University of Essex

Essex Business School

Essex Finance Centre

Working Paper Series

Working Paper No 91: 06-2025

"Nonparametric Detection of a Time-Varying Mean"

"Fabrizio Iacone and A. M. Robert Taylor"

Essex Business School, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ
Web site: http://www.essex.ac.uk/ebs/

Nonparametric Detection of a Time-Varying Mean*

Fabrizio Iacone[a] and A.M. Robert Taylor[b]

[a] Università degli Studi di Milano and University of York

[b] Essex Business School, University of Essex

June 2025

Abstract

We propose a nonparametric portmanteau test for detecting changes in the unconditional mean of a univariate time series which may display either long or short memory. Our approach is designed to have power against, among other things, cases where the mean component of the series displays abrupt level shifts, deterministic trending behaviour, or is subject to some form of time-varying, continuous change. The test we propose is simple to compute, being based on ratios of periodogram ordinates, has a pivotal limiting null distribution of known form which reduces to the multiple of a χ^2_2 random variable in the case where the series is short memory, and has power against a wide class of time-varying mean models. A Monte Carlo simulation study into the finite sample behaviour of the test shows it to have both good size properties under the null for a range of long and short memory series and to exhibit good power against a variety of plausible time-varying mean alternatives. Because of its simplicity, we recommend our periodogram ratio test as a routine portmanteau test for whether the mean component of a time series can reasonably be treated as constant.

JEL codes: C12; C22; C52.

Keywords: Time-varying Mean; Periodogram; Portmanteau Test; Trimmed Estimator.

*We thank the Co-Editor, Alex Aue, two anonymous referees, and Fabrizio Ghezzi for their helpful and constructive feedback. Iacone acknowledges financial support within the 'Fund for Departments of Excellence academic funding' provided by the Ministero dell'Università e della Ricerca (MUR), established by Stability Law, namely 'Legge di Stabilità n.232/2016, 2017' - Project of the Department of Economics, Management, and Quantitative Methods, University of Milan. Address correspondence to: Robert Taylor, Essex Business School, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, UK. Email: `robert.taylor@essex.ac.uk`.

1 Introduction

For many macroeconomic and financial time series the assumption that the mean of the series is constant is unrealistic, and incorrectly specifying the mean component of the series to be constant can have very serious consequences for the reliability of statistical modelling and inference and for forecasts generated by the fitted model. Well known early contributions include Perron (1989), who showed that an unmodelled abrupt level shift in the intercept or abrupt change in the drift term render the familiar Dickey-Fuller unit root test unreliable, resulting in spurious non-rejection of the unit root null hypothesis of unit root. Teverovsky and Taqqu (1997) showed that an unmodelled level shift can generate properties similar to long memory in a series that is otherwise weakly dependent. This phenomenon, also known as spurious long memory in the applied literature, is widely documented for stock market data in, *inter alia,* Lobato and Savin (1998), Mikosch and Stărică (2004), Diebold and Inoue (2001), Granger and Hyung (2004), Perron and Qu (2010). Recent evidence of possibly spurious long memory in macroeconomic and financial data is discussed in Iacone, Nielsen and Taylor (2022). Implications of unmodelled breaks in the mean for forecasting are considered in Clements and Hendry (1998).

These examples highlight the importance of testing whether the mean of a time series is stable over the sample or not. The exogenous level shift model, in which the mean of the process changes abruptly at some deterministic point in the sample, offers a very simple representation of the instability, and it has the advantage of being relatively easy to analyse. According to Aue and Horváth (2013) and Horváth and Rice (2014), tests for change points in the mean of a series date back to the 1940s, and have been applied across a wide range of fields, including climatology; see also Reeves *et al.* (2007). Wenger *et al.* (2019) provide a comparison of some of the techniques that have been proposed to detect the presence of a level shift in long memory series.

The abrupt, discontinuous, exogenous change in the mean implied by the level shift model may occasionally be justified: for example the drop of the discharge from the Nile at Aswan might be due to the construction of a dam that was completed in 1902, although Cobb (1978) warns that this change may also be due to other factors, such as a reduction in rainfall. In most cases, however, it is more plausible that changes in the mean occur gradually over time. This seems likely to be the case, for example, for the US inflation rate, which recorded a moderate increase until the early 1980s, when the Volcker-Greenspan era of inflation rate targeting by the U.S. Federal Reserve reversed the trend. A second limitation of the level shift model is that conventional tests for the null hypothesis of a constant mean against the alternative of a level break model have to assume a value for the maximum number of potential breaks, and have limiting null distributions, and hence critical values, which depend on this choice; see, for example, Bai (1999). A third drawback surrounds whether the breaks in observed time series can reasonably be assumed to have been generated exogenously.

Given the likely drawbacks of the simple level shift model, it is not surprising that tests have been developed in the literature which allow for a wider class of functional forms for the mean under the non-constant alternative. Two notable such tests are the VS test of Giraitis, *et al.* (2006), and the W test of Qu (2011). We contribute to this strand of the literature by developing a new test for the null hypothesis that the unconditional mean of a univariate time series process is constant, based on the ratio of selected ordinates of the periodogram of the series. In particular, our proposed test exploits a key feature of a time-varying mean in the frequency domain, namely, that its periodogram concentrates most of its power at the lowest spectral frequencies. This phenomenon was originally noted by Künsch (1986) in the context of small monotonic trends, but has also been discussed by Iacone (2010) for single level shifts, McCloskey and Perron (2013) for multiple level shifts, Perron and Qu (2010) for infrequent breaks, and Qu (2011) for smoothly varying trends. Taken together, these cases constitute a wide range of plausible models for the trend component of a series. In the presence of a time-varying mean of this kind, the periodogram diverges for some of the lowest frequencies, while for a constant mean it does not, and this is the key feature that is exploited in the diagnostic procedure we propose. While our null hypothesis of a constant mean is well specified, our alternative of a time-varying mean is necessarily more nebulous and, hence, we view this as a portmanteau test for non-constancy in the mean.

Our proposed test can be validly used for both short and long memory series, the latter provided the long memory parameter, denoted by δ, lies within the stationary and invertible region, $\delta \in (-1/2, 1/2)$. Two versions of our test are proposed, one for the case where the practitioner specifies a value of δ (e.g. $\delta = 0$ such that the series is weakly autocorrelated), or where δ is estimated from the data. In the latter case it is well known that a time-varying mean will tend to cause an upward bias in standard estimators of δ which assume a constant mean. To circumvent this we explore the use of trimmed estimates of δ in the construction of our statistic. We show that, regardless of whether δ is known or estimated, our test statistic has a well known pivotal limiting null distribution, which reduces to a multiple of a χ^2_2 random variable when the data are short memory. The theoretical power properties of the test are explored with theoretical conditions for consistency of the test provided. The finite sample power properties of our proposed tests against a range of plausible time-varying mean models is explored by Monte Carlo simulation. An empirical application of the tests to to U.S. CPI inflation over the period 1970 to 2022 is also reported.

The remainder of the paper is organised as follows. In Section 2 we present the model we consider for our testing problem and discuss a range of prototypical time-varying mean models that have been considered in the literature. In Section 3 we introduce our proposed portmanteau test for non-constancy of the mean. In Section 4 we derive its large sample properties under suitable regularity conditions and discuss its large sample power properties. Section 5 reports results from our Monte Carlo exercise exploring the finite sample size and power properties of our proposed test. An application to U.S. CPI data is reported in Section 6. Section 7 offers some conclusions. An on-line Supplementary Appendix contains proofs of our main results and additional Monte Carlo results.

