ESTIMATION AND INFECTION 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 \( \chi^2 \) 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-
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.\footnote{Although 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.}

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
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 (WNP), with leading examples being MI and $I(1/2)$ processes. To appreciate the significance of our results, consider the regression model

\[ y_t = m(x_{t-1}) + u_t \]

where $u_t$ is a martingale difference sequence; suppose for concreteness that $x_t \sim I(d)$ 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, 2011). 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 $x_t$.

\[ \text{Formally, the asymptotic size of a test } \phi_n \in \{0,1\} \text{ would be defined as } \lim \sup_{n \to \infty} \sup_{d \in (-1/2,3/2), m \in \mathcal{M}} \mathbb{P}_{d,m} \{ \phi_n = 1 \}, \text{ for } \mathcal{M} \text{ a class of functions respecting } \mathcal{H}_0. \]
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 \((\mu, \gamma)\) 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 \( m \) of 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 non-stationary. 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

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

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

---

\(^3\)By 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

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

where \(X^- \sim N[0, \sigma_2^-]\) (with possibly \(\sigma_2^- = 0\)) and \(\varphi_{\sigma_+^2}\) denotes the \(N[0, \sigma_+^2]\) density, and

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

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 Supplementary Material.

**Notation.** For deterministic sequences \(\{a_n\}\) and \(\{b_n\}\), \(a_n \sim b_n\) denotes \(\lim_{n \to \infty} a_n/b_n = 1\) and \(a_n \asymp b_n\) denotes \(\lim_{n \to \infty} |a_n/b_n| \in (0, \infty)\). For random variables \(X\) and \(Y\), \(X \sim Y\) denotes distributional equality. For a real number \(x\), \([x]\) denotes its integer part. \(1\{A\}\) denotes the indicator function for the set \(A\). \(\mathbb{R}, \mathbb{R}_+, \mathbb{R}_*^+\) are the extended, the nonnegative, and (strictly) positive real numbers respectively. For a function \(f(x)\), \(f^{(j)}(x)\) denotes its \(j\)th derivative. All limits are taken as \(n \to \infty\) unless otherwise indicated.

**2. Leading examples of WNPs: \(I(1/2)\) and MI processes.** In the next section, we will 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 will help to motivate our high-level conditions. To bring these processes into a common framework,
consider a linear process array of the form

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

\( t = 1, ..., n \in \mathbb{N}, \) 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) \( \xi_t \) is i.i.d. with \( E\xi_1 = 0 \) and \( Var(\xi_1) = \sigma_\xi^2 < \infty \).

(ii) \( \xi_1 \) has an absolutely continuous distribution, and a characteristic function \( \psi_{\xi}(\lambda) \) that satisfies \( \int_R |\psi_{\xi}(\lambda)|^\theta d\lambda < \infty \), for some \( \theta \in \mathbb{N} \).

2.1. \( I(1/2) \) processes. The definition of a ‘fractionally integrated process’ used 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 \( \beta_n^2 := Var(x_n) \). 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 \( d \in (1/2, 3/2) \), \( x_t \) is an \( I(d) \) process of

- type I: if \( \phi_j = 1, c_s \sim \ell(s)s^{d-2}, 1 \{d \in (1/2, 1)\} \cdot \sum_{s=0}^{\infty} c_s = 0; \) 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 \( \ell: \mathbb{R}_+ \to \mathbb{R}_+ \) is 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). \( \ell \) is SV such that \( L(n) := \int_1^n [\ell^2(x)/x]dx \to \infty \) as \( n \to \infty \), and either:
FRACTIONAL AND WEAKLY NONSTATIONARY PROCESSES 7

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 \( \phi_j \sim \ell(j)j^{-1/2}, \phi_0 \neq 0, \sum_{s=0}^{\infty} |c_s| < \infty, \sum_{s=0}^{\infty} c_s \neq 0, \) and either
\( (a) \ \lim_{j \to \infty} j|c_j| < \infty, \) or
\( (b) \ \lim_{n \to \infty} \ell(n)^{-1} \sum_{j=|n\delta|}^{\infty} j^{1/2}|c_j| = 0 \) for some \( 0 < \delta < 1. \)

Remark 2.1. (a) FR1 and FR2 respectively imply that \( x_t(n) \) is an \( I(1/2) \) process of types I and II. FR2 also specifies some additional regularity conditions that are used principally to ensure that \( \beta_n^{-1}x_t \) has a uniformly bounded density, the role of which will become clear in Section 3. (Similar conditions have appeared elsewhere in the literature, e.g. Jeganathan, 2008, and Hualde and Robinson, 2011.)

