x_{t_n}^+(n)] \quad \text{Var}[\gamma_n^{-1} x_{t_n}^-(n)]] \rightarrow [\mathcal{V}_\infty/2 \quad \mathcal{V}_\infty/2]$$

and for any $s \in (0, r]$,

$$\text{Var}\{\gamma_n^{-1}[x_{\lfloor rn \rfloor}^-(n) - x_{\lfloor sn \rfloor}^-(n)]\} \rightarrow 0.$$

(ii) *Under **FR2** or **MI**:*

$$\text{Var}[\gamma_n^{-1} x_{t_n}(n)] = \text{Var}[\gamma_n^{-1} x_{t_n}^+(n)] + o(1) \rightarrow \mathcal{V}_\infty$$

LEMMA CLT. *Under the assumptions of Lemma LVAR,*

$$\gamma_n^{-1} [x_{t_n}^+(n), x_{t_n}^-(n)] \xrightarrow{d} N(0, \Sigma),$$

where $\Sigma = \text{diag}\{\mathcal{V}_\infty/2, \mathcal{V}_\infty/2\}$ under **FR1**, and $\Sigma = \text{diag}\{\mathcal{V}_\infty, 0\}$ under **FR2** or **MI**.

B.3. Proofs of Theorems 3.2 and 3.4. We first state a preliminary lemma on regularly varying sequences, whose proof is given in Appendix C.2 below.

LEMMA B.2. *Let $\{\varphi_j\}_{j \in \mathbb{N}}$ be a real sequence, and ς a positive-valued, slowly varying function, that is locally integrable on $[1, \infty)$.*

- (i) Suppose $\varphi_j \sim j^l \varsigma(j)$ for $l > -1$. Then for all $0 \leq s < r < \infty$ as $n \rightarrow \infty$

$$\frac{1}{n^{1+l}\varsigma(n)} \sum_{j=\lfloor ns \rfloor + 1}^{\lfloor nr \rfloor} \varphi_j \rightarrow \int_s^r x^l dx.$$

- (ii) Suppose $\varphi_j \sim j^l \varsigma(j)$ for $l < -1$. Then as $n \rightarrow \infty$

$$\frac{1}{n^{1+l}\varsigma(n)} \sum_{j=n}^{\infty} \varphi_j \rightarrow \int_1^{\infty} x^l dx.$$

- (iii) Suppose $S(x) := \int_1^x u^{-1} \varsigma(u) du$ is such that $S(x) \rightarrow \infty$ as $x \rightarrow \infty$. Then for any $n_0 \geq 1$, as $n \rightarrow \infty$

$$\sum_{j=n_0}^n j^{-1} \varsigma(j) \sim S(n).$$

PROOF OF THEOREM 3.2. With the aid of Proposition B.1, we show that the conditions of Theorem 3.1 are satisfied for $X_t(n) = x_t(n)$. We first verify Assumptions **HL0–6**, before turning to the other conditions of the theorem.

HL0–1. Under either of **FR** or **MI**, $x_t(n)$ can be written in the form (5); that $\beta_{n,t} < \infty$ for each $n, t \in \mathbb{N}$ follows from $\mathbf{E} v_1^2 = \sigma_\xi^2 \sum_{k=0}^{\infty} c_k^2 < \infty$. Recall also from (S.9) and (S.11) that we may write $x_t(n) = \sum_{k=0}^{\infty} a_{k,t}(n) \xi_{t-k}$, where $a_{k,t}(n) = \sum_{j=0}^{(t-1) \wedge k} \phi_j(n) c_{k-j}$. Under **FR**, $\phi_0(n) = \phi_0 \neq 0$, while under **MI**, $\phi_0(n) = 1 - \kappa_n^{-1} \rightarrow 1$ and has $\phi_0(n) > 0$ for all n , since $\kappa_n > 1$. Now let k^* denote the smallest $k \in \mathbb{N}$ such that $c_k \neq 0$. Then

$$\beta_{n,t}^2 = \sum_{k=0}^{\infty} a_{k,t}^2(n) \geq a_{k^*,t}^2(n) = \phi_0^2(n) c_{k^*}^2 > 0$$

for all $n, t \in \mathbb{R}$. Thus **LP1** holds, whence **HL0–1** follows by Proposition B.1.

HL2. Once we have verified **LP6**, by Proposition B.1(i) this will hold with $(X^+, X^-) \sim N[0, \text{diag}\{\sigma_+^2, \sigma_-^2\}]$. Lemma LVAR implies that **LP6(a)** holds with $(\sigma_+^2, \sigma_-^2) = (1/2, 1/2)$ under **FR1** and $(\sigma_+^2, \sigma_-^2) = (1, 0)$ under **FR2/MI**. It further implies that $\beta_n \asymp \gamma_n \rightarrow \infty$, and thus **LP6(b)** follows by Lemma COEF. With respect to **LP6(c)**, it follows from (S.13) above that

$$\beta_n^{-2} \sum_{l=-\infty}^0 \left[a_{t_n-l, t_n}^-(n) - a_{n-l, n}^-(n) \right]^2 \sigma_\xi^2 \asymp \gamma_n^{-2} \text{Var} [x_{t_n}^-(n) - x_n^-(n)]$$

for $t_n \in \{\lfloor nr \rfloor, \dots, n\}$ for some $r \in (0, 1]$. By Lemma LVAR, the r.h.s. is immediately $o(1)$ under **FR1**. Under **FR2** or **MI**, that same result implies $\gamma_n^{-2} \text{Var}[x_{t_n}^-(n)] = o(1)$, and thus

$$\gamma_n^{-2} \text{Var}[x_{t_n}^-(n) - x_n^-(n)] \leq 2\gamma_n^{-2} \{ \text{Var}[x_{t_n}^-(n)] + \text{Var}[x_n^-(n)] \} = o(1).$$

HL3 (for some $t_0 \in \mathbb{N}$). By Proposition B.1(ii), it suffices to verify **LP3**. Consider **FR** first: in this case, $x_t(n)$ does not depend on n . Thus $\beta_{n,t}^+ = \beta_t^+ \asymp \gamma_t$ and $\beta_t \asymp \gamma_t$ as $t \rightarrow \infty$ by Lemma LVAR, whence **LP3(a)** holds. By Lemma COEF, $\max_{0 \leq k \leq t-1} |a_k(n)| \leq D_2$, which together with $\gamma_t = L^{1/2}(t) \rightarrow \infty$ delivers **LP3(b)**.

Next suppose **MI** holds. By Lemma COEF, we have

$$(\beta_{n,t}^+)^2 = \sum_{k=0}^{t-1} a_k^2(n) \geq D_1[(t-1) \wedge \kappa_n - k_0].$$

Since $C_1 := \sup_{n \geq 1, 1 \leq t \leq n} \kappa_t / \kappa_n < \infty$ by Assumption **MI**, where trivially $C_1 \geq 1$, we have

$$(t-1) \wedge \kappa_n - k_0 \geq (t-1) \wedge C_1^{-1} \kappa_t - k_0 \asymp \kappa_t,$$

since $\kappa_t = o(t)$ as $t \rightarrow \infty$. Thus there exists a $t_0 \in \mathbb{N}$ and $C \in (0, \infty)$ such that $\beta_{n,t}^+ \geq C\kappa_t^{1/2}$ for all $t_0 \leq t \leq n$. Since $\beta_t \asymp \gamma_t = \kappa_t^{1/2} \rightarrow \infty$ by Lemma LVAR, and $\max_{0 \leq k \leq t} |a_k(n)| \leq D_2$ by Lemma COEF, both parts of **LP3** follow immediately.

HL4. By Proposition B.1(iii), we need to only to verify **LP5**. First suppose **FR1** holds. Then recalling (S.11) above,

$$|a_k(n)| = \left| \sum_{j=0}^k \phi_j c_{k-j} \right| \leq \left| \sum_{j=0}^k c_j \right| =_{(1)} \left| \sum_{j=k+1}^{\infty} c_j \right| \asymp_{(2)} k^{-1/2} \ell(k)$$

as $k \rightarrow \infty$, where $=_{(1)}$ follows by $\sum_{k=0}^{\infty} c_k = 0$, and $\asymp_{(2)}$ by Lemma B.2(ii), since $c_j \sim k^{-3/2} \ell(k)$. Hence, by Lemma B.2(i)

$$\sum_{k=0}^n |a_k(n)| \leq Cn^{1/2} \ell(n)$$

under **FR1**. Alternatively, if **FR2** holds, then we also have

$$\sum_{k=0}^n |a_k(n)| \leq \sum_{k=0}^n \sum_{j=0}^k |\phi_j c_{k-j}| \leq \sum_{j=0}^n |\phi_j| \sum_{k=0}^{\infty} |c_k| \leq Cn^{1/2} \ell(n)$$

by $\sum_{k=0}^{\infty} |c_k| < \infty$ and Lemma B.2(i). Thus under **FR**, Lemma LVAR implies that

$$\frac{1}{n^{1/2}\beta_n} \sum_{k=0}^n |a_k(n)| \leq C \frac{n^{1/2}\ell(n)}{n^{1/2}L^{1/2}(n)} = C \left(\frac{\ell^2(n)}{\int_1^n x^{-1}\ell^2(x)dx} \right)^{1/2} \rightarrow 0$$

where the final convergence follows by Proposition 1.5.9a in BGT.

Next suppose **MI** holds. In this case $a_k(n) = \sum_{j=0}^k \rho_n^j c_{k-j}$ for $\rho_n := 1 - \kappa_n^{-1}$, and so for all n sufficiently large that $\rho_n > 0$,

$$\sum_{k=0}^n |a_k(n)| \leq \sum_{k=0}^n \sum_{j=0}^k |\rho_n^j c_{k-j}| \leq \sum_{j=0}^n |\rho_n^j| \sum_{k=j}^n |c_{k-j}| \leq_{(1)} \frac{C}{1 - \rho_n} = C\kappa_n$$

where $\leq_{(1)}$ holds since $\{c_i\}$ is absolutely summable. Hence by Lemma LVAR

$$\frac{1}{n^{1/2}\beta_n} \sum_{k=0}^n |a_k(n)| \leq \frac{C\kappa_n}{n^{1/2}\kappa_n^{1/2}} = C \left(\frac{\kappa_n}{n} \right)^{1/2} \rightarrow 0.$$

HL5. In view of Proposition B.1(iv), having already verified **LP6(b)**, we need to verify **LP7**. Suppose first that **FR** holds. Then since $x_t(n)$ does not depend on n , we have by Lemma LVAR that, as $t \rightarrow \infty$

$$\beta_{n,t}^2 = \beta_t^2 \asymp L(t) = \int_1^t x^{-1}\ell^2(x)dx$$

which is clearly monotone increasing. Thus there exist $C_1, C_2 \in (0, \infty)$ such that

$$\beta_{n,t} = \beta_t \leq C_1 L(t) \leq C_1 L(n) \leq C_2 \beta_n$$

for all $t \leq n$, whence **LP7** holds trivially.

Suppose next that **MI** holds. In this case, $x_t(n) = \sum_{k=0}^{\infty} a_{k,t}(n) \xi_{t-k}$ with $a_{k,t} = \sum_{j=0}^{(t-1) \wedge k} \phi_j(n) c_{k-j}$, where $\sum_{i=0}^{\infty} |c_i| < \infty$ and $\phi_j(n) = (1 - \kappa_n^{-1})^j > 0$ for all n sufficiently large. Now define a new process $\{\bar{x}_t(n)\}$ by $\bar{x}_t(n) = \sum_{k=0}^{\infty} \bar{a}_{k,t}(n) \xi_{t-k}$, where $\bar{a}_{k,t} = \sum_{j=0}^{(t-1) \wedge k} \phi_j(n) |c_{k-j}|$. Then $\{\bar{x}_t(n)\}$ also satisfies **MI**, and it is easily verified that

$$|a_{k,t}| \leq \bar{a}_{k,t} = \sum_{j=0}^{(t-1) \wedge k} \phi_j(n) |c_{k-j}| \leq \sum_{j=0}^{(n-1) \wedge k} \phi_j(n) |c_{k-j}| = \bar{a}_{n,t}$$

for all $1 \leq t \leq n$, whence

$$\beta_{n,t}^2 = \sum_{k=0}^{\infty} a_{k,t}^2 \leq \sum_{k=0}^{\infty} \bar{a}_{k,t}^2 =: \bar{\beta}_{n,t} \leq \bar{\beta}_{n,n} = \bar{\beta}_n.$$

It follows from Lemma LVAR that $\bar{\beta}_n^2 \asymp \kappa_n \asymp \beta_n^2$, whence

$$\beta_n^{-2} \beta_{n,t}^2 \leq \beta_n^{-2} \bar{\beta}_n^2 = O(1).$$

Thus by Proposition B.1(iv), either **HL5(a)** or **(b)** holds depending on the assumptions made on $\{\xi_t\}$ – i.e. depending on whether condition 2(a) or 2(b) of Theorem 3.2 is maintained.

HL6. By Lemma LVAR, $\beta_n \asymp \gamma_n$, which is regularly varying with index $\alpha \in [0, 1)$. Thus **HL6** follows by Lemma B.2(i).

We have thus verified that **HL0–6** hold under our assumptions, with $(X^+, X^-) \sim N[0, \text{diag}\{\sigma_+^2, \sigma_-^2\}]$. The conclusion of Theorem 3.1 accordingly holds for bounded f , with

$$\int_{\mathbb{R}} f(x + X^-) \Phi_{X^+}(x) dx = \int_{\mathbb{R}} f(x) \Phi_{X^+}(x - X^-) dx = \int_{\mathbb{R}} f(x) \varrho(x) dx$$

where ϱ is as defined in (13).

For more general f , we need to verify conditions (i)–(iii) of Theorem 3.1. With respect to condition (i), since it is assumed that f is locally integrable and satisfies $|f(x)| = O(e^{\lambda'|x|})$ for some $\lambda' < \infty$ as $|x| \rightarrow \infty$, we thus have

$$\int_{\mathbb{R}} |f(x + y)| \Phi_{X^+}(x) dx \leq C \int_{\mathbb{R}} |f(x)| e^{-(x+y)^2/2\sigma_+^2} dx < \infty$$

for every $y \in \mathbb{R}$, as required.

Condition (iii) of Theorem 3.1 is implied by condition 2 of Theorem 3.2, and the verification of **HL5** given above.

It thus remains to verify condition (ii) of Theorem 3.1. Under condition 1(a) of Theorem 3.2, there is nothing to left prove. Suppose instead that condition 1(b) of Theorem 3.2 is assumed, i.e. that **INN** holds with $\theta = 1$. In that case, we need to verify that **HL3** holds with $t_0 = 1$. By Proposition B.1(ii)(b), it suffices to verify **LP4**, since we have already verified **LP3**. Recall (S.9) and (S.11) above, and let k^* denote the smallest $k \in \mathbb{N}$ such that $c_{k^*} \neq 0$. Then

$$|a_{k^*,t}(n)| = |\phi_0(n)c_{k^*}| \geq |c_{k^*}| \inf_{n \in \mathbb{N}} |\phi_0(n)| > 0$$

where the final inequality follows from arguments given in the course of verifying **HL0–1** above. Thus **LP4** holds. \square

PROOF OF THEOREM 3.4. We verify the assumptions of Theorem 3.3 are satisfied for $X_t(n) = x_t(n)$.

HL0–2, HL4, HL6. These follow by the arguments given in the proof of Theorem 3.2.