We will use the following notational conventions throughout the paper: $A := B$ and $B =: A$ denote that A is defined by B; for a possibly random sequence, X_T, and a deterministic sequence, f_T, the notation $X_T = O_e(f_T)$ means that X_T/f_T convergences (either in distribution or in probability) to a non-degenerate, non-zero random variable. The operator $\lfloor . \rfloor$ denotes the integer part of its argument.

2 The Time-Varying Mean Model

We consider the univariate time series process, x_t, satisfying the following decomposition,

$$x_t = \mu_t + \xi_t, \quad t = 1, ..., T \tag{1}$$

where μ_t is a potentially time-varying mean component and ξ_t is a zero-mean, fractionally integrated process. The fractionally integrated component, ξ_t in (1), is defined by integrating a weakly dependent, or $I(0)$, process by the long memory parameter δ, $\delta \in (-1/2, 1/2)$. More formally, let η_t be a zero-mean, stationary process with spectral density $f_{\eta\eta}(\lambda)$ that is continuous, bounded, and bounded away from zero at all frequencies. Then,

$$\xi_t = \Delta^{-\delta}\eta_t, \; \delta \in (-1/2, 1/2) \tag{2}$$

is a fractionally integrated process of order δ, denoted $\xi_t \in I(\delta)$.

Denoting $\Delta_t^{(\delta)} := \Gamma(t+\delta) / (\Gamma(\delta)\Gamma(t+1))$, where $\Gamma(\cdot)$ is the Gamma function, such that $\Gamma(0) := \infty$ and $\Gamma(0)/\Gamma(0) := 1$, then $\xi_t = \sum_{s=-\infty}^{t} \Delta_{t-s}^{(\delta)} \eta_s$, $\delta \in (-1/2, 1/2)$. The absence of any truncation in the infinite MA representation entails that ξ_t is a Type 1 fractionally integrated process, see Marinucci and Robinson (1999), and that it is stationary with spectral density $f_{\xi\xi}(\lambda)$ such that $f_{\xi\xi}(\lambda) \to G\lambda^{-2\delta}$ as $\lambda \to 0^+$, for some $G \in (0, \infty)$. In the case of a series, z_t say, whose long memory parameter lay above 0.5, for example in the range $0.5 \leq \delta < 1.5$, this series could be differenced prior to analysis, such that $\Delta z_t := z_t - z_{t-1}$ was assumed to satisfy the model in (1)-(2).

For our purposes, it is convenient to characterise the level term, μ_t, in (1) in terms of the properties of its periodogram. To that end, for a generic sequence, $\{z_1, ..., z_T\}$, define the familiar Fourier transform,

$$w_z(\lambda) := \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} z_t e^{-i\lambda t},$$

where i is the complex operator and $\lambda \in (-\pi, \pi]$ is a user-chosen frequency, for which value we can define the *periodogram* as

$$I_{zz}(\lambda) := |w_z(\lambda)|^2.$$

Our focus is on testing the null hypothesis that μ_t is constant; that is,

$$H_0 : \mu_t = \mu, \quad t = 1, ..., T \tag{3}$$

where μ is a fixed constant. To that end, consider the periodogram of $\{\mu_1, ..., \mu_T\}$, $I_{\mu\mu}(\lambda)$, at the Fourier frequencies, $\lambda_j := \frac{2\pi j}{T}$ for $j = 1, ..., T-1$. Under H_0 of (3), we clearly have that $I_{\mu\mu}(\lambda_j) = 0$, for all λ_j. While our null hypothesis is well specified, our alternative is necessarily more nebulous, given our aim is to develop a portmanteau type test against non-constancy in μ_t. However, and in order to motivate a class of alternatives that we would ideally like our test to display power against, consider as motivation the class of time-varying μ_t such that, on at least some $j = o(T)$, the periodogram diverges as T goes to infinity,

$$I_{\mu\mu}(\lambda_{j^*}) \to \infty \text{ for some } j^* \text{ such that } j^* = o(T) \tag{4}$$

and, in general, satisfies the condition that

$$I_{\mu\mu}(\lambda_j) = O_p\left((\lambda_j)^{-2\phi} j^{-1}\right), \phi \in (0, 1/2]. \tag{5}$$

Taken together, these two conditions entail that the periodogram concentrates nearly all of its power in a band of frequencies which collapse towards the origin as the sample size diverges. This situation is characteristic of a number of prototypical time-varying mean models that have been considered in the literature. We now detail some leading examples of these.

Example 1: Deterministic Level Shifts

$$\mu_t = \mu + \sum_{k=1}^{K} \beta_k DU_t(\tau_k^*) \tag{6}$$

where $\|\beta\| > 0$ for $\beta := (\beta_1, ..., \beta_K)'$, and where $DU_t(\tau) := \mathbb{I}(t \geq \lfloor \tau T \rfloor)$, with $\mathbb{I}(.)$ denoting the indicator function, whose value is one when its argument is true and zero otherwise; the values $0 < \tau_1^* < \cdots < \tau_K^* < 1$ denote the location in the sample (as fractions of the sample size, T) where abrupt changes (of which there are at least one) in the mean occur. $\diamondsuit$

Example 2: Smoothly Varying Trend

$$\mu_t = \mu + \beta h(t/T) \tag{7}$$

where $h(s)$ is a Lipschitz continuous function on $[0, 1]$, $h(s) \neq h(r)$ for some $s \neq r$, and $\beta \neq 0$. $\diamondsuit$

Example 3: Power Trend

$$\mu_t = \mu + \beta t^{\varphi - 1/2} \tag{8}$$

for $\varphi \in (0, 1/2)$, $\beta \neq 0$. $\diamondsuit$

Example 4: Martingale Process

$$\mu_t = \mu_{t-1} + \beta_t \epsilon_t \tag{9}$$

with β_t i.i.d. Bernoulli(p), $p = O(T^{-1})$, ϵ_t i.i.d. N(0,σ_ϵ^2), β_t and ϵ_t mutually independent and $\sigma_\epsilon^2 > 0$. $\diamondsuit$

The Deterministic Level Shifts model in Example 1 is very popular in econometrics, possibly because it is relatively straightforward to establish asymptotic results for break location estimators and test statistics for the presence of breaks in this case. However, as discussed in the Introduction, abrupt exogenous changes in the mean are often not plausible in practice. The Smoothly Varying Trend model in Example 2 is a plausible model for many time series in economics, climatology and other fields. It also includes conventional trend models, such as linear and quadratic trend models as special cases. The definition we use in Example 2 is taken from Qu (2011, pp.424-425). The same model is also considered in Giraitis *et al.* (2006). Nonparametric, kernel-based, estimation of the smoothly varying trend model in presence of long memory disturbances is discussed in Robinson (1997), with forecasting from this model considered in Dalla *et al.* (2020). The Power Trend model in Example 3, and the capacity of vanishing power trends of this kind to generate spurious evidence of long memory, has been widely discussed in the literature; see, among others, Bhattacharya *et al.* (1983), Teverovsky *et al.* (1986), and Giraitis *et al.* (2006). The Martingale Process in Example 4, which allows changes in the mean to occur at random points in the sample, has been considered in Diebold and Inoue (2001), Perron and Qu (2010), McCloskey and Perron (2013) and Nyblom (1989), among others. Nyblom (1989), in particular, develops locally best invariant tests for the null hypothesis that μ_t in (1) is constant against the alternative that μ_t evolves according to Example 4, for the case where ξ_t in (1) is an IID Gaussian process.

