DEPARTMENT OF ECONOMICS
UNIVERSITY OF CYPRUS

NONPARAMETRIC PREDICTIVE REFRESSION

Ioannis Kasparis, Elena Andreou & Peter C.B. Phillips

Discussion Paper 14-2012

P.O. Box 20537, 1678 Nicosia, CYPRUS Tel.: +357-22893700, Fax: +357-22895028
Web site: http://www.econ.ucy.ac.cy

Nonparametric Predictive Regression*

Ioannis Kasparis†, Elena Andreou‡, Peter C. B. Phillips§

July 9, 2013

Abstract

A unifying framework for inference is developed in predictive regressions where the predictor has unknown integration properties and may be stationary or nonstationary. Two easily implemented nonparametric F-tests are proposed. The test statistics are related to those of Kasparis and Phillips (2012) and are obtained by kernel regression. The limit distribution of these predictive tests holds for a wide range of predictors including stationary as well as non-stationary fractional and near unit root processes. In this sense the proposed tests provide a unifying framework for predictive inference, allowing for possibly nonlinear relationships of unknown form, and offering robustness to integration order and functional form. Under the null of no predictability the limit distributions of the tests involve functionals of independent χ^2 variates. The tests are consistent and divergence rates are faster when the predictor is stationary. Asymptotic theory and simulations show that the proposed tests are more powerful than existing parametric predictability tests when deviations from unity are large or the predictive regression is nonlinear. Some empirical illustrations to monthly SP500 stock returns data are provided.

Keywords: Functional regression, Nonparametric predictability test, Nonparametric regression, Stock returns, Predictive regression

JEL classification: C22, C32

*Previous version of the paper: September 2012. The authors thank the coeditor and two referees for helpful comments and suggestions. Further thanks go to Timos Papadopoulos for substantial assistance with the simulations and to Eric Ghysels and Tassos Magdalinos for useful comments and suggestions. Andreou acknowledges support of the European Research Council under the European Community FP7/2008-2013 ERC grant 209116. Phillips acknowledges partial support from the NSF under Grant Nos. SES 09-56687 and SES 12-58258.

†University of Cyprus, corresponding author, email: kasparis@ucy.ac.cy

‡University of Cyprus and CERP

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

1 Introduction

The limit distributions of various estimators and tests are well known to be nonstandard in the presence of stochastic trends (e.g., Phillips, 1986, 1987; Chan and Wei, 1987). For instance, least squares cointegrating regression does not produce mixed-normal limit theory or pivotal tests unless strong conditions of long run orthogonality hold. Several early contributions (among others, Phillips and Hansen, 1990; Saikkonen, 1991; Phillips, 1995) developed certain modified versions of least squares for which mixed normality and standard methods of inference applied. While these approaches are now in widespread use in empirical research, some important obstacles to valid inference remain. First, modified statistics require for their validity some prior information about integration properties in order to choose appropriate tests. In consequence, the use of unit root and stationarity tests prior to parametric inference is common practice in applied work, exposing this approach to pre-test difficulties. Second, inference based on modified techniques is not robust to local deviations from the unit root model (Elliott, 1998) and modified tests can exhibit severe size distortions when there are local deviations from unity and significant correlations between the covariates and the equation error. Both of these problems arise in cointegrating and predictive regressions.

To address the second difficulty, several inferential methods that are robust to local deviations from unity have been proposed, including Wright (2000), Lanne (2002), Torus et. al. (2004), Campbell and Yogo (2006), Jansson and Moreira (2006), and Magdalinos and Phillips (2009). The methods have attracted particular attention in the predictive regression literature. Some of the techniques proposed are technically complicated and difficult to implement in practical work, which in part explains why some methods have never been used in empirical work. Most of these approaches also focus on regressions with nearly integrated (NI) covariates and some are invalid for stationary regressors. Implementation of the Campbell and Yogo (2006) method, for instance, typically imposes bounds on the near-to-unity parameter that rule out stable autoregressions. Further, if those bounds are relaxed, it has recently been shown that confidence intervals produced by this method have zero coverage probability in the limit when the predictive regressors are stationary (Phillips, 2012), so there is complete failure of robustness in this case. It is also unknown whether these techniques are valid when the regressors involve fractional processes or other types of nonstationarity. Extension of valid inference to fractional processes is particularly important. Unlike NI processes, fractional processes directly bridge the persistence gap between $I(0)$ and $I(1)$ processes, so that partial sums have a range of magnitudes of the form

$$\sum_{t=1}^{n} x_t = O_p(n^{\alpha}), \text{ for some } \alpha \in (1/2, 3/2)\,. \tag{1}$$

The approach of Magdalinos and Phillips (2009) holds for moderately integrated processes, whose partial sums are of the general form (1), and this method is robust to both NI and stationary regressors.

All of these methods are parametric and may not be robust to functional form misspecification. Functional form affects power in predictive tests under nonstationarity. For instance, fully modified t-tests are based on linear regression and for a near integrated predictor, the test statistic has divergence rate $O_p(n)$ under a linear alternative but may be inconsistent for certain nonlinear alternatives, as we discuss in the paper. In a related vein, Wang and Phillips (2012) found that nonparametric nonstationary specification tests have divergence rates under local alternatives that depend explicitly on the functional form and may be inconsistent for certain functional forms.

The present paper contributes to this literature in several ways. First, we adopt a nonparametric approach using recent theory for nonparametric regression in nonstationary settings by Wang and Phillips (2009a, hereafter WP). Nonparametric F-tests are proposed which have limit distributions that are invariant to integration order. The tests are easy to implement, rely on simple functionals of the Nadaraya-Watson kernel regression estimator, and have limit distributions that apply for a wide range of predictors including stationary as well as non-stationary fractional and near unit root process. In this sense the proposed tests provide a unifying framework for inference. Further, the tests are robust to functional form. The limit distribution of the tests, under the null hypothesis (no predictability), is determined by functionals of independent χ^2 variates. Under the alternative hypothesis (predictability), asymptotic power rates are obtained. The power rates of the nonparametric tests are affected by the bandwidth parameter and are slower than that of parametric tests against linear alternatives. Interestingly, however, the nonparametric tests may attain faster divergence rates than those of parametric tests in cases where parametric fits are misspecified in terms of functional form.

Simulation results suggest that in finite samples the proposed nonparametric tests have stable size properties and can be more powerful than existing parametric predictability tests even when the latter are based on correctly specified models. An empirical illustration of the proposed tests evaluates monthly S&P 500 returns for evidence of predictability using the Earnings Price and Dividend Price ratios over the period 1926-2010 and various subperiods.

The remainder of the paper is organized as follows. Section 2 provides the model, assumptions and some preliminary results. The nonparametric tests and limit theory is given in Section 3. Section 4 considers power. Simulations results are reported in Section 5. The empirical illustration is given in Section 6 and Section 7 concludes. Proofs are given in Appendices A and B.

Notation is standard. For instance, for two sequences a_n, b_n the notation $a_n \sim b_n$ denotes $\lim_{n\to\infty} a_n/b_n = 1$, and $=_d$ represents distributional equality. We use $\lfloor \cdot \rfloor$ to denote integer part, $1\{A\}$ as the indicator function of A, and $i = \sqrt{-1}$. For any sequence X_t, $\overline{X} = \frac{1}{n}\sum_{t=1}^{n} X_t$ and $\underline{X}_t := X_t - \overline{X}$. Similarly, for any functions f_r, $\overline{f} := \int_0^1 f_r dr$ and $\underline{f}_r := f_r - \overline{f}$. Integrals of the form $\int_0^1 G_r dr$ and $\int_0^1 G_r dV_r$ are often written as $\int_0^1 G$ and $\int_0^1 GdV$.

2 Model and Assumptions

We consider predictive regressions of the (possibly nonlinear) form

$$y_t = f(x_{t-\nu}) + u_t,\ f(x) = \mu + g(x), \tag{2}$$

where g is some unknown regression function, $\nu \geq 1$ is an integer valued lag term and u_t is a martingale difference term whose properties are specified below. When x_t is a stationary weakly dependent process, the limit theory of nonparametric regression estimators for models such as (2) is well known from early research (e.g., Robinson, 1983) and overviews in the literature (e.g. Li and Racine, 2007).The limit theory of the nonparametric tests proposed here follows readily from the standard theory in such cases.

The present work focuses on cases where x_t is nonstationary. We are particularly interested in models where $\{x_t\}_1^n$ is generated as a NI array of the commonly used form

$$x_t = \rho_n x_{t-1} + v_t,\ x_0 = 0, \tag{3}$$

with $\rho_n = 1 + \frac{c}{n}$, for some constant c. The error v_t may be a short-memory (SM) time series or an $ARFIMA(d)$, $d \in (-1/2, 1/2)$, process with either long memory (LM) or anti-persistence (AP). Both x_t and u_t are defined on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ with a filtration specified below. The regression function f in (2) is estimated by the Nadaraya-Watson estimator

$$\hat{f}(x) = \frac{\sum_{t=\nu+1}^{n} K_h(x_{t-\nu} - x) y_t}{\sum_{t=\nu+1}^{n} K_h(x_{t-\nu} - x)}, \tag{4}$$

where $K_h(.) = K(./h)$, $K(.)$ is a kernel function and h is a bandwidth with $h = h_n \to 0$ as $n \to \infty$.

To fix ideas and for subsequent analysis we introduce the following technical conditions. Assumptions 2.1 and 2.2 below are largely based on WP (2009a), to which we refer readers for discussion. The WP notation is used here for ease of cross-reference. First, it is convenient to standardise x_t in array form as $x_{t,n} = x_t/d_n$ for some suitable sequence $d_n \to \infty$ so that $x_{\lfloor ns \rfloor, n}$ is compatible with a functional law as $n \to \infty$. We introduce two companion sequences of real numbers c_n and $d_{l,k,n}$ with $d_{l,k,n} \sim C d_{l-k}/d_n$ for some constant C. We note that $(x_{l,n} - x_{k,n})/d_{l,k,n}$ has a limit distribution as $l - k \to \infty$. As in WP, it is convenient to use the set notation.

$$\Omega_n(\eta) = \{(l,k) : \eta n \leq k \leq (1-\eta)n,\ k + \eta n \leq l \leq n\}, 0 < \eta < 1/2.$$

Assumptions 2.1 and 2.2 deal with the density function properties of x_t and their relation to the function f.

Assumption 2.1

For all $0 \le k < l \le n, n \ge 1$, there exist a sequence of σ-fields $\mathcal{F}_{n,k-1} \subseteq \mathcal{F}_{n,k}$ such that, (u_k, x_k) is adapted to $\mathcal{F}_{n,k}$ and conditional on $\mathcal{F}_{n,k}$, $\left(x_{l,n} - \rho_n^{l-k} x_{k,n}\right)/d_{l,k,n}$ has density function $h_{l,k,n}(x)$ such that

(i) $\sup_{l,k,n} \sup_x h_{l,k,n}(x) < \infty$

(ii) *for some $m_o > 0$,*

$$\sup_{(l,k)\in\Omega_n\left(q_o^{1/(2m_o)}\right)} \sup_{|x|\le q_o} |h_{l,k,n}(x) - h_{l,k,n}(0)| = o_p(1),$$

when $n \to \infty$ first and then $q_o \to 0$.

(iii) *for some $m_o > 0$ and $C > 0$, as $n \to \infty$,*

$$\inf_{(l,k)\in\Omega_n(q_o)} d_{l,k,n} \ge q_o^{m_o}/C. \tag{5}$$

Further,

$$\lim_{\eta\to 0} \lim_{n\to\infty} \frac{1}{n} \sum_{l=1}^{\lfloor \eta n\rfloor} (d_{l,0,n})^{-1} = 0, \tag{6}$$

$$\lim_{\eta\to 0} \lim_{n\to\infty} \frac{1}{n} \sum_{l=\lfloor (1-\eta)\, n\rfloor}^{n} (d_{l,0,n})^{-1} = 0, \tag{7}$$

$$\lim_{\eta\to 0} \lim_{n\to\infty} \frac{1}{n} \max_{0\le k\le (1-\eta)\, n} \sum_{l=k+1}^{k+\lfloor \eta n\rfloor} (d_{l,k,n})^{-1} = 0, \tag{8}$$

$$\limsup_{n\to\infty} \frac{1}{n} \max_{0\le k\le n-1} \sum_{l=k+1}^{n} (d_{l,k,n})^{-1} < \infty; \tag{9}$$

Assumption 2.1(i)-(ii) modifies Assumption 2.3(b) of WP. WP consider the conditional density of the increment process $(x_{l,n} - x_{k,n})/d_{l,k,n}$, whereas here we consider the conditional density of $\left(x_{l,n} - \rho_n^{l-k} x_{k,n}\right)/d_{l,k,n}$. It is readily shown that Theorem 2.1 of WP continues to hold under Assumption 2.1 of the current paper.

Assumption 2.2

(a) *The process $x_{\lfloor ns\rfloor,n} := x_{\lfloor ns\rfloor}/d_n$ on the Skorohod space $D[0,1]$, converges weakly to a Gaussian process $G(s)$ that has a continuous local time process $L_G(s,.)$.*

(b) *On a suitably expanded probability space there exists a process $x_{t,n}^o$ such that $\left(x_{t,n}^o,\ 1 \le t \le n\right) =_d (x_{t,n},\ 1 \le t \le n)$ and $\sup_{0\le s\le 1} \left| x_{\lfloor n\eta\rfloor,n}^o - G(s) \right| = o_{a.s.}(1)$.*

Assumption 2.2 (or versions thereof) is standard in the nonstationary time series literature (e.g. Phillips, 1991; Park and Phillips, 1999, 2000, 2001; Berkes and

Horváth, 2006; Wang and Phillips, 2009a). Assumption 2.1 is the same as Assumption 2.3 of WP. In some cases it is more convenient to work with the Skorohod copy $x^o_{t,n}$, instead of $x_{t,n}$. The paper uses convergence results of the NW estimator to some well defined limit and limit distribution results for the NW estimator when x_t is the regression covariate. For our purposes, there is no loss of generality in taking $\left(x^o_{t,n},\ 1 \le t \le n\right) = (x_{t,n},\ 1 \le t \le n)$ instead of $\left(x^o_{t,n},\ 1 \le t \le n\right) =_d (x_{t,n},\ 1 \le t \le n)$. With this convention $\xrightarrow{p}$ (and $\xrightarrow{a.s.}$) convergence, for sample functionals of x_t, should be interpreted as $\xrightarrow{d}$ convergence unless the limit is deterministic.

WP showed that Assumption 2.1 holds when $\rho_n = 1$ and v_t is a long memory process (e.g. ARFIMA (d), $0 < d < 1/2$). The following lemma extends that result by showing that Assumption 2.1 also holds when $\rho_n = 1 + \frac{c}{n}$ and when v_t is anti-persistent $(-1/2 < d < 0)$. To be explicit, we make the following specific assumption on the innovation v_t in (3).

Assumption 2.3 *The time series v_t is a linear process*

$$v_t = \sum_{j=0}^{\infty} \phi_j \xi_{t-j}, \tag{10}$$

where $\xi_t \sim i.i.d.(0, \sigma_\xi^2)$ and $\mathbf{E}\,|\xi_t|^p < \infty$ with $p > 2$. The process ξ_t has characteristic function ψ satisfying $\int_{\mathbb{R}} |\psi(\lambda)|\, d\lambda < \infty$. The coefficients ϕ_j in (10) satisfy one of the following conditions:

SM (short memory). $\sum_{j=0}^{\infty} \left|\phi_j\right| < \infty$, $\sum_{j=0}^{\infty} \phi_j =: \phi \neq 0$;

LM (long memory). *for $j \ge 1$, $\phi_j \sim j^{-m}$, where $m \in (1/2, 1)$;*

AP (anti-persistence). $\sum_{j=0}^{\infty} \phi_j = 0$ *and for $j \ge 1$, $\phi_j \sim j^{-m}$, where $m \in (1, 3/2)$. When $c < 0$ the following additional requirement involving m and c holds. For all $r \in [0, 1)$ we have*

$$\overline{\Phi}_r < 0, \tag{11}$$

where

$$\overline{\Phi}_r := \frac{1}{1-m}(1-r)^{1-m} - \frac{c}{1-m}\int_0^{1-r} \exp(-cs)\left[(1-r)^{1-m} - s^{1-m}\right] ds.$$

Requirement (11) is a technical condition that we show suffices for the validity of the limit theory of Wang and Phillips (2009a) (c.f. Assumption 2.3(b) of Wang and Phillips (2009a) and Assumption 2.1 above). While the restrictions implied by (11) are not immediately clear, the following simple condition on the pair (c, m) for $c < 0$ is sufficient for its validity:

$$1 - ce^{-c}\frac{1-m}{2-m} > 0, \quad \text{or} \quad m > 1 + \frac{1}{1 - ce^{-c}} =: \bar{g}(c). \tag{12}$$

The function $\bar{g}(c)$ is monotonically increasing with $\bar{g}(c) \in (1, 2]$ for $c \in (-\infty, 0]$. Direct calculation shows that $\bar{g}(c) \in (1, 3/2)$ provided $c < -0.352$. Hence, the allowable range for m under **AP** increases as c decreases.

Lemma 1. *Suppose that Assumption 2.3 holds,* $\mathcal{F}_{n,k} \supset \sigma\left(..., \xi_{-1}, ..., \xi_k; 0 \leq k \leq n\right)$ *and* $V(s)$ *is a standard Brownian motion, then Assumptions 2.1 and 2.2(a) hold. If in addition* $3/2 - m > 1/p$ *(*p *is defined in Assumption 2.3), then Assumption 2.2(b) also holds. In particular, we have:*

(i) *under* **SM**, *the sequence* d_n *is* $d_n = n^{1/2}$ *and*

$$G(t) = \sigma_\xi \phi \int_0^t e^{c(t-s)} dV(s);$$

(ii) *under* **LM** *and* **AP**, *the sequence* d_n *is* $d_n = n^{\frac{3}{2}-m}$ *and*

$$G(t) = \sigma_\xi \int_0^t e^{c(t-s)} dB_m(s),$$

where B_m *is fractional Brownian Motion (with Hurst parameter* $H = 3/2 - m$*)*

$$B_m(t) = \frac{1}{1-m}\left\{\int_{-\infty}^0 \left[(t-s)^{1-m} - (-s)^{1-m}\right] dV(s) + \int_0^t (t-s)^{1-m}\, dV(s)\right\}.$$

Remark. The condition $3/2 - m > 1/p$ in Lemma 1 is needed for establishing a strong approximation result for the partial sum process $x_{\lfloor ns \rfloor}/d_n$. This condition requires that the innovation process ξ_t has increasingly higher moments (p) as the degree of anti-persistence increases (i.e. as $m \uparrow 3/2$). The proofs of the paper utilise the availability of the strong approximation in Assumption 2.2(b). Our results may however also be established under the new martingale central limit theory result for quantities like $x_{\lfloor ns \rfloor}/d_n$ under weak convergence of the martingale conditional variance shown in recent work of Wang (2013), thereby confirming the validity of Assumption 2.2(a).

We add the following two assumptions to complete the error specification and properties of the kernel function. Assumption 2.4 is standard in the prediction literature in financial applications and regularly appears in the local to unity regression literature (e.g. Jansson and Moreira, 2006) and nonparametric regression literature (Wang and Phillips, 2009a). Nonetheless, given the results in Wang and Phillips (2009b), there is reason to believe that the nonparametric predictive regression tests here may be extendable to structural regressions[1]. Assumption 2.5 is used in WP

[1] Simulation results (not reported) indicate that structural regression endogeneity results in some size distortion, which can be corrected by additional undersmoothing.

and provides technical conditions that facilitate the derivation of the limit distribution theory.

Assumption 2.4 $\{(\xi_t, u_t), \mathcal{F}_{n,t}\}$ *is a martingale difference sequence such that*

$$\mathbf{E}\left[(\xi_t, u_t)'(\xi_t, u_t)|\mathcal{F}_{n,t-1}\right] = \Psi = \begin{bmatrix} \sigma_\xi^2 & \sigma_{\xi,u} \\ \sigma_{\xi,u} & \sigma_u^2 \end{bmatrix} \ a.s.,$$

with $\|\Psi\| < \infty$ *a.s. Further, for some* $\omega > 2$, $\sup_{1\le t\le n} \mathbf{E}(u_t^\omega|\mathcal{F}_{n,t-1}) < \infty$ *a.s.*

Assumption 2.5. *The kernel function satisfies* $K(s) \geq 0$, $\int_{\mathbb{R}} K(s)ds < \infty$ *and* $\sup_s K(s) < \infty$. *The bandwidth* h_n *satisfies* $h_n \to 0$ *and* $d_n (nh_n)^{-1} \to 0$ *as* $n \to \infty$.

Assumption 2.6. *For given* x, *there exists a real function* $f_o(s, x)$ *and* $0 < \gamma \leq 1$ *such that, when* h *is sufficiently small,* $|f(hs + x) - f(x)| \leq h^\gamma f_o(s, x)$ *for all* $s \in \mathbb{R}$ *and* $\int_{\mathbb{R}} K(s) f_o(s, x)\, ds < \infty$. *Furthermore,* $nh_n^{1+2\gamma}/d_n \to 0$, *as* $n \to \infty$.

Suppose that y_t is generated by equations (2) and (3). Then we have the following result.

Lemma 2. *Suppose that Assumptions 2.1-2.6 hold and* $3/2 - m > 1/p$. *Then as* $n \to \infty$ *we have*

$$\sqrt{\frac{nh_n}{d_n}}\left(\hat{f}(x) - f(x)\right) \xrightarrow{d} MN\left(0, \frac{\sigma_u^2 \int_{-\infty}^{\infty} K(s)^2 ds}{L_G(1,0)\int_{-\infty}^{\infty} K(s)ds}\right) \tag{13}$$

and

$$\left(\sum_{t=\nu+1}^{n} K\left(\frac{x_{t-\nu} - x}{h_n}\right)\right)^{1/2} \left(\hat{f}(x) - f(x)\right) \xrightarrow{d} N\left(0, \sigma_u^2 \int_{-\infty}^{\infty} K(s)^2 ds\right).$$

It follows that in the predictive regression framework (2)-(3), the NW estimator is consistent and has a Gaussian limit distribution. Importantly, the limit distribution is free of the nuisance near to unity parameter c. As indicated earlier, when x_t is a stationary weakly dependent process such as a stable AR process, standard results confirm that the convergence in (13) still holds. Thus, (13) offers wide generality in the predictive regression context and this facilitates the development of a class of nonparametric predictability tests.

Remark. The convergence rate in (13) is $\sqrt{\frac{nh_n}{d_n}} = \sqrt{n^{m-\frac{1}{2}} h_n}$ for $m \in (1/2, 3/2)$ and depends on the memory parameter m. Faster convergence is attained under **AP** (with $m \in (1, 3/2)$). When the memory of the covariate increases, there is information loss in local methods of estimation like NW because there are fewer observations in local neighborhoods as the random wandering character of the series becomes more pronounced (i.e., as m decreases). As a result of this information loss, there is a reduction in the convergence rate of the NW estimator. In fact, as $m \downarrow 1/2$, the convergence rate becomes arbitrarily slow. The convergence rate of the NW estimator determines the asympotic power rates of the proposed tests, with faster convergence translating to higher power rates (see Remark(b) after Theorem 2 in the subsequent section).

3 Nonparametric Predictive Tests

The null hypothesis is no predictability in regression (2), so that under $H_0 : f(x) = \mu$ the regression function is constant and $y_t = \mu + u_t$. Hence, in view of (13), $\hat{f}(x) \xrightarrow{p} \mu$, which suggests a test based on

$$\hat{t}(x, \mu) := \left(\frac{\sum_{t=1+\nu}^{n} K\left(\frac{x_{t-\nu}-x}{h_n}\right)}{\hat{\sigma}_u^2 \int_{-\infty}^{\infty} K(s)^2 ds} \right)^{1/2} \left(\hat{f}(x) - \mu\right), \tag{14}$$

where $\hat{\sigma}_u^2 = \sum_{t=1+\nu}^{n} (y_t - \hat{\mu})^2 /n$ is a consistent estimator of σ_u^2. The idea is to compare the estimator $\hat{f}(x)$ with a constant function and, although μ is generally unknown, it can be consistently estimated by simple regression as $\hat{\mu} = \sum_{t=1+\nu}^{n} y_t/n$ under the null. Further, under H_0, it can be shown that $\hat{t}(x, \hat{\mu}) = \hat{t}(x, \mu) + o_p(1)$ and

$$\hat{t}(x, \hat{\mu}) \xrightarrow{d} N(0, 1). \tag{15}$$

Therefore, the feasible statistic $\hat{t}(x, \hat{\mu})$ involves a comparison of the nonparametric estimator $\hat{f}(x)$ with the parametric estimator $\hat{\mu}$. This statistic is similar to the linearity test of Kasparis and Phillips (2012) developed in the context of dynamic misspecification.

The predictive test statistics are based on making the comparison (14) over some point set. In particular, let X_s be a set of isolated points $X_s = \{\bar{x}_1, ..., \bar{x}_s\}$ in $\mathbb{R}$ for some fixed $s \in \mathbb{N}$. The tests we propose involve sum and sup functionals over this set, viz.,

$$\widehat{F}_{\text{sum}} := \sum_{x \in X_s} \hat{F}(x, \hat{\mu}) \text{ and } \widehat{F}_{\max} := \max_{x \in X_s} \hat{F}(x, \hat{\mu}), \text{ with } \hat{F}(x, \hat{\mu}) := \hat{t}(x, \hat{\mu})^2. \tag{16}$$

In practical work the set X_s can be chosen using uniform draws over some region of particular interest in the state space.

The no predictability hypothesis in (2) can be written as

$$H_0 : g(x) = 0, \; \textit{a.e. with respect to Lebesgue measure} \tag{17}$$

where $f = g + \mu$. The alternative hypothesis is

$$H_1 : g(x) \neq 0, \; \textit{on some set } S_g \textit{ of positive Lebesgue measure}$$

In some cases (see Theorem 2 and the subsequent Remark (a) below) for the tests to have power against H_1 it is important that the intersection of S_g and X_s be nonempty.

The following result gives the null limit distributions of the test statistics in (16).

Theorem 1. *Suppose that Assumptions 2.1-2.5 hold and $3/2 - m > 1/p$. Under H_0 as $n \to \infty$*

$$\widehat{F}_{\text{sum}} \xrightarrow{d} \chi_s^2 \text{ and } \widehat{F}_{\max} \xrightarrow{d} Y,$$

where the random variable Y has c.d.f. $F_Y(y) = P\left(X \le y\right)^s$ with $X \sim \chi_1^2$.