HL7. We first note that the proof of Proposition B.1(ii), together with the arguments used to verify **HL3** in the proof of Theorem 3.2, show that $\{\beta_t^{-1}x_t^+(n)\}$ has a density $\mathcal{D}_{t,n}^+$ that is uniformly bounded over all n and $t_0 \leq t \leq n$ for some $t_0 \in \mathbb{N}$. Now let $0 \leq s < t \leq n$ and note from (S.12) that we may write

$$x_t = \sum_{k=0}^{(t-s)-1} a_k \xi_{t-k} + \sum_{k=t-s}^{\infty} a_k \xi_{t-k} =: x_{s+1,t} + x'_{s,t}$$

where $x'_{s,t}$ is \mathcal{F}_s -measurable, and $x_{s+1,t}$ is independent of \mathcal{F}_s . Since

$$x_{s+1,t} = \sum_{k=0}^{(t-s)-1} a_k \xi_{t-k} \stackrel{d}{=} \sum_{k=0}^{(t-s)-1} a_k \xi_{(t-s)-k} = x_{t-s}^+$$

we must have that the density $\mathcal{D}_{t,s,n}$ of $\{\beta_{t-s}^{-1}[x_t(n) - x_s(n)]\}$, conditional on \mathcal{F}_s , satisfies

$$(S.16) \quad \mathcal{D}_{t,s,n}(x) = \mathcal{D}_{t-s,n}^+(x - \chi_{s,t}(n))$$

where $\chi_{s,t}(n) := \beta_{t-s}^{-1}[x_{s+1,t}(n) - x_s(n)]$ is \mathcal{F}_s -measurable (with the convention that $x_0(n) := 0$). It follows that

$$\begin{aligned} \sup_{0 \leq s < t \leq n, t-s \geq t_0} \sup_{x \in \mathbb{R}} |\mathcal{D}_{t,s,n}(x)| &\leq \sup_{0 \leq s < t \leq n, t-s \geq t_0} \sup_{x \in \mathbb{R}} |\mathcal{D}_{t-s,n}^+(x)| \\ &\leq \sup_{t_0 \leq r \leq n} \sup_{x \in \mathbb{R}} |\mathcal{D}_{r,n}^+(x)| < \infty, \end{aligned}$$

by the uniform boundedness of $\mathcal{D}_{r,n}^+$ noted above.

HL8. Let $\mathcal{D}_{r,n}^*$ denote the density of $\gamma_n^{-1}x_r^+(n) = (\gamma_n^{-1}\beta_r) \cdot \beta_r^{-1}x_r^+(n)$, where $\{\gamma_n\}$ is as defined in (S.15), and $\delta_\eta := q_0\eta^{q_1}$ for $\eta, q_0, q_1 > 0$. Let $s, t \in \Omega_n(\eta)$, and $r := t - s$. In view of (S.16),

$$\begin{aligned} \sup_{|x| \leq \delta_\eta} |\mathcal{D}_{t,s,n}(x) - \mathcal{D}_{t,s,n}(0)| &\leq \sup_{y \in \mathbb{R}} \sup_{|x| \leq \delta_\eta} |\mathcal{D}_{r,n}^+(y+x) - \mathcal{D}_{r,n}^+(y)| \\ &\leq (\beta_r^{-1}\gamma_n) \sup_{y \in \mathbb{R}} \sup_{|x| \leq \delta_\eta(\beta_r^{-1}\gamma_n)} |\mathcal{D}_{r,n}^*(y+x) - \mathcal{D}_{r,n}^*(y)|. \end{aligned}$$

By the definition of $\Omega_n(\eta)$, **HL8** will hold if we can show that when $r = r_n$ for any sequence $\{r_n\}$ with $r_n \in [\lfloor n\eta \rfloor, n]$, the r.h.s. converges to zero as $n \rightarrow \infty$ and then $\eta \rightarrow 0$.

By Lemma LVAR, $\beta_{r_n}^{-1}\gamma_n \asymp \gamma_{r_n}^{-1}\gamma_n \asymp 1$, since $\{\gamma_n\}$ is regularly varying. In particular, $\{\beta_{r_n}^{-1}\gamma_n\}$ is bounded above, and so (by redefinition of q_0) the result will follow if we can prove that

$$(S.17) \quad \sup_{y \in \mathbb{R}} \sup_{|x| \leq \delta_\eta} |\mathcal{D}_{r_n, n}^*(y+x) - \mathcal{D}_{r_n, n}^*(y)| \rightarrow 0$$

as $n \rightarrow \infty$ and then $\eta \rightarrow 0$. By Lemma CLT, $\gamma_n^{-1}x_{r_n}^+(n) \xrightarrow{d} N[0, \sigma_+^2]$ for some $\sigma_+^2 > 0$. By the argument given in the proof of Corollary 2.2 in Wang and Phillips (2009), it therefore suffices for (S.17) to show that the characteristic functions of $\{\gamma_n^{-1}x_{r_n}^+(n)\}$ are uniformly integrable: since in this case, $\mathcal{D}_{r_n, n}^*$ converges uniformly to the $N[0, \sigma_+^2]$ density. To that end, note that $\gamma_n^{-1}x_{r_n}^+(n)$ is a linear process with variance converging to $\sigma_+^2 > 0$ (as $n \rightarrow \infty$) by Lemma LVAR, and coefficients $\{\gamma_n^{-1}a_k(n)\}_{k=0}^{r_n-1}$ with $\max_{0 \leq k \leq r_n-1} \gamma_n^{-1}|a_k(n)| \rightarrow 0$ by Lemma COEF. Thus the required uniform integrability follows by Lemma B.1(i).

HL9. By Lemma LVAR, $\beta_n \asymp \gamma_n$, which is regularly varying with index $\alpha \in [0, 1/2)$ and so can be written as $\gamma_n = n^\alpha \varsigma(n)$, for ς a positive-valued, slowly varying function. Thus **HL9(a)–(c)** follow immediately from Lemma B.2(i). For **HL9(d)**, we note that

$$\begin{aligned} \beta_n^{-1} \inf_{(t,s) \in \Omega_n(\eta)} \beta_{t-s} &> C \gamma_n^{-1} \inf_{(t,s) \in \Omega_n(\eta)} \gamma_{t-s} \\ &= C n^{-\alpha} \varsigma^{-1}(n) \inf_{(t,s) \in \Omega_n(\eta)} (t-s)^\alpha \varsigma(t-s) \end{aligned}$$

for some $C > 0$. Since $(t, s) \in \Omega_n(\eta)$ implies $t - s > \lfloor \eta n \rfloor$, the r.h.s. can be bounded below (up to a positive constant) by

$$(\eta/2)^\alpha \inf_{\lambda \in [\eta/2, 1]} \frac{\varsigma(\lambda n)}{\varsigma(n)} = (\eta/2)^\alpha (1 + o(1))$$

where the equality follows by Theorem 1.2.1 in BGT. Thus **HL9(d)** holds. Finally, since $\beta_n \asymp \gamma_n$, and both $\beta_n > 0$ and $\gamma_n > 0$ for all n , there exist $\underline{C}, \overline{C} \in (0, \infty)$ such that $\beta_n \in [\underline{C}\gamma_n, \overline{C}\gamma_n]$ for all n . Thus

$$\sup_{n \geq 1} \sup_{1 \leq t \leq n} \beta_n^{-1} \beta_t \leq \underline{C}^{-1} \overline{C} \sup_{n \geq 1} \sup_{1 \leq t \leq n} \gamma_n^{-1} \gamma_t.$$

From (S.15), under **FR** $\gamma_n = L(n)$ is increasing, so $\sup_{n \geq 1} \sup_{1 \leq t \leq n} \gamma_n^{-1} \gamma_t \leq 1$ trivially. Under **MI**, $\gamma_n = \kappa_t$, for which $\sup_{n \geq 1} \sup_{1 \leq t \leq n} \kappa_n^{-1} \kappa_t < \infty$ is a maintained assumption. Thus **HL9(e)** holds. \square

APPENDIX C: PROOFS AUXILIARY TO APPENDIX B

This appendix provides proofs of the lemmas stated in Appendix B.

C.1. Proofs of Lemmas COEF, LVAR and CLT.

PROOF OF LEMMA COEF. **(i).** In all cases, $|\phi_j(n)| < C$ uniformly over $j, n \in \mathbb{N}$, and $\{c_s\}$ is absolutely summable. Hence from (S.11),

$$|a_{k,t}(n)| = \left| \sum_{j=0}^{(t-1) \wedge k} \phi_j c_{k-j} \right| \leq C \sum_{j=0}^{\infty} |c_j| =: D_2 < \infty.$$

(ii). Under **MI**, we have $\phi_j(n) = \rho_n^j$ for $\rho_n := 1 - \kappa_n^{-1}$, and thus

$$(S.18) \quad a_k(n) = \sum_{j=0}^k \rho_n^{k-j} c_j = \rho_n^k \left[\sum_{j=0}^k c_j + \sum_{j=0}^k (\rho_n^{-j} - 1) c_j \right].$$

Since $\rho_n \in (0, 1)$, we have for all $0 \leq k \leq \lfloor \kappa_n \rfloor$ that

$$(S.19) \quad 1 \geq \rho_n^k \geq \rho_n^{\kappa_n} = (1 - \kappa_n^{-1})^{\kappa_n} \rightarrow e^{-1}$$

from which follows that $\rho_n^k \in (e^{-1}/2, 1]$ for all $k \in \{0, \dots, \lfloor \kappa_n \rfloor\}$, for all $n \in \mathbb{N}$.

Thus to bound $a_k(n)$ away from zero, we need only to bound the bracketed term in (S.18) away from zero. To that end, note that for all $0 \leq k \leq \lfloor \kappa_n \rfloor$

$$\left| \sum_{j=0}^k (\rho_n^{-j} - 1) c_j \right| \leq \sum_{j=0}^{\infty} |\rho_n^{-j} - 1| |c_j| 1_{\{j \leq \kappa_n\}} \rightarrow 0$$

as $n \rightarrow \infty$ by the dominated convergence theorem, since $|\rho_n^{-j} - 1| |c_j| 1_{\{j \leq \kappa_n\}} \rightarrow 0$ for each $j \in \mathbb{R}$, and by an analogous argument to (S.19) is (for sufficiently large n) bounded above by $|c_j|(1+2e)$, which is summable. Since $\sum_{j=0}^{\infty} c_j \neq 0$, deduce that there exist $n_0, k_0 \in \mathbb{N}$ such that

$$\left| \sum_{j=0}^k (\rho_n^{-j} - 1) c_j \right| \leq \frac{1}{2} \sum_{j=0}^k c_j$$

for all $n \geq n_0$ and $k \in \{k_0, \dots, \lfloor \kappa_n \rfloor\}$. □

PROOF OF LEMMA LVAR. For convenience set $\sigma_\xi^2 = 1$. Under **FR** it is sufficient to prove the result with $t_n = n$; the result in the general case follows straightforwardly from the fact that $L(t_n)/L(n) \rightarrow 1$ as $n \rightarrow \infty$, since $L(n)$ is slowly varying.

FR1. Suppose we show that

$$(S.20) \quad \sum_{k=0}^{n-1} \left(\sum_{j=0}^k c_j \right)^2 \sim 4L(n) \quad \sum_{k=0}^{n-1} \left(\sum_{j=k+n+1}^{\infty} c_j \right)^2 = o[L(n)]$$

and that for any sequences $\{\delta_{1,n}\}, \{\delta_{2,n}\}$ with $\delta_{1,n} \asymp n \asymp \delta_{2,n}$,

$$(S.21) \quad \sum_{k=\delta_{1,n}}^{\infty} \left(\sum_{j=k+1}^{k+\delta_{2,n}} c_j \right)^2 = o[L(n)].$$

Since $\gamma_n^2 = L(n)$ and $a_{k,t} = \sum_{j=0}^{(t-1) \wedge k} c_{k-j}$, it follows from (S.20) that

$$\gamma_n^{-2} \text{Var}(x_n^+) = \gamma_n^{-2} \sum_{k=0}^{n-1} \left(\sum_{j=0}^k c_{k-j} \right)^2 = \gamma_n^{-2} \sum_{k=0}^{n-1} \left(\sum_{j=0}^k c_j \right)^2 \rightarrow 4.$$

Further, note that we may write

$$(S.22) \quad \begin{aligned} \text{Var}(x_n^-) &= \sum_{k=n}^{\infty} \left(\sum_{j=0}^{n-1} c_{k-j} \right)^2 = \sum_{k=0}^{\infty} \left(\sum_{j=k+1}^{k+n} c_j \right)^2 \\ &= \sum_{k=0}^{n-1} \left(\sum_{j=k+1}^{k+n} c_j \right)^2 + \sum_{k=n}^{\infty} \left(\sum_{j=k+1}^{k+n} c_j \right)^2. \end{aligned}$$

The second r.h.s. term is of the same form as (S.21) with $\delta_{1,n} = \delta_{2,n} = n$, and so is $o[L(n)]$. The first r.h.s. term expands as

$$\begin{aligned} \sum_{k=0}^{n-1} \left(\sum_{j=k+1}^{k+n} c_j \right)^2 &= \sum_{k=0}^{n-1} \left(\sum_{j=k+1}^{\infty} c_j - \sum_{j=k+n+1}^{\infty} c_j \right)^2 \\ &= \sum_{k=0}^{n-1} \left(\sum_{j=0}^k c_j + \sum_{j=k+n+1}^{\infty} c_j \right)^2, \end{aligned}$$

where we have used the fact that $\sum_{j=0}^{\infty} c_j = 0$, and so by (S.20) and the Cauchy-Schwarz inequality,

$$(S.23) \quad \sum_{k=0}^{n-1} \left(\sum_{j=k+1}^{k+n} c_j \right)^2 = \sum_{k=0}^{n-1} \left(\sum_{j=0}^k c_j \right)^2 + o[L(n)].$$

Thus it follows from (S.22) and (S.23) that $\gamma_n^{-2} \text{Var}(x_n^-) \rightarrow 4$, as required. Finally, noting that

$$x_t^- = \sum_{k=t}^{\infty} a_{k,t}^- \xi_{t-k} = \sum_{k=t}^{\infty} \sum_{j=0}^{t-1} c_{k-j} \xi_{t-k} = \sum_{i=0}^{\infty} \sum_{j=0}^{t-1} c_{i+t-j} \xi_{-i} = \sum_{i=0}^{\infty} \sum_{l=1}^t c_{i+l} \xi_{-i}$$

we have for $0 < s < r \leq 1$ that

$$\begin{aligned} \text{Var}(x_{[nr]}^- - x_{[ns]}^-) &= \mathbf{E} \left(\sum_{i=0}^{\infty} \sum_{l=1}^{[nr]} c_{i+l} \xi_{-i} - \sum_{i=0}^{\infty} \sum_{l=1}^{[ns]} c_{i+l} \xi_{-i} \right)^2 \\ &= \mathbf{E} \left(\sum_{i=0}^{\infty} \sum_{l=[ns]+1}^{[nr]} c_{i+l} \xi_{-i} \right)^2 \\ &= \sum_{i=0}^{\infty} \left(\sum_{l=[ns]+1}^{[nr]} c_{i+l} \right)^2 = \sum_{i=[ns]}^{\infty} \left(\sum_{l=i+1}^{i+[nr]-[ns]} c_l \right)^2, \end{aligned}$$

which is of the form (S.21) with $\delta_{1,n} = [ns]$ and $\delta_{2,n} = [nr] - [ns]$, and so is $o[L(n)]$ as required.