Despite their apparently different functional forms, the models in Examples 1-4 are all characterised by periodograms for which the bound in (5) holds: with $\phi = 1/2$ for the Level Break, Smoothly Varying Trend and Martingale Process examples, and $\phi = \varphi$ for the Power Trend. Detailed treatments of these bounds are given in, among others: Iacone (2010), Theorem 1, for the Level Shift and Power Trend models; Perron and Qu (2010), Proposition 3, for the Martingale Process; and Qu (2011), Lemma 1, for the Smoothly Varying Trend model. We also refer to McCloskey and Perron (2013) and Leschinski and Sibbertsen (2018) for a detailed discussion of these bounds. Some further properties for the periodograms of these trend models are given in Lemma S.1 in the Supplementary Appendix.

The diverging property in (4) is straightforwardly established whenever exact orders can be established. This is shown to hold for the Deterministic Level Shift model in Leschinski and Sibbertsen (2018) and also in Lemma S.1 in the Supplementary Appendix. For the Martingale Process, this is shown in Proposition 3 of Perron and Qu (2010). In both of these cases, the bound in (5) also holds as an exact, O_e, bound at at least some frequencies in λ_1, ..., λ_m, $m/T \to 0$, and in particular it holds as O_e for λ_1. For the Power Trend model, the exact expression for the real and complex part of the Fourier transform can be readily computed as a closed form formula for $\lambda + T\lambda^{-1} \to 0$ from the limit in Lemma 3.2 of Robinson and Marinucci (2001), which again implies that the bound in (5) holds exactly. Notice that the condition on λ does not allow for λ_1; we will discuss this case in Lemma S.1 in the Supplementary Appendix. In all of these cases, the periodogram therefore diverges at least for some frequencies. The case of the slowly varying trend is somewhat more delicate. Qu (2011, p.425) argues that for this model the periodogram is diverging at least at some ordinates; further discussion on this is also provided in Lemma S.1 in the Supplementary Appendix.

3 A Periodogram Ratio Test of a Constant Mean

As anticipated in Section 2, the periodogram of a time-varying mean component, μ_t, that satisfies the bounds in (4) and (5) has the property that it concentrates nearly all of its spectral power at the lowest frequencies. Indeed, it was exactly this property that led Künsch (1986) to propose ignoring the lowest frequencies when estimating features of ξ_t. The main advantage of this procedure, often referred to in the literature as *trimming*, is that it does not require the user to specify the nature of the contamination process, μ_t. In contrast, where the unconditional mean is constant, the periodogram of μ_t has zero power at all frequencies.

Applications of trimming to estimate δ, the memory parameter of ξ_t, include Iacone (2010) for the case of the local Whittle [LW] estimate, and McCloskey and Perron (2013) for the case of the log-periodogram regression estimate. McCloskey and Hill (2017) and Dalla *et al.* (2020) discuss trimmed LW estimation of a fully parametric model, and Christensen and Varneskov (2017) present an application of trimming in cointegration. Interestingly, given that the contamination due to μ_t, when the bound in (5) holds, only affects the lowest frequencies about zero, a judicious choice of the trimming parameter can result in little or no deterioration in the asymptotic properties of the estimate of δ. On the other hand, Monte Carlo simulation results in Iacone (2010) and McCloskey and Perron (2013) suggest that where the mean is constant, so that $\mu_t = \mu$ for all t, such that trimming is not necessary, the estimate of δ from the trimmed loss function can have markedly larger variance than its untrimmed counterpart.

In the light of the relative performance of the trimmed and untrimmed estimates of δ discussed above, a diagnostic for the presence of a time-varying mean component in the series would seem highly desirable for practitioners, and it is our aim in this paper to provide a portmanteau test to do just that. If our proposed test rejects the null hypothesis of a constant mean, then the practitioner should use a trimmed estimation procedure, while an untrimmed estimate might reasonably be used otherwise. Conveniently, the rates in (4) and in (5) provide a natural approach to design such a test based on the comparison of the value of the periodogram where the signal arising from the time-varying mean is strongest, j^* in (4), against a set of values of the periodogram at higher band frequencies (though still within a band that is degenerating to zero as T diverges), where the signal from the stochastic component ξ_t should dominate. As we will see later in Section 4 when we establish the consistency properties of our preferred test, for a wide range of the prototypical time-varying mean examples given in the Introduction, the periodogram diverges at λ_1.

Based on these considerations, our proposed test is based on the ratio of the first periodogram ordinate of $\{x_t\}$ to the sum of a range of higher ordinates, the latter used to standardise the numerator with respect to the long run variance of η_t, the $I(0)$ component of x_t. Specifically, for a generic memory parameter $d \in (-1/2, 1/2)$, our proposed portmanteau test of a constant mean against the alternative of a non-constant mean rejects for large values of the ratio statistic,

$$R(d) := \frac{(\lambda_1)^{2d} I_{xx}(\lambda_1)}{(m-l+1)^{-1}\sum_{j=l}^{m}(\lambda_j)^{2d} I_{xx}(\lambda_j)}. \tag{10}$$

If it were known that ξ_t was integrated of order δ, then our test could be based on the statistic, $R(\delta)$. For example, in many cases practitioners may have a plausible belief that the series under analysis is weakly dependent, such that $\delta = 0$. Where no such knowledge is either held or assumed about δ, our test can be based on evaluating $R(d)$ at an estimate of δ obtained from the data, $\widehat{\delta}$ say; in this case the test statistic of interest becomes $R(\widehat{\delta})$. As we will later show in Section 4, provided this estimate satisfies a minimal consistency rate, the large sample behaviour of $R(\delta)$ and $R(\widehat{\delta})$ will coincide.

Remark 1.

(i) The user-chosen tuning parameters, l and m in (10), satisfy the relation $1 < l < m < T/2$. We will refer to the set of periodogram ordinates used in the denominator of (10) as the range, $\{l, m\}$. Formal rate conditions needed to hold on l and m for establishing the large sample properties of the $R(d)$ statistic in (10) and will subsequently be detailed in Section 4. Notice that the denominator of $R(d)$,

$$\widehat{G}(d) := \frac{1}{m-l+1}\sum_{j=l}^{m}(\lambda_j)^{2d} I_{xx}(\lambda_j) \tag{11}$$

is an estimate of the long run variance of η_t, and it therefore standardises the numerator of the statistic with respect to this variance. In the context of long run variance estimation, the tuning parameters l and m are often referred to as the <u>trimming</u> and <u>bandwidth</u> parameters, respectively. Such long run variance standardisation is a common feature of other tests in this literature; see, for example, Lobato and Robinson (1998), Qu (2011), Iacone, Leyborne and Taylor (2017), and Giraitis et al. (2006). Our choice of long run variance estimate, motivated by our intention to create a portmanteau test for non-constancy of the mean, is, however, innovative compared to the approach taken in these other papers, because they focus on the behaviour of the long run variance estimators they propose only under the null and specific local alternative models for μ_t. These estimates may well be inconsistent, or at least *subject to a sizeable finite sample bias, for μ_t satisfying the bounds in (4) and (5). However, with judicious selection of the range $\{l, m\}$ in $\widehat{G}(\delta)$, we can estimate the long run variance of η_t consistently even in presence of a time-varying mean; see, for example, Iacone (2010). This is because the effect of the time-varying mean on $\widehat{G}(\delta)$ is reduced by the presence of the damping factor j^{-1} in (5).*