The components $\hat{t}(\bar{x}_1, \hat{\mu})$,..., $\hat{t}(\bar{x}_s, \hat{\mu})$ in the statistics $\widehat{F}_{\text{sum}}$ and $\widehat{F}_{\max}$ are asymptotically independent because the points $\{\bar{x}_j : j = 1, ..., s\}$ in X_s are isolated. As a result, $\widehat{F}_{\text{sum}}$ has a χ_s^2 limit and the limit distribution of $\widehat{F}_{\max}$ is determined as the maximum of s independently distributed χ_1^2 variates. Note that for a chi-square random variable, χ_s^2, we have the limiting Gaussian approximation $(2s)^{-1/2}\left(\chi_s^2 - s\right) \xrightarrow{d} N(0,1)$ as $s \to \infty$. Therefore, it seems possible to construct test statistics with standard limit distributions for the case where the number of grid points $s \to \infty$. Accordingly, we conjecture that under certain conditions $(2s)^{-1/2}\left(\widehat{F}_{\text{sum}} - s\right) \xrightarrow{d} N(0,1)$, as $n, s \to \infty$ (see for example de Jong and Bierens, 1994). We leave explorations in this direction for future work.

The properties of these tests under H_1 depend on the regression function. Under certain conditions, the scaled statistics $\frac{d_n}{h_n n}\widehat{F}_{\text{sum}}$ and $\frac{d_n}{h_n n}\widehat{F}_{\max}$ have well defined limits. These limits are determined by the nature of the regression function g for which it is convenient to use the following classification.

Definition. (H-regular regression functions): *The function g is H-regular (with respect to x_t) if*

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

where:

(i) $\sup_x |r_g(\lambda, x)| = o\left(\kappa_g(\lambda)\right)$ *as* $\lambda \to \infty$.
(ii) *for some* $0 < \alpha \le 1$, $|x|^{a-1} H_g(x)$ *is locally integrable and* $\int_0^1 \left(\mathbf{E}G(t)^2\right)^{-\alpha/2} dt < \infty$.
(iii) $\lim_{n\to\infty} n\left(d_{l,0,n}\right)^{\alpha} = \infty$ *for each* l.
(iv) $\lim\sup_{n\to\infty} \frac{1}{n}\sum_{l=1}^{n}\left(d_{l,0,n}\right)^{-\alpha} < \infty$.
(v) $x_{l,n}/d_{l,0,n}$ *has density* $h_{l,0,n}(x)$ *satisfying* $\sup_{l,n}\sup_x |x|^{1-\alpha} h_{l,0,n}(x) < \infty$;

Condition (i) above postulates that the regression function g is asymptotically homogeneous (see Park and Phillips 1999, 2001). Conditions (ii)-(v) are due to Berkes and Horváth (2006, Theorem 2.2) who extend the limit theory of Park and Phillips

(1999, 2001) to a more general class of nonlinear functions and processes such as ARFIMA models (see also de Jong (2004) and Pötscher (2004)).

Remark. Under Assumption 2.3, condition $\int_0^1 (\mathbf{E}G(t)^2)^{-\alpha/2} dt < \infty$ in (ii) of the definition is satisfied with $\alpha = 1$. To see this, set $\mathcal{C} = 1\{c \geq 0\} + e^{2c}\{c < 0\}$. Then, under **LM** or **AP**, we have for $t \in [0,1]$

$$\mathbf{E}G(t)^2 \geq \frac{\mathcal{C}}{(1-m)^2} \int_0^t (t-s)^{2(1-m)} ds = \frac{\mathcal{C}}{(1-m)^2 (3-2m)} t^{3-2m}.$$

Hence,

$$\int_0^1 \left(\mathbf{E}G(t)^2\right)^{-1/2} dt \leq \sqrt{\frac{(1-m)^2 (3-2m)}{\mathcal{C}}} \int_0^1 t^{m-3/2} dt = \frac{\sqrt{(1-m)^2 (3-2m)}}{(m-1/2)\sqrt{\mathcal{C}}} < \infty.$$

Similar arguments show that the above condition also holds under **SM**. Further, for $\alpha = 1$ condition (iii) is trivially satisfied, while conditons (iv) and (v) are special cases of (9) and Assumption 2.1(i) respectively.

Theorem 2. *Let Assumptions 2.1-2.5 hold and $3/2 - m > 1/p$. For $g(x)$ (and $g(x)^2$) H-regular, set $\sigma_*^2 = \int_0^1 \overline{H}_g(G(s))^2 ds$ and assume that $\sqrt{n}\kappa_g(d_n) \to \infty$ as $n \to \infty$.*[2] *Then under H_1 as $n \to \infty$ we have:*

$$\frac{d_n}{h_n n} \widehat{F}_{\text{sum}} \xrightarrow{p} \sum_{x \in X_s} D(x) \text{ and } \frac{d_n}{h_n n} \widehat{F}_{\max} \xrightarrow{p} \max_{x \in X_s} D(x),$$

where

(i) *for g H-regular with $\kappa_g(\lambda) = 1$*

$$D(x) = \frac{L_G(1,0) \int_{-\infty}^{\infty} K(s)\, ds}{(\sigma_*^2 + \sigma_u^2) \int_{-\infty}^{\infty} K(s)^2 ds} \left[g(x) - \int_0^1 H_g(G(s)) ds\right]^2.$$

(ii) *for g H-regular with $\lim_{\lambda \to \infty} \kappa_g(\lambda) = \infty$*

$$D(x) = \frac{L_G(1,0) \int_{-\infty}^{\infty} K(s)\, ds}{\sigma_*^2 \int_{-\infty}^{\infty} K(s)^2 ds} \left[\int_0^1 H_g(G(s)) ds\right]^2.$$

---

[2] Recall that for any functions f_r, $\overline{f} := \int_0^1 f_r dr$ and $\overline{f}_r := f_r - \overline{f}$. Therefore,

$$\int_0^1 \overline{H}_g(G(s))^2 ds := \int_0^1 H_g(G(s))^2 ds - \left[\int_0^1 H_g(G(s)) ds\right]^2.$$

(iii) *for g H-regular with* $\lim_{\lambda\to\infty}\kappa_g(\lambda)=0$ *or g integrable*

$$D(x)=\frac{L_G\left(1,0\right)\int_{-\infty}^{\infty}K\left(s\right)ds}{\sigma_u^2\int_{-\infty}^{\infty}K(s)^2ds}g(x)^2.$$

Remarks.

(a) The formulation of the test hypothesis is different than that of Kasparis and Phillips (2012). Kasparis and Phillips essentially require that the intersection of S_g and X_s be nonempty under H_1. Indeed, it follows from the form of the limit process $D\left(x\right)$ in Theorem 2(iii) that for g H-regular with $\lim_{\lambda\to\infty}\kappa_g(\lambda)=0$ or g integrable, the intersection of S_g and X_s must be nonempty for the tests to have power under H_1. Nevertheless, for g H-regular with $\lim_{\lambda\to\infty}\kappa_g(\lambda)=1$ or ∞, the tests have non trivial asymptotic power even if the intersection of S_g and X_s is empty. For example suppose that $g(x)=1\left\{x>0\right\}$, and the set X_s is the singleton $X_s=\{-1\}$. Then, using the arguments in the proof of Theorem 2, we have as $n\to\infty$

$$\begin{aligned}\hat{F}(x &= -1,\hat{\mu})\approx\frac{h_n n}{d_n}\frac{L_G(1,0)\int_{-\infty}^{\infty}K(\lambda)d\lambda}{\left(\int_0^1\left(\overline{1}\left\{G(r)>0\right\}\right)^2dr+\sigma_u^2\right)\int_{-\infty}^{\infty}K(\lambda)^2d\lambda}\\ &\times\left\{\left[\mu+\underbrace{g(-1)}_{=0}\right]-\left[\mu+\int_0^1 1\left\{G(r)>0\right\}dr\right]\right\}^2\xrightarrow{p}\infty.\end{aligned}$$

(b) If the term $\int_0^1 H_g(G(s))ds$ in Theorem 1(i,ii) is a continuous random variable, then $D(x)>0$ *a.s.* and the tests are consistent in this case. Further, if $g(x)\neq 0$ for some $x\in X_s$, then the term $D(x)$ in Theorem 1(iii) is $D(x)>0$ *a.s.*

(c) Theorem 2 shows that, under the alternative hypothesis and with $\rho_n=1+c/n$, the tests have the following order of magnitude

$$\widehat{F}_{\text{sum}},\,\widehat{F}_{\max}=O_p\left(h_n n^{m-1/2}\right)\ \text{with}\ m\in(1/2,3/2).$$

If h_n is chosen to vanish at a slowly varying rate (e.g. $h_n=1/\ln(n)$), then the statistics are divergent for all $m\in(1/2,3/2)$. But if $h_n\sim n^{-b}$ (with $b\in(0,1/2)$), then

$$\widehat{F}_{\text{sum}},\,\widehat{F}_{\max}=O_p\left(\frac{n^{m-1/2}}{n^b}\right).$$

In this case the statistics are divergent if $m-1/2>b$. This condition is satisfied under **SM** and **AP**, but not necessarily under **LM**. For the **LM** case some prior information about the parameter m is required to ensure that the bandwidth is appropriately chosen to yield consistent tests. For instance, a bandwidth of the form $h_n\sim n^{-\delta\left(m-\frac{1}{2}\right)}$, $\delta\in(0,1)$ yields consistent tests in all cases. Section 5 of the paper provides simulation results for bandwidths of the form $\hat{h}_n=n^{-\delta\left(\hat{m}-\frac{1}{2}\right)}$ where $\hat{m}$ is an estimator for m.

Memory estimators are common in inferential methods for fractional systems (e.g. Robinson and Hualde (2003), Marmol and Velasco (2004), Hualde and Robinson (2010)). Preliminary findings of the authors indicate that under certains conditions, Theorems 1 and 2 also hold when a stochastic bandwidth of the form $\hat{h}_n = n^{-\delta\left(\hat{m}-\frac{1}{2}\right)}$ is utilised.[3] We leave detailed theoretical explorations of this matter to future work.

(d) If the autoregressive parameter in (3) is fixed with $\rho_n = \rho$ and $|\rho| < 1$, then x_t is asymptotically stationary and weakly dependent. By standard limit theory in this case the proposed tests have divergence rate $O_p\left(h_n n\right)$.

(e) Kasparis and Phillips (2012) considered the consequences of dynamic misspecification on performance of nonparametric F-tests, in a similar context to that of the current paper. Misspecifing the lag order ν of the predictor variable in (2) is likely to reduce the power of the proposed tests in finite samples, under integrable alternatives or H-regular alternatives with regression functions of vanishing asymptotic order. Suppose for instance that the true lag order is ν and (i.e. (2) holds) but the NW estimator employed in the test statistics utilizes the predictor x_t with lag order $\nu_* \neq \nu$. Then extending the arguments of Kasparis and Phillips (2012) to the current framework it can be shown that under H_1 and for g H-regular with $\kappa_g(\lambda) \to 0$ or g integrable, Theorem 2(iii) holds with

$$D(x) = \frac{L_G\left(1,0\right)\int_{-\infty}^{\infty} K\left(s\right) ds}{\sigma_u^2 \int_{-\infty}^{\infty} K(s)^2 ds} \left[\mathbf{E}g\left(x \pm \sum_{i=1}^{|\nu_* - \nu|} v_i\right)\right]^2.$$

Therefore, the divergence rate is not affected by lag misspecification. Nevertheless, if $\lim_{x \to \pm\infty} g(x) = 0$, the term

$$\mathbf{E}g\left(x \pm \sum_{i=1}^{|\nu_* - \nu|} v_i\right)$$

vanishes in general as the degree of lag misspecification increases i.e. as $|\nu_* - \nu| \to \infty$.[4] A reduction in power as $|\nu_* - \nu| \to \infty$ is confirmed in this case by simulation results in Kasparis and Phillips (2012). For g H-regular with asymptotic order $\kappa_g(\lambda) \to \infty$, Theorem 2(ii) holds even if the lag order is misspecified. Finally, for g

[3] For instance, is can be shown that under some additional minor requirements on the kernel function and for MSE rate $\mathbf{E}(\hat{m} - m)^2 = O\left(\ln(n) n^{-\frac{4}{5}}\right)$, we have

$$\frac{d_n}{n\hat{h}_n} \sum_{t=1}^{n} K\left(\frac{x_t - x}{\hat{h}_n}\right) - \frac{d_n}{n h_n} \sum_{t=1}^{n} K\left(\frac{x_t - x}{h_n}\right) = o_p(1),$$

with $h_n = n^{-\frac{1}{5}\left(m-\frac{1}{2}\right)}$ and $\hat{h}_n = n^{-\frac{1}{5}\left(\hat{m}-\frac{1}{2}\right)}$, for $m \in (1/2, 3/2)$.

[4] See, for example, Kasparis, Phillips and Magdalinos (2012).

H-regular with asymptotic order $\kappa_g(\lambda) = 1$, Theorem 2(i) holds with

$$D(x) = \frac{L_G\left(1,0\right)\int_{-\infty}^{\infty} K\left(s\right)ds}{\left(\sigma_*^2 + \sigma_u^2\right)\int_{-\infty}^{\infty} K(s)^2 ds}\left[\mathbf{E}g\left(x \pm \sum_{i=1}^{|\nu_* - \nu|} v_i\right) - \int_0^1 H_g(G(s))ds\right]^2.$$

4 Divergence Rates of Parametric Predictive Tests under Functional Form Misspecification

Existing predictability tests are based on parametric linear fits of the form

$$y_t = \tilde{\mu} + \tilde{\beta} x_{t-\nu} + \hat{u}_t, \tag{18}$$

for certain intercept and slope coefficient estimators $\tilde{\mu}$, $\tilde{\beta}$. In this framework, the test hypothesis under consideration is $H_0 : \beta = 0$ (no predictability) against $H_1 : \beta \neq 0$ (predictability) where β is the assumed coefficient of the predictor. Parametric tests based on such linear fits may or may not have discriminatory power against various nonlinear alternatives such as

$$y_t = g(x_{t-\nu}) + u_t. \tag{19}$$

To explore the effects of nonlinearity under the alternative we consider the power properties of two parametric tests of predictability when the fitted model is linear and the predictive regression is non-linear. In particular, we examine the asymptotic behaviour of the fully modified t-statistic ($\hat{t}_{FM}$) (see Phillips and Hansen, 1990; Phillips, 1995) and the Jansson and Moreira (2006, hereafter JM) test statistic ($\hat{R}_\beta$). We assume that y_t is generated as in (19) where x_t is a (near) unit root process of the form (3) with short memory innovations[5].

When the regression function in (19) is linear, i.e. $g(x) = x$, it is readily shown that both test statistics attain a divergence rate of order n. For g non-linear and locally integrable (but not integrable), the divergence rate is slower. Finally for g integrable the test statistics are bounded in probability and therefore inconsistent. These results are demonstrated in Theorem 3 below.

Before presenting the results we introduce some notation. Define the covariance matrix

$$\Omega = \mathbf{E}\begin{bmatrix} u_t^2 & \sum_{k=-\infty}^{\infty} u_t v_{t+k} \\ \sum_{k=-\infty}^{\infty} v_t u_{t+k} & \sum_{k=-\infty}^{\infty} v_t v_{t+k} \end{bmatrix} = \begin{bmatrix} \Omega_{uu} & \Omega_{uv} \\ \Omega_{vu} & \Omega_{vv} \end{bmatrix}. \tag{20}$$

For simplicity in the following presentation, we assume that v_t is i.i.d.[6] The subsequent

[5]Note that the FM-OLS method of Phillips (1995) and the J&M tests are both developed for unit root processes driven by short memory innovations.

[6]For this case $\Omega = \mathbf{E}\begin{bmatrix} u_t^2 & u_t v_t \\ v_t u_t & v_t^2 \end{bmatrix}$.

results can be extended for the case where v_t is a short memory linear process[7]. Next, consider the FM-OLS estimator in (18):

$$\begin{aligned}\tilde{\beta} &= \frac{\sum_{t=1+\nu}^{n} y_t^+ x_{t-\nu} - \frac{1}{n}\sum_{t=1+\nu}^{n} y_t^+ \sum_{t=1+\nu}^{n} x_{t-\nu}}{\sum_{t=1+\nu}^{n} x_t^2 - \frac{1}{n}\left(\sum_{t=1+\nu}^{n} x_t\right)^2},\\ \tilde{a} &= \bar{y}^+ - \tilde{\beta}\bar{x},\end{aligned}$$

with $y_t^+ = y_t - \hat{v}_t\hat{\Omega}_{vv}^{-1}\hat{\Omega}_{vu}$, $\hat{v}_t = x_t - \hat{\rho}x_{t-1}$. Here, $\hat{\Omega}_{uu}$, $\hat{\Omega}_{vu}$ $\hat{\Omega}_{vv}$ are given by

$$\left[\hat{\Omega}_{uu}, \hat{\Omega}_{vv}, \hat{\Omega}_{vu}\right] := \frac{1}{n}\left[\sum_{t=1+\nu}^{n} \hat{u}_t^2, \sum_{t=2}^{n} \hat{v}_t^2, \sum_{t=1+\nu}^{n} \hat{v}_t\hat{u}_t\right].$$

Next, define the pseudo-true values[8]

$$a_* := \int_0^1 H_g(G(r))dr - \beta_*\int_0^1 G(r)dr,\ \beta_* := \frac{\int_0^1 \overline{H}_g(G(r))\overline{G}(r)dr}{\int_0^1 \overline{G}(r)^2 dr},$$

$$\beta_{**} := \frac{\int_0^1 G(r)dB_u(r) - \int_0^1 G(r)dr\left(\int_{-\infty}^{\infty} g(s)dsL_G(1,0) + B_u(1)\right)}{\int_0^1 \overline{G}^2(r)dr},$$

$$\Omega_{uu}^* := \int_0^1 \left[H_g(G(r)) - a_* - \beta_*G(r)\right]^2 dr \text{ and } \Omega_{uu}^{**} := \begin{cases} \Omega_{uu}^*, \text{ for } \kappa_g(\lambda) \to \infty \\ \Omega_{uu}^* + \Omega_{uu}, \text{ for } \kappa_g(\lambda) = 1 \\ \Omega_{uu}, \text{ for } \kappa_g(\lambda) \to 0 \end{cases}$$

The test statistics under consideration are

$$\hat{t}_{IV} = \frac{\tilde{\beta}}{\sqrt{\hat{\Omega}^+\left\{\sum_{t=1+\nu}^{n} x_{t-\nu}^2 - \frac{1}{n}\left(\sum_{t=1+\nu}^{n} x_{t-\nu}\right)^2\right\}}},$$

and

$$\hat{R}_\beta = \frac{1}{\sqrt{\hat{\Omega}_{vv}\hat{\Omega}^+}}\left\{\frac{1}{n}\sum_{t=1+\nu}^{n}\left(x_{t-\nu} - \frac{1}{n}\sum_{t=1+\nu}^{n} x_{t-\nu}\right)\left[y_t^+ - \hat{\beta}x_{t-\nu}\right]\right\},$$

where $\hat{\Omega}^+ = \hat{\Omega}_{uu} - \hat{\Omega}_{vv}^{-1}\hat{\Omega}_{vu}^2$, $\hat{\beta} = \left[\sum_t x_{t-\nu}^2 - \frac{1}{n}\left(\sum_t x_{t-\nu}\right)^2\right]^{-1}\left[\sum_t y_t x_{t-\nu} - \frac{1}{n}\sum_t y_t \sum_t x_{t-\nu}\right]$ and B_u is the Brownian motion limit of the partial sum process of u_t

[7] In order to obtain the limit properties of the parametric tests, when v_t is a linear process, we need to characterise the pseudo-true limits of various long run variance estimators under functional form misspecification, as in Kasparis (2008).

[8] The quantities a_*, β_* and Ω_{uu}^{**} are the random limits of the OLS coefficient and covariance estimators when the predictive regression is misspecified in terms of functional form.

Theorem 3. *Suppose that Assumption 2.3* ***SM*** *holds with v_t i.i.d. The fitted model is given by (18) and $\{y_t\}$ is generated by (19). Further, for g H-regular suppose that $\sqrt{n}\kappa_g(\sqrt{n}) \to \infty$. Then*

(a) *For $g(x)$ (and $xg(x)$, $g(x)^2$) H-regular and*

(i) $\kappa_g(\lambda) \to \infty$

$$\frac{1}{\sqrt{n}}\hat{t}_{FM} \xrightarrow{d} \frac{\int_0^1 \overline{H}_g(G(r))\overline{G}(r)dr}{\sqrt{\Omega_{uu}^{**}\int_0^1 \overline{G}(r)^2 dr}},$$

$$\frac{1}{\sqrt{n}}\hat{R}_\beta \xrightarrow{d} \frac{1}{\sqrt{\Omega_{vv}\Omega_{uu}^{**}}}\int_0^1 \overline{G}(r)\left[H_g(G(r)) - \beta_* G(r)\right]dr,$$

(ii) $\kappa_g(\lambda) = O(1)$

$$\frac{1}{\kappa_g(\sqrt{n})\sqrt{n}}\hat{t}_{FM} \xrightarrow{d} \frac{\int_0^1 \overline{H}_g(G(r))\overline{G}(r)dr}{\sqrt{\left\{\Omega_{uu}^{**} - \Omega_{vv}^{-1}\Omega_{vu}^2\right\}\int_0^1 \overline{G}(r)^2 dr}},$$

$$\frac{1}{\kappa_g(\sqrt{n})\sqrt{n}}\hat{R}_\beta \xrightarrow{d} \frac{1}{\sqrt{\Omega_{vv}\Omega_{uu}^{**} - \Omega_{vu}^2}}\int_0^1 \left\{\overline{G}(r)\left[H_g(G(r)) - \beta_* G(r)\right]\right\}dr.$$

(b) *For $g(x)$ integrable*

$$\hat{t}_{FM} \xrightarrow{d} \frac{1}{(\Omega^+)^{1/2}}\left[\left\{B_u(1) - V(1)\Omega_{vv}^{-1}\Omega_{vu}\right\} - c\Omega_{vv}^{-1}\Omega_{vu}\left\{\int_0^1 \overline{G}(r)^2 dr\right\}^{1/2}\right],$$

$$\hat{R}_\beta \xrightarrow{d} \mathcal{R}_\beta - \frac{\beta_{**}}{\sqrt{\Omega_{vv}\Omega^+}}\int_0^1 \overline{G}(r)^2 dr,$$

where

$$\mathcal{R}_\beta = \frac{1}{\sqrt{\Omega_{vv}\Omega^+}}\left\{\int_0^1 \overline{G}(r)d\left[B_u(r) - V(r)\Omega_{vv}^{-1}\Omega_{vu}\right] - c\Omega_{vv}^{-1}\Omega_{vu}\int_0^1 \overline{G}(r)^2 dr\right\}.$$

Remarks.

(a) As indicated above, when the fitted model is correctly specified in terms of a linear functional form, parametric tests attain a divergence rate of order n i.e.

$$\hat{t}_{FM}, \hat{R}_\beta = O_p(n).$$

But when functional form misspecification is committed, Theorem 3 suggests that parametric tests are either inconsistent or attain slower divergence rates. Divergence rates depend on the nature of the regression function. For locally integrable predictive functions (that are not integrable) the test statistics diverge at rates slower than n. For integrable g the test statistics are bounded in probability and therefore the tests are inconsistent. In particular, we have

$$\hat{t}_{FM}, \hat{R}_{\beta} = \begin{cases} O_p\left(\sqrt{n}\right), \; g \textit{ H-regular} \text{ with } \kappa_g(\lambda) \to \infty \\ O_p\left(\kappa_g(\sqrt{n})\sqrt{n}\right), \; g \textit{ H-regular} \text{ with } \kappa_g(\lambda) = O(1) \\ O_p(1), \; g \textit{ integrable} \end{cases}$$

Note that for g polynomial H-regular the divergence rate is of order $O_p(n^{\varsigma})$ with $0 < \varsigma \leq 1/2$.

(b) For g integrable we have the following outcomes.

(i) The limit distribution of the $\hat{t}_{FM}$ statistic is identical to that obtained under the null hypothesis. Therefore, in this case the asymptotic power of the test is identical to size. The simulation results presented in the subsequent section suggest that finite sample power is also close to size.

(ii) The limit distribution of the $\hat{R}_{\beta}$ statistic under the null hypothesis is given by $\mathcal{R}_{\beta}$. Under the alternative hypothesis an additional term features in the limit, viz.,

$$-\frac{\beta_{**}}{\sqrt{\Omega_{vv}\Omega^{+}}} \int_0^1 \overline{G}(r)^2 dr. \tag{21}$$

This additional term is random and its sign is determined by the (random) pseudo true value β_{**}. Power is correspondingly random, being influenced by the distribution of (21), and may therefore be greater or less than the size of the test. The test is inconsistent in this case.

(c) If Ω_{uu} is estimated by some HAC estimator, the divergence rates of $\hat{t}_{FM}$ and $\hat{R}_{\beta}$ will be adversely affected by the bandwidth term M_n ($M_n \to \infty$) employed in the HAC estimator[9]. In particular, it can be shown that

$$\hat{t}_{FM}, \hat{R}_{\beta} = \begin{cases} O_p\left(\sqrt{\frac{n}{M_n}}\right), \; M_n \kappa_g(\sqrt{n})^2 \to \infty \\ O_p\left(\kappa_g(\sqrt{n})\sqrt{n}\right), \; M_n \kappa_g(\sqrt{n})^2 = O(1) \\ O_p(1), \; g \textit{ integrable} \end{cases}$$

[9] If Ω_{vu} or Ω_{vv} are estimated by HAC procedures, the divergence rates are the same as those reported in part (a) of this Remark.

5 Simulations

This section reports simulation results for the finite sample properties of the F_{sum}, t_{FM} tests (2000 replications [10]) and the Jansson and Moreira (2006, JM) tests (500 replications[11]). As indicated in the previous footnotes, there is a substantial difference in computational time required for these two classes of tests and in our experience serious practical difficulties of convergence arise in implementing the JM procedure in some cases. Some simulation results are also provided for the $F_{\max}$ test. The F_{sum} test outperforms the $F_{\max}$ test in most cases, with the latter exhibiting better size under nearly explosive predictors and superior power against integrable alternatives. We consider two-sided versions of the t_{FM} and the JM tests.