It remains to prove (S.20) and (S.21). Recall that $c_j \sim j^{-3/2} \ell(j) =: g(j)$, which is regularly varying with index $-3/2$. Since ℓ is defined only up to an asymptotic equivalence, it is without loss of generality to take ℓ to be such that g is monotone decreasing (see Theorem 1.5.3 in BGT). For the first part of (S.20), using that $\sum_{j=0}^{\infty} c_j = 0$ we have

$$\sum_{k=0}^{n-1} \left(\sum_{j=0}^k c_j \right)^2 \sim \sum_{k=1}^{n-1} \left(\sum_{j=k+1}^{\infty} c_j \right)^2 = \sum_{k=1}^{n-1} k^2 g^2(k) \left(k^{-1} g(k)^{-1} \sum_{j=k+1}^{\infty} c_j \right)^2.$$

By Lemma B.2(ii),

$$k^{-1} g(k)^{-1} \sum_{j=k+1}^{\infty} c_j \rightarrow \int_1^{\infty} x^{-3/2} dx = 2$$

as $k \rightarrow \infty$, while by Lemma B.2(iii),

$$L(n)^{-1} \sum_{k=1}^{n-1} k^2 g^2(k) = L(n)^{-1} \sum_{k=1}^{n-1} \frac{\ell^2(k)}{k} \rightarrow 1,$$

whence by the Toeplitz lemma (Hall and Heyde, 1980, p. 31)

$$L(n)^{-1} \sum_{k=0}^{n-1} \left(\sum_{j=0}^k c_j \right)^2 \rightarrow 2^2 = 4$$

as required. For the second part of (S.20), we have

$$\begin{aligned} \sum_{k=0}^{n-1} \left(\sum_{j=k+n+1}^{\infty} c_j \right)^2 &= \sum_{k=0}^{n-1} \left(\sum_{j=0}^{k+n} c_j \right)^2 \\ &= \sum_{k=n}^{2n-1} \left(\sum_{j=0}^k c_j \right)^2 = \sum_{k=0}^{2n-1} \left(\sum_{j=0}^k c_j \right)^2 - \sum_{k=0}^{n-1} \left(\sum_{j=0}^k c_j \right)^2. \end{aligned}$$

By the first part of (S.20), the r.h.s. is asymptotically equivalent to

$$L(2n) - L(n) = L(2n) \left[1 - \frac{L(2n)}{L(n)} \right] = o[L(2n)] = o[L(n)]$$

by the slow variation of $L(n)$.

Finally, we turn to (S.21). Since $c_j \sim g(j)$, c_j is eventually strictly positive, and there exists a $C < \infty$ such that $0 < c_j \leq Cg(j)$ for all j sufficiently large. Hence, for all n sufficiently large

$$\sum_{k=\delta_{1,n}}^{\infty} \left(\sum_{j=k+1}^{k+\delta_{2,n}} c_j \right)^2 \leq C_1 \sum_{k=\delta_{1,n}}^{\infty} \left(\sum_{j=k+1}^{k+\delta_{2,n}} g(j) \right)^2 \leq C_1 \delta_{2,n}^2 \sum_{k=\delta_{1,n}}^{\infty} g^2(k),$$

where the second inequality holds by the monotonicity of $g(j)$. By Lemma B.2(ii),

$$\sum_{k=\delta_{1,n}}^{\infty} g^2(k) = \sum_{k=\delta_{1,n}}^{\infty} k^{-3} \ell^2(k) \sim \delta_{1,n}^{-2} \ell^2(\delta_{1,n}).$$

Since $\delta_{1,n} \asymp n \asymp \delta_{2,n}$, it follows that

$$L(n)^{-1} \sum_{k=\delta_{1,n}}^{\infty} \left(\sum_{j=k+1}^{k+\delta_{2,n}} c_j \right)^2 \leq C_2 \frac{\ell^2(n)}{L(n)} \frac{\ell^2(\delta_{1,n})}{\ell^2(n)} = o(1)$$

by the slow variation of ℓ^2 , and Theorem 1.5.9.a in BGT.

FR2. Consider x_n^+ first. We have

$$\begin{aligned}
 \text{Var}(x_n^+) &= \sum_{k=0}^{n-1} \left(\sum_{j=0}^k \phi_j c_{k-j} \right)^2 = \sum_{k=0}^{n-1} \left(\sum_{j=0}^k \phi_{k-j} c_j \right)^2 \\
 &= \sum_{k=0}^{n-1} \sum_{i=0}^k \sum_{j=0}^k c_i c_j \phi_{k-i} \phi_{k-j} \\
 (S.24) \quad &= \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} c_i c_j \sum_{k=i \vee j}^{n-1} \phi_{k-i} \phi_{k-j}
 \end{aligned}$$

Fix $i, j \in \mathbb{N}$, taking $i \geq j$ without loss of generality. Then

$$\sum_{k=i \vee j}^{n-1} \phi_{k-i} \phi_{k-j} = \sum_{k=i}^{n-1} \phi_{k-j} \phi_{k-i} = \sum_{k=0}^{n-i-1} \phi_{k+(i-j)} \phi_k = \sum_{k=0}^{n-i-1} \phi_k^2 \frac{\phi_{k+(i-j)}}{\phi_k}.$$

Since $\phi_k \sim k^{-1/2} \ell(k)$, which is regularly varying, $\phi_{k+(i-j)}/\phi_k \rightarrow 1$ as $k \rightarrow \infty$, while by Lemma B.2(iii),

$$\gamma_n^{-2} \sum_{k=0}^{n-i-1} \phi_k^2 \sim L(n)^{-1} \sum_{k=1}^n k^{-1} \ell^2(k) \rightarrow 1.$$

Hence by the Toeplitz Lemma (Hall and Heyde, 1980, p. 31),

$$(S.25) \quad \gamma_n^{-2} \sum_{k=i \vee j}^{n-1} \phi_{k-i} \phi_{k-j} = \gamma_n^{-2} \sum_{k=0}^{n-i-1} \phi_k^2 \frac{\phi_{k+(i-j)}}{\phi_k} \rightarrow 1$$

as $n \rightarrow \infty$, for each $i, j \in \mathbb{N}$ with $i \geq j$. Moreover, by the Cauchy-Schwarz inequality,

$$(S.26) \quad \gamma_n^{-2} \sum_{k=i \vee j}^{n-1} |\phi_{k-i} \phi_{k-j}| \leq \gamma_n^{-2} \sum_{k=0}^{n-1} \phi_k^2 < C < \infty$$

for some $C < \infty$ not depending on i or j . Since $\sum_{k=0}^{\infty} |c_k| < \infty$, it follows from (S.24)–(S.26) and the dominated convergence theorem that

$$\gamma_n^{-2} \text{Var}(x_n^+) \rightarrow \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} c_i c_j = \left(\sum_{i=0}^{\infty} c_i \right)^2.$$

With respect to x_n^- , we have:

$$\begin{aligned}
\text{Var}(x_n^-) &= \sum_{k=n}^{\infty} \left(\sum_{j=0}^{n-1} \phi_j c_{k-j} \right)^2 = \sum_{k=0}^{\infty} \left(\sum_{j=0}^{n-1} \phi_j c_{n+k-j} \right)^2 \\
&\leq 2 \sum_{k=0}^{\infty} \sum_{j=0}^{n-1} |\phi_j c_{n+k-j}| \sum_{i=j}^{n-1} |\phi_i c_{n+k-i}| \\
&\leq_{(1)} C \sum_{k=0}^{\infty} \sum_{j=0}^{n-1} \phi_j^2 |c_{n+k-j}| \sum_{i=j}^{n-1} |c_{n+k-i}| \\
&\leq_{(2)} C_1 \sum_{k=0}^{\infty} \sum_{j=0}^{n-1} \phi_j^2 |c_{n+k-j}|,
\end{aligned}$$

where $\leq_{(1)}$ holds since $\phi_k \sim k^{-1/2} \ell(k)$, which can be taken to be monotone without loss of generality (see Theorem 1.5.3 in BGT), while $\leq_{(2)}$ follows from $\sum_{i=j}^{n-1} |c_{n+k-i}| \leq \sum_{i=0}^{\infty} |c_i| < \infty$. Letting $\bar{c}_m := \sum_{k=m}^{\infty} |c_k|$, we thus have that $\text{Var}(x_n^-)$ is bounded, up to multiplicative constant, by

$$\begin{aligned}
\sum_{j=0}^{n-1} \phi_j^2 \sum_{k=0}^{\infty} |c_{n+k-j}| &= \sum_{j=0}^{n-1} \phi_j^2 \bar{c}_{n-j} \\
&\leq \left(\sum_{j=0}^{\lfloor n/2 \rfloor} + \sum_{j=\lfloor n/2 \rfloor+1}^{n-1} \right) \phi_j^2 \bar{c}_{n-j} \\
&\leq \bar{c}_{\lfloor n/2 \rfloor} \sum_{j=0}^{\lfloor n/2 \rfloor} \phi_j^2 + C \sum_{j=\lfloor n/2 \rfloor+1}^{n-1} \phi_j^2.
\end{aligned}$$

We claim that the r.h.s. is of smaller order than $\gamma_n^2 = L(n)$. For the first term, this follows from $\bar{c}_{\lfloor n/2 \rfloor} \rightarrow 0$ and $\sum_{j=0}^{\lfloor n/2 \rfloor} \phi_j^2 \sim L(\lfloor n/2 \rfloor) \sim L(n)$ by Lemma B.2(iii) and the slow variation of $L(n)$. By the same argument,

$$\begin{aligned}
\sum_{j=\lfloor n/2 \rfloor+1}^{n-1} \phi_j^2 &= \sum_{j=1}^{n-1} \phi_j^2 - \sum_{j=1}^{\lfloor n/2 \rfloor} \phi_j^2 \sim L(n) - L(\lfloor n/2 \rfloor) \\
&= L(n) \left[1 - \frac{L(\lfloor n/2 \rfloor)}{L(n)} \right] = o[L(n)].
\end{aligned}$$

III. By the same argument as which led to (S.24), and noting that

$\phi_j(n) = \rho_n^j$ for $\rho_n = 1 - \kappa_n^{-1}$ in this case, we have

$$\text{Var}(x_{t_n}^+(n)) = \sum_{i=0}^{t_n-1} \sum_{j=0}^{t_n-1} c_i c_j \sum_{k=i \vee j}^{t_n-1} \rho_n^{k-i} \rho_n^{k-j}.$$

Fix $i, j \in \mathbb{N}$, taking $i \geq j$ without loss of generality. Then

$$\sum_{k=i \vee j}^{t_n-1} \rho_n^{k-i} \rho_n^{k-j} = \sum_{k=i}^{t_n-1} \rho_n^{2(k-i)+(i-j)} = \rho_n^{i-j} \sum_{k=0}^{t_n-i-1} \rho_n^{2k} = \rho_n^{i-j} \frac{1 - \rho_n^{2(t_n-i)}}{1 - \rho_n^2}.$$

Since $1 - \rho_n^2 = (1 + \rho_n)(1 - \rho_n) \sim 2\kappa_n^{-1}$, it follows that

$$\kappa_n^{-1} \sum_{k=i \vee j}^{t_n-1} \rho_n^{k-i} \rho_n^{k-j} \sim 2^{-1} \rho_n^{i-j} (1 - \rho_n^{2(t_n-i)}) \rightarrow 2^{-1}$$

as $n \rightarrow \infty$ for each $i, j \in \mathbb{N}$ with $i \geq j$, since $\rho_n^{i-j} \rightarrow 1$ as $n \rightarrow \infty$ and

$$(S.27) \quad \rho_n^{t_n} = (1 - \kappa_n^{-1})^{t_n} = [(1 - \kappa_n^{-1})^{\kappa_n}]^{t_n/\kappa_n} \sim e^{-t_n/\kappa_n} \rightarrow 0$$

since $\kappa_n = o(t_n)$. Moreover, for all $i, j \in \{0, \dots, t_n\}$ with $i \geq j$, and all $n \in \mathbb{N}$,

$$\kappa_n^{-1} \sum_{k=i \vee j}^{t_n-1} \rho_n^{k-i} \rho_n^{k-j} \leq C |1 - \rho_n^{2(t_n-i)}| \leq C.$$

Hence it follows by the absolute summability of $\{c_s\}$ and the dominated convergence theorem that

$$\kappa_n^{-1} \text{Var}(x_{t_n}^+(n)) \rightarrow \frac{1}{2} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} c_i c_j = \frac{1}{2} \left(\sum_{i=0}^{\infty} c_i \right)^2.$$

Next, since the coefficients $\{a_{k,t}^-(n)\}$ are uniformly bounded by Lemma COEF, we have

$$\text{Var}(x_{t_n}^-(n)) = \sum_{k=t_n}^{\infty} [a_{k,t_n}^-(n)]^2 \leq C \sum_{k=t_n}^{\infty} |a_{k,t_n}^-(n)| \leq C \sum_{k=t_n}^{\infty} \sum_{j=0}^{t_n-1} |\rho_n^j c_{k-j}|.$$

Letting $\bar{c}_m := \sum_{k=m}^{\infty} |c_k|$, we can bound the r.h.s., up to a multiplicative constant, by

$$\sum_{j=0}^{t_n-1} \rho_n^j \sum_{k=t_n}^{\infty} |c_{k-j}| = \sum_{j=0}^{t_n-1} \rho_n^j \bar{c}_{t_n-j}$$

$$\begin{aligned}
&\leq \left(\sum_{j=0}^{\lfloor t_n/2 \rfloor - 1} + \sum_{j=\lfloor t_n/2 \rfloor}^{t_n-1} \right) \rho_n^j \bar{c}_{t_n-j} \\
&\leq_{(1)} C \left(\bar{c}_{\lfloor t_n/2 \rfloor} + \rho_n^{\lfloor t_n/2 \rfloor} \right) \sum_{j=0}^{t_n-1} \rho_n^j \\
&=_{(2)} o(\kappa_n)
\end{aligned}$$

where $\leq_{(1)}$ holds for all n sufficiently large that $\rho_n \in (0, 1)$, and $=_{(2)}$ from $\bar{c}_{\lfloor t_n/2 \rfloor} + \rho_n^{\lfloor t_n/2 \rfloor} \rightarrow 0$ (as per (S.27) above), and

$$\sum_{j=0}^{t_n-1} \rho_n^j = \frac{1 - \rho_n^{t_n}}{1 - \rho_n} = \kappa_n(1 - \rho_n^{t_n}) \sim \kappa_n. \quad \square$$

PROOF OF LEMMA CLT. By Lemma LVAR, we have

$$\text{Var}\{\gamma_n^{-1} x_{t_n}^+(n)\} \rightarrow \sigma_+^2 \quad \text{Var}\{\gamma_n^{-1} x_{t_n}^-(n)\} \rightarrow \sigma_-^2$$

for appropriate σ_+ and σ_- . Each of $\{\gamma_n^{-1} x_{t_n}^+(n)\}$ and $\{\gamma_n^{-1} x_{t_n}^-(n)\}$ are independent linear processes arrays with respective coefficients $\{\gamma_n^{-1} a_k(n)\}_{k=0}^{t_n-1}$ and $\{\gamma_n^{-1} a_{k,t_n}^-(n)\}_{k=t_n}^\infty$, which by Lemma COEF have the property that

$$\lim_{n \rightarrow \infty} \gamma_n^{-1} \left[\max_{0 \leq k \leq t_n-1} |a_k(n)| + \sup_{k \geq t_n} |a_{k,t_n}^-(n)| \right] = 0.$$

The postulated weak convergence thus follows by Lemma 2.1 in Abadir, Distaso, Giraitis and Koul (2014). \square

C.2. Proofs of Lemmas B.1 and B.2.

PROOF OF LEMMA B.1. Without loss of generality, we will prove the result as stated, but with \mathbb{N} in place of \mathbb{N}_0 , so that $\{\vartheta_k\}$ is indexed by $k \in \{1, 2, \dots\}$. Let $\eta := \sum_{k=1}^\infty \vartheta_k \xi_k$, and define

$$\psi_\eta(\lambda) := \mathbf{E} \exp(i\lambda\eta) = \prod_{k=1}^\infty \psi_\xi(\lambda\vartheta_k),$$

where the second equality holds by independence. Since $\mathbf{E} \xi_1^2 < \infty$, by a Taylor expansion of the log characteristic function, there exists a $\gamma \in (0, \infty)$ such that

$$(S.28) \quad |\psi_\xi(\lambda)| \leq e^{-\gamma\lambda^2} \vee e^{-\gamma}.$$

(i). Without loss of generality, suppose $\{\vartheta_k\}$ is ordered as so that $|\vartheta_1| \geq |\vartheta_2| \geq \dots$. Define

$$\mathcal{K} := \left\{ k \geq \theta + 1 \text{ and } k \in \mathbb{N} \mid \vartheta_k^2 \geq \frac{3\sigma_\vartheta^2}{2\pi} k^{-2} \right\}.$$

We claim that \mathcal{K} is nonempty. To see this, observe that by definition of \mathcal{K} and the stated condition on $\max_{k \in \mathbb{N}} \vartheta_k^2$,

$$\sum_{k \notin \mathcal{K}} \vartheta_k^2 = \sum_{k=1}^{\theta} \vartheta_k^2 + \sum_{\substack{k \notin \mathcal{K} \\ k \geq \theta+1}} \vartheta_k^2 \leq \theta \cdot \frac{\sigma_\vartheta^2}{2\theta} + \frac{3\sigma_\vartheta^2}{2\pi} \sum_{k=\theta+1}^{\infty} k^{-2} \leq \frac{3\sigma_\vartheta^2}{4}$$

where we have used that $\sum_{k=1}^{\infty} k^{-2} = \pi/6$. Hence

$$(S.29) \quad \sum_{k \in \mathcal{K}} \vartheta_k^2 \geq \frac{\sigma_\vartheta^2}{4} > 0,$$

and so \mathcal{K} is nonempty. Let k_0 denote its first element, noting that $k_0 \geq \theta + 1$ by construction.