(b) \( \text{Var}(x_n) \approx L(n) \) under FR, where \( L(n) \) is itself SV. The divergence or convergence of \( L(n) \) effectively demarcates the boundary between WNP's and stationary long memory processes. Either is possible, depending on \( \ell: \) e.g. \( \ell(n) = 1 \) gives \( L(n) \sim \ln n \) and \( \ell(n) = (\ln n)^{-1/2} \) gives \( L(n) \approx \ln \ln n, \) 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 \( L(n) \to \infty. \)

(c) FR encompasses both types of parametric ARFIMA models with \( d = 1/2. \) For example, consider the ARFIMA(1/2) type II model
\[ (1 - L)^{1/2}x_t = v_t 1 \{t > 0\}, \quad \text{where } a(L)v_t = b(L)\xi_t, \]
where \( a \) and \( b \) denote finite-order polynomials in the lag operator \( L. \) In this case, it is possible to write \( x_t = \sum_{j=0}^{\ell-1} \phi_j v_{t-j}, \) where \( \phi_j \geq 0 \) 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 both FR2(a) and FR2(b) are satisfied. Since \( \ell(x) = 1, \) we have trivially that \( L(n) \to \infty. \) Similar arguments show that FR1 is consistent with an ARFIMA(1/2) type I model, under which \( x_t \) is the partial sum of an 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
\[ x_t(n) = (1 - \kappa_n^{-1})x_{t-1}(n) + v_t, x_0(n) = 0, \]
where \( v_t \) is a stationary process and \( \kappa_n > 0 \) with \( \kappa_n \to \infty \), so that the autoregressive coefficient becomes increasingly proximate to unity as \( n \) grows. They can thus be encompassed within the framework of (5) if we allow \( \phi_j \) 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, 2019), 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 \( \kappa_n \). NI processes are defined by \( \kappa_n / n \to c \neq 0 \), with the consequence that \( n^{-1/2} x_{\lfloor nr \rfloor} \) converges weakly to an Ornstein-Uhlenbeck process. MI processes have \( \kappa_n / n \to 0 \), 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.

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 \( \kappa_n > 0 \), \( \kappa_n = n^{\alpha_n} \ell_\kappa(n) \) for \( \ell_\kappa \) SV and \( \alpha_\kappa \in [0,1) \), \( \kappa_n \to \infty \) and \( \sup_{n \geq 1} \sup_{1 \leq t \leq n} \kappa_t / \kappa_n < \infty \).

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

\(^4\)Previous 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.
be derived. These conditions, particularly HL0–HL2 and HL4 below, may be 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 nonnegative sequence with \( \beta_n \to \infty \).

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

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

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

\[
E \left[ h \left( \beta_n^{-1} X_t(n)^+ \right) \mid \mathcal{F}_0 \right] \xrightarrow{p} \int_{\mathbb{R}} h(x) \Phi_{X^+}(x) dx;
\]

(b) \( \beta_n^{-1} X_t(n)^- \xrightarrow{d} X^- \), and \( \beta_n^{-1} \left[ X_t(n)^- - X_t(n)^- \right] \xrightarrow{p} 0 \).

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

\[
\sup_{n \geq n_0, t_0 \leq t \leq n} \sup_{x} D_{n,t}(x) < \infty
\]

**HL4** For every bounded and Lipschitz continuous \( g : \mathbb{R} \to \mathbb{R} \)

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

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

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

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

**HL6** For some \( n_0 \geq t_0 \), \( \sup_{n \geq n_0} \frac{\beta_n}{n} \sum_{t=t_0}^{n} \beta_t^{-1} < \infty \), where \( t_0 \) is 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 \( \{ \xi_t \}_{t \leq 0} \) and \( \{ \xi_t \}_{t > 0} \).
(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 $f(x) = 1$. In this case, the application of a CLT for weighted sums of linear processes (see Abadir, Distaso, Giraitis, and Koul, 2014) yields that for every $r, s \in (0, 1]$,

$$
\beta_n^{-1}(x_{[nr]}^+, x_{[ns]}^+) \overset{d}{\rightarrow} (\eta_r, \eta_s)
$$