The model is generated from

$$y_t = f(x_{t-1}) + u_t,\ x_t = \left(1 + \frac{c}{n}\right) x_{t-1} + v_t,\ x_0 = 0$$

$$\begin{aligned} v_t &= \rho_x v_{t-1} + \eta_t,\ \rho_x = \{0, 0.3\} \quad \text{(SM)} \\ & \quad \textit{or} \\ (I - L)^d v_t &= \eta_t\ ,\ d = \{-0.25,\ 0.25,\ 0.35,\ 0.45\} \quad \text{(LM \& AP)} \\ \begin{bmatrix} u_t \\ \eta_t \end{bmatrix} &\sim iid\ N\left(0, \begin{bmatrix} 1 & R \\ R & 1 \end{bmatrix}\right),\ -1 < R < 1. \end{aligned} \tag{22}$$

[10] No simulation results are reported for the $F_{\max}$ test. Our findings indicate that the $F_{\max}$ test generally has more conservative size and power than the F_{sum} test. Preliminary simulation results show that the $F_{\max}$ test is more powerful than the F_{sum} only against integrable alternatives. In all the other cases, F_{sum} has superior power.

[11] Numerical computation of the JM test involves two dimensional quadrature and simulations were conducted using a modified version of the original Matlab program kindly supplied by Michael Jansson. Only 500 replications were used for this procedure because of the time involved in achieving convergence of the numerical procedure. The modified code allows for: (i) more general DGPs i.e. nonlinear models and fractional processes (ii) HAC estimation, (iii) parallelized execution of the computation and (iv) includes a Graphical User Interface front-end for the determination of the simulation parameters and the tabulation/visualization of the results. The computation was executed on the Milliped Cluster of the University of Groningen, the use of which is gratefully acknowledged. The Matlab installation on that cluster allows the use of a maximum of 8 cores per submitted job. By submitting a number of jobs at the same time we were able to utilize in the order of 50 cores in parallel for our computation. It should be noted that the time required for the computation of the double integral is heavily dependent on the value of the correlation parameter R (see (22) below) with absolute values of R close to 0 (i.e. $|R| \leq 0.2$) requiring excessively long computation time. We indicatively note that the results for the F_{sum}, t_{FM} tests presented in Figure 2(a) required a total CPU (core) time of approximately 4 minutes. On the other hand, the results for JM presented in Figure 2(a) required a total CPU time of approximately 353 hours which (given the 8-core parallelization) corresponds to actual computation time (wall time) of approximately 53 hours. Of the total CPU time (353 hours), the $R = 0$ job consumed 240 hours, the $R = \pm 0.2$ jobs consumed 76 hours and the $|R| > 0.2$ jobs consumed a total of 37 hours. It should also be noted that these computation times are strongly dependent on the initialisation seed of the random number generator, with different realisation requiring significantly varying computation times of the same order of magnitude.

The following regression functions are considered:

$f_0(x) = 0$ (null hypothesis)
$f_1(x) = 0.015x$, (linear)
$f_2(x) = \frac{1}{4} sign(x) |x|^{1/4}$ (polynomial)
$f_3(x) = \frac{1}{5} \ln(|x| + 0.1)$ (logarithmic)
$f_4(x) = (1 + e^{-x})^{-1}$ (logistic)
$f_5(x) = (1 + |x|^{0.9})^{-1}$ (reciprocal)
$f_6(x) = e^{-5x^2}$ (integrable)

The nonparametric test statistics $\hat{F}_{\text{sum}}$, $\hat{F}_{\max}$ employ the normal kernel and bandwidth is chosen as $h_n = n^{-b}$ with settings $b = 0.1, 0.2$. We also investigate the slowly varying bandwidth $h_n = 1/\ln(\ln(n))$ and data determined bandwidth $\hat{h}_n = n^{-\frac{1}{5}\left(\hat{m}-\frac{1}{2}\right)}$ where $\hat{m}$ is the local Whittle estimator (e.g. Robinson 1995) for the parameter m.[12] The last two bandwidths are optimal in the Wang and Phillips (2011) sense. For a twice continuously differentiable regression function f, using arguments similar to those in Wang and Phillips (2011) we get

$$\hat{f}(x) - f(x) = O_p\left(\sqrt{\frac{n^{\frac{3}{2}-m}}{nh_n}} + h_n^2 \frac{f''(x)}{2!} \int_{\mathbb{R}} s^2 K(s) ds\right),$$

where f'' is the second derivative of f. A bandwidth of the form $h_n = cons.n^{-\frac{1}{5}(m-\frac{1}{2})}$ is optimal in the sense that it minimises the order of magnitude of the approximation error shown above. Under the null hypothesis of the paper, the second derivative of the regression function, satisfies $f''(x) = 0$. In this case a slowly vanishing bandwidth (e.g. $h_n = 1/\ln(\ln(n))$) yields an approximation error of smaller order of magnitude than that of any $h_n = n^{-b}$, $b > 0$.

A wide range of values are considered for the correlation parameter: $R = \{0, \pm 0.2, \pm 0.4, \pm 0.6, \pm 0.8, \pm 0.99\}$. The grid X_s is chosen so that it comprises 25 equidistant points between the top and bottom observed percentiles of $\{x_t\}$. HAC estimators of the submatrices Ω_{uv} and Ω_{vv} of (20) were used in the $\hat{t}_{FM}$ and $\hat{R}_\beta$ test statistics employing a Bartlett kernel and lag truncation $n^{1/3}$. The variance Ω_{uu} was estimated parametrically and no HAC estimators were used in the JM statistic when $\rho_x = 0$. Nominal size was set to 5%.

The findings are summarized as follows:

1. Test size is in general stable and close to the nominal size for the nonparametric tests across all experiments, including both local to unity and long memory predictors. The bandwidth choice seems to have only a small effect on size (Figs 1(a) - (j) and Figs. 2(a) - (d)). The F_{sum} test exhibits some undersizing in the nearly explosive case whereas the $F_{\max}$ test has size very close to nominal.[13]

[12] The estimator $\hat{m}$ is obtained by applying local Whittle estimation to the differenced process Δx_t.

[13] The JM algorithm does not converge for $c > 1$, so no simulation results are reported for JM in this case.

2. Size distortions are considerable for the FM-OLS tests when $c \neq 0$ and when $d \neq 0$ (Figs 1(a) - (j) and Figs 2(a) - (d)).

3. The JM test shows size distortion when the endogeneity parameter $|R| \leq 0.2$. The distortion appears to be considerable when $R \approx 0$. No size computations have yet been done for the JM test when $|R| \leq 0.2$ and there is serial dependence because of the length of time (greater than 10 days) required.[14] When $|R| > 0.2$ we were able to complete 500 simulation runs and findings indicate that the JM statistic exhibits size distortions in the weakly dependent (Figs 1(g) - (j)) case when $|R| = \pm 0.99$ and in the fractional case (Figs 2(a) - (d)). The size distortion is particularly serious in the **LM** case with $d = 0.25$ (Figs 2(b) and (d)).

4. The nonparametric tests show higher power for the larger bandwidth which gives greater discriminatory capability in the test. (Figs 3(a) - (d) and Figs 4(a) - (f)).

5. Against linear alternatives, the nonparametric tests seem to perform reasonably well in comparison with the JM test (Figs 3(a) - (b)). Notably, the JM test has lower power than all the other tests when $c = -50$ (Fig 3(d)).

6. The nonparametric tests have good performance against the nonlinear alternatives (Figs. 4(a) - (f)).

7. The JM statistic has lower power than all the other tests in the case of reciprocal and integrable alternatives f (Figs 4(d) - (f)).

8. Figures 5(a) - (d) and 6(a) - (d) show size and power performance of the F_{sum} test for the following bandwidth choices: $h_n = n^{-0.1}$, $h_n = n^{-0.2}$, $h_n = 1/\ln(\ln(n))$ and $\hat{h}_n = n^{-\frac{1}{5}\left(\hat{m}-\frac{1}{2}\right)}$. Size and power are both reasonable. Better power performance is attained when $m \approx 3/2$. Power deteriorates when m approaches $1/2$. Both findings corroborate the asymptotic theory. The relative performance of $\hat{h}_n = n^{-\frac{1}{5}(\hat{m}-\frac{1}{2})}$ improves when m approaches $1/2$. Interestingly, in finite samples, $\hat{h}_n = n^{-\frac{1}{5}(\hat{m}-\frac{1}{2})}$ results in better power than $h_n = 1/\ln(\ln(n))$, despite the fact that the latter bandwidth yields faster divergence rates in the asymptotics. However, for sample size $n \approx 1000$, and $m \in (1/2, 1)$ we have $n^{\frac{1}{5}(m-\frac{1}{2})} < n^{0.1} \approx \ln(\ln(n))$, which partly explains the greater power in this case under the bandwidth setting $\hat{h}_n = n^{-\frac{1}{5}(\hat{m}-\frac{1}{2})}$.

[14] Simulations were attempted for this case without success. The job ran for 10 days in the MATLAB cluster (described in the earlier footnote) and had to be aborted because of administrative restrictions on the time permitted for each job. In consequence, we report simulation findings for cases where $|R| > 0.2$.

6 Stock Market Return Predictability

There is a large and continually developing literature on predictive regressions for equity returns. In spite of extensive research, the findings are still rather mixed (for a discussion and recent overview see, for instance, Welch and Goyal, 2008). The methods in this literature are almost completely dominated by linear or log-linear regression models in conjunction with assumptions that confine the predictors to stationary or near unit root processes.

The objective of this section is to briefly illustrate the use of nonparametric tests in the context of equity return predictive regressions. This application provides an opportunity to re-assess some earlier findings using our methods that do not require specific functional form, stationarity or memory properties for the predictor. Methodological extensions to a nonlinear framework are important in this application because the linear models in current use in predictive regressions for equity returns are typically developed or motivated in terms of linearized versions of underlying non-linear models of asset price determination.

In the context of stock return predictability, nonlinear specifications are also important from an econometric point of view. It is well known that for many data sets stock returns are weakly correlated whereas various popular predictors (e.g., the Dividend Price and Earnings Price ratios) are highly persistent. Therefore, a linear regression of stock returns on some persistent predictor can be potentially misbalanced (see for example Granger, 1995). Nonlinear regression functions, on the other hand, can provide balancing mechanisms between persistent financial predictors and less persistent stock return data. Certain nonlinear transformations (e.g. concave, bounded, integrable transformations) reduce the signal of the data (see Park and Phillips, 1998; Marmer, 2007; Kasparis, 2010; Berenguer-Rico and Gonzalo, 2012). The effects of such nonlinear transformations on the signal of a persistent process is demonstrated in Figures 7 and 8 which show the trajectories of nonlinear transforms of simulated random walks. The signal attenuation is evident in both cases. Equation misbalancing can be also addressed by introducing marginal departures from the null that shrink with the sample size at a particular rate and therefore allow for (near) balancing under the alternative. See the recent paper of Phillips and Lee (2013) where it is shown that marginal departures allow for (near) balancing under the alternative but also enable consistent tests of marginal predictability under some rate conditions.

We examine two predictors – the Dividend Price ratio and the Earnings Price ratio. These two valuation ratios are among the most frequently used predictors in the financial economics literature and serve as a good illustration of our methods. We leave to subsequent work an extensive analysis with a comprehensive set of predictors comparable to those in Welch and Goyal (2008). In addition, these two series are considered as highly persistent predictors in the empirical literature on stock return predictability (e.g. Campbell and Yogo (2006), Lewellen (2004), Torous et al. (2004)) and have been considered in a non-linear model in recent work (e.g. Gonzalo and

Pitarakis (2012)).

The dependent variable is the US monthly equity premium or excess return, i.e. the total rate of return on the stock market minus the short-term interest rate. We use S&P 500 index returns from 1926 to 2010 month-end values from CRSP. Stock returns are the continuously compounded returns on the S&P 500 index, including dividends. The short-term interest rate refers to the one month Treasury bill rate. The monthly dividend price ratio and the earnings price ratio obtained as follows:

(i) Dividend Price ratio, log(D/P), is the difference between the log of moving one-year average dividends and the log of S&P 500 prices found in Robert Shiller's webpage.

(ii) Earnings Price ratio, log(E/P), or smoothed Earnings Price ratio, is the difference between the log of moving ten-year average earnings and the log of S&P 500 prices. Data sources are CRSP, FRED and Welch and Goyal (2008) and Shiller's webpages.

The non-parametric tests are applied to monthly frequency data over the period 1926:M12-2010:M12 ($n = 1009$). Various subsamples are also considered following other studies in the literature such as: (i) the period 1929:M12-2002:M12 ($n = 913$) for which Campbell and Yogo (2006) find significant predictive ability of the monthly Earnings Price ratio but not the Dividend Price ratio, and (ii) the (relatively) tranquil period since 1952:M12 and ending either in 2005:M12 ($n = 606$) or before the recent financial crisis in 2007:M7 ($n = 625$), for which there is mixed evidence on the predictability of the Dividend Price ratio using alternative methods (e.g. Gonzalo and Pitarakis (2012), Campbell and Yogo (2006), Lewellen (2004) and Torous et al. (2004)).

Table 1 reports the significant predictability results (at the 0.05 level) from the Sum and Max nonparametric tests which evaluate the relationship between the S&P 500 stock market returns over the sample period 1926:M12-2010:M12 ($n = 1009$) and the two predictors at various lags (1 to 4 months) taken one at a time. Evidence of significant short-run predictability is reported for the alternative exponents b of the bandwidth, $h_n = \widehat{\sigma}_v n^{-b}$, of the nonparametric tests, where n denotes the sample size and $\widehat{\sigma}_v$ is an estimator of σ_v. In addition, we examine the null of predictability for the data driven bandwidth, $\hat{h}_n = \widehat{\sigma}_v n^{-\frac{1}{5}(\hat{m}-\frac{1}{2})}$. The reported results are evaluated for different equally spaced grid points (10, 25, 35, 50).

Summarizing the findings over the sample period 1926:M12-2010:M12, there is significant evidence of short-run S&P 500 returns predictability for the smoothed Earnings Price ratio using both the fixed and the stochastic bandwidths. Although there is partial evidence in favor of predictability for the Dividend Price ratio with the fixed bandwidth, those results are sensitive to the choice of the bandwidth exponent and the number of grid points. Moreover, the stochastic bandwidth yields no evidence of predictability for the Dividend Price ratio. Hence, our tests show the Earnings Price ratio to be a stronger and more robust predictor than the Dividend Price ratio. In particular, the findings confirm that predictability from the Earnings Price ratio is robust under: (i) alternative bandwidth exponents $b \in \{0.1, 0.2, 0.3, 0.4\}$ in $h_n = \widehat{\sigma}_v n^{-b}$; (ii) most lags using stochastic bandwidth, (iii) different equi-spaced grid point numbers (10, 25, 35, 50); and (iv) the sub-period 1929:M12-2002:M12 ($n = 913$). However, during the 'tranquil' sample period from 1952:M12 to 2005:M12 or to 2007:M12, evidence of predictability is weaker across the different lag lengths and bandwidths relative to the other two subsamples. Possible explanations may include the declining predictability for the D/P ratio reported in the literature for this period and/or the smaller sample size in the analysis.[15]

In evaluating these findings relative to those in the literature, the study by Campbell and Yogo (2006) is particularly relevant given that our methods are more comparable in terms of the allowance made for nonstationary predictors, than other studies. Our findings agree with those of Campbell and Yogo for the smoothed log E/P ratio for the monthly period 1929-2002, which we also extend in our updated sample to 2010. This empirical finding is consistent not only with Campbell and Yogo's tests for highly persistent regressors, but also with Bollerslev, Tauchen and Zhou (2008) who consider the more recent sample of 1990M1-2007:M12 but use Newey-West robust t-tests.

7 Conclusion

The use of nonparametric regression in prediction has some appealing properties in view of the robustness of this approach to the memory characteristics of the predictor and its endogeneity. As this paper shows, the asymptotic distributions of simple nonparametric F tests hold for a wide range of predictors that include stationary as well as non-stationary fractional and near unit root processes. This framework therefore helps to unify predictive inference in situations where both the model and the properties of the predictor are not known, allowing for nonlinearities and offering robustness to integration order. The finite sample performance of the procedure is promising in terms of both size and power. But, like many of the procedures in current use – particularly those that are based on local to unity limit theory – nonparametric regression is most likely to be useful in cases where the predictor is a scalar variable.

8 Appendix A: proofs of main results

In the following proofs, we use A as a generic constant whose value may change in each location. Further, let $0 < q_o, q_1 < 1$ and $\lfloor q_o n \rfloor \le l \le n$, $1 \le t \le \lfloor q_1 l \rfloor$. In the

[15] The results for the sub-periods 1929:M12-2002:M12 and 1952:M12-2007M7 can be found in the Working Paper version of this paper in Table 2.

subsequent proofs, we handle terms of the form

$$\frac{\sqrt{l}}{d_l}\sum_{j=0}^{l-t}\phi_j\rho_n^{l-j-t},\ m\in(1/2,1)$$

and

$$\frac{\sqrt{l}}{nd_l}\sum_{j=0}^{l-t}\rho_n^{-j}\sum_{k=j}^{\infty}\phi_k,\ m\in(1,3/2)$$

when $n\to\infty$. Set $\varepsilon>0$. In view of the assumption $\phi_j\sim j^{-m}$, for some $N_\varepsilon\in\mathbb{N}$ and all $j>N_\varepsilon$ we have $\left|\frac{\phi_j}{j^{-m}}-1\right|<\varepsilon$. Hence, as $n\to\infty$ first and then as $\varepsilon\to 0$ we get

$$\left|\frac{\sqrt{l}}{d_l}\sum_{j=0}^{l-t}\rho_n^{l-j-t}\left(\phi_j-j^{-m}\right)\right|\le\frac{\sqrt{l}}{d_l}\sum_{j=1}^{l-t}\rho_n^{l-j-t}\left|\phi_j-j^{-m}\right|+o(1)$$

$$=\frac{\sqrt{l}}{d_l}\sum_{j=1}^{l-t}\rho_n^{l-j-t}\left|\frac{\phi_j}{j^{-m}}-1\right|j^{-m}\le A\frac{\sqrt{l}}{d_l}\sum_{j=1}^{l-t}\left|\frac{\phi_j}{j^{-m}}-1\right|j^{-m}$$

$$=A\frac{\sqrt{l}}{d_l}\sum_{j=1}^{N_\varepsilon}\left|\frac{\phi_j}{j^{-m}}-1\right|j^{-m}+A\frac{\sqrt{l}}{d_l}\sum_{j=N_\varepsilon+1}^{l-t}\left|\frac{\phi_j}{j^{-m}}-1\right|j^{-m}$$

$$=o(1)+A\frac{\sqrt{l}}{d_l}\sum_{j=N_\varepsilon+1}^{l-t}\left|\frac{\phi_j}{j^{-m}}-1\right|j^{-m}\le\varepsilon A\int_0^1 s^{-m}ds+o(1)\overset{\varepsilon\to 0}{\to}0.$$

Therefore, as $n\to\infty$, $\frac{\sqrt{l}}{d_l}\sum_{j=0}^{l-t}\phi_j\rho_n^{l-j-t}=\frac{\sqrt{l}}{d_l}\sum_{j=1}^{l-t}\rho_n^{l-j-t}j^{-m}+o(1)$. Similarly, $\frac{\sqrt{l}}{nd_l}\sum_{j=0}^{l-t}\rho_n^{-j}\sum_{k=j}^{\infty}\phi_k=\frac{\sqrt{l}}{nd_l}\sum_{j=1}^{l-t}\rho_n^{-j}\sum_{k=j}^{\infty}k^{-m}+o(1)$. Approximations of this kind are used in the subsequent proofs, without further explanation.

The Propositions A1-A4 below, provide auxiliary results for the proof of Lemma 1. Propositions A1 and A2 provide upper bounds for the modulus of $\psi(\lambda)$. Recall that $\psi(\lambda)$ is the characteristic function of the innovation process ξ_t (see Assumption 2.3).

Proposition A1: *For some $\delta>0$ we have $\left|\psi\left(\frac{\lambda}{\sqrt{n}}\right)\right|\le e^{-\frac{\lambda^2}{4n}}$ when $\frac{|\lambda|}{\sqrt{n}}\le\delta$. Further, for all $\eta>0$ there is $0<\rho<1$ such that $\left|\psi\left(\frac{\lambda}{\sqrt{n}}\right)\right|\le\rho$, for $\frac{|\lambda|}{\sqrt{n}}\ge\eta$.*

Proof Proposition A1: See Feller (1971), Lemma 4 of p. 501 and eq. (5.6) of p. 516. ■

Proposition A2. *Let $\zeta_n\in\mathbb{N}$ such that for some $C_o>0$ and $n_o\in\mathbb{N}$*

$$\zeta_n\ge C_o n,\text{ for }n\ge n_o.$$

Then

(i) *for some* δ *there is* $A > 0$ *such that*

$$\left|\psi\left(\frac{\lambda}{\sqrt{n}}\right)\right|^{\zeta_n} \leq e^{-A\lambda^2},\ \frac{|\lambda|}{\sqrt{n}} \leq \delta,\ \textit{for all } n \in \mathbb{N}.$$

(ii) *for all* $\eta > 0$, there are $0 < \rho < 1$ and $B, C > 0$

$$\sup_{|\lambda|\geq\eta} |\psi(\lambda)|^{\zeta_n} \leq B\rho^{Cn},\ \textit{for all } n \in \mathbb{N}.$$

Proof Proposition A2: In view of of Proposition A1 the result can be proved using similar arguments to those used for the proof of Lemma 6 in Jeganathan (2008). ■

Proposition A3. *Define*

$$\mathcal{A}_{n,l,t} := \sum_{j=0}^{l-t} \rho_n^{l-t-j}\phi_j \text{ and } \Lambda_{l,n}^2 := \Lambda_l^2 := \sigma_\xi^2 \sum_{t=1}^{l}\left(\sum_{j=0}^{l-t}\rho_n^{l-t-j}\phi_j\right)^2.$$

Then, *for all* $0 < q_o < 1$, *some* $0 < q_1 < 1$, n *large enough* $\lfloor q_o n \rfloor \leq l \leq n$ *and* $1 \leq t \leq \lfloor q_1 l \rfloor$ *there are constants* D_1, D_2 *with* $0 < D_1 \leq D_2 < \infty$ *such that*

$$D_1 \leq \frac{\sqrt{l}\,|\mathcal{A}_{n,l,t}|}{\Lambda_l} \leq D_2. \tag{23}$$

Proof Proposition A3: Write

$$\frac{\sqrt{l}\,|\mathcal{A}_{n,l,t}|}{\Lambda_l} = \frac{\sqrt{l}\,|\mathcal{A}_{n,l,t}|}{d_l}\frac{d_l}{\Lambda_l}.$$

It can be shown that for all $0 < q_o < 1$, some $0 < q_1 < 1$, n large enough, $\lfloor q_o n \rfloor \leq l \leq n$ and $1 \leq t \leq \lfloor q_1 l \rfloor$ there are $0 < \alpha_1 \leq \alpha_2 < \infty$ and $0 < \beta_1 \leq \beta_2 < \infty$ such that

$$\alpha_1 \leq \frac{\sqrt{l}\,|\mathcal{A}_{n,l,t}|}{d_l} \leq \alpha_2 \tag{24}$$

and

$$\frac{1}{\beta_1} \geq \frac{\Lambda_l}{d_l} \geq \frac{1}{\beta_2}. \tag{25}$$

Then (23) follows from (24) and (25) with $D_1 = \alpha_1\beta_1$ and $D_2 = \alpha_2\beta_2$.