We want to bound the integral of $|\psi_\eta(\lambda)|$ over $[A, \infty)$. To that end, decompose

$$[A, \infty) = [A, A \vee \vartheta_{k_0}^{-1}] \cup [A \vee \vartheta_{k_0}^{-1}, \infty),$$

and consider the integral on each of these two pieces separately. We have

$$\int_{\{|\lambda| \in [A, A \vee \vartheta_{k_0}^{-1}]\}} |\psi_\eta(\lambda)| d\lambda \leq \int_{\{|\lambda| \in [A, A \vee \vartheta_{k_0}^{-1}]\}} \prod_{k \in \mathcal{K}} |\psi_\xi(\lambda \vartheta_k)| d\lambda.$$

Since $|\lambda| \leq \vartheta_{k_0}^{-1}$ on the domain of integration, we have $|\lambda \vartheta_k| \leq \vartheta_{k_0}^{-1} \vartheta_k \leq 1$ for all $k \in \mathcal{K}$, and so by (S.28) and then (S.29),

$$\begin{aligned} \int_{\{|\lambda| \in [A, A \vee \vartheta_{k_0}^{-1}]\}} \prod_{k \in \mathcal{K}} |\psi_\xi(\lambda \vartheta_k)| d\lambda &\leq \int_{\{|\lambda| \in [A, A \vee \vartheta_{k_0}^{-1}]\}} \exp \left(-\gamma \lambda^2 \sum_{k \in \mathcal{K}} \vartheta_k^2 \right) d\lambda \\ &\leq \int_{\{|\lambda| \geq A\}} \exp \left(-\gamma \lambda^2 \sigma_\vartheta^2 / 4 \right) d\lambda. \\ &=: G_1(A; \sigma_\vartheta^2), \end{aligned}$$

where G_1 depends on ψ_ξ through γ . By a change of variables $G_1(A; \sigma_\vartheta^2) \leq C_\gamma \sigma_\vartheta^{-1}$, for some C_γ depending only on γ . Moreover, $\lim_{A \rightarrow \infty} G_1(A; \sigma_\vartheta^2) = 0$ and $\sigma^2 \mapsto G_1(A; \sigma^2)$ is decreasing in σ^2 , as required,

Next, noting that the leading k_0 elements of $\{\vartheta_k\}$ (as ordered) must be nonzero, we have

$$\begin{aligned} \int_{\{|\lambda| \geq A\vartheta_{k_0}^{-1}\}} |\psi_\eta(\lambda)| d\lambda &\leq \int_{\{|\lambda| \geq A\vartheta_{k_0}^{-1}\}} \prod_{k=1}^{k_0} |\psi_\xi(\lambda\vartheta_k)| d\lambda \\ &\leq e^{-\gamma(k_0-\theta)} \int_{\{|\lambda| \geq A\vartheta_{k_0}^{-1}\}} \prod_{k=1}^{\theta} |\psi_\xi(\lambda\vartheta_k)| d\lambda \end{aligned}$$

where the second equality follows from (S.28) and the fact that $|\lambda\vartheta_k| \geq |\vartheta_{k_0}^{-1}\vartheta_k| \geq 1$ on the domain of integration, for all $k \geq k_0$. By Hölder's inequality,

$$\begin{aligned} \int_{\{|\lambda| \geq A\vartheta_{k_0}^{-1}\}} \prod_{k=1}^{\theta} |\psi_\xi(\lambda\vartheta_k)| d\lambda &\leq \int_{\{|\lambda| \geq A\}} \prod_{k=1}^{\theta} |\psi_\xi(\lambda\vartheta_k)| d\lambda \\ &\leq \prod_{k=1}^{\theta} \left(\int_{\{|\lambda| \geq A\}} |\psi_\xi(\lambda\vartheta_k)|^\theta d\lambda \right)^{1/\theta} \\ (S.30) \quad &\leq \max_{1 \leq k \leq \theta} \vartheta_k^{-1} \int_{\{|\lambda| \geq A\vartheta_k\}} |\psi_\xi(\lambda)|^\theta d\lambda. \end{aligned}$$

Now $\vartheta_k \geq \vartheta_{k_0} \geq c_0 \sigma_\vartheta k_0^{-1}$ for $k \leq k_0$, where $c_0 := (3/2\pi)^{1/2}$, whence

$$\max_{1 \leq k \leq \theta} \vartheta_k^{-1} \int_{\{|\lambda| \geq A\vartheta_k\}} |\psi_\xi(\lambda)|^\theta d\lambda \leq c_0^{-1} \sigma_\vartheta^{-1} k_0 \int_{\{|\lambda| \geq A c_0 k_0^{-1} \sigma_\vartheta\}} |\psi_\xi(\lambda)|^\theta d\lambda.$$

Thus,

$$\begin{aligned} \int_{\{|\lambda| \geq A\vartheta_{k_0}^{-1}\}} |\psi_\eta(\lambda)| d\lambda &\leq c_0^{-1} e^{\gamma\theta} \sigma_\vartheta^{-1} e^{-\gamma k_0} k_0 \int_{\{|\lambda| \geq A c_0 k_0^{-1} \sigma_\vartheta\}} |\psi_\xi(\lambda)|^\theta d\lambda \\ &=: G_{2,k_0}(A; \sigma_\vartheta^2). \end{aligned}$$

Since the final integral is bounded by $\int |\psi_\xi(\lambda)|^\theta d\lambda < \infty$, and $e^{-\gamma k} k \rightarrow 0$ as $k \rightarrow \infty$, it is evident that

$$G_2(A; \sigma_\vartheta^2) := \sup_{k \geq \theta+1} G_{2,k}(A; \sigma_\vartheta^2) \leq C_{\psi_\xi} \sigma_\vartheta^{-1},$$

for some $C_{\psi_\xi} < \infty$ depending only on ψ_ξ .

Since each $G_{2,k}(A; \sigma^2)$ is weakly decreasing in σ^2 , so too must be $G_2(A; \sigma^2)$. It remains therefore to show that $G_2(A; \sigma_\vartheta^2) \rightarrow 0$ as $A \rightarrow \infty$. To that end,

let $\epsilon > 0$ and note that since $e^{-\gamma k} k \rightarrow 0$ as $k \rightarrow \infty$, we may choose k^* such that $G_{2,k}(A; \sigma_{\vartheta}^2) \leq \epsilon$ for all $k \geq k^*$. Hence

$$G_2(A; \sigma^2) \leq \epsilon + \max_{\theta+1 \leq k \leq k^*} G_{2,k}(A; \sigma_{\vartheta}^2) \rightarrow \epsilon$$

as $A \rightarrow \infty$, since $G_{2,k}(A; \sigma_{\vartheta}^2) \rightarrow 0$ as $A \rightarrow \infty$ for each k fixed.

(ii). Without loss of generality, we may take $k_i = i$ for each $i \in \{1, \dots, \theta\}$. Then by the same argument as which led to (S.30) above,

$$\begin{aligned} \int_{\mathbb{R}} |\psi_{\eta}(\lambda)| d\lambda &\leq \int_{\mathbb{R}} \prod_{k=1}^{\theta} |\psi_{\xi}(\lambda \vartheta_k)| d\lambda \\ &\leq \max_{1 \leq k \leq \theta} \int_{\mathbb{R}} |\psi_{\xi}(\lambda \vartheta_k)|^{\theta} d\lambda \\ &= \max_{1 \leq k \leq \theta} \vartheta_k^{-1} \int_{\mathbb{R}} |\psi_{\xi}(\lambda)|^{\theta} d\lambda \\ &\leq \delta^{-1} \int_{\mathbb{R}} |\psi_{\xi}(\lambda)|^{\theta} d\lambda. \end{aligned} \quad \square$$

PROOF OF LEMMA B.2. **(i).** The result follows by similar arguments to those given in Giraitis, Koul and Surgailis (2012, p. 20).

(ii). Since $g(j) = j^l \varsigma(j)$ with $l < -1$, by Theorem 1.5.3 in BGT, we may without loss of generality take ς to be such that g is monotone decreasing. Since $\sum_{j=n}^{\infty} \varphi_j \sim \sum_{j=n}^{\infty} g(j)$ as $n \rightarrow \infty$, we have

$$\frac{1}{ng(n)} \sum_{j=n}^{\infty} \varphi_j \sim \frac{1}{ng(n)} \sum_{j=n}^{\infty} g(j).$$

By monotonicity of g ,

$$\frac{\int_n^{\infty} g(x) dx}{ng(n)} \leq \frac{\sum_{j=n}^{\infty} g(j)}{ng(n)} \leq \frac{\int_{n-1}^{\infty} g(x) dx}{ng(n)}.$$

By Theorem 1.5.11 in BGT,

$$\frac{\int_n^{\infty} g(x) dx}{ng(n)} \rightarrow -\frac{1}{l+1} = \int_1^{\infty} x^l dx.$$

while by the preceding and Theorem 1.5.2 in BGT,

$$\frac{\int_{n-1}^{\infty} g(x) dx}{ng(n)} = \frac{(n-1)g(n-1)}{ng(n)} \cdot \frac{\int_{n-1}^{\infty} g(x) dx}{(n-1)g(n-1)} \rightarrow -\frac{1}{l+1}.$$

(iii). Set $s(x) := x^{-1}\zeta(x)$, and note that

$$\underline{s}(x) := \inf_{1 \leq u \leq x} s(u) \leq s(x) \leq \sup_{u \geq x} \bar{s}(u) =: \bar{s}(x).$$

Let $\epsilon > 0$. By Theorem 1.5.3 of BGT, $\underline{s}(x) \sim s(x) \sim \bar{s}(x)$ as $x \rightarrow \infty$, and so we may choose $x_0 \in \mathbb{N}$ with $x_0 \geq 2$ such that

$$(S.31) \quad (1 - \epsilon)s(x) \leq \underline{s}(x) \leq s(x) \leq \bar{s}(x) \leq (1 + \epsilon)s(x)$$

for all $x \geq x_0$. Since \underline{s} and \bar{s} are monotone decreasing, we also have

$$(S.32) \quad \int_{x_0}^n \underline{s}(u) du \leq \sum_{j=x_0}^n \underline{s}(j) \leq \sum_{j=x_0}^n s(j) \leq \sum_{j=x_0}^n \bar{s}(j) \leq \int_{x_0-1}^n \bar{s}(u) du.$$

By (S.31) and (S.32),

$$\sum_{j=x_0}^n s(j) \geq \int_{x_0}^n \underline{s}(u) du \geq (1 - \epsilon) \int_{x_0}^n s(u) du \sim (1 - \epsilon)S(n)$$

as $n \rightarrow \infty$, noting that since $S(n) \rightarrow \infty$ by hypothesis, $\int_x^n s(u) du \sim S(n)$ for each $x \geq 1$. Similarly,

$$\sum_{j=x_0}^n s(j) \leq (1 + \epsilon) \int_{x_0-1}^n s(u) du \sim (1 + \epsilon)S(n)$$

Deduce

$$\sum_{j=1}^n s(j) \sim \sum_{j=x_0}^n s(j) \sim S(n)$$

as $n \rightarrow \infty$. □

APPENDIX D: PROOFS OF THEOREMS 4.1 AND 4.2

PROOF OF THEOREM 4.1. Let $\kappa_{gn} := \kappa_g(\beta_n)$. By standard arguments,

$$(S.33) \quad \begin{aligned} n^{1/2} \begin{bmatrix} \hat{\mu} - \mu \\ \kappa_{gn}(\hat{\gamma} - \gamma) \end{bmatrix} &= \left\{ \frac{1}{n} \sum_{t=2}^n \begin{bmatrix} 1 & \kappa_{gn}^{-1} g(x_{t-1}) \\ \kappa_{gn}^{-1} g(x_{t-1}) & \kappa_{gn}^{-2} g^2(x_{t-1}) \end{bmatrix} \right\}^{-1} \\ &\quad \cdot \frac{1}{n^{1/2}} \sum_{t=2}^n \begin{bmatrix} 1 \\ \kappa_{gn}^{-1} g(x_{t-1}) \end{bmatrix} u_t \\ &=: M_n^{-1} S_n. \end{aligned}$$

Consider M_n first. We claim that

$$(S.34) \quad \frac{1}{n} \sum_{t=2}^n [\kappa_{gn}^{-j} g^j(x_{t-1}) - H_{n,t-1}^j] = o_p(1)$$

for $j \in \{1, 2\}$, where $H_{n,t} := H_g(x_{n,t-1})$. When $j = 1$, this follows immediately from the definition of an AHF. When $j = 2$, the l.h.s. is bounded (in absolute value) by

$$(S.35) \quad \frac{1}{n\kappa_{gn}^2} \sum_{t=2}^n |H_{n,t-1} R_g(x_{t-1}, \beta_n)| + \frac{1}{n\kappa_{gn}^2} \sum_{t=2}^n R_g^2(x_{t-1}, \beta_n).$$

The second term is $o_p(1)$ by (18). Since H_g^2 satisfies the conditions of Theorem 3.2, we have

$$\frac{1}{n} \sum_{t=2}^n H_{n,t-1}^2 = O_p(1),$$

whence the first term in (S.35) is also $o_p(1)$, by the Cauchy-Schwarz inequality. Thus by (S.34) and Theorem 3.2,

$$(S.36) \quad M_n = \frac{1}{n} \sum_{t=2}^n \begin{bmatrix} 1 & \kappa_{gn}^{-1} g(x_{t-1}) \\ H_{n,t-1}^j & \kappa_{gn}^{-2} g^2(x_{t-1}) \end{bmatrix} \xrightarrow{d} \int_{\mathbb{R}} \begin{bmatrix} 1 & H_g(x) \\ H_g(x) & H_g^2(x) \end{bmatrix} \varrho(x) dx =: M_{H_g}.$$

which is a.s. nonsingular, since H_g is not a.e. equal from a constant function by assumption.

We turn next to S_n . We first note that

$$\mathbf{E} \left| \frac{1}{n^{1/2}} \sum_{t=2}^n [\kappa_{gn}^{-1} g(x_{t-1}) - H_{n,t-1}] u_t \right|^2 = \frac{\sigma_u^2}{n\kappa_{gn}^2} \sum_{t=2}^n \mathbf{E} R_g(x_{t-1}(n), \beta_n)^2 = o_p(1)$$

by (18). Thus

$$S_n = \frac{1}{n^{1/2}} \sum_{t=2}^n \begin{bmatrix} 1 \\ H_{n,t-1} \end{bmatrix} u_t + o_p(1)$$

where the leading r.h.s. term is a vector martingale with conditional variance matrix

$$\frac{\sigma_u^2}{n} \sum_{t=2}^n \begin{bmatrix} 1 & H_{n,t-1} \\ H_{n,t-1} & H_{n,t-1}^2 \end{bmatrix} \xrightarrow{d} \sigma_u^2 M_{H_g}$$

by Theorem 3.2. Thus, if we can verify the requirements of Wang's (2014) martingale CLT (his Theorem 2.1), the convergence

$$(S.37) \quad S_n \xrightarrow{d} \sigma_u M_{H_g}^{1/2} \zeta$$

will hold jointly with (S.36), where $\zeta \sim N[0, I_2]$ is independent of X^- (and of the possibly random density ϱ that depends on X^-) whence the result follows.