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

$$
\beta_n^{-1}(x_{[nr]}^-, x_{[ns]}^-) \overset{d}{\rightarrow} (\eta^-, \eta^-)
$$

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 WNNs, 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 $\beta_n^{-1}X_{[nr]}(n)$ converges weakly to a process with continuous sample paths.

Theorem 3.1. Suppose $f : \mathbb{R} \rightarrow \mathbb{R}$ is locally integrable, and HL0–HL4, HL6 hold. Further suppose $f$ is bounded or that the following hold:

(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 $\mathbb{P}(X^- \in \mathcal{Y}) = 1$;
(ii) $n^{-1} \sum_{t=1}^{n_0-1} f(\beta_n^{-1}X_t(n)) = o_p(1)$, for $t_0$ as in HL3;
(iii) For $\lambda' \in (0, \lambda)$, where $\lambda$ is as in HL5, either:
   (a) $|f(x)| = O(|x|^\lambda')$, as $|x| \rightarrow \infty$ and HL5(a) holds; or
   (b) $|f(x)| = O \left( \exp \left( \lambda' |x| \right) \right)$, as $|x| \rightarrow \infty$ and HL5(b) holds.
Then as \( n \to \infty \),
\[
\frac{1}{n} \sum_{t=1}^{n} f(\beta_n^{-1}X_t(n)) \xrightarrow{d} \int f(x + X^-) \Phi_X(x) dx.
\]

**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 \( \varphi_{x^2}(x) \) denotes the \( N[0, \sigma^2] \) density, and take \( \beta_n^2 = \text{Var}(x_t(n)) \). For \( X^- \sim N[0,1/2] \), let
\[
g(x) := \begin{cases} 
\varphi_{1/2}(x - X^-) & \text{under FR1} \\
\varphi_1(x) & \text{under FR2, MI},
\end{cases}
\]
noting that this is a density, albeit a random one under FR1.

**Theorem 3.2.** Suppose \( f : \mathbb{R} \to \mathbb{R} \) is locally Lebesgue integrable, and Assumptions INN and either FR or MI hold. Further suppose:

(i) 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 \( \theta = 1 \).

(ii) For some \( \lambda' \in (0, \infty) \), as \( |x| \to \infty \) either
   (a) \( |f(x)| = O(|x|^{\lambda'}) \), and \( \text{E} |\xi_1|^{2\lambda} < \infty \) for some \( \lambda > \lambda' \); or
   (b) \( |f(x)| = O(\exp(\lambda' |x|)) \), and \( \xi_1 \) has a finite m.g.f. in a neighborhood of zero.

Then
\[
\frac{1}{n} \sum_{t=1}^{n} f(\beta_n^{-1}x_t(n)) \xrightarrow{d} \int f g.
\]
Remark 3.3.  (a) Theorem 3.2 bridges existing asymptotic results for $I(d)$ processes of order $|d| < 1/2$ 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 $|d| < 1/2$. 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 ($\varphi_1$ 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 $\varrho$ 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 $\left( n^{-2} \sum_{t=1}^{n} x_t (n) \right)^2 \overset{P}{\to} 1$. 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 including $d = 1/2$.

(c) Assumption INN entails that $\xi_t$ has finite second moment, and so $\beta_n^{-1} x_{[nr]}(n)$ satisfies a CLT; the limiting distributions that appear in (14) are therefore Gaussian.\(^5\) Note that the process $r \mapsto \beta_n^{-1} x_{[nr]}(n)$ 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 $\beta_n^{-1} x_{[nr]}(n)$

\(^5\)Some preliminary work of the authors’ shows that Theorem 3.2 can be extended to the case where $\xi_t$ is in the domain of attraction of an $\alpha$-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