(ii) The role of the lowest frequency l used in constructing $\widehat{G}(d)$ is seen to be of great importance: a judicious choice of l exploits the contrasting characteristics of the periodogram of the time varying component μ_t and of the stationary counterpart ξ_t at different frequencies. A relatively small value for l is likely to be sufficient in practice, because the contribution of the periodogram $I_{\mu\mu}(\lambda_j)$ vanishes very quickly as j increases for μ_t satisfying the bounds in (4) and (5). The highest frequency, m, is more familiar in the literature, because it is widely discussed in the context of estimation of a long run variance: consistent estimation requires that $m \to \infty$ at some rate as $T \to \infty$, but in practice if m is chosen to be too large, the curvature of the spectral density of η_t may affect the precision of the estimate of its long run variance, so an inappropriate choice of m may result in a test with poor finite sample size performance; see, among others, Abadir et al. (2009). In our case, however, an additional consideration is necessary: in the presence of a time-varying mean the contribution of the periodogram of $I_{\mu\mu}(\lambda_j)$ may be small, but yet non-zero, at some of the frequencies in the range $\{\lambda_l, ..., \lambda_m\}$, and so a relatively large value for the bandwidth, m, is also important to render the average contribution of $I_{\mu\mu}(\lambda_j)$ to $\widehat{G}(d)$ asymptotically irrelevant. Intuitively, therefore, a tension between the finite sample size and power of the test based on $R(d)$ is to be anticipated in respect of the choice of m. This will be explored further in our Monte Carlo simulation study in Section 5.

Remark 2.

(i) The numerator in the $R(d)$ statistic uses only the first periodogram ordinate, λ_1. This is motivated by arguments of parsimony related to our previous observations that many members of the class of non-constant μ_t processes that we are looking to detect display their largest periodogram ordinate at λ_1. However, this is not always the case, and so a test which rejects for large values of generalised version of the $R(d)$ in (10) which includes the first $q < l$

periodogram ordinates in the numerator, i.e.,

$$\overline{R}(d) := \frac{q^{-1}\sum_{j=1}^{q}(\lambda_j)^{2d} I_{xx}(\lambda_j)}{(m-l+1)^{-1}\sum_{j=l}^{m}(\lambda_j)^{2d} I_{xx}(\lambda_j)}, \tag{12}$$

may potentially be more powerful than the test based on $R(d)$ in cases where λ_1 is not the largest ordinate of the periodogram of μ_t. Here the truncation parameter, q, is assumed to be independent of the data, although it could potentially be data-determined. The flip side, of course, is that a test based on $\overline{R}(d)$ would be expected to less powerful than a test based on $R(d)$ in cases where λ_1 is the largest ordinate of the periodogram of μ_t.
(ii) In the context of $\overline{R}(d)$ in (12), q is envisaged to be a relatively small fixed integer. One could also consider a version of $\overline{R}(d)$ with q chosen to be sufficiently large such that asymptotics in q could be justified; here we could consider a test which rejects for large values of the statistic

$$\widetilde{R}(d) := \frac{\overline{q}^{-1/2}\sum_{j=1}^{q}\nu_j(\lambda_j)^{2d} I_{xx}(\lambda_j)}{(m-l+1)^{-1}\sum_{j=l}^{m}(\lambda_j)^{2d} I_{xx}(\lambda_j)} \tag{13}$$

where $\nu_j := \ln(j) - \frac{1}{q}\sum_{k=1}^{q}\ln(k)$, $\overline{q} = \sum_{j=1}^{q}\nu_j^2$. When $l = 1$, $q = m$, the $\widetilde{R}(d)$ is the well known LM statistic used to test the null hypothesis that $x_t \in I(d)$; see Lobato and Robinson (1998) and Iacone, Nielsen and Taylor (2022). These choices of l and q would, however, be inappropriate (in that power would be expected to be very low) for the problem of testing for non-constancy in μ_t. Here choices of $l > 1$ and $q \ll m$ would be more appropriate.

Remark 3. *A related statistic, used to test the null that a series is a stationary long memory process against the alternative that it subject to spurious long memory induced by the presence of either a smoothly varying trend or stochastic level shifts, of the form given in Examples 2 and 4, respectively, in Section 2, is the W statistic proposed in Equation (8) of Qu (2011,p.426). The W statistic is based on a first order expansion of the LW loss function, and exploits the fact that the periodogram is diverging at a fast rate when μ_t is subject to such changes. The W statistic differs from those we propose in that while it is also formed from the periodogram ordinates of $\{x_t\}$ it is, in effect, a maximum CUSUM-type procedure based on sequential partial sums formed from the first $\bar{m}$ periodogram ordinates, where $\bar{m}/T^{1/2} \to 0$ as $T \to \infty$, with the first CUSUM in the sequence based on the sum of the first $\lfloor \bar{\varepsilon}\bar{m} \rfloor$ ordinates, where $\bar{\varepsilon}$ is a small number (Qu, 2011, recommends using $\bar{\varepsilon} = 0.05$*

if $T < 500$), and the last based on all $\bar{m}$. In common with the statistic $\widetilde{R}(d)$ in (12), the number of periodogram ordinates used in the numerator of W is an increasing function of T. Qu's W statistic needs to be evaluated based on the (untrimmed) LW estimate of the long memory index, δ, even in situations in which δ might reasonably be assumed known. Given that the LW estimate is known to suffer from potentially substantial upward biases for some (unmodelled) time-varying μ_t processes (see, among others, Mc Closkey and Perron, 2013, and Iacone, 2010), not exploiting information about the order δ, or not trimming the LW estimate, may incur a substantial loss in power in such cases. Moreover, the long run variance estimator used in Qu's W statistic uses all of the first $\bar{m}$ periodogram ordinates, rather than the trimmed version we use; cf. Remark 1.

4 Asymptotics for the Periodogram Ratio Statistic

In this section we will establish the large sample properties of the $R(d)$ statistic of (10) for both the case where d is a known index and where d is set equal to a consistent estimator of δ. We will first establish the limiting null distribution of the statistic, before establishing consistency against a class of time-varying μ_t processes.

To do so we need to set out some regularity conditions on the DGP given in (1)-(2). First, for the $I(0)$ component, η_t, we use the following set of conditions from Wu and Shao (2006):

Assumption A1. *Let $\eta_t := F(\ldots, \varepsilon_{t-1}, \varepsilon_t)$, $t \in \mathbb{Z}$, where ε_t are IID random variables and F is a measurable function such that η_t is well-defined. For a random variable X write $X \in \mathcal{L}^p$ if $\|X\|_p = [E|X|^p]^{1/p} < \infty$. Let $\mathcal{F} = (\ldots, \varepsilon_{t-1}, \varepsilon_t)$, and define the projections $\mathcal{P}_k$ by $\mathcal{P}_k X = E(X|\mathcal{F}_k) - E(X|\mathcal{F}_{k-1}), X \in \mathcal{L}^1$. Let η_t be such that $\eta_t \in \mathcal{L}^{p^*}$ with $p^* > \max(2, 2/(1+2\delta))$, $\|\sum_{k=0}^{\infty} \mathcal{P}_k \eta_t\|_{p^*} < \infty$ and $f_{\eta\eta}(0) > 0$.*