Regarding Wang's CLT, the only condition that is not trivially satisfied in our setting is his Assumption 2. Since the mixing variate X^- is the weak limit of $\beta_n^{-1}x_n^-(n)$, which is \mathcal{F}_0 -measurable, it can be seen from the proof of Wang's CLT that his condition (2.3) is unnecessary in our case. Thus we need only to verify that $n^{-1/2} \max_{2 \leq t \leq n} |H_{n,t-1}| = o_p(1)$, which is equivalent (see e.g. Hall and Heyde, 1980, p. 53) to the Lindeberg condition that for each $\eta > 0$,

$$(S.38) \quad \frac{1}{n} \sum_{t=1}^{n-1} H_g^2(\beta_n^{-1}x_t(n)) \mathbf{1} \left\{ |H_g(\beta_n^{-1}x_t(n))| > \eta n^{1/2} \right\} = o_p(1)$$

as $n \rightarrow \infty$. Recall that H_g^2 is assumed to satisfy the conditions of Theorem 3.2: therefore so too does $F_A(x) := H_g^2(x) \mathbf{1}\{|H_g(x)| > A\}$ for each $A \in \mathbb{R}$. Now

$$\begin{aligned} \Gamma_n(A) &:= \frac{1}{n} \sum_{t=1}^{n-1} F_A(\beta_n^{-1}x_t(n)) \xrightarrow{d} \int_{\mathbb{R}} F_A(x) \varrho(x) dx \\ &= \int_{\mathbb{R}} H_g^2(x) \mathbf{1}\{|H_g(x)| > A\} \varrho(x) dx =: \Gamma(A) \end{aligned}$$

by Theorem 3.2. By dominated convergence, $\Gamma(A) \xrightarrow{a.s.} 0$ as $A \rightarrow \infty$, since $H_g^2 \varrho$ is integrable. Therefore for a given $\epsilon > 0$, we may choose A_ϵ such that $\mathbf{P}\{\Gamma(A_\epsilon) \geq \epsilon\} \leq \epsilon$, whence

$$\begin{aligned} \limsup_{n \rightarrow \infty} \mathbf{P}\{\Gamma_n(\eta n^{1/2}) \geq \epsilon\} &\leq_{(1)} \limsup_{n \rightarrow \infty} \mathbf{P}\{\Gamma_n(A_\epsilon) \geq \epsilon\} \\ &\leq_{(2)} \mathbf{P}\{\Gamma(A_\epsilon) \geq \epsilon\} \leq \epsilon \end{aligned}$$

where $\leq_{(1)}$ holds since $A_\epsilon < \eta n^{1/2}$ for all n sufficiently large, and $\leq_{(2)}$ follows by the portmanteau theorem. Deduce $\Gamma_n(\eta n^{1/2}) \xrightarrow{p} 0$, i.e. (S.38) holds. \square

PROOF OF THEOREM 4.2. (i). We first prove (21). By Theorem 3.4,

$$(S.39) \quad \frac{\beta_n}{h_n n} \sum_{t=2}^n K^j \left(\frac{x_t(n) - x}{h_n} \right)^2 \xrightarrow{d} \varrho(0) \int_{\mathbb{R}} K^2(u) du =: V_{K^j}$$

jointly for $j \in \{1, 2\}$. Since m has bounded derivative, $|m(x_{t-1}) - m(x)| \leq C|x_{t-1} - x|$, and thus

$$\begin{aligned} \left| \sum_{t=2}^n K\left(\frac{x_{t-1} - x}{h_n}\right) [m(x_{t-1}) - m(x)] \right| &\leq Ch_n \sum_{t=2}^n \left| K\left(\frac{x_{t-1} - x}{h_n}\right) \frac{x_{t-1} - x}{h_n} \right| \\ &= O_p(h_n^2 n / \beta_n) \end{aligned}$$

also by Theorem 3.4. Hence

$$\begin{aligned} \text{(S.40)} \quad &\left(\frac{h_n n}{\beta_n}\right)^{1/2} [\hat{m}(x) - m(x)] \\ &= \frac{\left(\frac{\beta_n}{h_n n}\right)^{1/2} \sum_{t=2}^n K\left(\frac{x_{t-1} - x}{h_n}\right) u_t}{\frac{\beta_n}{h_n n} \sum_{t=2}^n K\left(\frac{x_{t-1} - x}{h_n}\right)} + O_p\left(\frac{h_n^3 n}{\beta_n}\right)^{1/2}, \end{aligned}$$

where the second r.h.s. term is $o_p(1)$ since $nh_n^3/\beta_n \rightarrow 0$ by assumption. The first r.h.s. term is a martingale with conditional variance

$$\sigma_u^2 \frac{\beta_n}{h_n n} \sum_{t=2}^n K^2\left(\frac{x_{t-1} - x}{h_n}\right) \xrightarrow{d} \sigma_u^2 V_{K^2}$$

by (S.39). Thus, similarly to the argument given in the proof of Theorem 4.1, if we can show the following Lindeberg condition holds, that

$$\text{(S.41)} \quad \frac{\beta_n}{h_n n} \sum_{t=2}^n K^2\left(\frac{x_{t-1} - x}{h_n}\right) \mathbf{1} \left\{ \left| K\left(\frac{x_{t-1} - x}{h_n}\right) \right| > \eta \left(\frac{h_n n}{\beta_n}\right)^{1/2} \right\} = o_p(1),$$

for each $\eta > 0$, then by Wang's (2014) martingale CLT,

$$\text{(S.42)} \quad \left(\frac{\beta_n}{h_n n}\right)^{1/2} \sum_{t=2}^n K\left(\frac{x_{t-1} - x}{h_n}\right) u_t \xrightarrow{d} \zeta \sigma_u V_{K^2}^{1/2}$$

jointly with (S.39), for $\zeta \sim N[0, 1]$ independent of $\varrho(0)$. (21) then follows from (S.39), (S.40) and (S.42).

Finally, to verify (S.41), we note that for n sufficiently large, the l.h.s of (S.41) is bounded by

$$\frac{\beta_n}{h_n n} \sum_{t=2}^n K^2\left(\frac{x_{t-1} - x}{h_n}\right) \mathbf{1} \left\{ \left| K\left(\frac{x_{t-1} - x}{h_n}\right) \right| > A \right\}$$

$$\xrightarrow{d}_{(1)} \varrho(0) \int_{\mathbb{R}} K^2(u) \mathbf{1}\{|K(u)| > A\} du \xrightarrow{a.s.}_{(2)} 0$$

where $\xrightarrow{d}_{(1)}$ holds by Theorem 3.4 as $n \rightarrow \infty$, and $\xrightarrow{a.s.}_{(2)}$ by dominated convergence as $A \rightarrow \infty$. (S.41) thus follows by similar arguments as were used to prove (S.38) above.

(ii). We next prove (22), using arguments similar to those given in part (i). Let $X'_t := (1, x_{t-1} - x)$, $K_{th} := K[h_n^{-1}(x_{t-1} - x)]$, $H_n := \sum_{t=2}^n K_{th} X_t X'_t$ and $\Lambda_n := (nh_n/\beta_n)^{1/2} \text{diag}\{1, h_n\}$. Then the LL estimator can be written as

$$\tilde{\mathbf{m}}(x) := \begin{bmatrix} \tilde{m}(x) \\ \tilde{m}^{(1)}(x) \end{bmatrix} = H_n^{-1} \sum_{t=2}^n X_t K_{th} y_t,$$

and by standard arguments decomposes as

$$\begin{aligned} \tilde{\mathbf{m}}(x) - \mathbf{m}(x) &= H_n^{-1} \left\{ \sum_{t=2}^n K_{th} X_t m(x_{t-1}) - H_n \mathbf{m}(x) \right\} + H_n^{-1} \sum_{t=2}^n K_{th} X_t u_t \\ &=: H_n^{-1} R_n + H_n^{-1} S_n. \end{aligned}$$

where $\mathbf{m}(x) := [m(x), m^{(1)}(x)]'$.

We consider each of H_n , R_n and S_n in turn. For H_n , we have

$$\begin{aligned} \Lambda_n^{-1} H_n \Lambda_n^{-1} &= \frac{\beta_n}{nh_n} \sum_{t=2}^n K \left(\frac{x_{t-1} - x}{h_n} \right) \begin{bmatrix} 1 & \frac{x_{t-1} - x}{h_n} \\ \frac{x_{t-1} - x}{h_n} & \left(\frac{x_{t-1} - x}{h_n} \right)^2 \end{bmatrix} \\ \text{(S.43)} \quad &\xrightarrow{d} \varrho(0) \int \begin{bmatrix} 1 & u \\ u & u^2 \end{bmatrix} K(u) du \end{aligned}$$

by Theorem 3.4. For R_n ,

$$\begin{aligned} R_n &= \sum_{t=2}^n K_{th} X_t \{m(x_{t-1}) - X'_t \mathbf{m}(x)\} \\ &= \sum_{t=2}^n K_{th} X_t \{m(x_{t-1}) - m(x) - m^{(1)}(x)(x_{t-1} - x)\} \\ &= \sum_{t=2}^n K_{th} X_t m^{(2)}(\bar{x}_{t-1})(x_{t-1} - x)^2 \end{aligned}$$

by Taylor's theorem, for some \bar{x}_{t-1} lying between x_{t-1} and x . Thus,

$$\text{(S.44)} \quad \Lambda_n^{-1} R_n = h_n^2 \left(\frac{\beta_n}{nh_n} \right)^{1/2} \sum_{t=2}^n K_{th} \begin{bmatrix} [h_n^{-1}(x_{t-1} - x)]^2 \\ [h_n^{-1}(x_{t-1} - x)]^3 \end{bmatrix} m^{(2)}(\bar{x}_{t-1}),$$

which since $m^{(2)}$ is bounded, is bounded (in norm) by a multiple of

$$(S.45) \quad h_n^2 \left(\frac{\beta_n}{nh_n} \right)^{1/2} \sum_{t=2}^n |K_{th}| \left\| \begin{bmatrix} [h_n^{-1}|x_{t-1} - x|^2] \\ [h_n^{-1}|x_{t-1} - x|^3] \end{bmatrix} \right\| \\ =_{(1)} O_p \left[h_n^2 \left(\frac{nh_n}{\beta_n} \right)^{1/2} \right] = O_p \left(\frac{nh_n^5}{\beta_n} \right)^{1/2} =_{(2)} o_p(1)$$

where $=_{(1)}$ is by Theorem 3.4, and $=_{(2)}$ by the assumption that $nh_n^5/\beta_n \rightarrow 0$. Finally, since $\Lambda_n^{-1}S_n$ is a martingale with conditional variance matrix

$$\sigma_u^2 \frac{\beta_n}{nh_n} \sum_{t=2}^n K_{th}^2 \begin{bmatrix} 1 & \frac{x_{t-1}-x}{h_n} \\ \frac{x_{t-1}-x}{h_n} & \left(\frac{x_{t-1}-x}{h_n} \right)^2 \end{bmatrix} \\ \xrightarrow{d} \varrho(0) \int \begin{bmatrix} 1 & u \\ u & u^2 \end{bmatrix} K^2(u) du =: \varrho(0)V$$

by Theorem 3.4, it follows by Wang's (2014) CLT and similar arguments as were used in part (i) that

$$(S.46) \quad \Lambda_n^{-1}S_n \xrightarrow{d} \sigma_u \varrho(0)^{1/2} V^{1/2} \zeta$$

jointly with (S.43), where $\zeta \sim N[0, I_2]$ is independent of $\varrho(0)$. Since

$$\Lambda_n[\tilde{\mathbf{m}}(x) - \mathbf{m}(x)] = [\Lambda_n^{-1}H_n\Lambda_n^{-1}]^{-1}\Lambda_n^{-1}R_n + \Lambda_n^{-1}S_n$$

the result now follows from (S.43)–(S.46), noting that the r.h.s. of (S.43) is invertible under condition (i). \square

APPENDIX E: PROOF OF THEOREM 5.1

The proof is divided into four parts. Part I derives the asymptotic distribution of \tilde{F} under \mathcal{H}_0 . Part II derives the asymptotics of $(\hat{\mu}, \hat{\gamma})$, and part III those of $\tilde{\sigma}_u^2(x)$, under \mathcal{H}_1 . Finally, part IV draws on the results of parts II and III to derive the asymptotics of \tilde{F} under \mathcal{H}_1 .

I. Asymptotics of \tilde{F} under \mathcal{H}_0 . Recall from (24) that \tilde{F} is constructed from an ensemble of t statistics of the form

$$\tilde{t}(x; \hat{\mu}, \hat{\gamma}) := \left[\frac{\sum_{t=2}^n K[(x_{t-1} - x)/h_n]}{\tilde{\sigma}_u^2(x)Q_{11}} \right]^{1/2} [\tilde{m}_g(x) - \hat{\mu} - \hat{\gamma}g(x)]$$

for x taking values in some finite set $\mathcal{X} \subset \mathbb{R}$. Let $x \in \mathcal{X}$ be fixed. We shall show that:

$$(S.47) \quad (nh_n/\beta_n)^{1/2}[\tilde{m}_g(x) - m(x)] \xrightarrow{d} \zeta_x \sigma_u \varrho(0)^{-1/2} Q_{11}^{1/2},$$

jointly over $x \in \mathcal{X}$, where $\zeta_x \sim N[0, 1]$ is independent of $\zeta_{x'} \sim N[0, 1]$ for each $x, x' \in \mathcal{X}$; and that

$$(S.48) \quad \tilde{\sigma}_u^2(x) \xrightarrow{p} \sigma_u^2$$

for each $x \in \mathcal{X}$. Since $(\beta_n/nh_n) \sum_{t=2}^n K[(x_{t-1} - x)/h_n] \xrightarrow{d} \varrho(0)$ by Theorem 3.4, and $\hat{\mu}$ and $\hat{\gamma}$ are consistent at rates $n^{-1/2}$ and $n^{-1/2}\kappa_g(\beta_n) = O_p(n^{-1/2})$ respectively by Theorem 4.1, it will then follow that

$$\begin{aligned} \tilde{t}(x; \hat{\mu}, \hat{\gamma}) &= \tilde{t}(x; \mu, \gamma) + O_p(h_n/\beta_n) \\ &= \left[\frac{\sum_{t=2}^n K[(x_{t-1} - x)/h_n]}{(\sigma_u^2 + o_p(1))Q_{11}} \right]^{1/2} [\tilde{m}_g(x) - m(x)] \xrightarrow{d} \zeta_x \end{aligned}$$

whence

$$\tilde{F} = \sum_{x \in \mathcal{X}} \tilde{t}(x; \hat{\mu}, \hat{\gamma})^2 \xrightarrow{d} \sum_{x \in \mathcal{X}} \zeta_x^2 \sim \chi_p^2$$

as required.

It thus remains to prove (S.47) and (S.48). Consider (S.47) first: the proof uses arguments similar to those used in part (ii) of the proof of Theorem 4.2. Define $X'_t := [1, g(x_{t-1}) - g(x)]$, $K_{th} := K[h_n^{-1}(x_{t-1} - x)]$, $\Lambda_n := (nh_n/\beta_n)^{1/2} \text{diag}\{1, h_n\}$ and $H_n := \sum_{t=2}^n K_{th} X_t X'_t$, noting how the definition of X_t has been modified. Recall that \tilde{m}_g is obtained by a K_{th} -weighted regression of y_t on a constant and $g(x_{t-1}) - g(x)$. Therefore under \mathcal{H}_0 , this estimator suffers from no approximation bias, and standard arguments give

$$(S.49) \quad \tilde{m}_g(x) - m(x) = e'_1 H_n^{-1} \sum_{t=2}^n K_{th} X_t u_t =: e'_1 H_n^{-1} S_n(x),$$

where $e'_1 := (1, 0)$.