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

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

**Assumption K (kernel).** \( K : \mathbb{R} \rightarrow \mathbb{R} \) 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 \( \beta_n \). 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 \( \beta_n \) of a WNP grows with \( n \), these visits accumulate only at rate \( nh_n/\beta_n \), 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

\[
\Omega_n(\eta) := \{\{s, t\} \in \mathbb{N} : [\eta n] \leq s \leq [1 - \eta]n, \; [\eta n] + s \leq t \leq n\},
\]

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

**Assumption HL (continued).**
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 $D_{t,s,n}(x)$ such that for some $n_0, t_0 \geq 1$ 

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

For all $q_0, q_1 > 0$ 

$$\lim_{\eta \downarrow 0} \lim_{n \to \infty} \sup_{(s,t) \in \Omega_n(\eta)} \sup_{|x| \leq q_0 \eta^{q_1}} |D_{t,s,n}(x) - D_{t,s,n}(0)| = 0$$  

For $t_0$ as in $\text{HL7}$: 

(a) $\lim_{\eta \downarrow 0} \lim_{n \to \infty} \frac{\beta_n}{n} \sum_{t = t_0}^{n} \beta_t^{-1} = 0$;  
(b) $\lim_{\eta \downarrow 0} \lim_{n \to \infty} \frac{\beta_n}{n} \sum_{t = \lfloor(1-\eta)n\rfloor}^{n} \beta_t^{-1} = 0$;  
(c) $\lim_{\eta \downarrow 0} \lim_{n \to \infty} \frac{\beta_n}{n} \sup_{0 \leq s \leq (1-\eta)n} \sum_{t = s + t_0}^{n} \beta_{t-s}^{-1} = 0$;  
(d) $\lim_{\eta \downarrow 0} \lim_{n \to \infty} \frac{\beta_n}{n} \sup_{0 \leq s \leq n-1} \sum_{t = s + t_0}^{n} \beta_{t-s}^{-1} < \infty$;  
(e) for each $\eta \in (0,1)$, there exist $l_0 > 0$ and $l_1 \in (0,1)$ such that 

$$\lim_{n \to \infty} \beta_n^{-1} \inf_{(s,t) \in \Omega_n(\eta)} \beta_{t-s}^{-1} \geq \eta^{l_1} / l_0.$$ 

$\text{HL7}$ and $\text{HL9}$ may be regarded as strengthened versions of $\text{HL3}$ and $\text{HL6}$, and are closely related to Assumptions 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 $\text{HL7}$, 

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

as $n \to \infty$ and then $\varepsilon \to 0$. 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 

$$\frac{1}{n} \sum_{t = 1}^{n} \varphi_{\mathbb{Z}}(\beta_n^{-1} X_t(n)) \overset{d}{\to} \varphi_{\mathbb{Z}}(x + X^-) \Phi_{X^+}(x) dx, \quad \text{as } n \to \infty$$  

$$\overset{p}{\to} \Phi_{X^+}(-X^-), \quad \text{as } \varepsilon \to 0.$$  

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

**Theorem 3.3.** Suppose that, in addition to Assumptions K, $\text{HL0}$–$\text{HL4}$ and $\text{HL6}$–$\text{HL9}$, the following hold: 

(i) $\{h_n\}$ is a positive sequence with $\beta_n^{-1} h_n + \beta_n (n h_n)^{-1} \to 0$; 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
\[
\frac{\beta_n}{nh_n} \sum_{t=1}^{n} K \left( \frac{X_t(n) - x}{h_n} \right) \overset{d}{\to} \Phi_X \left( -X^- \right) \int_{\mathbb{R}} K(u) du.
\]

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
\[
\frac{\beta_n}{nh_n} \sum_{t=1}^{n} K \left( \frac{x_t(n) - x}{h_n} \right) \overset{d}{\to} \varphi(0) \int K.
\]

**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 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 strictly 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 (2019) 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 \((\mu, \gamma)\) in the model

\[ y_t = \mu + \gamma g(x_{t-1}) + u_t \]

given by \((\hat{\mu}, \hat{\gamma}) := \arg\min_{a,b} \sum_{t=1}^{n} [y_t - a - bg(x_{t-1})]^2\), where \(g\) is a known nonlinear transformation. Since the regressor is predetermined (i.e. \(F_{t-1}\)-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 \(x_t(n)\) 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 \(\beta^2_n = \text{Var}(x_n(n))\). A function \(g : \mathbb{R} \to \mathbb{R}\) is **asymptotically homogeneous** for \(\{x_t(n)\}\), if for each \(\lambda > 0\), \(g\) admits the decomposition