Remark 4. *Assumption A1 includes a very wide class of processes; see the discussion in Wu and Shao (2006,pp.20-23), and the references therein. In particular, it includes linear processes of the form $\eta_t = \sum_{j=1}^{\infty} a_j \varepsilon_{t-j}$ as a special case. It also includes a large class of nonlinear time series models, including bilinear models, threshold models and GARCH-type models. The condition that $\|\sum_{k=0}^{\infty} \mathcal{P}_k \eta_t\|_{p^*} < \infty$ is discussed in some detail on page 23 of*

Wu and Shao (2006) and serves to restrict the amount of dependence allowed in $\{\eta_t\}$*; for example, in the linear process case it imposes the usual absolute summability condition that* $\sum_{j=0}^{\infty} |a_j| < \infty$*.*

We complete the set of necessary regularity conditions with some additional conditions related to the spectral density of ξ_t, $f_{\xi\xi}(\lambda)$:

Assumption A2.
(i) There exists a $G \in (0, \infty)$ *such that*

$$f_{\eta\eta}(\lambda) \sim G \text{ as } \lambda \to 0^+;$$

(ii) In a neighbourhood $(0, \epsilon)$ *of the origin,* $f_{\eta\eta}(\lambda)$ *is differentiable and*

$$\frac{d}{d\lambda} ln f_{\eta\eta}(\lambda) = O(\lambda^{-1}) \text{ as } \lambda \to 0^+.$$

Remark 5. *In view of* (2)*, Assumption A2 implies that* $f_{\xi\xi}(\lambda) \sim G\lambda^{-2\delta}$ *as* $\lambda \to 0^+$ *and* $f_{\xi\xi}(\lambda)$ *is differentiable in a neighbourhood* $(0, \epsilon)$ *of the origin, with* $\frac{d}{d\lambda} f_{\xi\xi}(\lambda) = O(\lambda^{-1-2\delta})$ *as* $\lambda \to 0$*. This is sufficient to allow us to establish bounds for the bias and covariances of the expected periodogram on a band of frequencies degenerating to zero.*

We are now in the position to state our main result, which provides the limiting null distribution of the $R(d)$ statistic of (10), both for the case where d is set equal to the true long memory parameter, δ, and where it is set equal to a consistent estimate thereof.

Theorem 1. *Let* $\{x_t\}$ *be defined as in* (1) *and* (2) *under* H_0 *of* (3)*, and let Assumptions A1 and A2 hold. Then provided the rate condition,* $\frac{l}{m} + \frac{m}{T} \to 0$ *as* $T \to \infty$*, holds:*
(i)

$$R(\delta) \xrightarrow{d} \left\{ \frac{L_1(\delta)}{2} + L_1^*(\delta) \right\} Z_1^2 + \left\{ \frac{L_1(\delta)}{2} - L_1^*(\delta) \right\} Z_2^2 =: \mathcal{R}_\infty(\delta) \tag{14}$$

where Z_1, Z_2 are independent, standard normal random variables, and where

$$L_j(d) := \frac{2}{\pi}\int_{-\infty}^{\infty} \frac{sin^2(\lambda/2)}{(2\pi j-\lambda)^2}\left|\frac{\lambda}{2\pi j}\right|^{-2d} d\lambda$$
$$L_j^*(d) := \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{sin^2(\lambda/2)}{(2\pi j-\lambda)(2\pi j+\lambda)}\left|\frac{\lambda}{2\pi j}\right|^{-2d} d\lambda;$$

(ii) For any $\widehat{\delta} = \delta + O_p(T^{-\epsilon})$ for some $\epsilon > 0$,

$$R(\widehat{\delta}) - R(\delta) = o_p(1). \tag{15}$$

The results in (14) and (15) also hold under the alternative rate condition, $\frac{1}{l} + \frac{l}{m} + \frac{m}{T} \to 0$ as $T \to \infty$.

Remark 6.

(i) The limiting distribution $\mathcal{R}_\infty(\delta)$ in (14) appears in Hurvich and Beltrao (1993), who established it for the case of linear Gaussian processes and this was extended to the case of non-Gaussian linear processes in Terrin and Hurvich (1994). The result in part (i) of Theorem 1 further extends this result to the case of non-linear processes satisfying Assumption A1.

(ii) The result in part (ii) of Theorem 1 imposes a necessary rate of convergence on $\widehat{\delta}$, although, in contrast to the W statistic in Qu (2011), it does not constrain the practitioner to use the LW estimator.

(iii) The limiting distribution $\mathcal{R}_\infty(\delta)$ in (14) depends on the true long memory parameter, δ; critical values from this distribution can be computed, for example, by using the formula in Moschopoulos (1985), Equation (2.10). In the special case of $\delta = 0$, the distribution in (14) is one-half times a χ^2_2 variate.

(iv) Under the conditions of Theorem 1, the limiting null distributions of $\overline{R}(\delta)$ and $\overline{R}(\widehat{\delta})$ coincide and can be straightforwardly obtained from the results given in Theorem 1 of Terrin and Hurvich (1994). This limiting distribution has a rather involved form, other than in the case where $\delta = 0$ where it simplifies to a $(2q)^{-1}\chi^2_{2q}$ random variable.

(v) It can also be shown that under H_0 of (3) and with some additional regularity conditions, that the $\widetilde{R}(\delta)$ statistic defined in (13) is such that $\widetilde{R}(\delta) \xrightarrow{d} N(0,1)$. The additional regularity

conditions required on ξ_t are given in, for example, Shao and Wu (2007a, 2007b). The result also requires the rate condition, $\frac{(\log(T))^3}{q} + \frac{q}{T^{2/3}} \to 0$ as $T \to \infty$, to hold on q. For $\widetilde{R}(\widehat{\delta})$ the same limiting result holds as for $\widetilde{R}(\delta)$, provided some further regularity conditions hold, primarily that $(\widehat{\delta} - \delta) = O_p(T^{-\alpha-\epsilon})$ and $T^{-\alpha} q^{1/2} \to 0$.

In Theorem 2 we next explore the class of models of time-variation in μ_t against which the test based on $R(\delta)$ will deliver consistent inference; a discussion on how these results extend to the case of the test based on the $R(\widehat{\delta})$ statistic is subsequently given in Remark 9. As we will show, consistency only holds for a subset of the set of μ_t processes that satisfy the bounds given in (4) and (5). However, it is important to stress that even where formal consistency is not established this does not mean that our proposed tests will have no power to detect departures from the null hypothesis. This will be further explored in our finite sample simulation study, the results from which are reported in Section 5.

Theorem 2. *Let $\{x_t\}$ be defined as in (1) and (2) and let Assumptions A1 and A2 hold. If the periodogram of μ_t satisfies the conditions that $I_{\mu\mu}(\lambda_1) = O_e(T^{2\phi})$, $I_{\mu\mu}(\lambda_j) = O_p(\lambda_j^{-2\phi} j^{-1})$ for $j \le m$, and $\sum_{j=l}^{m} \lambda_j{}^{2\delta} I_{\mu\mu}(\lambda_j) = O_e((T/l)^{2(\phi-\delta)})$, then provided the rate condition $\frac{1}{l} + \frac{l}{m} + \frac{m}{T} \to 0$ as $T \to \infty$, holds:*
(i) if $\phi > \delta$,

$$\text{if } \; T^{2(\phi-\delta)}\; m^{-1}\; l^{2(\delta-\phi)} \to \infty, \quad R(\delta) = O_e(m l^{2(\phi-\delta)}) \tag{16}$$

$$\text{if } \; T^{2(\phi-\delta)}\; m^{-1}\; l^{2(\delta-\phi)} \to 0, \quad R(\delta) = O_e(T^{2(\phi-\delta)}) \tag{17}$$

(ii) if $\phi < \delta$, $\quad R(\delta) \xrightarrow{d} \mathcal{R}_\infty(\delta)$.