We first prove (24). Note that, for $1 \leq t \leq n$ and n large enough we have ($\rho_n \neq 0$, for n large)

$$\begin{aligned} 0 &< \rho_n^{-1} \leq \rho_n^{-t} \leq \rho_n^{-n} < \infty,\ \text{if } c < 0 \\ 0 &< \rho_n^{-n} \leq \rho_n^{-t} \leq \rho_n^{-1} < \infty,\ \text{if } c > 0 \end{aligned} \tag{26}$$

LM case: Under **LM** Euler summation gives

$$\sup_{1\le t\le n}\left|\frac{\sqrt{n}}{d_n}\sum_{j=1}^{n-t}\rho_n^{n-j}\phi_j-\int_0^{1-\frac{t}{n}}r^{-m}e^{c(1-r)}dr\right|=o(1). \tag{27}$$

Next, consider the term

$$\frac{\sqrt{l}}{d_l}\sum_{j=0}^{l-t}\phi_j\rho_n^{l-j-t}=\rho_n^{-t}\frac{1}{l}\sum_{j=1}^{l-t}\left(\frac{j}{l}\right)^{-m}\rho_n^{l-j}+o(1).$$

Then for $\lfloor q_o n\rfloor\le l\le n$ as $n\to\infty$

$$\frac{1}{l}\sum_{j=1}^{l-t}\left(\frac{j}{l}\right)^{-m}\rho_n^{l-j}=\frac{1}{l}\sum_{j=1}^{l-t}\left(\frac{j}{l}\right)^{-m}\exp\left\{(l-j)\ln\left(1+\frac{c}{n}\right)\right\}$$

$$\begin{aligned}&= \frac{1}{l}\sum_{j=1}^{l-t}\left(\frac{j}{l}\right)^{-m}\exp\left\{\frac{l}{l}(l-j)\left[\frac{c}{n}+O\left(\frac{1}{n^2}\right)\right]\right\}=\frac{1}{l}\sum_{j=1}^{l-t}\left(\frac{j}{l}\right)^{-m}\exp\left\{\frac{l}{n}c\left(1-\frac{j}{l}\right)\right\}+o(1)\\ &= :T_{l,n}\end{aligned}$$

Next,

$$\left.\begin{array}{c}\frac{1}{l}\sum_{j=1}^{l-t}\left(\frac{j}{l}\right)^{-m}\exp\left\{q_o c\left(1-\frac{j}{l}\right)\right\},\ c>0\\ \frac{1}{l}\sum_{j=1}^{l-t}\left(\frac{j}{l}\right)^{-m}\exp\left\{c\left(1-\frac{j}{l}\right)\right\},\ c<0\end{array}\right\}\le T_{l,n} \tag{28}$$

and

$$T_{l,n}\le\left\{\begin{array}{c}\frac{1}{l}\sum_{j=1}^{l-t}\left(\frac{j}{l}\right)^{-m}\exp\left\{c\left(1-\frac{j}{l}\right)\right\},\ c>0\\ \frac{1}{l}\sum_{j=1}^{l-t}\left(\frac{j}{l}\right)^{-m}\exp\left\{q_o c\left(1-\frac{j}{l}\right)\right\},\ c<0\end{array}\right. \tag{29}$$

Hence, in view of (28), (29) and the uniform convergence in (27) we have

$$\begin{aligned}&\left.\begin{array}{c}\inf_{1\le t\le\lfloor q_1 l\rfloor}\int_0^{1-\frac{t}{l}}r^{-m}e^{\{q_o c(1-r)\}}dr,\ c>0\\ \inf_{1\le t\le\lfloor q_1 l\rfloor}\int_0^{1-\frac{t}{l}}r^{-m}e^{\{c(1-r)\}}dr,\ c<0\end{array}\right\}\le T_{l,n}+o(1)\\ &\le\left\{\begin{array}{c}\sup_{1\le t\le\lfloor q_1 l\rfloor}\int_0^{1-\frac{t}{l}}r^{-m}e^{\{c(1-r)\}}dr,\ c>0\\ \sup_{1\le t\le\lfloor q_1 l\rfloor}\int_0^{1-\frac{t}{l}}r^{-m}e^{\{q_o c(1-r)\}}dr,\ c<0\end{array}\right.\end{aligned}$$

Therefore, in view of the above and (26), for n large enough, all $0<q_o<1$, some $0<q_1<1$ and $\lfloor q_o n\rfloor\le l\le n$, $1\le t\le\lfloor q_1 l\ \rfloor$ there are $0<\alpha_1\le\alpha_2<\infty$ such that

$$\alpha_1\le\rho_n^{-t}\frac{\sqrt{l}}{d_l}\sum_{j=0}^{l-t}\rho_n^{l-j}\phi_j\le\alpha_2.$$

SM case: Suppose that $\sum_{j=0}^{\infty} \phi_j = \phi \neq 0$. Then, for l large enough and $1 \leq t \leq \lfloor q_1 l \rfloor$.

$$0 < |\phi|/2 \leq \left| \sum_{j=0}^{l-t} \phi_j \right| \leq A < \infty. \tag{30}$$

To see this, fix $\varepsilon = |\phi|/2$. Then, there is $N_\varepsilon \in \mathbb{N}$ such that for $l - t > N_\varepsilon$, $\left| \sum_{j=0}^{l-t} \phi_j - \phi \right| < \varepsilon$. Hence, for $l - \lfloor q_1 l \rfloor > N_\varepsilon$ we have

$$\sup_{1 \leq t \leq \lfloor q_1 l \rfloor} \left| \sum_{j=0}^{l-t} \phi_j - \phi \right| < \varepsilon.$$

The above postulates that, for l large enough and $1 \leq t \leq \lfloor q_1 l \rfloor$, the term $\sum_{j=0}^{l-t} \phi_j$ is bounded and bounded away from zero. Next, let

$$\tilde{\phi}_{j,s} := \begin{cases} \sum_{k=j}^{s} \phi_k, \text{ for } 0 \leq j \leq s \\ 0, \text{ otherwise} \end{cases}$$

Then, in view of the fact that $\tilde{\phi}_{l-t+1,l-t} = 0$, summation by parts gives

$$\sum_{j=0}^{l-t} \rho_n^{l-t-j} \phi_j = \sum_{j=0}^{l-t} \rho_n^{l-t-j} \left(\tilde{\phi}_{j,l-t} - \tilde{\phi}_{j+1,l-t} \right) =$$

$$= \left[\rho_n^{l-t} \tilde{\phi}_{0,l-t} - \rho_n^0 \tilde{\phi}_{l-t+1,l-t} - \tilde{\phi}_{1,l-t} \left(\rho_n^{n-t} - \rho_n^{n-t-1} \right) - \tilde{\phi}_{2,l-t} \left(\rho_n^{n-t-1} - \rho_n^{n-t-2} \right) - \ldots - \tilde{\phi}_{l-t,l-t} \left(\rho_n^1 - \rho_n^0 \right) \right]$$

$$= \rho_n^{l-t} \tilde{\phi}_{0,l-t} - \rho_n^0 \tilde{\phi}_{n-t+1,l-t} - \sum_{j=0}^{l-t} \tilde{\phi}_{j,l-t} \left(\rho_n^{l-t-j+1} - \rho_n^{l-t-j} \right) = \rho_n^{l-t} \tilde{\phi}_{0,l-t} - \sum_{j=0}^{l-t} \tilde{\phi}_{j,l-t} \left(\rho_n^{l-t-j+1} - \rho_n^{l-t-j} \right)$$

$$= \rho_n^{l-t} \left[\tilde{\phi}_{0,l-t} - (\rho_n - 1) \sum_{j=0}^{l-t} \tilde{\phi}_{j,l-t} \rho_n^{-j} \right] = \rho_n^{l-t} \left[\tilde{\phi}_{0,l-t} - \frac{c}{n} \sum_{j=0}^{l-t} \tilde{\phi}_{j,l-t} \rho_n^{-j} \right]$$

Hence,

$$\sum_{j=0}^{l-t} \rho_n^{l-t-j} \phi_{j,l-t} = \rho_n^{l-t} \left[\tilde{\phi}_{0,l-t} - \frac{c}{n} \sum_{j=0}^{l-t} \tilde{\phi}_{j,l-t} \rho_n^{-j} \right]. \tag{31}$$

Further, for $\lfloor q_o n \rfloor \leq l \leq n$ as $n \to \infty$ we have

$$\sup_{1 \leq t \leq l} \left| \rho_n^{l-t} \frac{c}{n} \sum_{j=0}^{l-t} \tilde{\phi}_{j,l-t} \rho_n^{-j} \right| = o(1). \tag{32}$$

The asymptotic negligibility of the term shown above is justified by the following. First, for n large, $\sup_{1 \leq j \leq n} |\rho_n^{-j}| \leq A < \infty$. Next, the term

$$\left| \frac{c}{n} \sum_{j=0}^{l-t} \tilde{\phi}_{j,l-t} \rho_n^{-j} \right| \leq \frac{1}{n} A \sum_{j=0}^{l} \sum_{k=j}^{l} |\phi_k| \leq \frac{1}{l} A \sum_{j=0}^{l} \sum_{k=j}^{\infty} |\phi_k| = o(1),$$

where the last approximation is due to Césaro's Lemma. Finally, note that $\tilde{\phi}_{0,l-t} \to \phi$ as $l-t \to \infty$. In view of this, (26) and (30) $\rho_n^{l-t}\tilde{\phi}_{0,l-t}$ is bounded and, bounded away from zero for n large enough, $\lfloor q_o n \rfloor \le l \le n$ and $1 \le t \le \lfloor q_1 l \rfloor$.

AP case: By (31)

$$\Phi_{n,l,t} := \frac{\sqrt{l}}{d_l}\sum_{j=0}^{l-t}\rho_n^{l-t-j}\phi_j = \rho_n^{l-t}\left[\frac{\sqrt{l}}{d_l}\tilde{\phi}_{0,l-t} - \frac{c}{n}\frac{\sqrt{l}}{d_l}\sum_{j=0}^{l-t}\tilde{\phi}_{j,l-t}\rho_n^{-j}\right]$$

$$\rho_n^{l-t}\left[\frac{\sqrt{l}}{d_l}\left(1-\frac{c}{n}\right)\tilde{\phi}_{0,l-t} - \frac{c}{n}\frac{\sqrt{l}}{d_l}\sum_{j=1}^{l-t}\tilde{\phi}_{j,l-t}\rho_n^{-j}\right] =: \rho_n^{l-t}\left[\mathcal{B}_{n,l,t} - \mathcal{C}_{n,l,t}\right].$$

Now for n large enough $\mathcal{B}_{n,l,t}$

$$\mathcal{B}_{n,l,t} = \frac{\sqrt{l}}{d_l}\tilde{\phi}_{0,l-t} + o(1) = \frac{\sqrt{l}}{d_l}\sum_{k=0}^{l-t}\phi_k = \frac{\sqrt{l}}{d_l}\sum_{k=0}^{\infty}\phi_k - \frac{\sqrt{l}}{d_l}\sum_{k=l-t+1}^{\infty}\phi_k = -\frac{\sqrt{l}}{d_l}\sum_{k=l-t+1}^{\infty}\phi_k$$

$$= -\int_{1-\frac{t}{l}}^{\infty} s^{-m}ds + o(1) = -\frac{1}{1-m}\left[s^{1-m}\right]_{1-\frac{t+1}{l}}^{\infty} = \frac{1}{1-m}\left(1-\frac{t}{l}\right)^{1-m}. \tag{33}$$

Next, for n large enough the term $\mathcal{C}_{n,l,t}$ is

$$\mathcal{C}_{n,l,t} \le \frac{|c|}{n}\frac{\sqrt{l}}{d_l}\sum_{j=1}^{l-t}\rho_n^{-j}\sum_{k=j}^{\infty}|\phi_k| = \frac{|c|}{n}\frac{\sqrt{l}}{d_l}\sum_{j=1}^{l-t}\rho_n^{-j}\sum_{k=j}^{\infty}k^{-m} + o(1)$$

$$\le \frac{|c|}{n}\frac{\sqrt{l}}{d_l}\sum_{j=1}^{l-t}\rho_n^{-j}\left(j^{-m} + \int_{j+1}^{\infty}(x-1)^{-m}dx\right)$$

$$= \frac{|c|}{(1-m)n}\frac{\sqrt{l}}{d_l}\sum_{j=1}^{l-t}\rho_n^{-j}\left[(x-1)^{1-m}\right]_{j+1}^{\infty} + o(1) = -\frac{l\,|c|}{(1-m)n}\frac{1}{l}\sum_{j=1}^{l-t}\rho_n^{-j}\left(\frac{j}{l}\right)^{1-m}$$

$$= -\frac{l\,|c|}{(1-m)n}\frac{1}{l}\sum_{j=1}^{l-t}\exp\left\{-j\left[\frac{c}{n}+O\left(\frac{1}{n^2}\right)\right]\right\}\left(\frac{j}{l}\right)^{1-m} = \frac{l\,|c|}{(m-1)n}\frac{1}{l}\sum_{j=1}^{l-t}\exp\left\{-j\frac{c}{n}\right\}\left(\frac{j}{l}\right)^{1-m}.$$

In view of this, for n large enough $\lfloor q_o n \rfloor \le l \le n$ and $1 \le t \le \lfloor q_1 l \rfloor$

$$|\mathcal{C}_{n,l,t}| \le \begin{cases} \left|\frac{c}{(m-1)}\right|\frac{1}{l}\sum_{j=1}^{l-t}\exp\left\{-c\frac{j}{l}\right\}\left(\frac{j}{l}\right)^{1-m}, & c<0 \\ \frac{c}{(m-1)}\frac{1}{l}\sum_{j=1}^{l-t}\exp\left\{-q_o c\frac{j}{l}\right\}\left(\frac{j}{l}\right)^{1-m}, & c>0 \end{cases}$$

$$= \begin{cases} \left|\frac{c}{(m-1)}\right|\int_0^{1-\frac{t}{l}}\exp\{-cs\}s^{1-m}ds + o(1), & c<0 \\ \frac{c}{(m-1)}\int_0^{1-\frac{t}{l}}\exp\{-q_o cs\}s^{1-m}ds + o(1), & c>0 \end{cases} \tag{34}$$

Hence, in view of (33) and (34) for n large enough $\lfloor q_o n \rfloor \le l \le n$ and $1 \le t \le \lfloor q_1 l \rfloor$, there is some $0 \le \alpha_2 < \infty$ such that $|\Phi_{n,l,t}| \le \alpha_2$.

Next, we show that for n large $\Phi_{n,l,t}$ is bounded away from zero. We start with the case $c \ge 0$. Note that $\tilde{\phi}_{0,l-t}$ is negative and for l large enough $\sum_{j=1}^{l-t} \tilde{\phi}_{j,l-t} \rho_n^{-j}$ is positive.[16] Hence, in view of (33) for n large enough, $\lfloor q_o n \rfloor \le l \le n$ and $1 \le t \le l$ we have

$$\Phi_{n,l,t} = \rho_n^{l-t} \left[\mathcal{B}_{n,l,t} - \mathcal{C}_{n,l,t}\right] \le \rho_n^{l-t} \left(1 - \frac{c}{n}\right) \frac{\sqrt{l}}{d_l} \tilde{\phi}_{0,l-t} \le A \left(1 - \frac{c}{n}\right) \frac{\sqrt{l}}{d_l} \tilde{\phi}_{0,l-t}$$

$$= A \frac{1}{1-m} \left(1 - \frac{t}{l}\right)^{1-m} + o(1), \tag{35}$$

where $0 < A < \infty$ and $\frac{1}{1-m} \left(1 - \frac{t}{l}\right)^{1-m} \le \frac{1}{1-m} \left(1 - q_1\right)^{1-m} < 0$, when $1 \le t \le \lfloor q_1 l \rfloor$. This shows that $\Phi_{n,l,t}$ is bounded away from zero, for a suitable choice of l and t and n large.

Next, suppose that $c < 0$. We shall show that under (11) and n large enough,

$$\sup_{n \ge l \ge \lfloor q_o n \rfloor, 1 \le t \le \lfloor q_1 l \rfloor} \Phi_{n,l,t} < 0.$$

Using arguments similar to those used for the derivation of (34), for n large we have

$$\Phi_{n,l,t} / \rho_n^{l-t} = \left[\mathcal{B}_{n,l,t} - \mathcal{C}_{n,l,t}\right] = \left(1 - \frac{c}{n}\right) \frac{\sqrt{l}}{d_l} \tilde{\phi}_{0,l-t} - \frac{c}{n} \frac{\sqrt{l}}{d_l} \sum_{j=1}^{l-t} \rho_n^{-j} \sum_{k=j}^{l-t} k^{-m} + o(1)$$

$$= \frac{\sqrt{l}}{d_l} \tilde{\phi}_{0,l-t} - \frac{c}{n} \frac{\sqrt{l}}{d_l} \sum_{j=1}^{l-t} \rho_n^{-j} \left(j^{-m} + \sum_{k=j+1}^{l-t} k^{-m}\right) \le \frac{\sqrt{l}}{d_l} \tilde{\phi}_{0,l-t} - \frac{c}{n} \frac{\sqrt{l}}{d_l} \sum_{j=1}^{l-t} \rho_n^{-j} \left(j^{-m} + \int_{j+1}^{l-t} (x-1)^{-m} \, dx\right)$$

$$= \frac{\sqrt{l}}{d_l} \tilde{\phi}_{0,l-t} - \frac{c}{n(1-m)} \frac{\sqrt{l}}{d_l} \sum_{j=1}^{l-t} \exp\left\{-j \frac{c}{n}\right\} \left[(l-t)^{1-m} - j^{1-m}\right] + o(1)$$

$$= \frac{\sqrt{l}}{d_l} \tilde{\phi}_{0,l-t} - \frac{c}{n(1-m)} \frac{\sqrt{l}}{d_l} \sum_{j=1}^{l-t} \exp\left\{-\frac{j}{l} \frac{cl}{n}\right\} \left[(l-t)^{1-m} - j^{1-m}\right]$$

$$\le \frac{\sqrt{l}}{d_l} \tilde{\phi}_{0,l-t} - \frac{c}{l(1-m)} \frac{\sqrt{l}}{d_l} \sum_{j=1}^{l-t} \exp\left\{-c \frac{j}{l}\right\} \left[(l-t)^{1-m} - j^{1-m}\right]$$

$$= \frac{\sqrt{l}}{d_l} \tilde{\phi}_{0,l-t} - \frac{c}{1-m} \frac{1}{l} \sum_{j=1}^{l-t} \exp\left\{-c \frac{j}{l}\right\} \left[\left(1 - \frac{t}{l}\right)^{1-m} - \left(\frac{j}{l}\right)^{1-m}\right]$$

---

[16] Note that under **AP**, $\tilde{\phi}_{0,l-t} = \sum_{k=0}^{l-t} \phi_k = \phi_0 + \sum_{k=1}^{\infty} j^{-m} + o(1) = 0$. Hence, $\phi_0 = -\sum_{k=1}^{\infty} j^{-m} < 0$.

$$= \frac{1}{1-m}\left(1-\frac{t}{l}\right)^{1-m} - \frac{c}{1-m}\int_0^{1-\frac{t}{l}} \exp\left(-cs\right)\left[\left(1-\frac{t}{l}\right)^{1-m} - s^{1-m}\right] ds + o(1)$$

$$=: \overline{\Phi}_{t/l} + o(1) \tag{36}$$

Therefore, for n large enough $\sup_{n\geq l\geq \lfloor q_o n\rfloor, 1\leq t\leq \lfloor q_1 l\rfloor} \Phi_{n,l,t} < 0$ if for all l and some $0 < q_1 < l$

$$\sup_{1\leq t\leq \lfloor q_1 l\rfloor} \overline{\Phi}_{t/l} < 0. \tag{37}$$

Note that the requirement $\overline{\Phi}_r < 0$, $r \in [0,1)$ is sufficient for (37). Next, we shall obtain an upper bound for $\overline{\Phi}_{t/l}$ that justifies (12). We have

$$\overline{\Phi}_{t/l} \leq \frac{1}{1-m}\left(1-\frac{t}{l}\right)^{1-m} - \frac{c}{1-m}e^{-c}\int_0^{1-\frac{t}{l}} \left[\left(1-\frac{t}{l}\right)^{1-m} - s^{1-m}\right] ds$$

$$= \frac{1}{1-m}\left(1-\frac{t}{l}\right)^{1-m} - \frac{c}{1-m}e^{-c}\left[\left(1-\frac{t}{l}\right)^{2-m} - \frac{\left(1-\frac{t}{l}\right)^{2-m}}{2-m}\right]$$

$$= \frac{1}{1-m}\left(1-\frac{t}{l}\right)^{1-m} - \frac{c}{1-m}e^{-c}\left(1-\frac{t}{l}\right)^{2-m}\left[1-\frac{1}{2-m}\right]$$

$$= \frac{1}{1-m}\left(1-\frac{t}{l}\right)^{1-m} - \frac{c}{1-m}e^{-c}\left(1-\frac{t}{l}\right)^{2-m}\frac{1-m}{2-m}$$

$$= \frac{\left(1-\frac{t}{l}\right)^{1-m}}{1-m}\left\{1-ce^{-c}\left(1-\frac{t}{l}\right)\frac{1-m}{2-m}\right\} = \left(1-\frac{t}{l}\right)^{1-m}\left\{\frac{1}{1-m} - \frac{c}{1-m}e^{-c}\left(1-\frac{t}{l}\right)\frac{1-m}{2-m}\right\}$$

$$\leq \frac{1}{1-m}\left(1-\frac{t}{l}\right)^{1-m}\left\{1-ce^{-c}\frac{1-m}{2-m}\right\} \leq \frac{1}{1-m}\left(1-q_1\right)^{1-m}\left\{1-ce^{-c}\frac{1-m}{2-m}\right\}.$$

In view of the above, (11) and (37) are satisfied for $1 - ce^{-c}\frac{1-m}{2-m} > 0$.

Next, we show that (25) holds. Using similar arguments as those used above it can be easily be shown that $\Lambda_l/d_l \leq 1/\beta_1$, for n large enough and $l \geq \lfloor q_o n\rfloor$. We shall show that $1/\beta_2 \leq \Lambda_l/d_l$ holds.

LM case: By (26), (28) and Euler summation for $\lfloor q_o n\rfloor \leq l \leq n$, as $n\to\infty$ we get

$$\Lambda_l^2/d_l^2 \geq \begin{cases} \frac{1}{l}\sum_{t=1}^{l}\left(\rho_n^{-n}\frac{1}{l}\sum_{j=1}^{l-t}\left(\frac{j}{l}\right)^{-m}\exp\left\{q_o c\left(1-\frac{j}{l}\right)\right\}\right)^2 + o(1), & c>0 \\ \frac{1}{l}\sum_{t=1}^{l}\left(\rho_n^{-1}\frac{1}{l}\sum_{j=1}^{l-t}\left(\frac{j}{l}\right)^{-m}\exp\left\{c\left(1-\frac{j}{l}\right)\right\}\right)^2 + o(1), & c<0 \end{cases}$$

$$= \left\{ \begin{array}{ll} e^{-c} \int_0^1 \left(\int_0^{1-s} r^{-m} e^{q_o c(1-r)} dr \right)^2 ds + o(1), \ c > 0 \\ \int_0^1 \left(\int_0^{1-s} r^{-m} e^{c(1-r)} dr \right)^2 ds + o(1), \ c < 0 \end{array} \right. > 0$$

as required.

SM case: For n large, by (31), (32) and Césaro's Lemma we get

$$\Lambda_l^2/d_l^2 = \frac{1}{l} \sum_{t=1}^{l} \left(\sum_{j=0}^{l-t} \rho_n^{l-t-j} \phi_j \right)^2 = \frac{1}{l} \sum_{t=1}^{l} \left(\rho_n^{l-t} \tilde{\phi}_{0,l-t} \right)^2 + o(1)$$

$$\geq \left\{ \begin{array}{l} \frac{1}{l} \sum_{t=1}^{l} \left(\rho_n^{-n} \tilde{\phi}_{0,l-t} \right)^2, \ c > 0 \\ \frac{1}{l} \sum_{t=1}^{l} \left(\rho_n^{-1} \tilde{\phi}_{0,l-t} \right)^2, \ c < 0 \end{array} \right. = \left\{ \begin{array}{l} \frac{1}{l} \sum_{t=1}^{l} \left(\rho_n^{-n} \sum_{k=0}^{l-t} \phi_k \right)^2, \ c > 0 \\ \frac{1}{l} \sum_{t=1}^{l} \left(\rho_n^{-1} \sum_{k=0}^{l-t} \phi_k \right)^2, \ c < 0 \end{array} \right.$$

$$\to \left\{ \begin{array}{l} \left(e^{-c} \sum_{k=0}^{\infty} \phi_k \right)^2, \ c > 0 \\ \left(\sum_{k=0}^{\infty} \phi_k \right)^2, \ c < 0 \end{array} \right. > 0,$$

as required.

AP case: First, suppose that $c > 0$. Then by (35)

$$\Phi_{n,l,t}^2 \geq \left[A \frac{1}{1-m} \left(1 - \frac{t}{l} \right)^{1-m} \right]^2 + o(1),$$

uniformly in $1 \leq t \leq l$, where as before $0 < A < \infty$. In view of the above for $\lfloor q_1 n \rfloor \leq l \leq n$ and as $n \to \infty$, we have

$$\Lambda_l^2/d_l^2 = \frac{1}{l} \sum_{t=1}^{l} \Phi_{n,l,t}^2 \geq \frac{1}{l} \sum_{t=1}^{l} \left[\frac{A}{1-m} \left(1 - \frac{t}{l} \right)^{1-m} \right]^2 + o(1)$$

$$\to \int_0^1 \left[\frac{A}{1-m} (1-s)^{1-m} \right]^2 ds > 0.$$

Next, suppose that $c < 0$ and $\sup_{1 \leq t \leq \lfloor q_1 l \rfloor} \overline{\Phi}_{t/l} < 0$. Then recall that by (36) for $1 \leq t \leq \lfloor q_1 l \rfloor$ and n large enough we have

$$\Phi_{n,l,t} \leq \rho_n^{t-l} \overline{\Phi}_{t/l} < 0.$$

The above implies that

$$\Phi_{n,l,t}^2 \geq \left(\rho_n^{t-l} \overline{\Phi}_{t/l} \right)^2 > 0.$$

Hence, as $n \to \infty$ we have

$$\Lambda_l^2/d_l^2 = \frac{1}{l} \sum_{t=1}^{l} \Phi_{n,l,t}^2 \geq \frac{1}{l} \sum_{t=1}^{l} \left(\rho_n^{t-l} \overline{\Phi}_{t/l} \right)^2 + o(1)$$

$$\geq \frac{1}{l}\sum_{t=1}^{l}\left(\rho_n^0\overline{\Phi}_{t/l}\right)^2 \geq \frac{1}{l}\sum_{t=1}^{\lfloor q_1 l\rfloor}\overline{\Phi}_{t/l}^2 \to \int_0^{q_1}\overline{\Phi}_s^2 ds > 0,$$

as required. ■

Proposition A4. (CLT for a truncated Linear Process) *Consider the process*[17]

$$\tilde{x}_l := \sum_{t=1}^{l}\sum_{j=0}^{l-t}\rho_n^{l-t-j}\phi_j\xi_t.$$

For all $0 < q_0 < 1$*, as* $n \to \infty$ *we have*

$$\mathbf{E}\exp\left(i\lambda\frac{1}{\Lambda_l}\tilde{x}_l\right) \to e^{-\lambda^2/2}, \text{ uniformly in } \lfloor q_o n\rfloor \leq l \leq n. \tag{38}$$

Proof of Proposition A4. The uniform convergence result of (38) follows from a straightforward modification of a CLT for triangular arrays e.g. Hall and Heyde (1980), Corollary 3.1 (see also Hall and Heyde (1980), Theorem 3.1 and Lemma 3.1). In particular, a modification of Hall and Heyde (1980), Corollary 3.1 shows that the two following requirements are sufficient for (38)

$$\sup_{\lfloor q_o n\rfloor \leq l \leq n}\left|\sum_{t=1}^{l}\mathbf{E}\left\{\left[\frac{1}{\Lambda_l}\left(\sum_{j=0}^{l-t}\rho_n^{l-t-j}\phi_j\right)\xi_t\right]^2 \mid \mathcal{F}_{t-1}\right\} - 1\right| \to 0,$$

as $n \to \infty$. Further, for $\delta > 0$, as $n \to \infty$

$$\sup_{\lfloor q_o n\rfloor \leq l \leq n}\sum_{t=1}^{l}\mathbf{E}\left\{\left[\frac{1}{\Lambda_l}\left(\sum_{j=0}^{l-t}\rho_n^{l-t-j}\phi_j\right)\xi_t\right]^2 I\left\{\left|\frac{1}{\Lambda_l}\left(\sum_{j=0}^{l-t}\rho_n^{l-t-j}\phi_j\right)\xi_t\right| > \delta\right\} \mid \mathcal{F}_{t-1}\right\} \to 0.$$

The first condition holds trivially from the fact that

$$\sum_{t=1}^{l}\mathbf{E}\left\{\left[\frac{1}{\Lambda_l}\left(\sum_{j=0}^{l-t}\rho_n^{l-t-j}\phi_j\right)\xi_t\right]^2 \mid \mathcal{F}_{t-1}\right\} = \frac{\sigma_\xi^2}{\Lambda_l^2}\sum_{t=1}^{l}\left\{\left(\sum_{j=0}^{l-t}\rho_n^{l-t-j}\phi_j\right)^2\right\} = 1. \tag{39}$$