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

where \(\kappa_g : \mathbb{R}^*_+ \to \mathbb{R}^*_+\), \(H_g : \mathbb{R} \to \mathbb{R}\), \(R_g : \mathbb{R} \times \mathbb{R}^*_+ \to \mathbb{R}\) and for \(j = \{1, 2\},\)

\[ R_{g,n}^j := \frac{1}{\kappa_g(\beta_n)^j n} \sum_{t=1}^{n} \mathbb{E} |R_g(x_t(n), \beta_n)|^j = o(1). \]

AHFs encompass a wide range of commonly used regression functions, such as polynomial functions, cumulative distribution functions (with \(H_g(u) = 1\{u > 0\}\)), and logarithmic functions (with \(H_g(u) = 1\)); see Park and Phillips
(1999, 2001) for some further examples. Such a condition as
\[
\lim_{\lambda \to \infty} \kappa_g(\lambda)^{-1} \sup_x |R_g(x, \lambda)| = 0
\]
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 \( \gamma \) satisfies
\[
\hat{\gamma}_n - \gamma = \frac{\sum_{t=2}^{n} g(x_{t-1})u_t}{\sum_{t=2}^{n} g^2(x_{t-1})} = (1 + o_p(1)) \frac{\sum_{t=2}^{n} H_g(\beta_n^{-1} x_{t-1}) u_t}{\sum_{t=2}^{n} H^2_g(\beta_n^{-1} x_{t-1})}.
\]

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
Thm 3.2 in Hall and Heyde, 1980, or Wang, 2014); here Theorem 3.2 may be
applied to verify the stability condition pertaining to its conditional variance.

Reasoning along these lines yields our main result on parametric OLS.

To state it, let \( MN[0, \varsigma^2] \) denote a mixed normal distribution with mix-
ing variate \( \varsigma^2 \) (i.e. which has characteristic function \( u \mapsto E e^{-\varsigma^2 u^2/2} \)), and
\( \sigma(\{\xi_s, u_s\}_{s \leq t}) \) denote the \( \sigma \)-field generated by \( \{\xi_s, u_s\}_{s \leq t} \).

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

(i) \( x_t(n) \) satisfies (5), **Assumption INN** and either **FR** or **MI**;
(ii) \( \{u_t, F_t\}_{t \geq 1} \) is a martingale difference sequence such that \( E[u_t^2 | F_{t-1}] = \sigma^2_u \) a.s. for some constant \( \sigma^2_u \in \mathbb{R}^*_+ \);
(iii) \( \sup_{1 \leq t \leq n} E[u_t^2 1\{|u_t| > A_n\} | F_{t-1}] = o_p(1) \), for any \( 0 < A_n \to \infty \); and
(iv) \( g(x) \) is AHF for the array \( \{x_t(n)\} \), with limit homogeneous component
\( H_g \) such that \( H^2_g \) satisfies the conditions of Theorem 3.2.

Then
\[
n^{1/2} \left[ \hat{\gamma}_n - \gamma \right] \overset{d}{\to} MN \left[ 0, \sigma_u^2 \left( \int \left[ \begin{array}{cc} 1 & -H_g / 2_H_g \end{array} \right] \otimes \right)^{-1} \right].
\]

**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 \( g(x) = \varphi(x - X^-) \) is random. In either case, the \( t \) statistics for testing hypotheses about \( \mu \) or \( \gamma \) will be asymptotically standard normal, so that
inferences may be drawn in the usual manner. This is in contrast to the
case where regressors are \( I(d) \) for \( d > 1/2 \), e.g. see Phillips (1995), Park and
(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 $\gamma$ 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 $L(n)/n^{1/2}$ (see Remark 2.1(b)).