Remark 7.
(i) The requirement that $\sum_{j=l}^{m} \lambda_j{}^{2\delta} I_{\mu\mu}(\lambda_j) = O_e((T/l)^{2(\phi-\delta)})$ controls the allowable rate of divergence of the denominator of $R(\delta)$, $\widehat{G}(\delta)$ of (11). This assumption automatically holds if $I_{\mu\mu}(\lambda_j) = O_e(\lambda_j^{-2\phi} j^{-1})$ for $j \le m$, as is the case for the martingale process in (9). The more involved formulation of the rate conditions needed is to accommodate the level shift model in (6). This is seen most clearly through the example of a single level shift at $\tau^* = 0.5$: here it is

straightforward to verify that $I_{\mu\mu}(\lambda_j) = O_p(\lambda_j^{-2\phi} j^{-1})$ holds, but with $I_{\mu\mu}(\lambda_j) = 0$ for j even. For the power trend (8) this requirement holds (with $\phi = \varphi$) in view of the limit in Lemma 3.2 of Robinson and Marinucci (2001).

(ii) All of the stated rate conditions on $I_{\mu\mu}(\lambda_j)$ hold for both the level shift model in (6) and for the martingale process in (9), in both cases with $\phi = 1/2$, and also for the power trend model in (8) (with $\phi = \varphi$), see Lemma S.1.

(iii) In the case of the smooth trend model, the required order condition that $I_{\mu\mu}(\lambda_1) = O_e(T)$ is not exhaustive of all models of the form in (7). For example, the deterministic cosine trend model $\mu_t = \cos(2\pi 2t/T)$ has a spectral peak at λ_2, but has zero spectral power at λ_1. In order to have power against such processes one could use tests based on either the $\overline{R}(\delta)$ or $\widetilde{R}(\delta)$ statistics of (12) and (13), respectively.

(iv) In view of the low frequency approximation $f_{\xi\xi}(\lambda) \sim G\lambda^{-2\delta}$ as $\lambda \to 0^+$, it is clear that the power of the test based on the $R(\delta)$ statistic depends on the interplay between the bound $O\left((\lambda_j)^{-2\phi} j^{-1}\right)$ and $\lambda_j^{-2\delta}$. Larger values of δ, relative to ϕ, may serve to mask the presence of time-variation in the mean, making its detection more difficult. This feature is common for this class of testing problems; see, for example Theorem 1 of Iacone et al. (2017), and Theorem 1 of Iacone et al. (2014) for the case of the single level break model. Indeed, if $\delta > \phi$, then as part (ii) of Theorem 2 establishes, the asymptotic power of the test is equal to its asymptotic size. For example the power trend model of (8), is not detectable whenever $\delta > \varphi$. However, low power in such situations is arguably less of a concern, because the distorting effect of μ_t when estimating features of ξ_t is likely to be relatively weak.

Remark 8. *The bounds in (16) and (17) are suggestive that power is increasing in both l and m, the trimming and bandwidth parameters, respectively, used in the estimation of the long run variance which forms the denominator of $R(\delta)$. Because the term $I_{\mu\mu}(\lambda_j)$ is decreasing in j, eliminating the lowest frequencies by setting $l > 1$ reduces or removes contamination from μ_t in this estimate. In view of (16) and (17), it might seem advisable to choose l and m as large as possible. In practice, however, there are other factors to take into consideration. First, the results are based on approximating the spectral density $f_{\eta\eta}(\lambda)$ as a constant, G, at low frequencies but in reality $f_{\eta\eta}(\lambda)$ is unlikely to be constant for $\lambda \neq 0$, and so the approximation may be less reliable as λ moves away from the origin. Curvature in $f_{\eta\eta}(\lambda)$ may generate a bias in the estimation of G, and this becomes more relevant when larger values of m are* *considered. A detailed discussion of this issue is given in Abadir et al. (2009), who recommend using $m = \lfloor T^{0.8} \rfloor$ for linear processes and, following Dalla et al. (2006), a smaller rate, such as $m = \lfloor T^{0.7} \rfloor$, for nonlinear processes. The trimming parameter, l, by removing the very frequencies for which $f_{\eta\eta}(\lambda)$ is usually closest to G, may also amplify this bias. The impact of the choice of l and m on the finite sample behaviour of the test will be explored further in Section 5.*

Remark 9. *In the case where δ is estimated, provided $\widehat{\delta} = \delta + O_p(T^{-\epsilon})$ for some $\epsilon > 0$, $R(\widehat{\delta}) - R(\delta) = o_p(1)$, such that the results stated in Theorem 2 will still hold. Giraitis et al. (2006) observe that the LW estimate satisfies this condition in the case of the smooth trend model in (7) when δ is estimated based on the residuals from a nonparametric regression. In cases where the characterisation in (5) allows for abrupt discontinuities, such as the level break model in (6), the sufficient rate of convergence for $\widehat{\delta}$ can be established using either trimmed log-periodogram regression estimation (see McCloskey and Perron, 2013,pp.1201,1205) or trimmed LW estimation (see, Iacone, 2010). We will explore the properties of the tests based on untrimmed and trimmed LW estimates in our simulation study in Section 5.*

In the case where $\{x_t\}$ is weakly dependent, such that $\delta = 0$, if it holds that $I_{\mu\mu}(\lambda_1) = O_e(T)$, an example of which is the Level Shift model of Example 1, then Theorem 2 yields the corollary:

Corollary 1. *Let $\{x_t\}$ be defined as in (1) and (2) with $\delta = 0$, and let Assumptions A1 and A2 hold. If μ_t is such that $I_{\mu\mu}(\lambda_1) = O_e(T)$ and $I_{\mu\mu}(\lambda_j) = O_e(\lambda_j^{-1} j^{-1})$, then provided the rate condition $\frac{1}{l} + \frac{l}{m} + \frac{m}{T} \to 0$ as $T \to \infty$, holds:*

$$\text{if } \frac{T}{lm} \to \infty, \quad R(0) = O_e(lm) \tag{18}$$

$$\text{if } \frac{T}{lm} \to 0, \quad R(0) = O_e(T) \tag{19}$$

It is also interesting to consider the power of $R(\delta)$ against level shifts of small magnitude. For example, for weakly dependent processes ($\delta = 0$) the Sup Wald test of Andrews (1993) can detect deterministic level shifts of dimension $\beta_T = \theta T^{-1/2+\epsilon}$ for $\epsilon > 0$, and this result can be generalised to $\beta_T = \theta T^{-1/2+\delta+\epsilon}$ for generic fractionally integrated processes; see for example Shao (2011) and Iacone *et al.* (2017). In this case, $I_{\mu\mu}(\lambda_1) = O_e(T^{2(\delta-1/2+\epsilon)}\lambda_1^{-1})$, $I_{\mu\mu}(\lambda_j) = O_p(T^{2(\delta-1/2+\epsilon)}\lambda_j^{-1}\ j^{-1})$ for some $\epsilon > 0$.