We shall consider each of H_n and S_n in turn. We claim that

$$(S.50) \quad \Lambda_n^{-1} H_n \Lambda_n^{-1} \xrightarrow{d} \varrho(0) \int \begin{bmatrix} 1 & g^{(1)}(x)u \\ g^{(1)}(x)u & g^{(1)}(x)^2 u^2 \end{bmatrix} K(u) du =: \varrho(0) \tilde{H}(x),$$

and recall that $g^{(1)}(x) \neq 0$ by condition (ii). Consider e.g. the (2, 2) element of the l.h.s. matrix: this is equal to

$$(S.51) \quad \frac{\beta_n}{nh_n^3} \sum_{t=2}^n K_{ht} [g(x_{t-1}) - g(x)]^2$$

$$= \frac{\beta_n}{nh_n^3} \sum_{t=2}^n K_{ht} [g^{(1)}(x)(x_{t-1} - x) + g^{(2)}(\bar{x}_{t-1})(x_{t-1} - x)^2]^2$$

for some \bar{x}_{t-1} lying between x_{t-1} and x , by a Taylor expansion. By Theorem 4.2,

$$(S.52) \quad \frac{\beta_n}{nh_n^3} \sum_{t=2}^n K_{ht}(x_{t-1} - x)^2 \\ = \frac{\beta_n}{nh_n} \sum_{t=2}^n K \left(\frac{x_{t-1} - x}{h_n} \right) \left(\frac{x_{t-1} - x}{h_n} \right)^2 \xrightarrow{d} \varrho(0) \int u^2 K(u) du,$$

whereas, since $K(u)u^4$ is integrable by condition (i)

$$\frac{\beta_n}{nh_n^3} \sum_{t=2}^n K_{ht}(x_{t-1} - x)^4 = O_p(h_n^2) = o_p(1),$$

whence by the Cauchy–Schwarz inequality,

$$\frac{\beta_n}{nh_n^3} \sum_{t=2}^n K_{ht} [g(x_{t-1}) - g(x)]^2 = g^{(1)}(x)^2 \frac{\beta_n}{nh_n^3} \sum_{t=2}^n K_{ht}(x_{t-1} - x)^2 + o_p(1) \\ \xrightarrow{d} g^{(1)}(x)^2 \varrho(0) \int u^2 K(u) du.$$

This gives the claimed convergence for the $(2, 2)$ element in (S.50); the result for the other elements follows by analogous arguments.

We next turn to S_n . For each $x \in \mathcal{X}$, $\Lambda_n^{-1} S_n(x)$ is a vector martingale with conditional variance

$$\sigma_u^2 \frac{\beta_n}{nh_n} \sum_{t=2}^n K_{th}^2 \begin{bmatrix} 1 & \frac{g(x_{t-1}) - g(x)}{h_n} \\ \frac{g(x_{t-1}) - g(x)}{h_n} & \left(\frac{g(x_{t-1}) - g(x)}{h_n} \right)^2 \end{bmatrix} \\ \xrightarrow{d}_{(1)} \varrho(0) \int \begin{bmatrix} 1 & g^{(1)}(x)u \\ g^{(1)}(x)u & g^{(1)}(x)^2 u^2 \end{bmatrix} K^2(u) du =: \varrho(0) V(x)$$

where $\xrightarrow{d}_{(1)}$ follows by arguments identical to those used to prove (S.50), with K^2 in place of K . Moreover, the conditional covariation of $\Lambda_n^{-1} S_n(x)$ with $\Lambda_n^{-1} S_n(x')$ for $x' \neq x$ is a 2×2 matrix with (i, j) element

$$(S.53) \quad \sigma_u^2 \frac{\beta_n}{nh_n} \sum_{t=2}^n K_{th}(x) K_{th}(x') \left(\frac{g(x_{t-1}) - g(x)}{h_n} \right)^{i-1} \left(\frac{g(x_{t-1}) - g(x')}{h_n} \right)^{j-1},$$

where $K_{th}(x) := K[h_n^{-1}(x_{t-1} - x)]$. We claim this is $o_p(1)$, in which case it will follow by an application of Wang's (2014) CLT that $\Lambda_n^{-1}S_n(x)$ and $\Lambda_n^{-1}S_n(x')$ are asymptotically independent, with

$$(S.54) \quad \Lambda_n^{-1}S_n(x) \xrightarrow{d} \varrho^{1/2}(0)V^{1/2}(x)\zeta_x$$

jointly with (S.50) over all $x \in \mathcal{X}$, where $\zeta_x \sim N[0, I_2]$ is independent of $\zeta_{x'} \sim N[0, I_2]$ for $x' \neq x$.

To verify that (S.53) is indeed $o_p(1)$, note that by similar arguments to those given in (S.51) above, we can bound (S.53) by linear combinations of functionals of the form

$$\sigma_u^2 \frac{\beta_n}{nh_n} \sum_{t=2}^n L_a\left(\frac{x_{t-1} - x}{h_n}\right) L_b\left(\frac{x_{t-1} - x'}{h_n}\right)$$

where $L_a(u) = |K(u)u^a|$ for $a \in \{0, 1, 2\}$, which is bounded and integrable by condition (i). Finally, note that by the proof of Theorem 3.2, $\{x_t(n)\}$ satisfies **HL3** (for some $t_0 \in \mathbb{N}$) and **HL6**. Hence for $t \geq t_0$ and $n \geq n_0$,

$$(S.55) \quad \begin{aligned} & \frac{1}{h_n} \mathbf{E} L_a\left(\frac{x_t - x}{h_n}\right) L_b\left(\frac{x_t - x'}{h_n}\right) \\ &= \frac{1}{h_n} \int_{\mathbb{R}} L_a\left(\frac{\beta_t u - x}{h_n}\right) L_b\left(\frac{\beta_t u - x'}{h_n}\right) \mathcal{D}_{n,t}(u) du \\ &\leq \sup_{n \geq n_0, t_0 \leq t \leq n} \sup_{v \in \mathbb{R}} |\mathcal{D}_{n,t}(v)| \beta_t^{-1} \int_{\mathbb{R}} L_a(u) L_b(u - h_n^{-1}(x' - x)) du. \end{aligned}$$

Since L_4 is bounded by condition (i), it must be the case that $L_b(u) \rightarrow 0$ as $|u| \rightarrow \infty$ for $b \in \{0, 1, 2\}$, whence the r.h.s. integral converges to zero by the dominated convergence theorem as $h_n \rightarrow 0$. Thus

$$\begin{aligned} & \frac{\beta_n}{nh_n} \sum_{t=2}^n \mathbf{E} L_a\left(\frac{x_{t-1} - x}{h_n}\right) L_b\left(\frac{x_{t-1} - x'}{h_n}\right) \\ &= \frac{\beta_n}{nh_n} \sum_{t=t_0}^{n-1} \int_{\mathbb{R}} L_a\left(\frac{\beta_t^{-1}x_t - x}{\beta_t^{-1}h_n}\right) \left(\frac{\beta_t^{-1}x_t - x}{\beta_t^{-1}h_n}\right) + o(1) \\ &= o\left(\frac{\beta_n}{n} \sum_{t=t_0}^{n-1} \beta_t^{-1} + 1\right) = o(1), \end{aligned}$$

as was required, where the final equality follows since **HL6** holds, as was noted above.

(S.47) now follows from (S.49), (S.50), (S.54) and the fact that

$$Q_{11} = e_1' \tilde{H}(x)^{-1} V(x) \tilde{H}(x)^{-1} e_1$$

as may be verified by direct calculation.

We turn therefore to (S.48). We have

$$(S.56) \quad \tilde{\sigma}_u^2(x) = \frac{\sum_{t=2}^n [(\mu - \hat{\mu}) + (\gamma - \hat{\gamma})g(x_{t-1}) + u_t]^2 K_{th}}{\sum_{t=2}^n K_{th}}.$$

Recognising

$$\begin{aligned} & |g^2(x_{t-1}) - g^2(x)| \\ &= |[g(x_{t-1}) - g(x)]^2 + 2[g(x_{t-1}) - g(x)]g(x)| \\ &\leq 2|g^{(1)}(x)|^2(x_{t-1} - x)^2 + 2|g^{(2)}(\bar{x}_{t-1})|^2(x_{t-1} - x)^4 \\ &\quad + 2|g(x)|\{|g^{(1)}(x)||x_{t-1} - x| + |g^{(2)}(\bar{x}_{t-1})||x_{t-1} - x|^2\}, \end{aligned}$$

for some \bar{x}_{t-1} lying between x_{t-1} and x , and that $g^{(2)}$ is bounded, it follows from Theorem 4.2 (arguing similarly as in (S.52) above) that

$$(S.57) \quad \frac{\beta_n}{nh_n} \sum_{t=2}^n K_{th} g^2(x_{t-1}) = g^2(x) \frac{\beta_n}{nh_n} \sum_{t=2}^n K_{th} + O_p(h_n) \xrightarrow{d} g^2(x) \varrho(0).$$

Since $\{u_t^2 - \sigma_u^2, \mathcal{F}_t\}$ is a martingale difference sequence and K_{th} is \mathcal{F}_{t-1} -measurable, $(\beta_n/nh_n) \sum_{t=2}^n K_{th}(u_t^2 - \sigma_u^2)$ is a martingale with conditional variance

$$\begin{aligned} & \left(\frac{\beta_n}{nh_n} \right)^2 \sum_{t=2}^n K_{th}^2 \mathbf{E}[(u_t^2 - \sigma_u^2)^2 | \mathcal{F}_{t-1}] \leq_{(1)} Z \left(\frac{\beta_n}{nh_n} \right)^2 \sum_{t=t_0}^n K_{th}^2 + o_p(1) \\ &=_{(2)} O_p \left(\frac{\beta_n}{nh_n} \right) + o_p(1) \\ &\rightarrow 0 \end{aligned}$$

where $\leq_{(1)}$ holds by $Z := \sup_t \mathbf{E}[u_t^4 | \mathcal{F}_{t-1}] < \infty$ a.s., and $=_{(2)}$ by Theorem 3.2. Thus $(\beta_n/nh_n) \sum_{t=2}^n K_{th}(u_t^2 - \sigma_u^2) = o_p(1)$ by Corollary 3.1 in Hall and Heyde (1980), whence

$$(S.58) \quad \frac{\sum_{t=2}^n u_t^2 K_{th}}{\sum_{t=2}^n K_{th}} = \sigma_u^2 + \frac{\sum_{t=2}^n K_{th}(u_t^2 - \sigma_u^2)}{\sum_{t=2}^n K_{th}} \xrightarrow{p} \sigma_u^2.$$

Since $(\beta_n/nh_n) \sum_{t=2}^n K_{th} \xrightarrow{d} \varrho(0)$ by Theorem 3.2, (S.57), (S.58), the consistency of $(\hat{\mu}, \hat{\gamma})$ (from Theorem 4.1) and applications of the Cauchy–Schwarz inequality to the cross-product terms on the r.h.s. of (S.56) yield

$$\tilde{\sigma}_u^2(x) = \frac{\sum_{t=2}^n u_t^2 K_{th}}{\sum_{t=2}^n K_{th}} + o_p(1) \xrightarrow{p} \sigma_u^2,$$

i.e. (S.48) holds.

II. Asymptotics of $(\hat{\mu}, \hat{\gamma})$ under \mathcal{H}_1 . Recall that

$$m(x) = \mu + \gamma g(x) + r_n g_1(x)$$

under \mathcal{H}_1 . Let $\kappa_{gn} := \kappa_g(\beta_n)$ and $D_n := n^{1/2} \text{diag}\{1, \kappa_{gn}\}$.

(a). Suppose condition (iii.a) holds. Let $R'_n := r_n^{-1} n^{-1/2} \beta_n$, which is $o(1)$ by assumption. By (S.33) and subsequent arguments given in the proof of Theorem 4.1,

$$\begin{aligned} R'_n D_n \begin{bmatrix} \hat{\mu} - \mu \\ \hat{\gamma} - \gamma \end{bmatrix} &= M_n^{-1} \frac{R'_n}{n^{1/2}} \sum_{t=2}^n \begin{bmatrix} 1 \\ \kappa_{gn}^{-1} g(x_{t-1}) \end{bmatrix} [u_t + r_n g_1(x_{t-1})] \\ &=: M_n^{-1} [R'_n S_n + R'_n S_n^\dagger] \end{aligned}$$

where $M_n \xrightarrow{d} M_{H_g}$ and $S_n = O_p(1)$ by (S.36) and (S.37) respectively, and

$$R'_n S_n^\dagger = \frac{\beta_n}{n} \sum_{t=2}^n \begin{bmatrix} g_1(x_{t-1}) \\ \kappa_{gn}^{-1} g(x_{t-1}) g_1(x_{t-1}) \end{bmatrix} \xrightarrow{d} \begin{bmatrix} \varrho(0) \int g_1 \\ 0 \end{bmatrix}$$

by Theorem 3.4, since g_1 and $g \cdot g_1$ are integrable by assumption, and $\kappa_{gn} \rightarrow \infty$. Since $R'_n \rightarrow 0$, it follows that

$$R'_n D_n \begin{bmatrix} \hat{\mu} - \mu \\ \hat{\gamma} - \gamma \end{bmatrix} \xrightarrow{d} \left\{ \int \begin{bmatrix} 1 & H_g \\ H_g & H_{g^2} \end{bmatrix} \varrho \right\}^{-1} \begin{bmatrix} \varrho(0) \int g_1 \\ 0 \end{bmatrix}.$$

(b). Suppose condition (iii.b) holds. Let $\kappa_{g_1 n} := \kappa_{g_1}(\beta_n)$ and $R''_n := r_n^{-1} n^{-1/2} \kappa_{g_1 n}^{-1}$; the latter is $o(1)$ by assumption. By analogous arguments to those given in part (a),

$$\begin{aligned} R''_n D_n \begin{bmatrix} \hat{\mu} - \mu \\ \hat{\gamma} - \gamma \end{bmatrix} &= M_n^{-1} \frac{R''_n}{n^{1/2}} \sum_{t=2}^n \begin{bmatrix} 1 \\ \kappa_{gn}^{-1} g(x_{t-1}) \end{bmatrix} r_n g_1(x_{t-1}) + o_p(1) \\ &= M_n^{-1} \frac{1}{n} \sum_{t=2}^n \begin{bmatrix} \kappa_{g_1 n}^{-1} g_1(x_{t-1}) \\ \kappa_{gn}^{-1} g(x_{t-1}) \kappa_{g_1 n}^{-1} g_1(x_{t-1}) \end{bmatrix} + o_p(1) \end{aligned}$$

$$\begin{aligned}
&=_{(1)} \frac{1}{n} \sum_{t=2}^n \left[\frac{H_{g_1}(x_{t-1}/\beta_n)}{H_g(x_{t-1}/\beta_n)H_{g_1}(x_{t-1}/\beta_n)} \right] + o_p(1) \\
&\xrightarrow{(2)} \left\{ \int \begin{bmatrix} 1 & H_g \\ H_g & H_{g^2} \end{bmatrix} \varrho \right\}^{-1} \begin{bmatrix} \int H_{g_1} \varrho \\ \int H_g H_{g_1} \varrho \end{bmatrix} =: - \begin{bmatrix} \mu_* \\ \gamma_* \end{bmatrix},
\end{aligned}
\tag{S.59}$$

where $=_{(1)}$ follows by straightforward calculations, since g and g_1 are AHF, and $\xrightarrow{(2)}$ follows by Theorem 3.2.