(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

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

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 limit properties of the OLS estimator when $x_t \sim I(d)$ 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

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

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

$$\begin{bmatrix} \hat{m}(x) \\ \hat{m}^{(1)}(x) \end{bmatrix} := \arg\min_{a,b} \sum_{t=2}^{n} [y_t - a - b(x_{t-1} - x)]^2 K_{th}(x),$$ 

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

$$Q := \left\{ \int \begin{bmatrix} 1 & x \\ x & x^2 \end{bmatrix} K \right\}^{-1} \left\{ \int \begin{bmatrix} 1 & x \\ x & x^2 \end{bmatrix} K^2 \right\} \left\{ \int \begin{bmatrix} 1 & x \\ x & x^2 \end{bmatrix} K \right\}^{-1}.$$
Theorem 4.2. Let \( \{y_t\}_{t=1}^{\infty} \) be generated by (20) and suppose that:

(i) conditions (i)–(iii) of Theorem 4.1 hold, and \( \int K = 1 \);
(ii) \( x^j[K(x) + K^2(x)] \) are integrable for \( j \in \{0, 1, 2\} \);
(iii) \( h_n + \beta_n/nh_n \rightarrow 0 \);
(iv) \( m \) has bounded second derivative and \( nh_n^4/\beta_n \rightarrow 0 \);

Then

\[
\sqrt{nh_n} \left( \frac{\tilde{m}(x) - m(x)}{h_n \left( \tilde{m}(1)(x) - m(1)(x) \right)} \right) \xrightarrow{d} MN \left[ 0, \sigma_u^2 \varphi(0)^{-1} Q \right].
\]

If instead of (iv), \( m \) has a bounded derivative and \( nh_n^3/\beta_n \rightarrow 0 \), then

\[
\sqrt{nh_n} \left( \frac{\hat{m}(x) - m(x)}{\tilde{\sigma}^2_u} \right) \xrightarrow{d} MN \left[ 0, \sigma_u^2 \int K^2 \varphi(0)^{-1} \right].
\]

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 it 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 \( \sigma_u^2 \), and consider the nonparametric \( t \) statistic for the hypothesis \( H_0 : m(x) = m_0(x) \) based on the local linear estimator, as given by

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

It follows directly from Theorem 4.2 that \( \bar{t}(x; m_0) \xrightarrow{d} N[0, 1] \), 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 result 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 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

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

where \( g \) is a known function; the alternative is that no such \( \mu \) and \( \gamma \) 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.\footnote{Gao 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.}

We propose to test $H_0$ by comparing parametric OLS and kernel nonparametric estimates of $m$. The model specified in (23) can be estimated parametrically by OLS regression, and also nonparametrically at each $x$ as

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

which under $H_0$ has no asymptotic bias, even if $h$ remains fixed as $n \to \infty$. If $g(x) = x$, so that the null of linearity is being tested, $\hat{m}_g(x)$ specialises to the local linear regression estimator; but in general $\hat{m}_g$ 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 $\hat{m}_g(x)$ has the same limiting distribution as displayed in (21), under $H_0$. Letting $(\hat{\mu}, \hat{\gamma})$ denote the OLS estimates of $(\mu, \gamma)$, 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

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

where $\hat{\sigma}_g^2(x) := \frac{1}{n-1} \sum_{t=2}^{n} [y_t - \hat{\mu} - \hat{\gamma} g(x_{t-1})] K_{th}(x)$. Under $H_0$, 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 $X \subset \mathbb{R}$, constructed as per

(24) $\hat{F} := \sum_{x \in X} \tilde{t}(x; \hat{\mu}, \hat{\gamma})^2$. 
\(\tilde{F}\) is related to the ‘non-predictability sum test’ developed by Kasparis, Andreou and Phillips (2015), who are 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 WNPs. Our next result gives the limiting distribution of the statistic in (24) under the null, and under a sequence of local alternatives of the form

\begin{equation}
H_1 : m(x) = \mu + \gamma g(x) + r_n g_1(x)
\end{equation}

for some \(g_1\), where \(r_n \to 0\). 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 \(\kappa_{g_1}\); this helps to characterise the limiting behaviour of the test statistic under \(H_1\). Define \(\mu_* := \left(-\frac{\int H_{g_2} \varrho \int H_{g_1} \varrho - \int H_{g} \varrho \int H_{g} H_{g_1} \varrho}{\int H_{g} \varrho - (\int H_{g} \varrho)^2}\right)^{-1}(\hat{\mu} - \mu)\) when \(g_1\) is AHF.