Corollary 2. *Let $\{x_t\}$ be defined as in (1) and (2) with $\mu_t = \mu + \beta_T DU_t(\tau^*)$ such that $\beta_T = \theta T^{-1/2+\delta+\epsilon}$, and ϵ so that $m^{-1}\ T^{2\epsilon}\ l^{2(\delta-1/2)} \to 0$. Provided either $\frac{1}{l} + \frac{l}{m} + \frac{m}{T} \to 0$ as $T \to \infty$, or $\frac{1}{m} + \frac{m}{T} \to 0$ as $T \to \infty$, then $R(\delta) = O_e(T^{2\epsilon})$.*

Remark 10. *The rate at which the test statistic $R(\delta)$ diverges (the same rate applies to $R(\widehat{\delta})$, under the condition that $\widehat{\delta} = \delta + O_p(T^{-\epsilon})$ for some $\epsilon > 0$) in the presence of local breaks coincides with that of the tests designed specifically for detecting level breaks in Shao (2011) and Iacone et al. (2017).*

5 Monte Carlo Study

In this section we report the results from a Monte Carlo study investigating the finite sample size and power properties of the $R(\delta)$ and $R(\widehat{\delta})$ statistics. In both our size and power studies we consider samples of size $T = 128$ and $T = 512$. All reported situations were based on $10,000$ repetitions for $T = 128$, and $1,000$ replications for $T = 512$, and were performed in Gauss 23 using the RNDN random number generator.

We consider both the case where the practitioner makes a choice for the value δ, say $\delta^\dagger$, and where δ is estimated from the data. In the former case, where the practitioner uses the true value of δ, such that $\delta^\dagger = \delta$ (eg if they rightly assume the data are weakly dependent, where $\delta = 0$), this allows us to explore the properties of our proposed tests uncontaminated by the finite sample error in the estimation of δ. We also explore the impact of the practitioner choosing a wrong value for δ such that $\delta^\dagger \neq \delta$. In the case where δ is estimated we used the trimmed LW estimate (restricted to the interval $[-0.49, 0.49]$),

$$\widehat{\delta} := \arg \min_{d\in[-0.49,0.49]} \ln \left\{ \frac{1}{m^* - l^* + 1} \sum_{j=l^*}^{m^*} \lambda_j^{2d} I_{xx}(\lambda_j) \right\} - 2d \frac{1}{m^* - l^* + 1} \sum_{j=l^*}^{m^*} \lambda_j \tag{20}$$

where l^* and m^* denote the trimming and bandwidth parameters used in the LW estimation.

We set $m^* = \lfloor T^{0.65} \rfloor$, in view of the recommendations of Dalla *et al.* (2006) and Abadir *et al.* (2007). We considered three possible values of l^*: $l^* = 1$ (no trimming), $l^* = 2$, and $l^* = 3$. We found that trimming can increase the variance of the estimate of δ under the null, causing some finite sample size deterioration in the test. To control for this effect we use a parametric bootstrap to generate critical values for our test, based on $B = 999$ bootstrap replications.[1]

Section 5.1 investigates the finite sample size properties of the $R(\delta)$ and $R(\widehat{\delta})$ tests for a variety of different ξ_t processes. Then in Section 5.2 we will investigate the finite sample power properties of these tests in the case of a single deterministic level break. Following up on the discussion in Section 4, in these first two sections we will explore, for a variety of simulation DGPs, the size-power trade from the choices made for the trimming and bandwidth parameters, l and m respectively, which feature in our proposed test statistic, providing some empirical guidelines for choosing l and m in practice. Finally, Section 5.3 summarises the results from a comparison of the finite sample size and power properties of our proposed tests, based on our recommended tuning parameters, with some relevant tests in the literature.

5.1 Size Study

We generate simulation data according to (1)-(2) with $\mu_t = \mu$, $t = 1, ..., T$, setting $\mu = 0$, without loss of generality. In addition, for the stochastic component, ξ_t, we consider the following range of simulation DGPs:

- DGP1: (NIID) $\xi_t \sim IID\, N(0, 1)$
- DGP2: (AR(1)) $\xi_t = 0.5\xi_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim IID\, N(0, 1)$,
- DGP3: (ARCH(1)) $\xi_t \sim NIID(0, \sigma_t^2)$, where $\sigma_t^2 = 1 + 0.5\xi_{t-1}^2$.

[1] This was done as follows: (i) setting $\bar{\delta}$ to be equal to either $\delta^\dagger$ in the case where the practitioner specifies a value for δ, or $\widehat{\delta}$ the (trimmed or untrimmed, as relevant) LW estimate from the original data, generate B T-dimensional $I(\bar{\delta})$ series, $\{x_{t,i}^*\}$, $i = 1, ..., B$, $t = 1, ..., T$, according to (1)-(2) with $\xi_t \sim IID\ N(0, 1)$ and, $\mu_t = \mu$, $t = 1, ..., T$, setting $\mu = 0$ (without loss of generality); (ii) for each bootstrap series, $\{x_{t,i}^*\}$, $i = 1, ..., B$, calculate either the statistic $R(\delta^\dagger)_i$ if $\bar{\delta} = \delta^\dagger$, or $R(\widetilde{\delta}_i^*)_i$ if $\bar{\delta} = \widehat{\delta}$, where $\widetilde{\delta}_i^*$ is the LW estimate obtained from $\{x_{t,i}^*\}$, in each case using the same values of l, m and, where relevant, l^*, m^* as for the original statistic, $R(\bar{\delta})$; (iii) arrange the B bootstrap statistics from step (ii) in ascending order, and denote by $\widetilde{cv}^*(0.05)$ the upper 5% quantile of this ordered sequence; (iv) reject H_0 if $R(\bar{\delta}) > \widetilde{cv}^*(0.05)$.

- DGP4: (FGN(0.3)) $\xi_t = (1-L)^{-0.3}\varepsilon_t, \quad \text{with } \varepsilon_t \sim IID\,N(0,1)$

- DGP5: (FGN(-0.3)) $\xi_t = (1-L)^{0.3}\varepsilon_t, \quad \text{with } \varepsilon_t \sim IID\,N(0,1)$

DGP1 may be considered as a benchmark case where ξ_t is uncorrelated. DGP2, where ξ_t follows an AR(1) process, allows us to investigate the effect of curvature in the spectrum, relative to the benchmark case. DGP3, where ξ_t follows an ARCH(1) process, allows us to study the impact of condtional heteroskedasticity. DGP1-DGP3 are all cases where $\delta = 0$. DGP4 and DGP5 are cases where ξ_t is a fractional Gaussian white noise processes with $\delta = 0.3$ (persistent long memory) and $\delta = -0.3$ (anti-persistent long memory), respectively.

We report results for situations in which: (i) Table 2 - the user correctly sets $\delta^\dagger = \delta$ the true order of integration; (ii) Table 3 - the order of integration is estimated by (trimmed) LW; (iii) Table 2 - the practitioner incorrectly sets $\delta^\dagger = 0$ in the cases of DGP4 and DGP5 where $\delta = 0.3$ and $\delta = -0.3$, respectively. Results are reported for following values of l and m: $l = 2$, $l = 4$ and $l = 6$ for the $T = 128$ case, and $l = 2$, $l = 4$ and $l = 8$ for the $T = 512$ case (except for DGP3-DGP5 where, in the interests of brevity, we only report results for $l = 4$); $m = \lfloor T^{0.5} \rfloor$, $m = \lfloor T^{0.65} \rfloor$ and $m = \lfloor T^{0.8} \rfloor$.