III. Asymptotics of $\tilde{\sigma}_u^2(x)$ under \mathcal{H}_1 . We have

$$\tilde{\sigma}_u^2(x) = \frac{\sum_{t=2}^n [(\mu - \hat{\mu}) + (\gamma - \hat{\gamma})g(x_{t-1}) + r_n g_1(x_{t-1}) + u_t]^2 K_{th}}{\sum_{t=2}^n K_{th}}.
\tag{S.60}$$

(a). Suppose condition (iii.a) holds. The argument here is similar to the proof of (S.48) given in part I. Recall (S.57), noting that this also holds with g_1 in place of g , since g_1 is also assumed to have bounded second derivative. Thus

$$\frac{\sum_{t=2}^n f^2(x_{t-1}) K_{th}}{\sum_{t=2}^n K_{th}} \xrightarrow{p} f^2(x)$$

for $f \in \{g, g_1\}$. By the analysis of part II(a) of the proof, and noting that both diagonal elements of

$$R'_n D_n = (r_n^{-1} n^{-1/2} \beta_n) n^{1/2} \text{diag}\{1, \kappa_{gn}\} = r_n^{-1} \beta_n \text{diag}\{1, \kappa_{gn}\}
\tag{S.61}$$

are divergent, it follows that $(\hat{\mu}, \hat{\gamma}) \xrightarrow{p} (\mu, \gamma)$. In view of $r_n = o(1)$ and (S.58), the preceding facts and applications of the Cauchy–Schwarz inequality to the cross-product terms on the r.h.s. of (S.60) yield

$$\tilde{\sigma}_u^2(x) = \frac{\sum_{t=2}^n u_t^2 K_{th}}{\sum_{t=2}^n K_{th}} + o_p(1) \xrightarrow{p} \sigma_u^2.$$

(b). Suppose condition (iii.b) holds. Recall the analysis given in part II(b) of the proof, and note that

$$\begin{aligned}
R''_n D_n &= (r_n^{-1} n^{-1/2} \kappa_{g_1 n}^{-1}) n^{1/2} \text{diag}\{1, \kappa_{gn}\} \\
&= \text{diag}\{r_n^{-1} \kappa_{g_1 n}^{-1}, r_n^{-1} \kappa_{g_1 n}^{-1} \kappa_{gn}\}.
\end{aligned}
\tag{S.62}$$

The behaviour of these sequences, and thus that of $\tilde{\sigma}_u^2(x)$, will depend on $\kappa_{**} := \lim_{n \rightarrow \infty} r_n \kappa_{g_1 n}$; we need to separately consider the cases where: $\kappa_{**} = 0$; $\kappa_{**} \in (0, \infty)$; or $\kappa_{**} = \infty$. By a suitable rescaling of r_n , it is without loss of generality to normalise $\kappa_{**} = 1$ in the second of these cases.

Suppose $\kappa_{**} = 0$. Then both diagonal element of $R_n'' D_n$ are divergent, and $(\hat{\mu}, \hat{\gamma}) \xrightarrow{p} (\mu, \gamma)$. The same arguments given in part III(a) thus imply that $\tilde{\sigma}_u^2(x) \xrightarrow{p} \sigma_u^2$.

Suppose $\kappa_{**} = 1$. Only the second element diagonal element of $R_n'' D_n$ diverges, so $\hat{\gamma} \xrightarrow{p} \gamma$; and since the first diagonal element converges to $\kappa_{**}^{-1} = 1$, $\hat{\mu} - \mu \xrightarrow{d} -\mu_*$ by (S.59). Thus in this case, we have

$$\begin{aligned} \tilde{\sigma}_u^2(x) &= \frac{\sum_{t=2}^n [(\mu - \hat{\mu}) + u_t]^2 K_{th}}{\sum_{t=2}^n K_{th}} + o_p(1) \\ &= \mu_*^2 + \sigma_u^2 + 2 \frac{(\mu - \hat{\mu}) \sum_{t=2}^n K_{th} u_t}{\sum_{t=2}^n K_{th}} + o_p(1) \\ &\xrightarrow{d}_{(1)} \mu_*^2 + \sigma_u^2 \end{aligned}$$

where $\xrightarrow{d}_{(1)}$ follows by $\sum_{t=2}^n K_{th} u_t = O_p(\beta_n/nh_n)^{1/2}$, as follows e.g. from (S.42) in the proof of Theorem 4.2.

Suppose $\kappa_{**} = \infty$. In this case, (S.62) implies that $\hat{\mu} - \mu$ is divergent at rate $r_n \kappa_{g_1 n}$, and dominates $\hat{\gamma} - \gamma$. Thus by (S.59)

$$(r_n \kappa_{g_1 n})^{-2} \tilde{\sigma}_u^2(x) = (r_n \kappa_{g_1 n})^{-2} (\mu - \hat{\mu})^2 + o_p(1) \xrightarrow{d} \mu_*^2.$$

In summary, in each case $\varsigma_n^{-1} \tilde{\sigma}_u^2(x) \xrightarrow{d} \sigma_*^2$, where

$$(S.63) \quad [\varsigma_n, \sigma_*^2] := \begin{cases} [1, \sigma_u^2] & \text{if } \kappa_{**} = 0 \\ [r_n^2 \kappa_{g_1 n}^2, \mu_*^2 + \sigma_u^2] & \text{if } \kappa_{**} = 1 \\ [r_n^2 \kappa_{g_1 n}^2, \mu_*^2] & \text{if } \kappa_{**} = \infty. \end{cases}$$

Note that in the case where $\kappa_{**} = 1$, we have $\lim_{n \rightarrow \infty} r_n \kappa_{g_1 n} = \kappa_{**} = 1$, and so we may equivalently take either $\varsigma_n = 1$ or $\varsigma_n = r_n^2 \kappa_{g_1 n}^2$; we present (S.63) in terms of the latter to facilitate the next part of the proof.

IV. Asymptotics of \tilde{F} under \mathcal{H}_1 . Since g_1 has bounded second derivative (by condition (v)), arguments similar to those given in part I of this proof and in the proof of Theorem 4.2 yield that, so long as $\tilde{\sigma}_u^2(x) \xrightarrow{p} \sigma_u^2$ (and $nh_n^5/\beta_n \rightarrow 0$ as assumed),

$$(S.64) \quad \begin{aligned} t^*(x; \mu, \gamma) &:= \left[\frac{\sum_{t=2}^n K_{th}}{\tilde{\sigma}_u^2(x) Q_{11}} \right]^{1/2} [\tilde{m}_g(x) - \mu - \gamma g(x) - r_n g_1(x)] \\ &\xrightarrow{d} N[0, 1]. \end{aligned}$$

For future reference, we note here that

$$\begin{aligned}
\text{(S.65)} \quad \tilde{t}(x; \hat{\mu}, \hat{\gamma}) - t^*(x; \mu, \gamma) &= \left[\frac{\sum_{t=2}^n K_{th}}{\tilde{\sigma}_u^2(x) Q_{11}} \right]^{1/2} [(\mu - \hat{\mu}) + (\gamma - \hat{\gamma})g(x) + r_n g_1(x)]
\end{aligned}$$

and that if $\tilde{\sigma}_u^2(x) \xrightarrow{p} \sigma_u^2$, then by Theorem 3.4

$$\text{(S.66)} \quad U_n := \frac{\frac{\beta_n}{nh_n} \sum_{t=2}^n K[(x_{t-1} - x)/h_n]}{\tilde{\sigma}_u^2(x) Q_{11}} \xrightarrow{d} \frac{\varrho(0)}{\sigma_u^2 Q_{11}} =: U,$$

which is a.s. nonzero.

We show below that in each of the cases contemplated by condition (iii), $\delta_n \tilde{t}(x; \hat{\mu}, \hat{\gamma}) \xrightarrow{d} \omega(x)$, where $\delta_n \rightarrow 0$, and $\omega(x)$ is nonzero for at least one $x \in \mathcal{X}$. That $\tilde{F} \xrightarrow{p} \infty$ then follows immediately from (24).

(a). Suppose condition (iii.a) holds. By the results of parts II(a) and III(a), and (S.61) in particular,

$$\begin{aligned}
(\mu - \hat{\mu}) + (\gamma - \hat{\gamma})g(x) + r_n g_1(x) &= O_p(r_n \beta_n^{-1}) + O_p(r_n \beta_n^{-1} \kappa_{gn}^{-1}) + r_n g_1(x) \\
&= r_n [g_1(x) + o_p(1)].
\end{aligned}$$

Thus by (S.65) and recalling from part III(a) that $\tilde{\sigma}_u^2(x) \xrightarrow{p} \sigma_u^2$

$$\begin{aligned}
\tilde{t}(x; \hat{\mu}, \hat{\gamma}) &= t^*(x; \mu, \gamma) + o_p(1) \\
&\quad + \left[\frac{\sum_{t=2}^n K_{th}}{\tilde{\sigma}_u^2(x) Q_{11}} \right]^{1/2} r_n [g_1(x) + o_p(1)]
\end{aligned}$$

whence, noting that $\beta_n/r_n^2 nh_n \rightarrow 0$ by condition (iv),

$$\left(\frac{\beta_n}{r_n^2 nh_n} \right)^{1/2} \tilde{t}(x; \hat{\mu}, \hat{\gamma}) = o_p(1) + U_n^{1/2} g_1(x) \xrightarrow{d} U^{1/2} g_1(x)$$

as required, by (S.66). By condition (vi), $g_1(x) \neq 0$ for at least one $x \in \mathcal{X}$.

(b). Suppose condition (iii.b) holds. Recall $\kappa_* := \lim_{n \rightarrow \infty} \kappa_{g_1 n}$ and $\kappa_{**} := \lim_{n \rightarrow \infty} \kappa_{g_1 n} r_n$. We consider each of the cases contemplated by condition (iv), in turn.

Suppose $\kappa_* \in [0, \infty)$, which implies $\kappa_{**} = 0$. By the results of part II(b) and (S.62),

$$\begin{aligned}
r_n^{-1}(\hat{\mu} - \mu) &= \kappa_{g_1 n} (r_n \kappa_{g_1 n})^{-1} (\hat{\mu} - \mu) \xrightarrow{d} -\kappa_* \mu_* \\
r_n^{-1}(\hat{\gamma} - \gamma) &= O_p(\kappa_{g_1 n} \kappa_{gn}^{-1}) = o_p(1).
\end{aligned}$$

By part III(b), $\tilde{\sigma}_u^2(x) \xrightarrow{p} \sigma_u^2$, and thus, similarly to the argument given in part IV(a), noting that $\beta_n/r_n^2 nh_n \rightarrow 0$ also in the present case,

$$\begin{aligned} \left(\frac{\beta_n}{r_n^2 nh_n}\right)^{1/2} \tilde{t}(x; \hat{\mu}, \hat{\gamma}) &= o_p(1) + U_n^{1/2}[r_n^{-1}(\hat{\mu} - \mu) + g_1(x)] \\ &\xrightarrow{d} U^{1/2}[\kappa_* \mu_* + g_1(x)], \end{aligned}$$

and note that by condition (vi), $\kappa_* \mu_* + g_1(x) \neq 0$ for at least one $x \in \mathcal{X}$.

Suppose $\kappa_* = \infty$ and $\kappa_{**} = 0$. In this case, it remains true that $\tilde{\sigma}_u^2(x) \xrightarrow{p} \sigma_u^2$, by part III(b) of the proof. Moreover, by the results of part II(b) and (S.62),

$$(S.67) \quad r_n^{-1} \kappa_{g_1 n}^{-1}(\hat{\mu} - \mu) \xrightarrow{d} -\mu_*$$

$$(S.68) \quad r_n^{-1} \kappa_{g_1 n}^{-1}(\hat{\gamma} - \gamma) = O_p(\kappa_{g_1 n}^{-1}) = o_p(1).$$

Thus under the assumption that $\beta_n/r_n^2 \kappa_{g_1 n}^2 nh_n \rightarrow 0$, similar arguments as were used in the previous case yield

$$\begin{aligned} \left(\frac{\beta_n}{r_n^2 \kappa_{g_1 n}^2 nh_n}\right)^{1/2} \tilde{t}(x; \hat{\mu}, \hat{\gamma}) &= o_p(1) + U_n^{1/2}\{r_n^{-1} \kappa_{g_1 n}^{-1}(\hat{\mu} - \mu) + o_p(1)\} \\ &\xrightarrow{d} U^{1/2} \mu_* \end{aligned}$$

which is nonzero a.s. by condition (vi).

Suppose $\kappa_{**} \in (0, \infty]$, which implies $\kappa_* = \infty$; as in part III(b), normalise $\kappa_{**} = 1$ if $\kappa_{**} \in (0, \infty)$. We have

$$(r_n \kappa_{g_1 n})^{-2} \tilde{\sigma}_u^2(x) \xrightarrow{d} \sigma_*^2,$$

for σ_*^2 as in (S.63). Thus $\tilde{\sigma}_u^2(x)$ is bounded away from zero, and possibly divergent, whence $t^*(x; \mu, \gamma) = O_p(1)$. In view of (S.67)–(S.68) and the fact that $\beta_n/nh_n \rightarrow 0$, it follows that

$$\begin{aligned} \left(\frac{\beta_n}{nh_n}\right)^{1/2} \tilde{t}(x; \hat{\mu}, \hat{\gamma}) &= o_p(1) + \left[\frac{\frac{\beta_n}{nh_n} \sum_{t=2}^n K_{th}}{(r_n \kappa_{g_1 n})^{-2} \tilde{\sigma}_u^2(x) Q_{11}} \right]^{1/2} \\ &\quad \cdot \{r_n^{-1} \kappa_{g_1 n}^{-1}(\mu - \hat{\mu}) + o_p(1)\} \\ &\xrightarrow{d} \left[\frac{\varrho(0)}{\sigma_*^2 Q_{11}} \right]^{1/2} \mu_*, \end{aligned}$$

which is nonzero a.s. by condition (vi). □

APPENDIX F: ADDITIONAL SIMULATIONS FOR SECTION 6

This section provides the results of some additional simulations not reported in the main text. Table 4 is the counterpart of Table 2 in the main text, reporting size-adjusted power of the procedure for a type II fractional process, relative to Wang and Phillips (2012), for the cases where $d \in \{0.25, 0.75\}$.

We also repeated the simulation exercise described in Section 6, but now allowing x_t to be mildly (or nearly) integrated, that is of the form

$$x_t(n) = (1 - \kappa_n^{-1})x_{t-1}(n) + \xi_t \quad \kappa_n = n^{\alpha_\kappa}$$

with $x_0(n) = 0$. All other aspects of the simulation design are exactly as in Section 6. We generated data for $\alpha_\kappa \in \{0.25, 0.50, 0.75, 1.00\}$; this last corresponds to a nearly integrated process. Size results are given in Table 3, while size-adjusted power relative to Wang and Phillips (2012) is reported in Tables 5 and 6.