**Theorem 5.1.** Suppose that \(X\) has \(p\) elements and

(i) conditions (i)-(iii) of Theorem 4.2 hold, and \(x^j[K(x) + K^2(x)]\) are bounded for \(j \in \{0, 1, 2\}\); and

(ii) \(g(x)\) is AHF for the array \(\{x_t(n)\}\), with limit homogeneous component \(H_g\) such that \(\kappa_g(\beta_n) \to \infty\) and \(H_g^2\) satisfies the requirements of Theorem 3.2. Further, \(g\) and \(g^2\) have bounded second derivatives.

\(^7\)The 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 $H_0$,

$\tilde{F} \xrightarrow{d} \chi^2_p$

Suppose that in addition:

(iii) Either

(a) $g_1, g \cdot g_1$ are bounded and integrable, $r_n^{-1} n^{-1/2} \beta_n \to 0$; or

(b) $g_1(x)$ is AHF for the array $\{ x_1(n) \}$, with limit homogeneous function $H_{g_1}$ such that $H_{g_1} H_g$ satisfies the requirements of Theorem 3.2, and $r_n^{-1} n^{-1/2} \kappa g_1^{-1}(\beta_n) \to 0$. Further, each of the following limits exist (allowing ‘convergence’ to $\infty$):

\[ \kappa_* := \lim_{n \to \infty} \kappa g_1(\beta_n) \quad \kappa_{**} := \lim_{n \to \infty} \kappa g_1(\beta_n) r_n. \]

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

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

(v) $g_1$ and $g \cdot g_1$ have bounded derivatives, and $n h_n^3 r_n^2 / \beta_n \to 0$.

(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 $H_1$,

$\tilde{F} \xrightarrow{P} \infty.$

Remark 5.1. (a) The requirement $\kappa g(\beta_n) \to \infty$ 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 $H_1$, 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 ($\kappa g_1(\beta_n) \to \infty$) and $\kappa g_1(\beta_n) r_n \to \infty$. In such cases, the divergence rate is otherwise unaffected by $r_n$. When $\kappa g_1(\beta_n) \to \infty$ but $\kappa g_1(\beta_n) r_n \to 0$, the...
divergence rate reduces to \( \kappa^2 g_1(\beta_n) r_n^2 n h_n / \beta_n \). The divergence rate is smallest (i.e. \( r_n^2 n h_n / \beta_n \)) in the cases where \( g_1 \) is integrable or AHF of vanishing asymptotic order (i.e. \( \kappa g_1(\beta_n) \to 0 \)).

(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 \( (2p)^{-1/2} (\tilde{F} - p) \overset{d}{\to} (2p)^{-1/2} (\chi^2_p - p) \overset{d}{\to} N(0, 1) \) as \( n \to \infty \) and then \( p \to \infty \).

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 (see Appendix E), it is clear that (26) holds if

(i) the OLS estimator for \( (\mu, \gamma) \) 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 \( x \neq x' \),

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

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 \((n h_n)^{-1/2}\)-consistency and asymptotic normality of the nonparametric estimators follows e.g. from the results of Wu and Mielniczuk.
Thus conditions (i)–(ii) hold. Finally, if $x_t$ has a bounded density $D_x(u)$ (see Wu and Mielniczuk, 2002, Lemma 1) and under standard regularity conditions on $K$, we have

$$
\frac{1}{nh_n} \sum_{t=1}^{n} \mathbb{E}|K_j[(x_t - x) / h_n]K_j[(x_t - x') / h_n]| \\
\leq \sup_u D_x(u) \int_{\mathbb{R}} |K_j[z]K_j[z + (x - x') / h_n]| \, dz \to 0,
$$

so that condition (iii) holds, since $\beta_n \approx 1$ 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 Wang and Phillips (2012), 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 \( d \in \{0.5, 1.0\}; \alpha = 0.1 \)