Next, we show that the uniform Lindeberg condition holds. Set $\pi_{n,l,t} := \frac{\sigma_\xi^2}{\delta^2\Lambda_l^2}\left(\sum_{j=0}^{l-t}\rho_n^{l-t-j}\phi_j\right)^2$, let ζ be as in Assumption 2.3 and fix $\delta > 0$. Then using Hölder's and Markov's inequalities we get

$$\sum_{t=1}^{l}\mathbf{E}\left\{\left[\frac{1}{\Lambda_l}\left(\sum_{j=0}^{l-t}\rho_n^{l-t-j}\phi_j\right)\xi_t\right]^2 I\left\{\left|\frac{1}{\Lambda_l}\left(\sum_{j=0}^{l-t}\rho_n^{l-t-j}\phi_j\right)\xi_t\right| > \delta\right\} \mid \mathcal{F}_{t-1}\right\}$$

[17] Note that $\tilde{x}_l$ is a truncated version of the x_l process of eq. (43).

$$\leq \sum_{t=1}^{l} \left\{ \mathbf{E} \left[\frac{1}{\Lambda_l} \left(\sum_{j=0}^{l-t} \rho_n^{l-t-j} \phi_j \right) \xi_t \right]^{2(1+\zeta)} \right\}^{\frac{1}{1+\zeta}} \left\{ \mathbf{P} \left(\left| \frac{1}{\Lambda_l} \left(\sum_{j=0}^{l-t} \rho_n^{l-t-j} \phi_j \right) \xi_t \right| > \delta \right) \right\}^{\frac{\zeta}{1+\zeta}}$$

$$= \left\{ \mathbf{E} \left[\xi_t \right]^{2(1+\zeta)} \right\}^{\frac{1}{1+\zeta}} \sum_{t=1}^{l} \left[\frac{1}{\Lambda_l} \left(\sum_{j=0}^{l-t} \rho_n^{l-t-j} \phi_j \right) \right]^2 \left\{ \mathbf{P} \left(\left| \frac{1}{\Lambda_l} \left(\sum_{j=0}^{l-t} \rho_n^{l-t-j} \phi_j \right) \xi_t \right| > \delta \right) \right\}^{\frac{\zeta}{1+\zeta}}$$

$$\leq \left\{ \mathbf{E} \left[\xi_t \right]^{2(1+\zeta)} \right\}^{\frac{1}{1+\zeta}} \sum_{t=1}^{l} \left[\frac{1}{\Lambda_l} \left(\sum_{j=0}^{l-t} \rho_n^{l-t-j} \phi_j \right) \right]^2 \left\{ \pi_{n,l,t} \right\}^{\frac{\zeta}{1+\zeta}}$$

$$\leq \left\{ \mathbf{E} \left[\xi_t \right]^{2(1+\zeta)} \right\}^{\frac{1}{1+\zeta}} \sup_{1 \leq t \leq n, \lfloor q_o n \rfloor \leq l \leq n} \pi_{n,l,t} \sum_{t=1}^{l} \left[\frac{1}{\Lambda_l} \left(\sum_{j=0}^{l-t} \rho_n^{l-t-j} \phi_j \right) \right]^2$$

$$= \frac{1}{\sigma_\xi^2} \left\{ \mathbf{E} \left[\xi_t \right]^{2(1+\zeta)} \right\}^{\frac{1}{1+\zeta}} \sup_{1 \leq t \leq n, \lfloor q_o n \rfloor \leq l \leq n} \pi_{n,l,t}. \tag{40}$$

We shall show that

$$\sup_{1 \leq t \leq n, \lfloor q_o n \rfloor \leq l \leq n} \pi_{n,l,t} = O(1/ \lfloor q_o n \rfloor). \tag{41}$$

In view of (41), (39) and (40) are sufficient for (38). By (25) for $\lfloor q_o n \rfloor \leq l \leq n$ and n large enough we have

$$\frac{\delta^2}{\sigma_\xi^2} \pi_{n,l,t} = \frac{1}{\Lambda_l^2} \left(\sum_{j=0}^{l-t} \rho_n^{l-t-j} \phi_j \right)^2 = \frac{1}{l} \frac{1}{\Lambda_l^2/d_l^2} \left(\frac{\sqrt{l}}{d_l} \sum_{j=0}^{l-t} \rho_n^{l-t-j} \phi_j \right)^2 \leq \frac{(\beta_2)^2}{\lfloor q_o n \rfloor} \left(\frac{\sqrt{l}}{d_l} \sum_{j=0}^{l-t} \rho_n^{l-t-j} \phi_j \right)^2.$$

Hence, to get (41) it suffices to show that

$$\sup_{1 \leq t \leq n} \left(\frac{\sqrt{l}}{d_l} \sum_{j=0}^{l-t} \rho_n^{l-t-j} \phi_j \right)^2 = O(1). \tag{42}$$

First, suppose that **LM** is satisfied. Then by (29) it can be easily seen that (42) holds. Next, under **SM** it can be easily seen that (42) follows from the arguments following (31). Finally, under **AP** (42) follows easily from (33) and (34). ■

Proposition A5 (strong approx.). *Suppose that* $\xi_t \sim i.i.d.(0, \sigma_\xi^2)$ *with* $\mathbf{E} \left| \xi_t \right|^p < \infty$, $p > 2$. *Then on an expanded prob. space there are* $\xi_t^* \sim i.d.N(0, \sigma_\xi^2)$ *such that as* $n \to \infty$

$$\sup_{1 \leq k \leq n} \left| \sum_{t=1}^{k} \xi_t - \sum_{t=1}^{k} \xi_t^* \right| = o_{a.s.}(n^{1/p}).$$

Proof of Proposition A5: See Kolmós et al. (1976). ■

Proposition A6 (strong approx. for linear process). *Suppose that* ξ_t *and* ξ_t^* *are as in Proposition A5. Suppose that one of the following holds:*

(i) $\phi_j \sim j^{-m}$ $m \in (1/2, 1)$*;*

(ii) $\sum_{j=0}^{\infty} |\phi_j| < \infty$ *and* $\frac{3}{2} - m > 1/p$*,* $m \in [1, 3/2)$*.*[18]

Then as $n \to \infty$

$$n^{-\left(\frac{3}{2}-m\right)} \sup_{1\leq k\leq n} \left| \sum_{t=1}^{k} \sum_{j=0}^{\infty} \phi_j \xi_{t-j} - \sum_{t=1}^{k} \sum_{j=0}^{\infty} \phi_j \xi_{t-j}^* \right| = o_{a.s.}(1).$$

Proof Proposition A6: (i) ($\phi_j \sim j^{-m}$, $m \in (1/2, 1)$): see Wang et al. (2003).

(ii) ($\sum_{j=0}^{\infty} |\phi_j| < \infty$ *and* $\frac{3}{2} - m > 1/p$, $m \in [1, 3/2)$): Consider $x_n = \sum_{t=1}^{n} \sum_{j=0}^{\infty} \phi_j \xi_{t-j}$. Write

$$x_n = \sum_{j=0}^{n} \phi_j \sum_{t=1}^{n-j} \xi_t + \sum_{t=1}^{n} \left(\sum_{j=0}^{\infty} \phi_{t+j} \right) \xi_{-t} =: A_n + B_n,$$

$$x_n^* := \sum_{j=0}^{n} \phi_j \sum_{t=1}^{n-j} \xi_t^* + \sum_{t=1}^{n} \left(\sum_{j=0}^{\infty} \phi_{t+j} \right) \xi_{-t}^* =: A_n^* + B_n^*,$$

Following Wang et al. (2003) (see for example eq. (24) in Wang et al. (2003)) we shall show that

$$n^{-\left(\frac{3}{2}-m\right)} \sup_{1\leq k\leq n} |x_k - x_k^*| = o_{a.s.}(1).$$

Consider first

$$n^{-\left(\frac{3}{2}-m\right)} \sup_{1\leq k\leq n} |A_k^* - A_k| := n^{-\left(\frac{3}{2}-m\right)} \sup_{1\leq k\leq n} \left| \sum_{j=0}^{k} \phi_j \sum_{t=1}^{k-j} \xi_t - \sum_{j=0}^{k} \phi_j \sum_{t=1}^{k-j} \xi_t^* \right|$$

$$\leq n^{-\left(\frac{3}{2}-m\right)} \sup_{1\leq k\leq n} \sum_{j=0}^{k} \left| \phi_j \left(\sum_{t=1}^{k-j} \xi_t - \sum_{t=1}^{k-j} \xi_t^* \right) \right| \leq n^{-\left(\frac{3}{2}-m\right)} \sup_{1\leq k\leq n} \sum_{j=0}^{k} |\phi_j| \sup_{1\leq j\leq k} \left| \sum_{t=1}^{k-j} \xi_t - \sum_{t=1}^{k-j} \xi_t^* \right|$$

$$\leq n^{-\left(\frac{3}{2}-m\right)} \sup_{1\leq k\leq n} \sum_{j=0}^{k} |\phi_j| \times \sup_{1\leq k\leq n} \sup_{1\leq j\leq k} \left| \sum_{t=1}^{k-j} \xi_t - \sum_{t=1}^{k-j} \xi_t^* \right|$$

$$\leq \sum_{j=0}^{\infty} |\phi_j| \times n^{-\left(\frac{3}{2}-m\right)} \sup_{1\leq k\leq n} \left| \sum_{t=1}^{k} \xi_t - \sum_{t=1}^{k} \xi_t^* \right| = o_{a.s.}\left(\frac{n^{1/p}}{n^{\frac{3}{2}-m}} \right) = o_{a.s.}(1).$$

[18] Note that the requirement $\frac{3}{2} - m > 1/p$ is always satisfied for $m = 1$.

Next consider the term B_n. By Lemma 4.1 of Wang et al. (2003)

$$B_n = \sum_{j=0}^{\infty} \left(\phi_{j+1} - \phi_{n+j+1}\right) \sum_{i=0}^{j} \xi_{-i}, \ a.s.$$

Hence,

$$\sup_{1\le k\le n} |B_k - B_k^*| = \sup_{1\le k\le n} \left| \sum_{j=0}^{\infty} \left(\phi_{j+1} - \phi_{k+j+1}\right) \sum_{i=0}^{j} \xi_{-i} - \sum_{j=0}^{\infty} \left(\phi_{j+1} - \phi_{k+j+1}\right) \sum_{i=0}^{j} \xi_{-i}^* \right|$$

$$= \sup_{1\le k\le n} \left| \sum_{j=0}^{\infty} \left(\phi_{j+1} - \phi_{k+j+1}\right) \sum_{i=0}^{j} \left(\xi_{-i} - \xi_{-i}^*\right) \right|$$

$$\le \sup_{1\le k\le n} \left| \sum_{j=0}^{n} \left(\phi_{j+1} - \phi_{k+j+1}\right) \sum_{i=0}^{j} \left(\xi_{-i} - \xi_{-i}^*\right) \right| + \sup_{1\le k\le n} \left| \sum_{j=n+1}^{\infty} \left(\phi_{j+1} - \phi_{k+j+1}\right) \sum_{i=0}^{j} \left(\xi_{-i} - \xi_{-i}^*\right) \right|$$

$$\le \sup_{1\le k\le n} \sum_{j=0}^{n} \left|\phi_{j+1} - \phi_{k+j+1}\right| \left| \sum_{i=0}^{j} \left(\xi_{-i} - \xi_{-i}^*\right) \right| + \sup_{1\le k\le n} \sum_{j=n+1}^{\infty} \left|\phi_{j+1} - \phi_{k+j+1}\right| \underbrace{\left| \sum_{i=0}^{j} \left(\xi_{-i} - \xi_{-i}^*\right) \right|}_{\bar{\xi}_j}$$

$$\le \sup_{1\le j\le n} \left(\bar{\xi}_j\right) 2 \sum_{j=0}^{\infty} \left|\phi_j\right| + 2 \sum_{j=n+1}^{\infty} \left|\phi_j\right| j^{1/p} \frac{1}{j^{1/p}} \bar{\xi}_j.$$

Note that for all $\epsilon > 0$ there is N_ϵ such that $j^{-1/p}\bar{\xi}_j < \epsilon \ a.s.$ when $j \ge N_\epsilon$. Therefore, for n large enough

$$n^{-\left(\frac{3}{2}-m\right)} \sup_{1\le k\le n} |B_k - B_k^*| \le o_{a.s.}\left(\frac{n^{1/p}}{n^{\frac{3}{2}-m}}\right) + \epsilon 2 \left(\frac{1}{n^{\frac{3}{2}-m}}\right) \sum_{j=n+1}^{\infty} \left|\phi_j\right| j^{1/p} \ a.s.$$

$$\le o_{a.s.}\left(\frac{n^{1/p}}{n^{\frac{3}{2}-m}}\right) + \epsilon 2 \left(\frac{n^{1/p}}{n^{\frac{3}{2}-m}}\right) \sum_{j=n+1}^{\infty} \left|\phi_j\right| = o_{a.s.}(1),$$

and the result follows. ■

Proof of Lemma 1: The proof has four parts. Parts (i)-(iii) show that parts (i)-(iii) of Assumption 2.1 hold respectively. Part (iv) shows that Assumption 2.2 holds.

(i) (Proof that Assumption 2.1(i) holds) First, we shall show that there is some $n_o \in \mathbb{N}$ and some $0 < q_o < 1$ such that the density function $h_l(x)$ of x_l/Λ_l is $\sup_{l\ge\lfloor q_o n_o\rfloor} \sup_x h_l(x) < \infty$. Subsequently we shall show that $\sup_{l<\lfloor q_o n_o\rfloor} \sup_x h_l(x) < \infty$.

Note that we can decompose x_l as follows

$$x_l = \sum_{t=1}^{l} \sum_{j=0}^{l-t} \rho_n^{l-t-j} \phi_j \xi_t + \sum_{t=-\infty}^{0} \sum_{j=1}^{l} \rho_n^{l-j} \phi_{j-t} \xi_t =: \sum_{t=-\infty}^{l} \theta_{l,n}(t) \xi_t. \tag{43}$$

We will show that for $1 \leq l \leq n$ and all n, the characteristic function of x_l/d_l has L_1-norm bounded by finite constant. The subsequent manipulations are similar to those of Jeganathan (2008, Lemma 7) (see also Pötscher, 2004). Choose b such that $D_2 b = \delta$, where δ is as in Proposition A1 and D_2 as in Lemma A3. Then in view of (23), for n large enough, $n \geq n_o$ say, $\lfloor q_o n \rfloor \leq l \leq n$ and $1 \leq t \leq \lfloor q_1 l \rfloor$ we have

$$\int_{|\lambda| \leq b\sqrt{l}} \left| \mathbf{E}\left(e^{i\lambda x_l/\Lambda_l}\right) \right| d\lambda \leq \int_{|\lambda| \leq b\sqrt{l}} \left| \mathbf{E}\left[\exp\left(\frac{i\lambda}{\Lambda_l} \sum_{t=1}^{l} \sum_{j=0}^{l-t} \rho_n^{l-t-j} \phi_j \xi_t \right) \right] \right| d\lambda$$

$$= \int_{|\lambda| \leq b\sqrt{l}} \prod_{t=1}^{l} \left| \mathbf{E}\left[\exp\left(\frac{i\lambda}{\Lambda_l} \sum_{j=0}^{l-t} \rho_n^{l-t-j} \phi_j \xi_t \right) \right] \right| d\lambda = \int_{|\lambda| \leq b\sqrt{l}} \prod_{t=1}^{l} \left| \psi\left(\frac{\lambda}{\Lambda_l} \sum_{j=0}^{l-t} \rho_n^{l-t-j} \phi_j \right) \right| d\lambda$$

$$\leq \int_{|\lambda| \leq b\sqrt{l}} \prod_{t=1}^{\lfloor q_1 l \rfloor} \left| \psi\left(\frac{\lambda}{\Lambda_l} \sum_{j=0}^{l-t} \rho_n^{n-t-j} \phi_j \right) \right| d\lambda \leq \prod_{t=1}^{\lfloor q_1 l \rfloor} \left\{ \int_{|\lambda| \leq b\sqrt{l}} \left| \psi\left(\frac{\lambda}{\Lambda_l} \sum_{j=0}^{l-t} \rho_n^{l-t-j} \phi_j \right) \right|^{\lfloor q_1 l \rfloor} d\lambda \right\}^{\frac{1}{\lfloor q_1 l \rfloor}}$$

$$= \prod_{t=1}^{\lfloor q_1 l \rfloor} \left| \Lambda_l / \sum_{j=0}^{l-t} \rho_n^{l-t-j} \phi_j \sqrt{l} \right| \left\{ \int_{|\mu| \leq b\sqrt{l} \left| \sum_{j=0}^{l-t} \rho_n^{l-t-j} \phi_j \sqrt{l} \right| / \Lambda_l} \left| \psi\left(\frac{\mu}{\sqrt{l}} \right) \right|^{\lfloor q_1 l \rfloor} d\mu \right\}^{\frac{1}{\lfloor q_1 l \rfloor}}$$

$$\leq D_1^{-1} \int_{|\mu| \leq D_2 b\sqrt{l}} \left| \psi\left(\frac{\mu}{\sqrt{l}} \right) \right|^{\lfloor q_1 l \rfloor} d\mu \leq D_1^{-1} \int_{|\mu| \leq \delta\sqrt{l}} e^{-A\mu^2} d\mu \leq D_1^{-1} \int_{\mathbb{R}} e^{-A\mu^2} d\mu < \infty.$$

Next, for $n < n_o$ we have

$$\int_{|\lambda| \leq b\sqrt{l}} \left| \mathbf{E}\left(e^{i\lambda x_l/\Lambda_l}\right) \right| d\lambda \leq \int_{|\lambda| \leq b\sqrt{n_o}} \left| \mathbf{E}\left(e^{i\lambda x_l/\Lambda_l}\right) \right| d\lambda \leq 2b\sqrt{n_o} < \infty.$$

Hence, $\int_{|\lambda| \leq b\sqrt{l}} \left| \mathbf{E}\left(e^{i\lambda x_l/\Lambda_l}\right) \right| d\lambda < \infty$ for all $1 \leq l \leq n \in \mathbb{N}$.

Next, in view of Proposition A2(ii) for $n \geq n_o$ and $\lfloor q_o n \rfloor \leq l \leq n$ we get

$$\int_{|\lambda| > b\sqrt{l}} \left| \mathbf{E}\left(e^{i\lambda x_l/\Lambda_l}\right) \right| d\lambda \leq \prod_{t=1}^{\lfloor q_1 l \rfloor} \left\{ \int_{|\lambda| > b\sqrt{l}} \left| \psi\left(\frac{\lambda}{\Lambda_l} \sum_{j=1}^{l-t} \rho_n^{l-t-j} \phi_j \right) \right|^{\lfloor q_1 l \rfloor} d\lambda \right\}^{\frac{1}{\lfloor q_1 l \rfloor}}$$

$$= \prod_{t=1}^{\lfloor q_1 l \rfloor} \left| \Lambda_l / \sum_{j=0}^{l-t} \rho_n^{l-t-j} \phi_j \sqrt{l} \right| \left\{ \int_{|\mu| > b\sqrt{l} \left| \sum_{j=0}^{l-t} \rho_n^{l-t-j} \phi_j \sqrt{l} \right| / \Lambda_l} \left| \psi\left(\frac{\mu}{\sqrt{l}} \right) \right|^{\lfloor q_1 l \rfloor} d\mu \right\}^{\frac{1}{\lfloor q_1 l \rfloor}}$$

$$= D_1^{-1} \prod_{t=1}^{\lfloor q_1 l \rfloor} \left\{ \int_{|\mu|>D_1 b\sqrt{l}} \left| \psi\left(\frac{\mu}{\sqrt{l}}\right) \right|^{\lfloor q_1 l \rfloor} d\mu \right\}^{\frac{1}{\lfloor q_1 l \rfloor}} = D_1^{-1} \int_{|\mu|>D_1 b\sqrt{l}} \left| \psi\left(\frac{\mu}{\sqrt{l}}\right) \right|^{\lfloor q_1 l \rfloor} d\mu$$

$$= D_1^{-1} \int_{|\mu|>D_1 b\sqrt{l}} \left| \psi\left(\frac{\mu}{\sqrt{l}}\right) \right|^{\lfloor q_1 l \rfloor -1} \left| \psi\left(\frac{\mu}{\sqrt{l}}\right) \right| d\mu$$

$$\leq \sqrt{l} \sup_{|\lambda|>D_1 b\sqrt{l}} \left| \psi\left(\frac{\lambda}{\sqrt{l}}\right) \right|^{\lfloor \alpha l \rfloor} \int_{\mathbb{R}} |\psi(\lambda)|\, d\lambda \leq \sqrt{l} B\rho^{Cl} \int_{\mathbb{R}} |\psi(\lambda)|\, d\lambda,$$

where $0 < \rho < 1$, $\alpha > 0$ is such that $\lfloor q_1 l \rfloor - 1 \geq \lfloor \alpha l \rfloor$, for l large enough, and the last inequality follows from Proposition A2(ii). Note that last term above is bounded because $\sqrt{l}\rho^{Cl} \to 0$, as $l \to \infty$.

Next, we show $\int_{|\lambda|>b\sqrt{l}} \left| \mathbf{E}\left(e^{i\lambda x_l/dl}\right) \right| d\lambda < \infty$, for $1 \leq l \leq n < n_o$. Note that under Assumption 2.3, for all $1 \leq l \leq n \in \mathbb{N}$, there is some $t^* \leq l$, $t^* \in \mathbb{Z}$ such that the coefficients in (43) satisfy

$$\theta_{l,n}(t^*) \neq 0. \tag{44}$$

The proof of (44) is provided later. In view of (44),

$$\int_{\mathbb{R}} \left| \mathbf{E}\left(e^{i\lambda x_l/\Lambda_l}\right) \right| d\lambda = \int_{\mathbb{R}} \left| \mathbf{E} \exp\left[\frac{i\lambda}{\Lambda_l} \left(\sum_{t=-\infty}^{l} \theta_{l,n}(t)\xi_t \right) \right] \right| d\lambda \leq \int_{\mathbb{R}} \left| \psi\left(\frac{\lambda}{\Lambda_l} \theta_{l,n}(t^*) \right) \right| d\lambda$$

$$= \int_{\mathbb{R}} \left| \psi\left(\lambda \left| \frac{1}{\Lambda_l} \theta_{l,n}(t^*) \right| \right) \right| d\lambda = \frac{|\Lambda_l|}{|\theta_{l,n}(t^*)|} \int_{\mathbb{R}} |\psi(\lambda)|\, d\lambda < \infty, \text{ for all } 1 \leq l \leq n < n_o,$$

as required.

Next, we show that (44) holds. Suppose that $\theta_{l,n}(t) = 0$ for all $t \leq l$, $t \in \mathbb{Z}$. Then we have

$$\left. \begin{array}{l} \theta_{l,n}(l) = \phi_0 \\ \theta_{l,n}(l-1) = \rho_n \phi_0 + \phi_1 \\ \theta_{l,n}(l-2) = \rho_n^2 \phi_0 + \rho_n \phi_1 + \phi_2 \\ . \\ \theta_{l,n}(1) = \rho_n^{l-1}\phi_0 + \rho_n^{l-2}\phi_1 + ... + \phi_{l-1} \\ \theta_{l,n}(0) = \rho_n^{l-1}\phi_1 + \rho_n^{l-2}\phi_2 + +... + \phi_l \\ \theta_{l,n}(-1) = \rho_n^{l-1}\phi_2 + \rho_n^{l-1}\phi_3 + ... + \phi_{l+1} \\ . \end{array} \right\} = 0,$$

which in turn implies that $\phi_j = 0$ for all $j \in \mathbb{Z}_+$. Under **SM** this contradicts the fact that $\sum_{j=0}^{\infty} \phi_j \neq 0$. Therefore, (44) holds. Under **LM** or **AP**, $\phi_j = 0$ for all $j \in \mathbb{Z}_+$ contradicts the fact that $\phi_j \sim j^{-m}$.

Hence, the above shows that x_l/Λ_l has density $h_l(x)$ satisfying $\sup_{n\geq 1} \sup_{1\leq l\leq n} \sup_x h_l(x) < \infty$. Next, set $d_{l,k,n} = \Lambda_{l-k}/d_n$. In view of this the result follows from the fact that conditionally on $\mathcal{F}_{k,n}$, $\left(x_{l,n} - \rho_n^{l-k} x_{k,n}\right)/d_{l,k,n} = \left(x_l^* + x_l^{**}\right)/\Lambda_{l-k}$ has density

$h_{l-k}\left(x - x_l^{**}/\Lambda_{l-k}\right)$, where x_l^* and x_l^{**} are defined in part (ii) of the proof below. Hence, $h_{l-k}\left(x - x_l^{**}/\Lambda_{l-k}\right) \leq \sup_{n\geq 1}\sup_{1\leq l\leq n}\sup_x h_l(x) < \infty$, as required.

(ii) Proof that Assumption 2.1(ii) holds: First, by part (i) of the current proof, Proposition A4 and using the same arguments as those used in WP (page 729-730) it follows that for $\lfloor q_o n\rfloor \leq l \leq n,\ h_l(x)$, the density of x_l/Λ_l, satisfies

$$\sup_{\lfloor q_o n\rfloor \leq l\leq n}\sup_x \left| h_l(x) - \frac{1}{\sqrt{2\pi}}e^{-x^2/2}\right| \to 0,$$

as $n \to \infty$. Write

$$x_l = \sum_{t=1}^{l}\rho_n^{l-t}v_t = \rho_n^{l-k}\sum_{t=1}^{k}\rho_n^{k-t}v_t + \sum_{t=k+1}^{l}\rho_n^{l-t}v_t = \rho_n^{l-k}x_k + \sum_{t=k+1}^{l}\rho_n^{l-t}v_t$$

$$= \rho_n^{l-k}x_k + \sum_{t=k+1}^{l}\sum_{j=0}^{l-t}\rho_n^{l-t-j}\phi_j\xi_t + \sum_{t=-\infty}^{0}\sum_{j=k+1}^{l}\rho_n^{l-j}\phi_{j-t}\xi_t$$

$$:= \rho_n^{l-k}x_k + x_l^* + x_l^{**}$$

Next, note that $\tilde{x}_{l-k} \overset{d}{=} x_l^*$. Set $d_{l,k,n} = \Lambda_{l-k}/d_n$. Hence, conditionally on $\mathcal{F}_{k,n}$, $\left(x_{l,n} - \rho_n^{l-k}x_{k,n}\right)/d_{l,k,n} = \left(x_l^* + x_l^{**}\right)/\Lambda_{l-k}$ has density $h_{l-k}\left(x - x_l^{**}/\Lambda_{l-k}\right)$. In view of this, the result follows easily from WP page 731.