TABLE 3
*MI processes: $x_t = (1 - \kappa_n^{-1})x_{t-1} + \xi_t$; $\kappa_n = n^{\alpha_\kappa}$
 Size: maximum rejection frequency over $\rho \in \{-0.5, 0, 0.5\}$; $\alpha = 0.1$*

α_κ	n	WP (2012)			$p = 17$			$p = 25$		
$h = n^b, b =$		-0.2	-0.1	-0.05	-0.2	-0.1	-0.05	-0.2	-0.1	-0.05
0.25	100	0.02	0.01	0.00	0.06	0.03	0.02	0.07	0.03	0.02
	200	0.03	0.01	0.00	0.06	0.04	0.02	0.08	0.04	0.03
	500	0.04	0.01	0.00	0.07	0.04	0.03	0.08	0.05	0.03
0.50	100	0.04	0.02	0.01	0.07	0.05	0.04	0.09	0.06	0.04
	200	0.06	0.03	0.02	0.08	0.06	0.05	0.10	0.08	0.06
	500	0.08	0.05	0.03	0.08	0.07	0.07	0.10	0.09	0.08
0.75	100	0.06	0.03	0.02	0.08	0.06	0.05	0.10	0.08	0.06
	200	0.08	0.05	0.04	0.07	0.07	0.06	0.10	0.09	0.08
	500	0.09	0.07	0.06	0.08	0.08	0.08	0.09	0.10	0.10
1.00	100	0.06	0.04	0.03	0.08	0.07	0.06	0.11	0.09	0.08
	200	0.08	0.05	0.04	0.08	0.08	0.08	0.10	0.10	0.10
	500	0.09	0.07	0.07	0.08	0.08	0.09	0.09	0.10	0.11

TABLE 4
Fractional type II processes
Size-adjusted power when $d \in \{0.25, 0.75\}$; $\alpha = 0.1$

n		WP (2012)			$p = 17$			$p = 25$		
$h = n^b, b =$		-0.2	-0.1	-0.05	-0.2	-0.1	-0.05	-0.2	-0.1	-0.05
$d = 0.25$		Size adj. power			Size adjusted, relative to WP					
$\varphi_1(x)$	100	0.07	0.05	0.04	0.11	0.10	0.09	0.14	0.12	0.10
	200	0.16	0.14	0.12	0.15	0.17	0.16	0.19	0.20	0.19
	500	0.47	0.53	0.51	0.20	0.20	0.20	0.24	0.23	0.23
$\varphi_1(2x)$	100	0.08	0.05	0.03	0.10	0.08	0.07	0.13	0.10	0.08
	200	0.15	0.12	0.08	0.14	0.13	0.11	0.18	0.15	0.13
	500	0.47	0.45	0.37	0.18	0.17	0.16	0.22	0.21	0.20
$ x ^{-2} \wedge 1$ ($\times 0.5$)	100	0.06	0.03	0.02	0.14	0.13	0.13	0.18	0.16	0.15
	200	0.11	0.07	0.04	0.20	0.22	0.21	0.24	0.25	0.24
	500	0.35	0.27	0.20	0.14	0.12	0.13	0.16	0.14	0.14
$ x ^{-1} \wedge 1$ ($\times 0.5$)	100	0.05	0.02	0.01	0.09	0.09	0.08	0.12	0.11	0.09
	200	0.06	0.04	0.03	0.13	0.15	0.15	0.16	0.18	0.17
	500	0.14	0.13	0.11	0.19	0.22	0.22	0.23	0.25	0.25
$ x ^{1.5}$ ($\times 0.02$)	100	0.10	0.09	0.07	0.03	0.04	0.04	0.05	0.05	0.04
	200	0.25	0.25	0.23	0.03	0.04	0.04	0.06	0.06	0.05
	500	0.73	0.78	0.77	0.04	0.07	0.06	0.06	0.09	0.08
x^2 ($\times 0.02$)	100	0.06	0.04	0.03	0.05	0.06	0.05	0.07	0.07	0.06
	200	0.12	0.11	0.10	0.06	0.08	0.07	0.09	0.10	0.09
	500	0.39	0.44	0.44	0.10	0.14	0.16	0.13	0.18	0.19
$d = 0.75$		Size adj. power			Size adjusted, relative to WP					
$\varphi_1(x)$	100	0.07	0.06	0.05	0.06	0.08	0.08	0.10	0.11	0.10
	200	0.12	0.11	0.10	0.07	0.10	0.11	0.11	0.14	0.14
	500	0.20	0.23	0.23	0.07	0.12	0.14	0.12	0.16	0.17
$\varphi_1(2x)$	100	0.07	0.05	0.04	0.05	0.06	0.06	0.08	0.08	0.08
	200	0.10	0.08	0.06	0.05	0.07	0.07	0.09	0.10	0.09
	500	0.14	0.13	0.12	0.04	0.07	0.08	0.08	0.11	0.11
$ x ^{-2} \wedge 1$ ($\times 0.5$)	100	0.07	0.04	0.03	0.09	0.11	0.11	0.12	0.14	0.14
	200	0.09	0.06	0.05	0.10	0.13	0.14	0.15	0.18	0.18
	500	0.11	0.10	0.08	0.07	0.12	0.13	0.12	0.15	0.16
$ x ^{-1} \wedge 1$ ($\times 0.5$)	100	0.07	0.04	0.03	0.07	0.08	0.08	0.09	0.11	0.11
	200	0.08	0.07	0.06	0.07	0.11	0.12	0.11	0.14	0.15
	500	0.11	0.11	0.11	0.08	0.13	0.15	0.13	0.16	0.18
$ x ^{1.5}$ ($\times 0.02$)	100	0.10	0.09	0.07	0.03	0.05	0.05	0.06	0.07	0.07
	200	0.18	0.18	0.18	0.04	0.06	0.08	0.07	0.09	0.10
	500	0.37	0.42	0.44	0.01	0.08	0.11	0.06	0.12	0.14
x^2 ($\times 0.02$)	100	0.07	0.06	0.05	0.08	0.11	0.13	0.12	0.14	0.16
	200	0.13	0.13	0.12	0.04	0.09	0.10	0.08	0.11	0.12
	500	0.26	0.31	0.32	0.00	0.00	0.00	0.01	0.00	0.00

TABLE 5
*MI processes: $x_t = (1 - \kappa_n^{-1})x_{t-1} + \xi_t$; $\kappa_n = n^{\alpha_\kappa}$
Size-adjusted power when $\alpha_\kappa \in \{0.25, 0.50\}$; $\alpha = 0.1$*

n		WP (2012)			$p = 17$			$p = 25$		
$h = n^b, b =$		-0.2	-0.1	-0.05	-0.2	-0.1	-0.05	-0.2	-0.1	-0.05
$\alpha_\kappa = 0.25$		Size adj. power			Size adjusted, relative to WP					
$\varphi_1(x)$	100	0.06	0.03	0.02	0.11	0.09	0.07	0.13	0.11	0.08
	200	0.15	0.13	0.09	0.17	0.17	0.15	0.20	0.20	0.17
	500	0.50	0.54	0.50	0.21	0.22	0.23	0.24	0.24	0.25
$\varphi_1(2x)$	100	0.09	0.04	0.02	0.13	0.10	0.07	0.16	0.12	0.08
	200	0.22	0.16	0.10	0.19	0.17	0.13	0.23	0.20	0.16
	500	0.61	0.58	0.48	0.19	0.18	0.19	0.22	0.21	0.22
$ x ^{-2} \wedge 1$ ($\times 0.5$)	100	0.07	0.03	0.01	0.12	0.12	0.10	0.15	0.13	0.11
	200	0.17	0.10	0.05	0.20	0.22	0.21	0.23	0.24	0.24
	500	0.50	0.39	0.26	0.15	0.15	0.16	0.17	0.16	0.17
$ x ^{-1} \wedge 1$ ($\times 0.5$)	100	0.03	0.01	0.00	0.08	0.06	0.05	0.09	0.08	0.06
	200	0.06	0.03	0.02	0.12	0.12	0.11	0.15	0.15	0.12
	500	0.15	0.12	0.09	0.20	0.24	0.24	0.23	0.26	0.27
$ x ^{1.5}$ ($\times 0.02$)	100	0.07	0.05	0.03	0.04	0.02	0.01	0.05	0.03	0.02
	200	0.22	0.20	0.16	0.04	0.03	0.02	0.05	0.04	0.03
	500	0.72	0.76	0.74	0.04	0.04	0.03	0.06	0.05	0.04
x^2 ($\times 0.02$)	100	0.03	0.02	0.01	0.04	0.03	0.02	0.05	0.03	0.02
	200	0.09	0.07	0.05	0.05	0.04	0.03	0.06	0.05	0.04
	500	0.33	0.35	0.33	0.06	0.07	0.07	0.09	0.09	0.08
$\alpha_\kappa = 0.50$		Size adj. power			Size adjusted, relative to WP					
$\varphi_1(x)$	100	0.08	0.06	0.04	0.11	0.12	0.11	0.15	0.14	0.13
	200	0.18	0.18	0.16	0.16	0.17	0.18	0.20	0.21	0.21
	500	0.47	0.54	0.54	0.17	0.19	0.19	0.22	0.22	0.22
$\varphi_1(2x)$	100	0.08	0.05	0.03	0.11	0.09	0.07	0.14	0.12	0.09
	200	0.15	0.12	0.09	0.14	0.13	0.12	0.18	0.17	0.15
	500	0.34	0.33	0.29	0.15	0.17	0.16	0.20	0.21	0.20
$ x ^{-2} \wedge 1$ ($\times 0.5$)	100	0.07	0.04	0.02	0.15	0.15	0.15	0.19	0.18	0.17
	200	0.12	0.08	0.05	0.19	0.22	0.22	0.24	0.25	0.25
	500	0.24	0.20	0.15	0.12	0.11	0.10	0.14	0.12	0.11
$ x ^{-1} \wedge 1$ ($\times 0.5$)	100	0.05	0.03	0.01	0.10	0.11	0.10	0.13	0.13	0.11
	200	0.08	0.06	0.05	0.15	0.17	0.17	0.19	0.21	0.21
	500	0.17	0.18	0.17	0.17	0.18	0.19	0.21	0.21	0.21
$ x ^{1.5}$ ($\times 0.02$)	100	0.12	0.10	0.09	0.03	0.03	0.03	0.05	0.04	0.04
	200	0.31	0.33	0.31	0.04	0.06	0.06	0.07	0.08	0.08
	500	0.76	0.83	0.85	0.06	0.11	0.14	0.10	0.15	0.17
x^2 ($\times 0.02$)	100	0.07	0.05	0.04	0.05	0.06	0.05	0.07	0.07	0.07
	200	0.16	0.17	0.16	0.09	0.12	0.13	0.13	0.15	0.17
	500	0.50	0.60	0.61	0.04	0.07	0.07	0.08	0.09	0.09

TABLE 6
MI processes: $x_t = (1 - \kappa_n^{-1})x_{t-1} + \xi_t$; $\kappa_n = n^{\alpha_\kappa}$
Size-adjusted power when $\alpha_\kappa \in \{0.75, 1.00\}$; $\alpha = 0.1$

n		WP (2012)			$p = 17$			$p = 25$		
$h = n^b, b =$		-0.2	-0.1	-0.05	-0.2	-0.1	-0.05	-0.2	-0.1	-0.05
$\alpha_\kappa = 0.75$		Size adj. power			Size adjusted, relative to WP					
$\varphi_1(x)$	100	0.08	0.06	0.05	0.08	0.09	0.09	0.11	0.12	0.11
	200	0.13	0.13	0.12	0.09	0.12	0.13	0.14	0.16	0.16
	500	0.25	0.29	0.29	0.09	0.15	0.17	0.14	0.19	0.21
$\varphi_1(2x)$	100	0.07	0.05	0.03	0.06	0.07	0.06	0.10	0.09	0.08
	200	0.11	0.09	0.07	0.06	0.08	0.08	0.10	0.11	0.11
	500	0.17	0.17	0.15	0.06	0.09	0.09	0.10	0.13	0.14
$ x ^{-2} \wedge 1$ ($\times 0.5$)	100	0.07	0.04	0.03	0.11	0.13	0.13	0.15	0.16	0.15
	200	0.09	0.07	0.05	0.13	0.17	0.18	0.18	0.20	0.21
	500	0.13	0.11	0.09	0.09	0.14	0.15	0.15	0.18	0.18
$ x ^{-1} \wedge 1$ ($\times 0.5$)	100	0.06	0.03	0.02	0.08	0.09	0.10	0.11	0.12	0.11
	200	0.09	0.06	0.05	0.10	0.13	0.14	0.14	0.16	0.17
	500	0.13	0.14	0.14	0.10	0.16	0.17	0.15	0.20	0.21
$ x ^{1.5}$ ($\times 0.02$)	100	0.11	0.09	0.08	0.03	0.04	0.05	0.06	0.06	0.06
	200	0.21	0.23	0.22	0.04	0.07	0.08	0.06	0.10	0.11
	500	0.45	0.52	0.54	0.00	0.07	0.09	0.06	0.11	0.12
x^2 ($\times 0.02$)	100	0.08	0.06	0.05	0.08	0.11	0.11	0.11	0.13	0.13
	200	0.14	0.15	0.14	0.06	0.11	0.12	0.10	0.14	0.14
	500	0.33	0.39	0.42	0.00	0.00	0.00	0.00	0.00	0.00
$\alpha_\kappa = 1.00$		Size adj. power			Size adjusted, relative to WP					
$\varphi_1(x)$	100	0.08	0.06	0.05	0.06	0.08	0.07	0.09	0.10	0.09
	200	0.11	0.10	0.09	0.06	0.10	0.11	0.10	0.13	0.14
	500	0.17	0.19	0.19	0.04	0.09	0.10	0.08	0.13	0.14
$\varphi_1(2x)$	100	0.07	0.05	0.04	0.05	0.05	0.06	0.08	0.08	0.07
	200	0.10	0.08	0.06	0.04	0.06	0.06	0.07	0.09	0.09
	500	0.13	0.12	0.11	0.03	0.05	0.06	0.05	0.08	0.10
$ x ^{-2} \wedge 1$ ($\times 0.5$)	100	0.07	0.04	0.03	0.09	0.11	0.11	0.13	0.14	0.14
	200	0.09	0.06	0.05	0.10	0.13	0.14	0.14	0.17	0.17
	500	0.11	0.09	0.08	0.05	0.10	0.13	0.10	0.14	0.16
$ x ^{-1} \wedge 1$ ($\times 0.5$)	100	0.06	0.04	0.03	0.06	0.08	0.08	0.10	0.10	0.10
	200	0.09	0.07	0.06	0.07	0.11	0.12	0.11	0.14	0.15
	500	0.12	0.12	0.11	0.05	0.10	0.13	0.09	0.14	0.17
$ x ^{1.5}$ ($\times 0.02$)	100	0.09	0.08	0.07	0.03	0.04	0.05	0.05	0.06	0.06
	200	0.16	0.17	0.16	0.04	0.07	0.08	0.07	0.10	0.11
	500	0.29	0.33	0.34	-0.04	0.03	0.05	0.01	0.06	0.07
x^2 ($\times 0.02$)	100	0.07	0.06	0.05	0.08	0.11	0.13	0.12	0.14	0.15
	200	0.11	0.12	0.11	0.04	0.06	0.07	0.07	0.08	0.08
	500	0.22	0.26	0.27	0.00	0.00	0.00	0.00	0.00	0.00

REFERENCES

- [1] Abadir, K.M., Distaso, W., Giraitis, L., and Koul, H.L. (2014). Asymptotic normality for weighted sums of linear processes. *Econometric Theory*, **30**(1): 252–284.
- [2] Andrews, D.W.K. (1992). Generic uniform convergence. *Econometric Theory*, **8**(2): 241–257.
- [3] Billingsley, P. (1968). *Convergence of probability measures*. New York, Wiley.
- [4] Bingham, N.H., Goldie, C.M. and Teugels, J.L. (1987). *Regular Variation*. Cambridge University Press.
- [5] Feller, W., 1971, *An introduction to probability theory and its applications*. Vol. II, Wiley Series in Probability and Mathematical Statistics, Wiley, USA.
- [6] Folland, G.B. (1999). *Real Analysis*. Wiley.
- [7] Giraitis, L., Koul, H.L. and Surgailis D. (2012) *Large Sample Inference for Long Memory Processes*. Imperial College Press.
- [8] Hall, P. and Heyde, C.C. (1980). *Martingale limit theory and its application*. Academic Press, New York.
- [9] Jeganathan, P. (2008). Limit theorems for functionals of sums that converge to fractional Brownian and stable motions. *Cowles Foundation Discussion Paper* No. 1949.
- [10] Kallenberg, O. (2001). *Foundations of Modern Probability*. 2nd ed., Springer.
- [11] Wainwright, M. J. (2019) *High-Dimensional Statistics: a Non-Asymptotic Viewpoint*. Cambridge University Press.
- [12] Wang, Q. (2014). Martingale limit theorem revisited and nonlinear cointegrating regression. *Econometric Theory*, **30**(3): 509–535.
- [13] Wang, Q. and Phillips P.C.B. (2009). Asymptotic theory for local time density estimation and nonparametric cointegrating regression. *Econometric Theory*, **25**(3): 710–738.
- [14] Wang, Q. and Phillips P.C.B. (2012). A specification test for nonlinear nonstationary models. *Annals of Statistics*, **40**(2): 727–758.
- [15] Whittle, P. (1960). Bounds for the moments of linear and quadratic forms in independent variables. *Theory of Probability and its Applications*, **5**(3): 302–305.
- [16] Wu, W.B. and Mielniczuk, J. (2002). Kernel density estimation for linear processes. *Annals of Statistics*, **30**(5): 1441–1459.

DEPARTMENT OF ECONOMICS
 UNIVERSITY OF OXFORD
 OXFORD OX1 3UQ
 UNITED KINGDOM
 E-MAIL: james.duffy@economics.ox.ac.uk

DEPARTMENT OF ECONOMICS
 UNIVERSITY OF CYPRUS
 P.O. Box 20537
 CY-1878, NICOSIA CYPRUS
 E-MAIL: kasparis@ucy.ac.cy