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

<table>
<thead>
<tr>
<th>( d = 1.0 )</th>
<th>\text{Size adj. power}</th>
<th>\text{Size adjusted, relative to WP}</th>
<th>\text{Size adjusted, relative to WP}</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \varphi_1(x) )</td>
<td>100</td>
<td>0.08 0.07 0.06</td>
<td>0.04 0.06 0.06</td>
</tr>
<tr>
<td>200</td>
<td>0.10 0.09 0.08</td>
<td>0.02 0.05 0.07</td>
<td>0.06 0.09 0.10</td>
</tr>
<tr>
<td>500</td>
<td>0.12 0.12 0.12</td>
<td>0.02 0.05 0.05</td>
<td>0.05 0.07 0.09</td>
</tr>
<tr>
<td>( \varphi_1(2x) )</td>
<td>100</td>
<td>0.08 0.06 0.05</td>
<td>0.02 0.04 0.05</td>
</tr>
<tr>
<td>200</td>
<td>0.09 0.08 0.07</td>
<td>0.01 0.03 0.04</td>
<td>0.05 0.06 0.07</td>
</tr>
<tr>
<td>500</td>
<td>0.10 0.09 0.09</td>
<td>0.02 0.02 0.03</td>
<td>0.03 0.04 0.05</td>
</tr>
<tr>
<td>(</td>
<td>x</td>
<td>^{-2} \lor 1 ) ((\times 0.5))</td>
<td>100</td>
</tr>
<tr>
<td>200</td>
<td>0.09 0.07 0.07</td>
<td>0.04 0.08 0.10</td>
<td>0.09 0.12 0.14</td>
</tr>
<tr>
<td>500</td>
<td>0.09 0.09 0.08</td>
<td>0.02 0.06 0.08</td>
<td>0.05 0.09 0.11</td>
</tr>
<tr>
<td>(</td>
<td>x</td>
<td>^{-1} \lor 1 ) ((\times 0.5))</td>
<td>100</td>
</tr>
<tr>
<td>200</td>
<td>0.08 0.07 0.07</td>
<td>0.03 0.07 0.08</td>
<td>0.08 0.11 0.12</td>
</tr>
<tr>
<td>500</td>
<td>0.11 0.11 0.11</td>
<td>0.02 0.06 0.08</td>
<td>0.06 0.09 0.12</td>
</tr>
<tr>
<td>(</td>
<td>x</td>
<td>^{1.5} ) ((\times 0.02))</td>
<td>100</td>
</tr>
<tr>
<td>200</td>
<td>0.13 0.13 0.13</td>
<td>-0.02 0.04 0.07</td>
<td>0.05 0.08 0.11</td>
</tr>
<tr>
<td>500</td>
<td>0.19 0.21 0.22</td>
<td>-0.10 -0.01 0.00</td>
<td>-0.04 -0.01 0.02</td>
</tr>
<tr>
<td>( x^2 ) ((\times 0.02))</td>
<td>100</td>
<td>0.09 0.07 0.06</td>
<td>0.04 0.06 0.08</td>
</tr>
<tr>
<td>200</td>
<td>0.11 0.11 0.11</td>
<td>0.00 0.01 0.01</td>
<td>0.01 0.01 0.01</td>
</tr>
<tr>
<td>500</td>
<td>0.16 0.19 0.19</td>
<td>0.00 0.00 0.00</td>
<td>0.00 0.00 0.00</td>
</tr>
</tbody>
</table>
For all simulation exercises, the null hypothesis is $H_0 : m(x) = \mu + \beta x$, 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 $(u_t, \xi_t)$ are i.i.d. bivariate Gaussian with unit variances and correlation $\rho$. 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 $h^{-b}$ 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 Wang and Phillips’s (2012), against alternatives that are either integrable ($g_1(x) = \varphi_1(x), \varphi_1(2x), \text{or } x^{-2} \wedge 1$), non-integrable but vanishing at infinity ($g_1(x) = x^{-1} \wedge 1$), or polynomials ($g_1(x) = |x|^v$ with $v \in \{1.5, 2.0\}$).\footnote{By 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).} 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 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 Wang and Phillips (2012), 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 Wang and Phillip’s
Indeed, the 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 \( d = 1 \).

**SUPPLEMENTARY MATERIAL**


References.


Corpus Christi College
University of Oxford
Oxford OX1 4JF
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: kaspars@ucy.ac.cy