(iii) Eq. (5) follows using arguments similar to those used in the proof of Proposition A3. For instance, suppose that **LM** holds and $c > 0$. Then

$$\inf_{(l,k)\in\Omega(q_o)} d_{l,k,n} = \sqrt{\frac{1}{d_n^2}\inf_{(l,k)\in\Omega(q_o)}\Lambda_{l-k}^2} = \sqrt{\frac{1}{d_n^2}\inf_{\lfloor q_o n\rfloor\leq l\leq n}\Lambda_l^2}$$

$$= \sqrt{\frac{1}{d_n^2}\inf_{\lfloor q_o n\rfloor\leq l\leq n}\sum_{t=1}^{l}\left(\rho_n^{-t}\sum_{j=0}^{l-t}\phi_j\rho_n^{l-j}\right)^2} = \sqrt{\inf_{\lfloor q_o n\rfloor\leq l\leq n}\frac{1}{n}\sum_{t=1}^{l}\left(\rho_n^{-t}\frac{1}{n}\sum_{j=1}^{l-t}\left(\frac{j}{n}\right)^{-m}\rho_n^{l-j}\right)^2}+o(1)$$

$$\geq \sqrt{\inf_{\lfloor q_o n\rfloor\leq l\leq n}\frac{1}{n}\sum_{t=1}^{l}\left(\rho_n^{-t}\frac{1}{n}\sum_{j=1}^{l-t}\left(\frac{j}{n}\right)^{-m}\right)^2} \geq \sqrt{\rho_n^{-2n}\inf_{\lfloor q_o n\rfloor\leq l\leq n}\frac{1}{n}\sum_{t=1}^{l}\left(\frac{1}{n}\sum_{j=1}^{l-t}\left(\frac{j}{n}\right)^{-m}\right)^2}$$

$$= \sqrt{\rho_n^{-2n}\frac{1}{n}\sum_{t=1}^{\lfloor q_o n\rfloor}\left(\frac{1}{n}\sum_{j=1}^{\lfloor q_o n\rfloor - t}\left(\frac{j}{n}\right)^{-m}\right)^2} \to \sqrt{e^{-2c}\int_0^{q_o}\left(\int_0^{q_o-r}s^{-m}ds\right)^2 dr} = \frac{e^{-c}q_o^{(3-2m)/2}}{\sqrt{(1-m)^2(3-2m)}}.$$

Finally, (6)-(9) can be shown to hold using arguments similar to those used for the proof of (25). For instance suppose that **LM** holds and $c > 0$. We shall show that (8) holds. Without loss of generality set $\sigma_\xi^2 = 1$. As $n \to \infty$ we have

$$\frac{1}{n} \max_{0 \le k \le (1-\eta)\, n} \sum_{l=k+1}^{k+\lfloor \eta n \rfloor} (d_{l,k,n})^{-1} = \frac{1}{n} \max_{0 \le k \le (1-\eta) n} \sum_{l=k+1}^{k+\lfloor \eta n \rfloor} \left[\frac{1}{d_n} \sum_{t=1}^{l-k} \left(\sum_{j=0}^{l-k-t} \phi_j \rho_n^{l-k-t-j} \right)^2 \right]^{-1/2}$$

$$\frac{1}{n} \max_{0 \le k \le (1-\eta) n} \sum_{l=k+1}^{k+\lfloor \eta n \rfloor} \left[\frac{1}{n} \sum_{t=1}^{l-k} \left(\frac{1}{n} \sum_{j=1}^{l-k-t} \left(\frac{j}{n} \right)^{-m} \rho_n^{l-k-t-j} \right)^2 \right]^{-1/2} + o(1)$$

$$\le \frac{1}{n} \max_{0 \le k \le (1-\eta) n} \sum_{l=k+1}^{k+\lfloor \eta n \rfloor} \left[\rho_n^{-2n} \frac{1}{n} \sum_{t=1}^{l-k} \left(\frac{1}{n} \sum_{j=1}^{l-k-t} \left(\frac{j}{n} \right)^{-m} \right)^2 \right]^{-1/2} + o(1)$$

$$\max_{0 \le k \le (1-\eta) n} \frac{1}{n} \sum_{l=k+1}^{k+\lfloor \eta n \rfloor} \left[\frac{(1-m)^2 (3-2m) e^{2c}}{\int_0^{\frac{l-k}{n}} \left(\int_0^{\frac{l-k}{n}-r} s^{-m} ds \right)^2 dr} \right]^{1/2} + o(1) = \max_{0 \le k \le (1-\eta) n} \frac{1}{n} \sum_{l=k+1}^{k+\lfloor \eta n \rfloor} \frac{A}{\left(\frac{l}{n} - \frac{k}{n} \right)^{3/2-m}}$$

Next, Euler summation gives (see for example (27))

$$\max_{0 \le k \le (1-\eta) n} \left| \frac{1}{n} \sum_{l=k+1}^{k+\lfloor \eta n \rfloor} \left(\frac{l}{n} - \frac{k}{n} \right)^{-(3/2-m)} - \int_{\frac{k+1}{n}}^{\frac{k}{n}+\eta} \left(s - \frac{k}{n} \right)^{-(3/2-m)} ds \right| \to 0.$$

Hence, as $n \to \infty$,

$$\max_{0 \le k \le (1-\eta) n} \frac{1}{n} \sum_{l=k+1}^{k+\eta n} \frac{A}{\left(\frac{l}{n} \right)^{3/2-m}} \to \max_{0 \le k \le (1-\eta) n} A \int_{\frac{k+1}{n}}^{\frac{k}{n}+\eta} \left(s - \frac{k}{n} \right)^{-3/2+m+1} ds$$

$$= A \max_{0 \le k \le (1-\eta) n} \left\{ \left(\frac{k}{n} + \eta - \frac{k}{n} \right)^{m-1/2} - \left(\frac{k+1}{n} - \frac{k}{n} \right)^{m-1/2} \right\}$$

$$= A \left\{ \eta^{m-1/2} - \left(\frac{1}{n} \right)^{m-1/2} \right\} \overset{n \to \infty}{\to} A \eta^{m-1/2} \overset{\eta \to 0}{\to} 0,$$

as required.

(iv) We next show that Assumption 2.2 holds. Write

$$x_t = \sum_{j=1}^{n} \rho_n^{t-j} v_j.$$

Let $S_t = \sum_{j=1}^{t} v_j$. Then for $s \in [0,1]$ summation by parts gives

$$x_{\lfloor ns \rfloor} = \sum_{j=1}^{\lfloor ns \rfloor} \rho_n^{t-j} v_j = \sum_{j=1}^{\lfloor ns \rfloor} \rho_n^{t-j} \Delta S_j = \rho_n^{-1} S_{\lfloor ns \rfloor} - \sum_{j=1}^{\lfloor ns \rfloor} \left(\rho_n^{\lfloor ns \rfloor - j - 1} - \rho_n^{\lfloor ns \rfloor - j} \right) S_j$$

$$= \rho_n^{-1} \left[S_{\lfloor ns \rfloor} - (1 - \rho_n) \sum_{j=1}^{\lfloor ns \rfloor} \rho_n^{\lfloor ns \rfloor - j} S_j \right]. \tag{45}$$

Next, consider the term

$$\sum_{j=1}^{\lfloor ns \rfloor} \rho_n^{\lfloor ns \rfloor - j} S_j = \int_1^{\lfloor ns \rfloor} \rho_n^{\lfloor ns \rfloor - \lfloor x \rfloor} S\left(\lfloor x \rfloor \right) dx = n \int_{1/n}^{\lfloor ns \rfloor / n} \rho_n^{\lfloor ns \rfloor - \lfloor ny \rfloor} S\left(\lfloor ny \rfloor \right) dy.$$

The term

$$\rho_n^{\lfloor ns \rfloor - [ny]} = \exp\left\{ \left(\lfloor ns \rfloor - \lfloor ny \rfloor \right) \ln\left(1 + \frac{c}{n} \right) \right\} = \exp\left\{ \left(\lfloor ns \rfloor - \lfloor ny \rfloor \right) \left[\frac{c}{n} + O\left(n^{-2} \right) \right] \right\}$$

$$= \exp\left\{ \left([ns] - \lfloor ny \rfloor \right) \frac{c}{n} + O\left(n^{-1} \right) \right\} = \exp\left\{ \left(\lfloor ns \rfloor - \lfloor ny \rfloor \right) \frac{c}{n} \right\} + o(1),$$

uniformly in $s, y \in [0,1]$. Hence, (45) and the invariance principle for fractional processes (e.g. Jeganathan, 2008) gives

$$\frac{1}{d_n} x_{\lfloor ns \rfloor} = \rho_n^{-1} \left[\frac{1}{d_n} S\left(\lfloor ns \rfloor \right) + c\left(1 + O(n^{-1}) \right) \int_{1/n}^{\lfloor ns \rfloor / n} \rho_n^{\lfloor ns \rfloor - \lfloor ny \rfloor} \frac{1}{d_n} S\left(\lfloor ny \rfloor \right) dy \right] \Longrightarrow$$

$$\sigma_\xi \left[B_m(s) + c \int_0^s \exp\left[c\left(s - y \right) \right] B_m(y) dy \right] = \sigma_\xi \int_0^t e^{c(t-s)} dB_m(s).$$

In the **LM** case the strong approximation result of Assumption 2.2(b) can be obtained using the same arguments as those above together with the limit theory of Wang, Lin and Gulati (2003). Phillips (2007) provides a strong approximation result in the short memory case under the stronger summability requirement $\sum_{j=0}^{\infty} j \left| \phi_j \right| < \infty$. In the **AP** and **SM** cases, of the current paper, the result follows from Proposition A6 (see also Wang et al. (2003) and Oodaira (1976)). ■

Proof of Lemma 2: In view of Lemma 1, the result follows easily from Theorem 3.1 of WP. ■

Proof of Theorem 1: By Assumption 2.1 x_t posseses a density. Therefore, under the null hypothesis, $f(x_t) = \mu$ $a.s.$ Hence, the result follows by arguments similar to those used in the proof of Theorem 4 of Kasparis and Phillips (2012). ■

Proof of Theorem 2: We first determine the limit behaviour of the parametric estimators $\hat{\mu}$ and $\hat{\sigma}_u^2$ under H_1. By Berkes and Horváth (2006, Theorem 2.2) we get

$$\frac{1}{\kappa_g(d_n)}\hat{\mu} = \frac{1}{n\kappa_g(d_n)}\sum_{t=1+\nu}^{n} y_t = \frac{\mu}{\kappa_g(d_n)} + \frac{1}{n\kappa_g(d_n)}\sum_{t=1+\nu}^{n} g\left(x_{t-\nu}\right) + O_p\left(\frac{1}{\sqrt{n}\kappa_g(d_n)}\right)$$

$$= \begin{cases} \mu + \int_0^1 H_g(G(s))ds + o_p(1), \ \kappa_g(\lambda) = 1 \\ \int_0^1 H_g(G(s))ds + o_p(1), \ \lim_{\lambda\to\infty}\kappa_g(\lambda) = \infty \\ \frac{\mu}{\kappa_g(d_n)} + \int_0^1 H_g(G(s))ds + o_p(1), \ \lim_{\lambda\to\infty}\kappa_g(\lambda) = 0 \end{cases}$$

$$= \frac{\mu}{\kappa_g(d_n)} + \int_0^1 H_g(G(s))ds + o_p(1).$$

Further, for integrable g we have $\hat{\mu} = \mu + o_p(1)$.[19]

Next, the variace estimator is

$$\frac{1}{\kappa_g(d_n)^2}\hat{\sigma}_u^2 = \frac{1}{n\kappa_g(d_n)^2}\sum_{t=1+\nu}^{n}\left(y_t - \hat{\mu}\right)^2$$

$$= \frac{1}{n\kappa_g(d_n)^2}\left\{\sum_{t=1+\nu}^{n}\left[(\mu-\hat{\mu}) + g\left(x_{t-\nu}\right)\right]^2 + 2\left[(\mu-\hat{\mu}) + g\left(x_{t-\nu}\right)\right]u_t + u_t^2\right\}$$

$$= \frac{1}{n\kappa_g(\sqrt{d_n})^2}\left\{\sum_{t=1+\nu}^{n}\left(\mu-\hat{\mu}\right)^2 + \sum_{t=1+\nu}^{n} g^2\left(x_{t-\nu}\right) + 2\left(\mu-\hat{\mu}\right)\sum_{t=1+\nu}^{n} g\left(x_{t-\nu}\right) + \sum_{t=1+\nu}^{n} u_t^2\right\} + o_p(1)$$

$$= \frac{\left(\mu-\hat{\mu}\right)^2}{\kappa_g(d_n)^2} + \frac{1}{n\kappa_g(d_n)^2}\sum_{t=1+\nu}^{n} g^2\left(x_{t-\nu}\right) + 2\left(\mu-\hat{\mu}\right)\frac{1}{n\kappa_g(d_n)^2}\sum_{t=1+\nu}^{n} g\left(x_{t-\nu}\right)$$
$$+ \frac{1}{n\kappa_g(d_n)^2}\sum_{t=1+\nu}^{n} u_t^2 + o_p(1)$$

$$= \begin{cases} \left[\int_0^1 H_g(G(s))ds\right]^2 + \int_0^1 H_g(G(s))^2 ds - 2\left[\int_0^1 H_g(G(s))ds\right]^2 + \sigma_u^2 + o_p(1), \ \kappa_g(\lambda) = 1 \\ \left[\int_0^1 H_g(G(s))ds\right]^2 + \int_0^1 H_g(G(s))^2 ds - 2\left[\int_0^1 H_g(G(s))ds\right]^2 + o_p(1), \ \lim_{\lambda\to\infty}\kappa_g(\lambda) = \infty \\ \frac{1}{\kappa_g(d_n)^2}\sigma_u^2 + \sigma_*^2 + o_p(1/\kappa_g(d_n)^2) + o_p(1), \ \lim_{\lambda\to\infty}\kappa_g(\lambda) = 0 \end{cases}$$

---

[19]Note that the above postulates that $\kappa_g(d_n)^{-1}(\hat{\mu} - \mu) = \int_0^1 H_g(G(s))ds + o_p(1)$ for $\lim_{\lambda\to\infty}\kappa_g(\lambda) = 1$ or ∞. Further, $(\hat{\mu}-\mu) = o_p(1)$ for g H-regular with $\lim_{\lambda\to\infty}\kappa_g(\lambda) = 0$ or g integrable.

$$= \begin{cases} \overbrace{\int_0^1 H_g(G(s))^2 ds - \left[\int_0^1 H_g(G(s))ds\right]^2}^{\sigma_*^2} + \sigma_u^2, \ \kappa_g(\lambda) = 1 \\ \int_0^1 H_g(G(s))^2 ds - \left[\int_0^1 H_g(G(s))ds\right]^2, \ \lim_{\lambda\to\infty} \kappa_g(\lambda) = \infty \\ \frac{1}{\kappa_g(d_n)^2}\sigma_u^2 + \sigma_*^2 + o_p(1/\kappa_g(d_n)^2), \ \lim_{\lambda\to\infty} \kappa_g(\lambda) = 0 \end{cases}$$

$$= \frac{1}{\kappa_g(d_n)^2}\sigma_u^2 + \sigma_*^2 + o_p(1/\kappa_g(d_n)^2) + o_p(1).$$

Moreover, for g integrable we have $\hat{\sigma}_u^2 = \sigma_u^2 + o_p(1)$.[20]

Hence, in view of the above and WP (Theorem 2.1) we have

$$\left(\frac{d_n}{h_n n}\right)^{1/2} \hat{t}(x,\hat{\mu}) = \left(\frac{d_n}{h_n n}\right)^{1/2} \left(\frac{\sum_{t=1+\nu}^n K\left(\frac{x_{t-\nu}-x}{h_n}\right)}{\hat{\sigma}_u^2 \int_{-\infty}^{\infty} K(\lambda)^2 d\lambda}\right)^{1/2} \left(\hat{f}(x) - \hat{\mu}\right)$$

$$= \left(\frac{d_n}{h_n n}\right)^{1/2} \left(\frac{\sum_{t=1+\nu}^n K\left(\frac{x_{t-\nu}-x}{h_n}\right)}{\hat{\sigma}_u^2 \int_{-\infty}^{\infty} K(\lambda)^2 d\lambda}\right)^{1/2} \left(\hat{f}(x) - (\mu + g(x))\right)$$

$$+ \left(\frac{d_n}{h_n n}\right)^{1/2} \left(\frac{\sum_{t=1+\nu}^n K\left(\frac{x_{t-\nu}-x}{h_n}\right)}{\hat{\sigma}_u^2 \int_{-\infty}^{\infty} K(\lambda)^2 d\lambda}\right)^{1/2} (g(x) + \mu - \hat{\mu})$$

$$= \begin{cases} \left(\frac{L_G(1,0)\int_{-\infty}^{\infty} K(s)ds}{(\sigma_*^2+\sigma_u^2)\int_{-\infty}^{\infty} K(s)^2 ds}\right)^{1/2} \left[g(x) - \int_0^1 H_g(G(s))ds\right] + o_p(1), \ \kappa_g(\lambda) = 1 \\ -\left(\frac{L_G(1,0)\int_{-\infty}^{\infty} K(s)ds}{\sigma_*^2\int_{-\infty}^{\infty} K(s)^2 ds}\right)^{1/2} \int_0^1 H_g(G(s))ds + o_p(1), \ \lim_{\lambda\to\infty} \kappa_g(\lambda) = \infty \\ \left(\frac{L_G(1,0)\int_{-\infty}^{\infty} K(s)ds}{\sigma_u^2\int_{-\infty}^{\infty} K(s)^2 ds}\right)^{1/2} g(x) + o_p(1), \ g \text{ integrable or } g \text{ H-regular with } \lim_{\lambda\to\infty} \kappa_g(\lambda) = 0 \end{cases}.$$

The result follows easily from the above and the fact that $\hat{F}(x,\hat{\mu}) = \hat{t}(x,\hat{\mu})^2$. ■

9 Appendix B: power rates of parametric tests

Proof of Theorem 3. The proof is organised in three parts. We first derive the limit properties of the parametric estimators $\hat{a}$ and $\hat{\beta}$, under functional form misspecification. Subsequently, we obtain the limit properties of the variance estimators $\hat{\Omega}_{uu}, \hat{\Omega}_{vv}$ and $\hat{\Omega}_{vu}$. Finally, we analyse the test statistics $\hat{t}_{FM}$ and $\hat{\mathcal{R}}_\beta$ under H_1 when functional form misspecification is committed.

[20] The above postulates that $\kappa_g(d_n)^{-2}\hat{\sigma}_u^2 = \kappa_g(d_n)^{-2}\sigma_u^2 + \sigma_*^2 + o_p(1)$ for $\lim_{\lambda\to\infty} \kappa_g(\lambda) = 1$ or ∞. Further, $\hat{\sigma}_u^2 = \sigma_u^2 + o_p(1)$ for g H-regular with $\lim_{\lambda\to\infty} \kappa_g(\lambda) = 0$ or g integrable.

Limit behaviour of OLS estimators:

Case I (H-regular $g(\lambda x) \approx \kappa_g(\lambda) H_g(x)$)

$$\frac{\sqrt{n}}{\kappa_g(\sqrt{n})}\hat{\beta} = \frac{\frac{1}{\kappa_g(\sqrt{n})}\left\{\sum_t y_t x_{t-\nu} - \frac{1}{n}\sum_t y_t \sum_t x_{t-\nu}\right\}}{\frac{1}{\sqrt{n}}\left\{\sum_t x_{t-\nu}^2 - \frac{1}{n}\left(\sum_t x_{t-\nu}\right)^2\right\}} = \frac{\frac{1}{\kappa_g(\sqrt{n})n^{3/2}}\left\{\sum_t y_t x_{t-\nu} - \frac{1}{n}\sum_t y_t \sum_t x_{t-\nu}\right\}}{\frac{1}{\sqrt{n}n^{3/2}}\left\{\sum_t x_{t-\nu}^2 - \frac{1}{n}\left(\sum_t x_{t-\nu}\right)^2\right\}}$$

$$= \frac{\frac{1}{\kappa_g(\sqrt{n})n^{3/2}}\sum_t H_g(x_{t-\nu})x_{t-\nu} - \frac{1}{n\kappa_g(\sqrt{n})n^{3/2}}\sum_t H_g(x_{t-\nu})\sum_t x_{t-\nu}}{\frac{1}{\sqrt{n}n^{3/2}}\left\{\sum_t x_{t-\nu}^2 - \frac{1}{n}\left(\sum_t x_{t-\nu}\right)^2\right\}} + o_p(1)$$

$$= \frac{\frac{1}{\kappa_g(\sqrt{n})n^{3/2}}\sum_t H_g(x_{t-\nu})x_{t-\nu} - \frac{1}{n\kappa_g(\sqrt{n})n^{3/2}}\sum_t H_g(x_{t-\nu})\sum_t x_{t-\nu}}{\frac{1}{n^2}\sum_t x_{t-\nu}^2 - \left(\frac{1}{n^{3/2}}\sum_t x_{t-\nu}\right)^2}$$

$$\xrightarrow{p} \frac{\int_0^1 H_g(G)G - \left(\int_0^1 H_g(G)\right)\left(\int_0^1 G\right)}{\int_0^1 G^2 - \left(\int_0^1 G\right)^2} =: \beta_*$$

Hence,

$$\hat{\beta} \approx \frac{\kappa_g(\sqrt{n})}{\sqrt{n}}\beta_*. \tag{46}$$

Similarly,

$$\frac{1}{\kappa_g(\sqrt{n})}\hat{a} = \frac{1}{\kappa_g(\sqrt{n})}\left(\bar{y} - \hat{\beta}\bar{x}\right) = \frac{1}{\kappa_g(\sqrt{n})}\left(\frac{1}{n}\sum_t y_t - \hat{\beta}\frac{1}{n}\sum_t x_{t-\nu}\right) =$$

$$= \frac{1}{\kappa_g(\sqrt{n})}\left(\frac{1}{n}\sum_t H_g(x_{t-\nu}) - \left(\frac{\sqrt{n}}{\kappa_g(\sqrt{n})}\hat{\beta}\right)\frac{\kappa_g(\sqrt{n})}{\sqrt{n}n}\sum_t x_{t-\nu}\right) + o_p(1)$$

$$= \left(\frac{1}{n\kappa_g(\sqrt{n})}\sum_t H_g(x_{t-\nu}) - \beta_*\frac{1}{n^{3/2}}\sum_t x_{t-\nu}\right) + o_p(1) \xrightarrow{p} \int_0^1 H_g(G) - \beta_*\left(\int_0^1 G\right) =: a_*.$$

Hence,

$$\hat{a} \approx \kappa_g(\sqrt{n})a_*. \tag{47}$$

Case II (I-regular $g(x)$):

$$n\hat{\beta} = n\frac{\sum_t y_t x_{t-\nu} - \frac{1}{n}\sum_t y_t \sum_t x_{t-\nu}}{\sum_t x_{t-\nu}^2 - \frac{1}{n}\left(\sum_t x_{t-\nu}\right)^2}$$

$$= \frac{\frac{1}{n}\left[\sum_t u_t x_{t-\nu} - \frac{1}{n}\sum_t (g_t + u_t)\sum_t x_{t-\nu}\right]}{\frac{1}{n^2}\left[\sum_t x_{t-\nu}^2 - \frac{1}{n}\left(\sum_t x_{t-\nu}\right)^2\right]} + o_p(1) = \frac{\left[\frac{1}{n}\sum_t u_t x_{t-\nu} - \frac{1}{\sqrt{n}}\sum_t (g_t + u_t)\frac{1}{n^{3/2}}\sum_t x_{t-\nu}\right]}{\left[\frac{1}{n^2}\sum_t x_{t-\nu}^2 - \left(\frac{1}{n^{3/2}}\sum_t x_{t-\nu}\right)^2\right]}$$

$$= \frac{\int_0^1 GdB_u - \left(\int_{-\infty}^{\infty} g(s)dsL_G + B_u(1)\right)\left(\int_0^1 G\right)}{\int_0^1 G^2 - \left(\int_0^1 G\right)^2} =: \beta_{**},$$

where as before B_u is the BM limit of the partial sum of $n^{-1/2}u_t$. Hence,

$$\hat{\beta} \approx \frac{1}{n}\beta_{**}. \tag{48}$$

Next,

$$\sqrt{n}\hat{a} = \sqrt{n}\left(\bar{y} - \hat{\beta}\bar{x}\right) = \sqrt{n}\left(\frac{1}{n}\sum_t y_t - \hat{\beta}\frac{1}{n}\sum_t x_{t-\nu}\right) = \sqrt{n}\left(\frac{1}{n}\sum_t y_t - \left(n\hat{\beta}\right)\frac{1}{n^2}\sum_t x_{t-\nu}\right)$$

$$= \frac{1}{\sqrt{n}}\sum_t y_t - \beta_* \frac{1}{n^{3/2}}\sum_t x_{t-\nu} + o_p(1) \xrightarrow{p} \left(\int_{-\infty}^{\infty} g(s)dsL_G + B_u(1)\right) - \beta_*\left(\int_0^1 G\right) =: a_{**}$$

Hence,

$$\hat{a} \approx \frac{1}{\sqrt{n}}a_{**}. \tag{49}$$

Limit behaviour of variance estimators and parametric tests:

Case I (H-regular $g(\lambda x) \approx \kappa_g(\lambda)H_g(x)$): Consider first $\hat{\rho} := \left[\sum_{t=2}^n x_{t-1}^2\right]^{-1}\sum_{t=2}^n x_{t-1}x_t$. Then

$$n\left(\hat{\rho} - \rho_n\right) = \left[\frac{1}{n^2}\sum_{t=2}^n x_{t-1}^2\right]^{-1}\frac{1}{n}\sum_{t=2}^n x_{t-1}v_t \xrightarrow{p} \left[\int_0^1 G(r)^2 dr\right]^{-1}\int_0^1 G(r)dV(r) =: \gamma_*.$$

Next,

$$\hat{\Omega}_{vu} = \frac{1}{n}\sum_{t=1+\nu}^n \hat{v}_t\hat{u}_t = \frac{1}{n}\sum_{t=1+\nu}^n \left[\left(\hat{\rho} - \rho_n\right)x_{t-1} + v_t\right]\left[y_t - \hat{a} - \hat{\beta}x_{t-\nu}\right]$$

$$= \frac{1}{n}\sum_{t=1+\nu}^n \left[\left(\hat{\rho} - \rho_n\right)x_{t-1} + v_t\right]\left[g\left(x_{t-\nu}\right) + u_t - \hat{a} - \hat{\beta}x_{t-\nu}\right]$$

$$= \left\{n\left(\hat{\rho} - \rho_n\right)\frac{1}{n^2}\sum_{t=1+\nu}^n x_{t-1}g\left(x_{t-\nu}\right) + n\left(\hat{\rho} - \rho\right)\frac{1}{n^2}\sum_{t=1+\nu}^n x_{t-1}u_t\right.$$

$$\left. - \hat{a}n\left(\hat{\rho} - \rho\right)\frac{1}{n^2}\sum_{t=1+\nu}^n x_{t-1} - \hat{\beta}n\left(\hat{\rho} - \rho_n\right)\frac{1}{n^2}\sum_{t=1+\nu}^n x_{t-1}x_{t-\nu}\right\}$$

$$+ \frac{1}{n}\sum_{t=1+\nu}^n \left\{g\left(x_{t-\nu}\right)v_t + v_tu_t - \hat{a}v_t - \hat{\beta}x_{t-\nu}v_t\right\}$$

Then using (46), (47) and the limit results of Park and Phillips (2001) we have :

(i) for $\sqrt{n}/\kappa_g(\sqrt{n}) \to 0$

$$\frac{\sqrt{n}}{\kappa_g(\sqrt{n})}\hat{\Omega}_{vu} = \gamma_* \int_0^1 [G(r)\{H_g(G(r)) - a_* - \beta_* G(r)\}]\, dr$$

$$+ \int_0^1 [H_g(G(r)) - a_* - \beta_* G(r)]\, dV(r) \Big\} + o_p(1).$$

(ii) for $\sqrt{n}/\kappa_g(\sqrt{n}) \to 1$

$$\hat{\Omega}_{vu} = \gamma_* \int_0^1 [G(r)\{H_g(G(r)) - a_* - \beta_* G(r)\}]\, dr$$

$$+ \int_0^1 [H_g(G(r)) - a_* - \beta_* G(r)]\, dV(r) \Big\} + \Omega_{vu} + o_p(1).$$

(iii) for $\sqrt{n}/\kappa_g(\sqrt{n}) \to \infty$

$$\hat{\Omega}_{vu} = \Omega_{vu} + o_p(1).$$

Next, consider

$$\hat{\Omega}_{uu} = \frac{1}{n}\sum_{t=1+\nu}^{n} \hat{u}_t^2 = \frac{1}{n}\sum_{t=1+\nu}^{n} \left[H_g(x_{t-\nu}) + u_t - \hat{a} - \hat{\beta} x_{t-\nu}\right]^2$$

Using (46), (47) and the limit results of Park and Phillips (2001) we have

(i) for $\kappa_g(\sqrt{n}) \to \infty$

$$\frac{1}{\kappa_g(\sqrt{n})^2}\hat{\Omega}_{uu} = \int_0^1 [H_g(G(r)) - a_* - \beta_* G(r)]^2\, dr + o_p(1),$$

(ii) for $\kappa_g(\sqrt{n}) = 1$

$$\hat{\Omega}_{uu} = \int_0^1 [H_g(G(r)) - a_* - \beta_* G(r)]^2\, dr + \Omega_{uu} + o_p(1)$$

(iii) for $\kappa_g(\sqrt{n}) \to 0$

$$\hat{\Omega}_{uu} = \Omega_{uu} + o_p(1).$$

Next, let $\hat{\Omega}^+ = \hat{\Omega}_{uu} - \hat{\Omega}_{uv}\hat{\Omega}_{vv}^{-1}\hat{\Omega}_{vu}$. As $n \to \infty$ we get

(i) For $\kappa_g(\sqrt{n}) \to \infty$, $\frac{\kappa_g(\sqrt{n})}{\sqrt{n}} \to \infty$

$$\frac{1}{\kappa_g(\sqrt{n})^2}\hat{\Omega}^+ = \frac{\hat{\Omega}_{uu}}{\kappa_g(\sqrt{n})^2} - \hat{\Omega}_{vv}^{-1}\frac{\hat{\Omega}_{vu}^2}{\kappa_g(\sqrt{n})^2}$$

$$= \Omega^*_{uu} - \frac{1}{n}\hat{\Omega}^{-1}_{vv}\frac{n\hat{\Omega}^2_{vu}}{\kappa_g(\sqrt{n})^2} = \Omega^*_{uu} + O_p\left(\frac{1}{n}\right) = \Omega^*_{uu} + o_p(1).$$

(ii) For $\kappa_g(\sqrt{n}) \to \infty$, $\frac{\kappa_g(\sqrt{n})}{\sqrt{n}} = O(1)$

$$\frac{1}{\kappa_g(\sqrt{n})^2}\hat{\Omega}^+ = \frac{\hat{\Omega}_{uu}}{\kappa_g(\sqrt{n})^2} - \hat{\Omega}^{-1}_{vv}\frac{\hat{\Omega}^2_{vu}}{\kappa_g(\sqrt{n})^2}$$

$$= \Omega^*_{uu} + O_p\left(\frac{1}{\kappa_g(\sqrt{n})^2}\right) = \Omega^*_{uu} + o_p(1).$$

(iii) For $\kappa_g(\sqrt{n}) = O(1)$, (in this case we necessarily have $\frac{\kappa_g(\sqrt{n})}{\sqrt{n}} = o(1)$)

$$\hat{\Omega}^+ = \hat{\Omega}_{uu} - \hat{\Omega}^{-1}_{vv}\hat{\Omega}^2_{vu} = \Omega^{**}_{uu} + \Omega^{-1}_{vv}\Omega^2_{vu} + o_p(1).$$

Next, consider the FMLS t-statistic:

$$\frac{1}{\sqrt{n}}\hat{t}_{IV} = \frac{\frac{1}{\sqrt{n}}\tilde{\beta}}{\sqrt{\hat{\Omega}^+\left\{\sum_t x^2_{t-\nu} - \frac{1}{n}\left(\sum_t x_{t-\nu}\right)^2\right\}^{-1}}} = \frac{\frac{1}{n^{3/2}}\left\{\sum_t y^+_t x_{t-\nu} - \frac{1}{n}\sum_t y^+_t \sum_t x_{t-\nu}\right\}}{\sqrt{\hat{\Omega}^+\frac{1}{n^2}\left\{\sum_t x^2_{t-\nu} - \frac{1}{n}\left(\sum_t x_{t-\nu}\right)^2\right\}}}.$$

Hence, for $\kappa_g(\sqrt{n}) \to \infty$ we have

$$\frac{1}{\sqrt{n}}\hat{t}_{IV} = \frac{\frac{1}{\kappa_g(\sqrt{n})n^{3/2}}\left\{\sum_t H_g(x_{t-\nu})x_{t-\nu} - \frac{1}{n}\sum_t H_g(x_{t-\nu})\sum_t x_{t-\nu}\right\}}{\sqrt{\frac{\hat{\Omega}^+}{\kappa_g(\sqrt{n})^2}\frac{1}{n^2}\left\{\sum_t x^2_{t-\nu} - \frac{1}{n}\left(\sum_t x_{t-\nu}\right)^2\right\}}} + o_p(1)$$

$$= \frac{\left\{\int_0^1 H_g(G(r))G(r)dr - \int_0^1 H_g(G(r))dr\int_0^1 G(r)dr\right\}}{\sqrt{\Omega^*_{uu}\left\{\int_0^1 G(r)^2dr - \left[\int_0^1 G(r)dr\right]^2\right\}}} + o_p(1).$$

For $\kappa_g(\sqrt{n}) = O(1)$ we have

$$\frac{1}{\kappa_g(\sqrt{n})\sqrt{n}}\hat{t}_{IV} = \frac{\frac{1}{\kappa_g(\sqrt{n})n^{3/2}}\left\{\sum_t H_g(x_{t-\nu})x_{t-\nu} - \frac{1}{n}\sum_t H_g(x_{t-\nu})\sum_t x_{t-\nu}\right\}}{\sqrt{\hat{\Omega}^+\frac{1}{n^2}\left\{\sum_t x^2_{t-\nu} - \frac{1}{n}\left(\sum_t x_{t-\nu}\right)^2\right\}}} + o_p(1)$$

$$= \frac{\left\{\int_0^1 H_g(G(r))G(r)dr - \int_0^1 H_g(G(r))dr\int_0^1 G(r)dr\right\}}{\sqrt{\left(\Omega^{**}_{uu} + \Omega^{-1}_{vv}\Omega^2_{vu}\right)\left\{\int_0^1 G(r)^2dr - \left[\int_0^1 G(r)dr\right]^2\right\}}} + o_p(1)$$

Similarly, for $\kappa_g(\sqrt{n}) \to \infty$ the $\hat{R}_\beta$ statistic

$$\frac{1}{\sqrt{n}}\hat{R}_\beta = \frac{1}{\sqrt{\hat{\Omega}_{vv}\hat{\Omega}^+/\kappa_g(\sqrt{n})^2}}\left\{\frac{1}{n^{3/2}\kappa_g(\sqrt{n})}\sum_t\left(x_{t-\nu}-\frac{1}{n}\sum_t x_{t-\nu}\right)\left[y_t^+ - \hat{\beta}x_{t-\nu}\right]\right\}$$

$$= \frac{1}{\sqrt{\Omega_{vv}\Omega_*^+}}\left\{\frac{1}{n^{3/2}\kappa_g(\sqrt{n})}\sum_t\left(x_{t-\nu}-\frac{1}{n}\sum_t x_{t-\nu}\right)\left[y_t^+ - \hat{\beta}x_{t-\nu}\right]\right\} + o_p(1)$$

$$= \frac{1}{\sqrt{\Omega_{vv}\Omega_*^+}}\left\{\frac{1}{n^{3/2}\kappa_g(\sqrt{n})}\sum_t\left(x_{t-\nu}-\frac{1}{n}\sum_t x_{t-\nu}\right)\left[H_g(x_{t-\nu}) - \hat{\beta}x_{t-\nu}\right]\right\} + o_p(1)$$

$$= \frac{1}{\sqrt{\Omega_{vv}\Omega_*^+}}\int_0^1\left\{\left(G(r)-\int_0^1 G(s)ds\right)\left[H_g(G(r)) - \beta_* G(r)\right]\right\}dr + o_p(1)$$

The proof for $\kappa_g(\sqrt{n}) = O(1)$ is similar and therefore omitted.

Case II (g integrable): Using the limit theory of Park and Phillips (2001) or Wang and Phillips (2009) and in view of (48) and (49) it can be shown that

$$\hat{\Omega}_{vu} = \Omega_{vu} + o_p(1) \text{ and } \hat{\Omega}_{uu} = \Omega_{uu} + o_p(1).$$

In view of the above, standard arguments show that the $\hat{t}_{IV}$ test statistic is

$$\hat{t}_{IV} = \frac{\tilde{\beta}}{\sqrt{\hat{\Omega}^+\left\{\sum_t x_{t-\nu}^2 - \frac{1}{n}\left(\sum_t x_{t-\nu}^2\right)^2\right\}^{-1}}} = \frac{\frac{1}{n}\left\{\sum_t y_t^+ x_{t-\nu} - \frac{1}{n}\sum_t y_t^+\sum_t x_{t-\nu}\right\}}{\sqrt{\hat{\Omega}^+\left[\frac{1}{n^2}\left\{\sum_t x_{t-\nu}^2 - \frac{1}{n}\left(\sum_t x_{t-\nu}\right)^2\right\}\right]}}$$

$$= \frac{\frac{1}{n}\left\{\sum_t u_t^+ x_{t-\nu} - \frac{1}{n}\sum_t u_t^+\sum_t x_{t-\nu}\right\}}{\sqrt{\hat{\Omega}^+\left[\frac{1}{n^2}\left\{\sum_t x_{t-\nu}^2 - \frac{1}{n}\left(\sum_t x_{t-\nu}\right)^2\right\}\right]}} + o_p(1)$$

$$= \frac{\int_0^1\left[G(r)-\left(\int_0^1 G(s)ds\right)\right]d\left\{B_u(r) - V(r)\Omega_{vv}^{-1}\Omega_{vu}\right\} - c\int_0^1\left[G(r)-\left(\int_0^1 G(s)ds\right)\right]^2 dr\Omega_{vv}^{-1}\Omega_{vu}}{\left\{\Omega^+\int_0^1\left[G(r)-\left(\int_0^1 G(s)ds\right)\right]^2 dr\right\}^{1/2}} + o_p(1)$$

$$= \frac{1}{(\Omega^+)^{1/2}}\left[\left\{B_u(1) - V(1)\Omega_{vv}^{-1}\Omega_{vu}\right\} - c\Omega_{vv}^{-1}\Omega_{vu}\left\{\int_0^1\left[G(r)-\left(\int_0^1 G(s)ds\right)\right]^2 dr\right\}^{1/2}\right].$$

Further, the $\hat{R}_\beta$ statistic is asymptotically

$$\hat{R}_\beta = \frac{1}{\sqrt{\hat{\Omega}_{vv}\hat{\Omega}^+}}\left\{\frac{1}{n}\sum_t\left(x_{t-\nu}-\frac{1}{n}\sum_t x_{t-\nu}\right)\left[y_t^+ - \hat{\beta}x_{t-\nu}\right]\right\}$$

$$= \frac{1}{\sqrt{\Omega_{vv}\Omega^{+}}} \left\{ \int_0^1 \left[G(r) - \left(\int_0^1 G(s)ds \right) \right] d\left[B_u(r) - V(r)\Omega_{vv}^{-1}\Omega_{vu} \right] \right.$$

$$\left. - \left(c\Omega_{vv}^{-1}\Omega_{vu} + \beta_* \right) \int_0^1 \left[G(r) - \left(\int_0^1 G(s)ds \right) \right]^2 dr \right\} + o_p(1).$$

■

10 References

Berenguer-Rico V. and J. Gonzalo (2011), Summability of stochastic processes: a generalization of integration and co-integration valid for non-linear processes. Mimeo.

Berkes I, and L. Horváth (2006), Convergence of integral functionals of stochastic processes. *Econometric Theory*, 2, 304 - 322.

Bollerslev, T., Tauchen, G. and H. Zhou (2009), Expected stock returns and variance risk premia. *Review of Financial Studies*, 22, 4463-4492.

Campbell, J. Y. and M. Yogo (2006). Efficient tests of stock return predictability. *Journal of Financial Economics*, 81, 27–60.

Chan, N.H. and C.Z. Wei (1987). Asymptotic inference for nearly nonstationary AR(l) processes. *Annals of Statistics*, 15, 1050-1063.

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

de Jong, R. and H. Bierens (1994). On the limit behavior of a chi-square type test if the number of conditional moments tested approaches infinity. *Econometric Theory*, 10, 70-90.

Elliott, G. (1998). On the robustness of cointegration methods when regressors almost have unit roots. *Econometrica*, 66, 149-158.

Feller, W. (1971). *An introduction to probability theory and its applications*. Vol. II, Wiley Series in Probability and Mathematical Statistics.

Ghysels E., P. Santa-Clara and R. Valkanov (2005). There is a risk-return trade-off afterall. *Journal of Financial Economics*, 76, 509-548.

Gonzalo, J. and J-Y. Pitarakis (2012). Regime Specific Predictability in Predictive Regressions, *Journal of Business and Economic Statistics*, 30, 229-241.

Granger, C. W. J. (1995). Modelling nonlinear relationships between extended-memory variables, *Econometrica*, 63, 265-279.

Hualde, J and P.M. Robinson (2010). Semiparametric inference in multivariate fractionally cointegrated systems. *Journal of Econometrics*, 157, 492-511.

Jansson, M. and M. J. Moreira (2006). Optimal inference in regression models with nearly integrated time series. *Econometrica*, 74, 681–714.

Jaganathan, P. (2008). Limit theorems for functionals of sums that converge to fractional Brownian and stable motions. Cowles Foundation Discussion Paper No. 1649

Kasparis, I. (2008). Detection of functional form misspecification in cointegrating relationships. *Econometric Theory*, 24, 1373-1403.

Kasparis, I. (2010) The Bierens test for certain nonstationary models. *Journal of Econometrics*, 158, 221-230.

Kasparis, I. and P. C. B. Phillips (2012). Dynamic misspecification in nonparametric cointegrating regression. *Journal of Econometrics*, 168, 270-284.

Kasparis, I., Phillips P. C. B. and T. Magdalinos (2012). Non-linearity Induced Weak Instrumentation. *Econometric Reviews*, accepted.

Komlós, J., Major, P., and Tusnády, G. (1976). An approximation of partial sums of independent R.V.'s and sample DF. II. *Z. Wahrsch. verw. Gebiete* 34, 33–58.

Lewellen, J. (2004). Predicting returns with financial ratios. *Journal of Financial Economics* 74, 209–235.

Li, Q. and J. S. Racine (2007). *Nonparametric Econometrics: Theory and Practice.* Princeton University Press.

Magdalinos, T. and P.C.B. Phillips (2009). Inference in the vicinity of the unity. Mimeo

Marmer, V., (2007) Nonlinearity, nonstationarity and spurious forecasts. *Journal of Econometrics*, 142, 1-27.

Marmol, F. and C. Velasco, (2004). Consistent testing of cointegrating relationships, *Econometrica*, 72, 1809-1844.

Oodaira, H. (1976). Some limit theorems for the maximum of normalized sums of weakly dependent random variables. In Maruyama, G., and Prokhorov, J. V. (eds.), Proceedings of the Third Japan-USSR Symposium on Probability Theory, Lecture Notes in Math.,Vol. 550, Springer, Berlin, Heidelberg, New York, pp. 467–474.

Park, J.Y. and P.C.B. Phillips (1998). Unit roots in nonlinear transformations of integrated time series. Mimeo.

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

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

Phillips P. C. B. (1987) Time Series Regression with a Unit Root. *Econometrica*, 55, 277-301.

Phillips P. C. B. (1987) Towards a Unified Asymptotic Theory for Autoregression. *Biometrika*, 74, 535-547.

Phillips P. C. B. (1995). Fully Modified Least Squares and Vector Autoregression. *Econometrica*, 63, 1023-1078.

Phillips P. C. B. (2007). Unit root log periodogram regression. *Journal of Econometrics,* 138, 104–124.

Phillips P. C. B. (2012). On Confidence Intervals for Autoregressive Roots and Predictive Regression. Unpublished Paper, Yale University.

Phillips P. C. B. and B. Hansen (1990). Statistical Inference in Instrumental Variables, Regression with I(1) Processes. *Review of Economic Studies*, 57, 99-125.

Phillips P. C. B. and J. H. Lee (2012) Predictive regression under various degrees of persistence and robust long-horizon regression", *Journal of Econometrics*, forthcoming.

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

Robinson, P.M. (1983). Nonparametric estimators for time series. *Journal of Time Series Analysis*, 4, 185-207.

Robinson, P.M. (1995). Gaussian semiparametric estimation of long rangedependence, *Annals of Statistics*, 23, 1630-1661.

Robinson, P.M. and J. Hualde (2003). Cointegration in fractional systems with unknown integration orders. *Econometrica*, 71, 1727-1766.

Torous, W., Valkanov, R., and Yan, S., On predicting stock returns with nearly integrated explanatory variables. *Journal of Business*, 77, 937-966.

Wang, Q. (2013). Martingale limit theorem revisited and nonlinear cointegration. Mimeo.

Wang, Q., Lin, Y.-X., Gulati, C. M. (2003). Strong Approximation for Long Memory Processes with Applications. *Journal of Theoretical Probability,* 16, 377–389.

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

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

Wang, Q. and P.C.B. Phillips (2012). A specification test for nonlinear nonstationary models. *Annals of Statistics,* 40, 727–758.

Welch I. and A. Goyal (2008). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, 21, 1455-1508

Wright J. H. (2000) Confidence sets for cointegrating coefficients based on stationarity tests. *Journal of Business & Economic Statistics*, 18, 211-222.

11 Simulations results

Size (5%): $n = 500$, $\rho_x = 0$ (No HAC estimation used in the JM statistic)

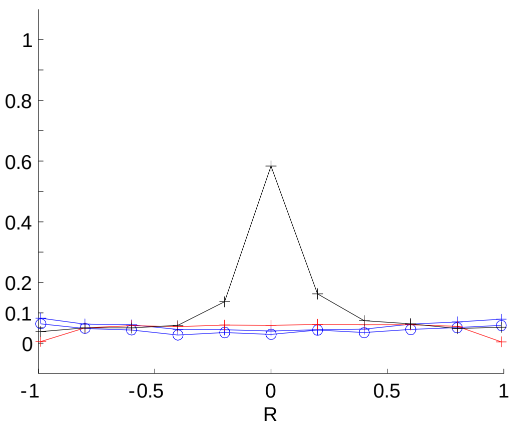


Fig. 1(a) $c = 0$

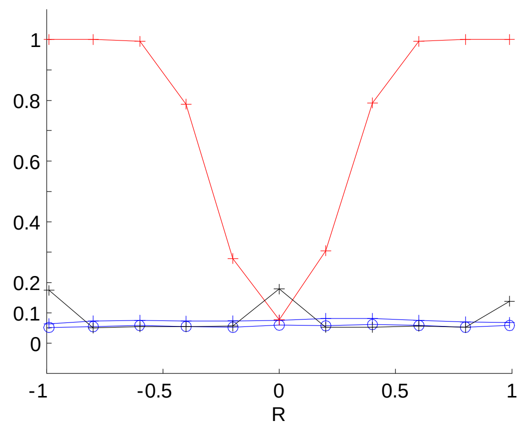


Fig. 1(b) $c = -50$

Size (5%): $n = 500$, $\rho_x = 0$ (No HAC estimation used in the JM statistic)

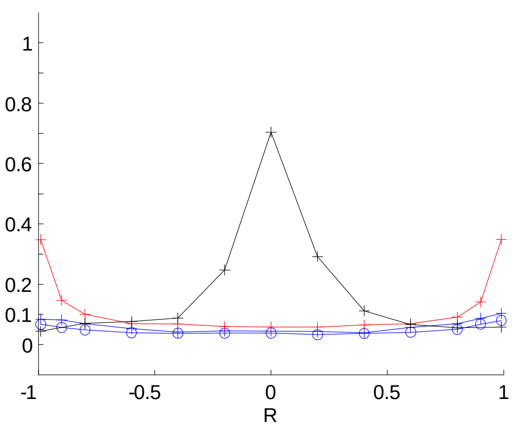


Fig. 1(c) $c = 1$

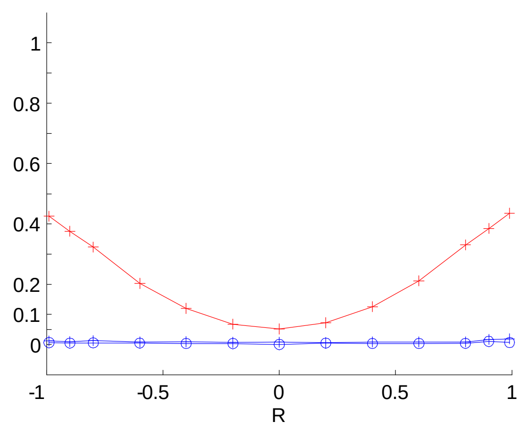
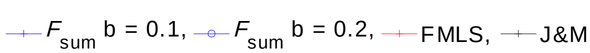


Fig. 1(d) $c = 5$

Note: No simulations results were obtained for JM for $c = 5$.



Size (5%): $n = 500$, $\rho_x = 0$ (No HAC estimation used in the JM statistic)

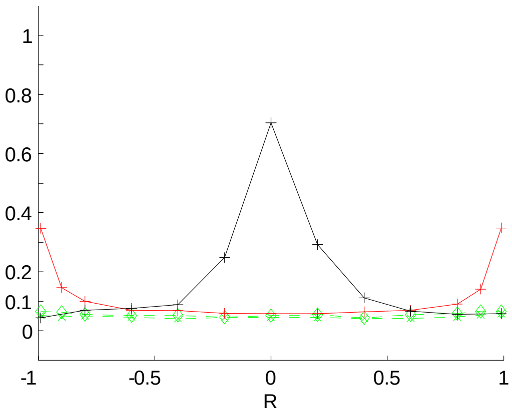


Fig. 1(e) $c = 1$

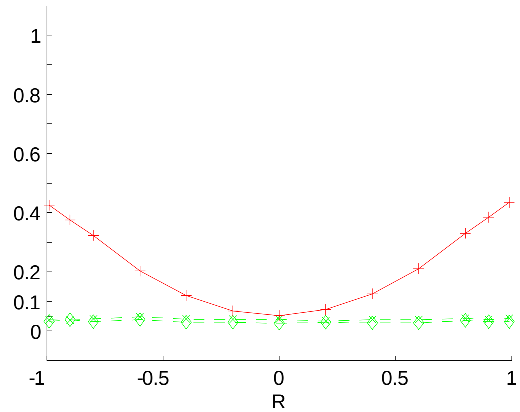


Fig. 1(f) $c = 5$

Note: No simulations results were obtained for JM for $c = 5$.

Size (5%): $n = 500$, $\rho_x = 0.3$ (HAC estimation used in the JM statistic)

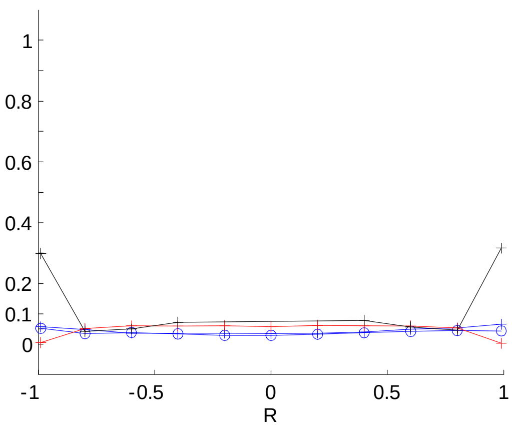


Fig. 1(g) $c = 0$

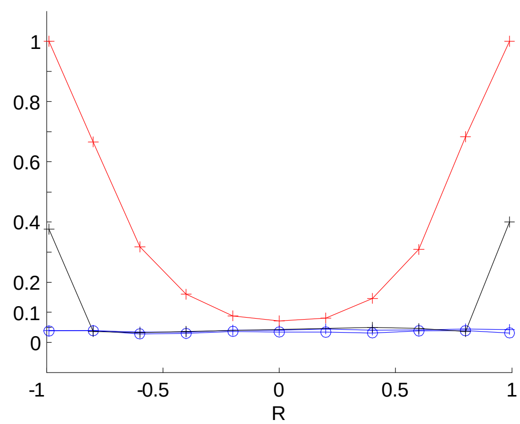
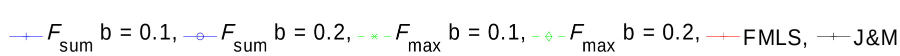


Fig. 1(h) $c = -10$

Note: No simulations results were obtained for JM when $R = -0.2, 0, 0.2$ and the lines shown are interpolated over this interval using results for $R = \pm.4$.



Size (5%): $n = 500$, $\rho_x = 0.3$ (HAC estimation used in the JM statistic)

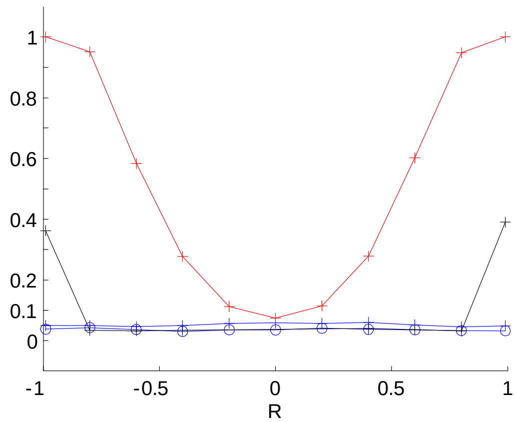


Fig. 1(i) $c = -20$

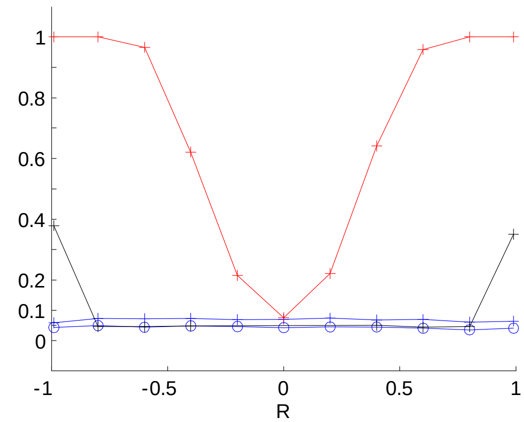


Fig. 1(j) $c = -50$

Note: No simulation results were obtained for JM when $R = -0.2, 0, 0.2$ and the lines shown are interpolated over this interval using results for $R = \pm .4$.

Size (5%) $n = 500$, $\Delta x_t \sim ARFIMA(d)$ (no HAC estimation used in JM statistic)

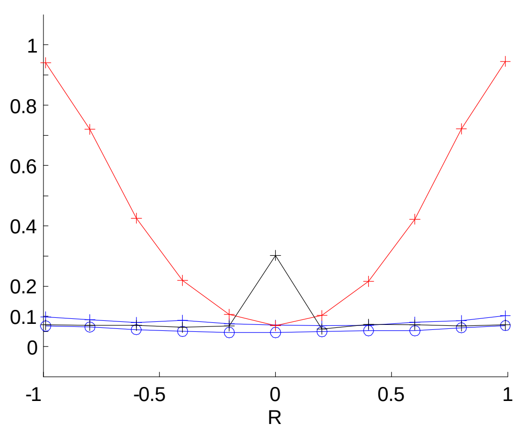


Fig. 2(a) $d = -0.25$

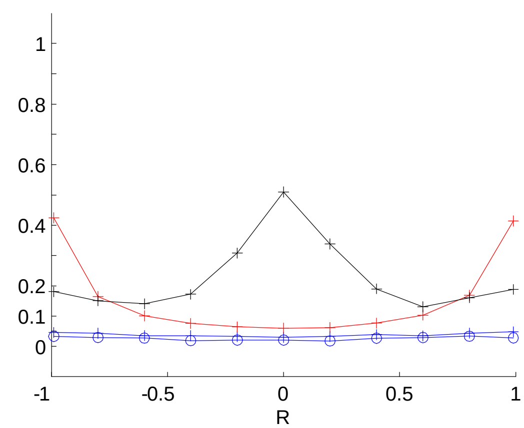


Fig. 2(b) $d = 0.25$

F_{sum} b = 0.1, F_{sum} b = 0.2, FMLS, J&M

Size (5%): $n = 500$, $\Delta x_t \sim ARFIMA(d)$ (HAC estimation used in JM statistic)

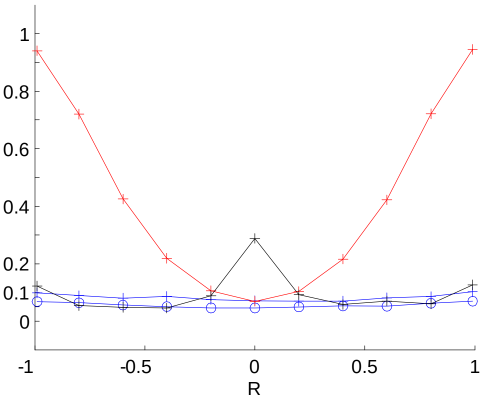


Fig. 2(c) $d = -0.25$

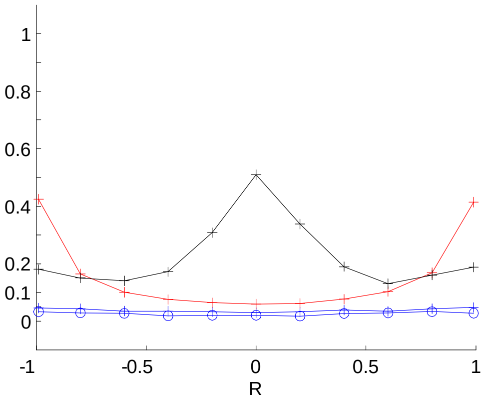


Fig. 2(d) $d = -0.25$

Power $(f(x)) = 0.015x)$: $n = 1000$, $\rho_x = 0.3$ (HAC estimation used in JM statistic)

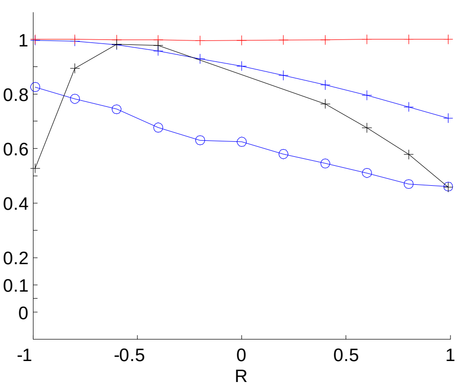


Fig. 3(a) $c = 0$

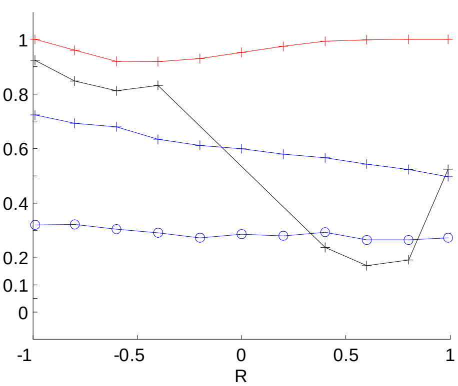
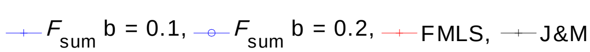


Fig. 3(b) $c = -10$

Note: No simulation results were obtained for JM when $R = -0.2, 0, 0.2$ and the lines shown are interpolated over this interval using results for $R = \pm.4$.



Power $(f(x)) = 0.015x)$: $n = 1000$, $\rho_x = 0.3$ (HAC estimation used in JM statistic)

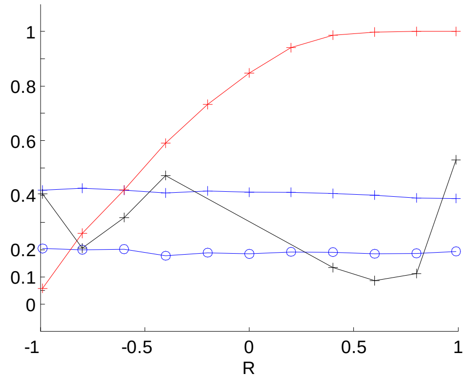


Fig. 3(c) $c = -20$

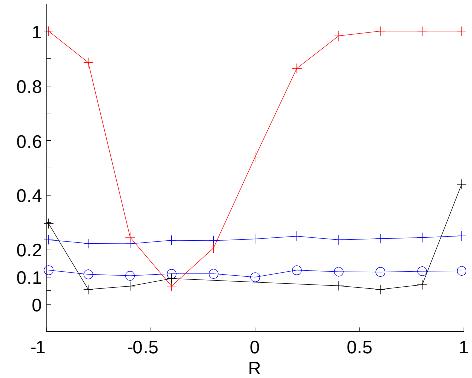


Fig. 3(d) $c = -50$

Note: No simulation results were obtained for JM when $R = -0.2, 0, 0.2$ and the lines shown are interpolated over this interval using results for $R = \pm.4$.

Power: $n = 1000$, $c = 0$, $\rho_x = 0.3$ (HAC estimation used in JM statistic)

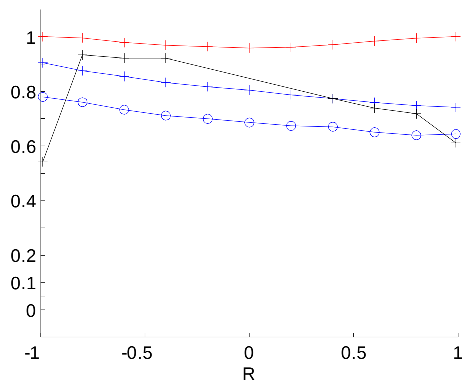


Fig. 4(a) $f(x) = \frac{1}{4} sign(x) |x|^{1/4}$

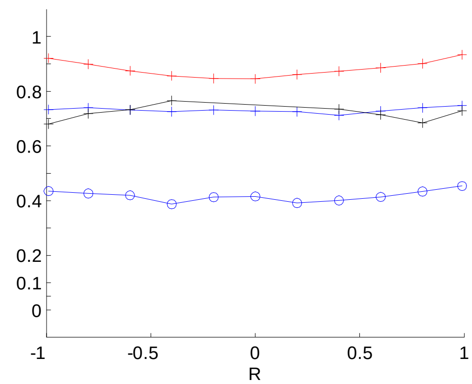


Fig. 4(b) $f(x) = \frac{1}{5} \ln(|x| + 0.1)$

Note: No simulation results were obtained for JM when $R = -0.2, 0, 0.2$ and the lines shown are interpolated over this interval using results for $R = \pm.4$.

F_{sum} b = 0.1, F_{sum} b = 0.2, FMLS, J&M

Power: $n = 1000,\ c = 0,\ \rho_x = 0.3$ (HAC estimation used in JM statistic)

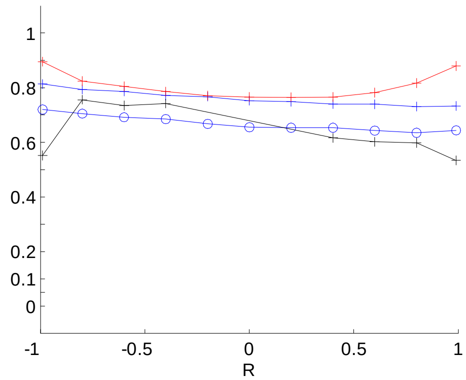


Fig. 4(c) $f(x) = (1 + e^{-x})^{-1}$

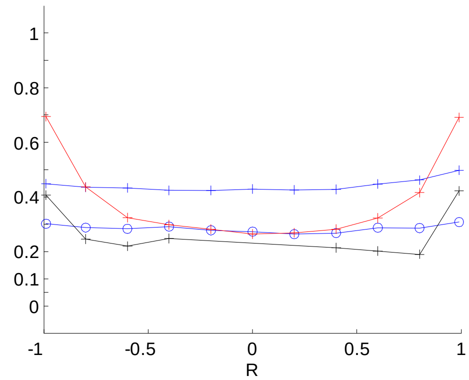


Fig. 4(d) $f(x) = (1 + |x|^{0.9})^{-1}$

Note: No simulation results were obtained for JM when $R = -0.2, 0, 0.2$ and the lines shown are interpolated over this interval using results for $R = \pm.4$.

Power: $n = 1000,\ c = 0,\ \rho_x = 0.3$ (HAC estimation used in JM statistic)

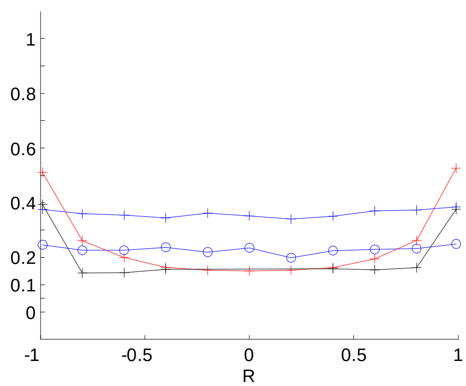


Fig. 4(e) $f(x) = e^{-5x^2}$

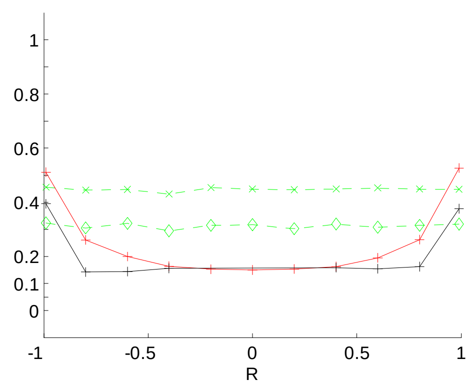
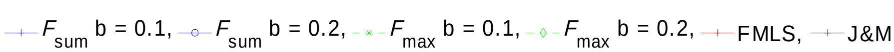


Fig. 4(f) $f(x) = e^{-5x^2}$

Note: No simulation results were obtained for JM when $R = -0.2, 0, 0.2$ and the lines shown are interpolated over this interval using results for $R = \pm.4$.



Size (5%): $n = 500$, $\Delta x_t \sim ARFIMA(d)$

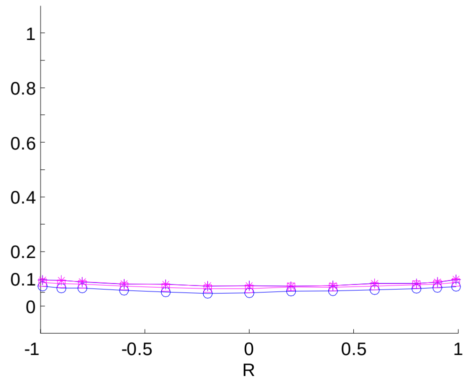


Fig. 5(a) $d = -0.25$

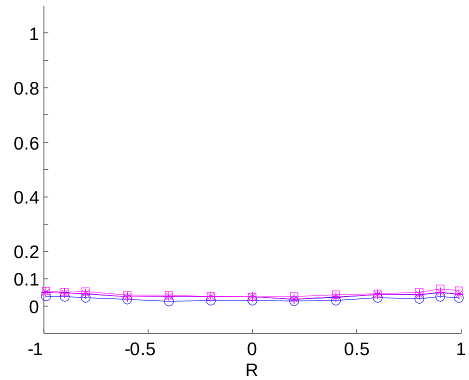


Fig. 5(b) $d = 0.25$

Size (5%): $n = 500$, $\Delta x_t \sim ARFIMA(d)$

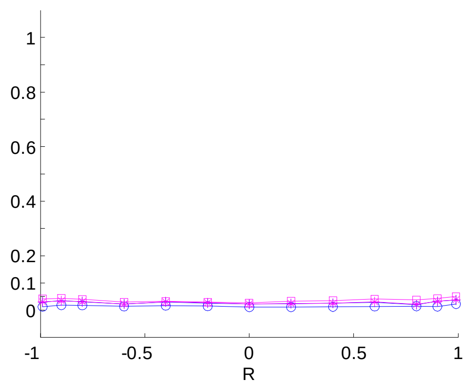


Fig. 5(c) $d = 0.35$

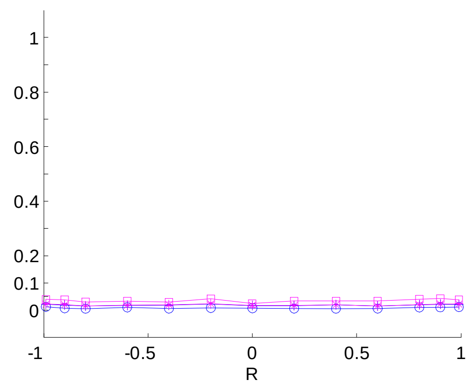
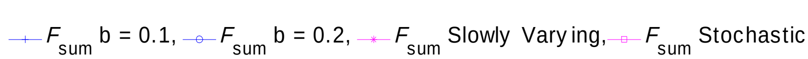


Fig. 5(d) $d = 0.45$

Note: The simulation experiment of Figure 5 involves the following bandwidth choices: $h_n = n^{-b}$, $b = 0.1, 0.2$; $h_n = 1/\ln(\ln(n))$ (slowly varying); $h_n = n^{-\frac{1}{5}(\hat{m}-\frac{1}{2})}$ (stochastic).



Power $(f(x) = (1 + \exp(-x))^{-1})$: $n = 1000$, $\Delta x_t \sim ARFIMA(d)$

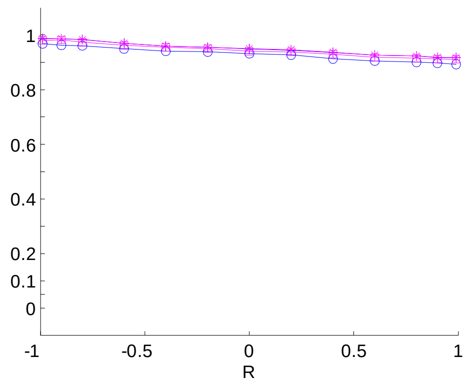


Fig. 6(a) $d = -0.25$

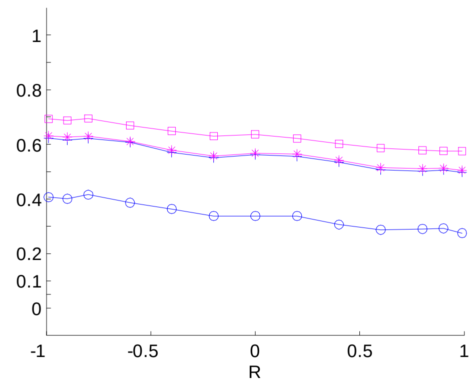


Fig. 6(b) $d = 0.25$

Power $(f(x) = (1 + \exp(-x))^{-1})$: $n = 1000$, $\Delta x_t \sim ARFIMA(d)$

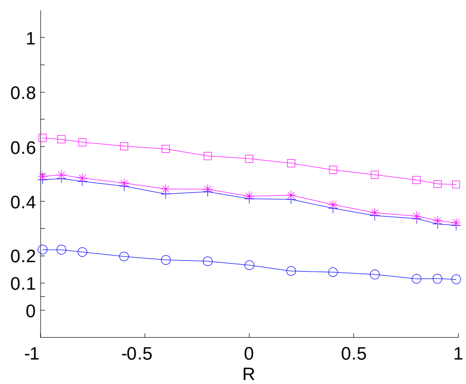


Fig. 6(c) $d = 0.35$

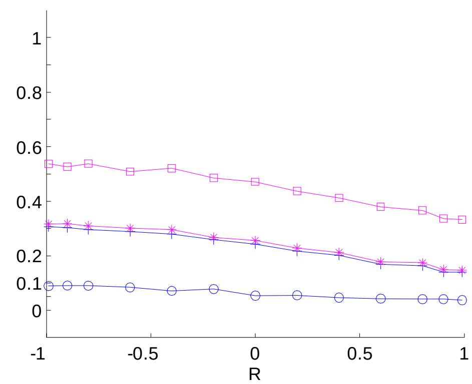


Fig. 6(d) $d = 0.45$

Note: The simulation experiment of Figure 6 involves the following bandwidth choices: $h_n = n^{-b}$, $b = 0.1, 0.2$; $h_n = 1/\ln(\ln(n))$ (slowly varying); $h_n = n^{-\frac{1}{5}(\hat{m}-\frac{1}{2})}$ (stochastic).

F_{sum} b = 0.1, F_{sum} b = 0.2, F_{sum} Slowly Varying, F_{sum} Stochastic

Paths of a simulated random walk (RW) and transformations

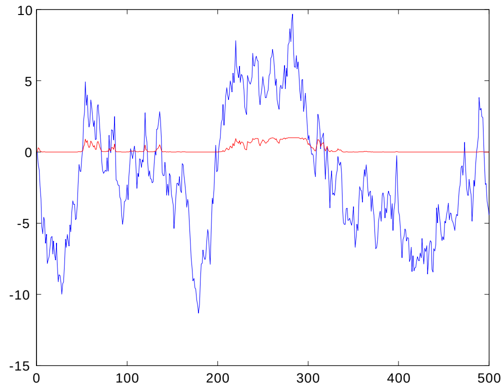


Fig. 7(a) RW (blue) vs logistic transformation of RW (red)

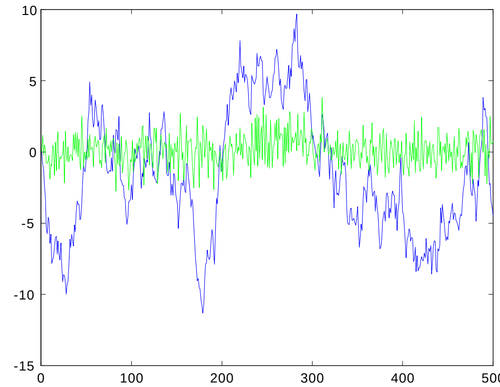


Fig. 7(b) Logistic transformation of RW + white noise (green)

Paths of a simulated random walk (RW) and transformations

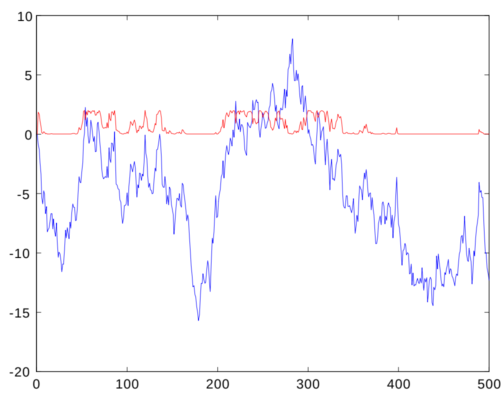


Fig. 8(a) RW vs integrable transformation of RW (red)

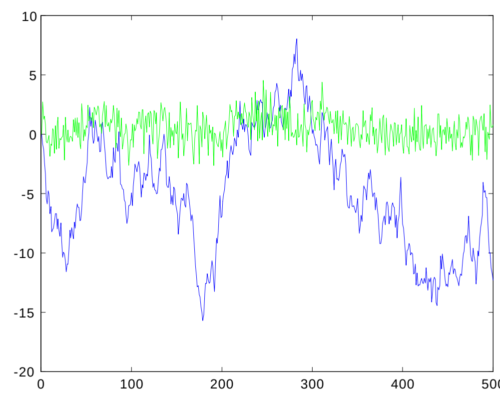


Fig. 8(b) Integrable transformation of RW + white noise (green)

Table 1: Significant Nonparametric Predictability Test Results for the S&P 500 Returns using Different Predictors, Various Grid Points, and Alternative Bandwidths, over 1926:M12-2010:M12 ($n = 1009$).

Predictor:			Dividend Price ratio Log(D/P)		Earnings Price ratio Log(E/P)	
Tests	**Grid pts**	**Lag**	Fixed bandwidth, b	Stochastic bandwidth	Fixed bandwidth, b	Stochastic bandwidth
Sum	10	1	-	-	0.1	✓
		2	-	-	-	✓
		3	-	-	0.1	✓
		4	0.3,0.4	-	0.1	✓
Max		1	-	-	0.1,0.2	✓
		2	-	-	0.1	✓
		3	-	-	0.1	✓
		4	0.1,0.2,0.3,0.4	-	0.1	✓
Sum	25	1	0.2,0.3,0.4	-	-	-
		2	-	-	0.1	✓
		3	-	-	0.1	✓
		4	-	-	-	-
Max		1	0.1,0.2,0.3,0.4	-	-	-
		2	0.2	-	0.1,0.4	✓
		3	0.4	-	0.1,0.2,0.3	✓
		4	0.2	-	0.3,0.4	-
Sum	35	1	-	-	0.1	✓
		2	0.2	-	0.1,0.2,0.3	✓
		3	-	-	0.1,0.2,0.3,0.4	✓
		4	0.2,0.3	-	0.1,0.2	✓
Max		1	0.2,0.3,0.4	-	0.1	✓
		2	0.1,0.2,0.3	-	0.1,0.2,0.3,0.4	✓
		3	0.4	-	0.1,0.2,0.3,0.4	✓
		4	0.2,0.3,0.4	-	0.1,0.2,0.3,0.4	✓
Sum	50	1	0.2,0.3,0.4	-	0.1,0.2	✓
		2	-	-	0.1,0.2,0.3	✓
		3	-	-	0.1,0.2,0.3	✓
		4	-	-	0.1,0.2,0.3,0.4	✓
Max		1	0.2,0.3,0.4	-	0.1,0.2,0.3,0.4	✓
		2	0.2	-	0.1,0.2,0.3,0.4	✓
		3	0.1,0.2,0.4	✓	0.1,0.2,0.3,0.4	✓
		4	0.3,0.4	-	0.1,0.2,0.3,0.4	✓

Notes: The table reports significant predictability results (at the 0.05 level) for the Sum and Max nonparametric tests of the relationship between S&P 500 returns and alternative predictors at various lags. Evidence of significant predictability is reported for alternative exponents b used in the fixed bandwidth case, $h_n = \hat{\sigma}_v n^{-b}$, and for the stochastic bandwidth case, $\hat{h}_n = \widehat{\sigma}_v n^{-\frac{1}{5}(\hat{m}-\frac{1}{2})}$, for which we report a (✓) when there is evidence not rejecting the null hypothesis and (-) otherwise. The reported results use various equi-spaced grids taken over an interval between the 1st and 99th percentiles of the predictor's sample range at (10, 25, 35, 50) points. The empirical results refer to the following predictors: the Dividend Price ratio, log(D/P) and the Earnings Price ratio, Log(E/P).