The main findings of these results can be summarised as follows:

(i) When the spectral density of η_t (recall $\eta_t = \Delta^\delta \xi_t$) has no curvature, the test is well sized in all cases. This includes the ARCH case (as well as the fractional Gaussian noise cases). Knowledge of δ (or estimation of it) does not affect this result. The only evident size distortions are seen in the case where the sample size is small ($T = 128$), when $\delta = -0.3$ and δ is estimated with trimming, in which case some mild oversize is seen. These distortions are, however, not present for the larger sample size, $T = 512$.

(ii) When the spectral density of η_t has a curvature, DGP1, empirical size is seen to depend on whether δ is known and on the value m chosen. In the case of known δ, size distortions are larger when the value of m used in the test statistic is large; conversely, in the case of estimated δ, the size distortion is smaller if m in the test statistic is large.

Positive autocorrelation is associated with size inflation if δ is known, and size deflation if δ is estimated. Other things equal, the size distortions are reduced for $T = 512$ *vis-à-vis* $T = 128$.

(iii) Incorrectly setting $\delta^{\dagger} = 0$ when $\delta = 0.3$ results in very large positive size distortions, which get worse as T or m increase. Conversely, incorrectly setting $\delta^{\dagger} = 0$ when $\delta = -0.3$ results in significant distortions below the nominal significance level.

(iv) Estimating δ using the LW estimate, or its trimmed version, results in broadly similar size properties, other things equal. In other words, trimming does not appear to adversely impact the finite sample size performance of the test relative to the untrimmed case.

5.2 Power Study: Single Level Shift Model

In this part of our exercise we generate simulation data according to (1)-(2) with ξ_t again generated according to DGP1-DGP5. For the level component, μ_t, we now introduce a single level break: $\mu_t = 0.5 \times DU_t(0.5)$, so that the unconditional mean of x_t abruptly changes from 0 to 0.5 half way through the sample.

We report results for situations in which: (i) Table 4 - the user correctly sets $\delta^{\dagger} = \delta$ the true order of integration; (ii) Table 5 - the order of integration is estimated by (trimmed) LW; (iii) Table 4 - the practitioner incorrectly sets $\delta^{\dagger} = 0$ in the cases of DGP4 and DGP5 where $\delta = 03$ and $\delta = -0.3$, respectively. The same values of l, m and l^*, m^* are used as in the results in Section 5.1, except that, to avoid repeating redundant information, we now only report results for $l = 4$ for DGP1.

The main findings of these results can be summarised as follows:

(i) For given choices of m and l, the power of the tests are, other things equal, increased for the larger sample size, $T = 512$, relative to the smaller sample size, $T = 128$, but are decreasing with δ. The former reflects the consistency property of the tests established in Section 4. The latter is also reflective of the consistency rate given in part (i) of

Theorem 2 which predicts that power is an increasing function of $(\phi - \delta)$ when $\phi > \delta$: heuristically, one can think of this as a signal plus noise model, where the periodogram of μ_t is the signal, and the spectral density of ξ_t is the noise, and so the larger is δ, the larger is the confounding effect of ξ_t on the spectrum of x_t.

(ii) For given T, the choice of m has a significant impact on power. When δ is known, power is increasing in m. However, where δ is estimated, power appears to be highest between $m = 0.5$ and $m = 0.65$ and then declines for larger values of m. The latter is perhaps not surprising, as in these cases $\widehat{\delta}$ is likely to be upward biased in finite samples and, as the periodogram ordinates in the denominator are scaled by $j^2\widehat{\delta}$, a larger value for $\widehat{\delta}$ pushes the $R(\widehat{\delta})$ towards zero, other things equal, thereby making it more difficult to reject the null hypothesis. This effect can also be seen in the case of DGP5, where $\delta = -0.3$, by comparing the power when the user sets $\delta^\dagger = 0$ with the case where they (correctly) set $\delta^\dagger = -0.3$.

(iii) The choice of l does not appear to have a significant impact on power.

(iv) Using trimming in connection with the LW estimate of δ is strongly improving for power in cases where larger values of m are used in the $R(\widehat{\delta})$ statistic, but less so otherwise. Using a trimming parameter of $l^* = 2$ delivers slightly superior power to using $l^* = 3$.

Based on the finite sample size and power simulation results presented so far, we can make some tentative recommendations on the values of the tuning parameters l, m, l^*, and m^* which feature in our proposed $R(\delta)$ and $R(\widehat{\delta})$ statistics. First, in general, we recommend the use of the test based on $R(\widehat{\delta})$ rather than $R(\delta^\dagger)$, given the uncontrolled size distortions that can occur when $\delta^\dagger \neq \delta$. Second, for $\widehat{\delta}$ we recommend using a trimmed LW estimate with trimming parameter $l^* \geq 2$, and bandwidth $m^* = \lfloor T^{0.65} \rfloor$, the latter as recommended by Dalla *et al.* (2006) and Abadir *et al.* (2007). For the numerator of the $R(\widehat{\delta})$ statistic, balancing size and power considerations, overall we recommend a bandwidth m of somewhere in the range $[\lfloor T^{0.50} \rfloor, \lfloor T^{0.65} \rfloor]$. The choice of the trimming parameter, l, seems less crucial, and we suggest considering a range of values of l: the simulations results presented suggest using $l = 2$ for $T = 128$ and $l = 4$ for $T = 512$.

5.3 Additional Monte Carlo Results

In the last part of our Monte Carlo exercise, we ran a comparative study of the finite sample size and power properties of the $R(\delta)$ and $R(\widehat{\delta})$ tests against a set of benchmark tests from the literature. For $R(\delta)$ and $R(\widehat{\delta})$ we follow the recommended settings for the tuning parameters given at the end of Section 5.2, setting $m = \lfloor T^{0.55} \rfloor$. The comparator tests we considered are the W test of Qu (2011), and the $T_n(\widehat{\delta})$ test [denoted VS in what follows] of Giraitis *et al.* (2006), both of which are designed to detect general forms of non-constancy in μ_t, allowing for long memory in x_t. We also consider the SW test of Iacone *et al.* (2014) which is designed to detect a single deterministic level break in the presence of long memory. All of the tests were run using the recommended settings given by the authors of the tests; further details on these tests can be found in Section S.2 of the Supplementary Appendix where details of the Monte Carlo designs considered and the results of the experiments can also be found.

Size properties against DGP1-DGP5 from Section 5.1 were investigated together with power results against a variety of non-constant designs for μ_t. Here we provide a summary of those findings as follows:

(i) The SW test has good size properties, both for the known and estimated δ cases. The only exceptions occurs, as would be expected, in cases where the user specifies a value for $\delta^{\dagger}$ which is different from the true value of δ. Even in such cases, however, the size distortion is the smallest of the three tests, suggesting that the distorting effect due to an imprecise estimate of δ is lowest for this test. This is confirmed by the performances of the VS and W tests: the VS test is subject to some potentially large size distortion even in the larger sample size, when δ is assumed known, at least when the spectral density of η_t is subject to some curvature, as in the AR(1) case or when $\delta = 0.3$. The size performance of the VS test is improved if the LW estimate is used, but is still significantly over-sized for DGP4 where $\delta = 0.3$. Finally, the results verify the invalidity of the W test when based on an assumed (rather than estimated) value of δ, even where the correct value of δ is assumed.

(ii) For any given combination of μ_t process, value of δ, and test, replacing δ with its estimate always results in a loss of power. Increasing the sample size increases the empirical power; however, the relative performance of the tests is not affected by T, in the sense that the power ranking of the tests, for a given scenario, is essentially the same for $T = 128$ and $T = 512$.

(iii) For abrupt level break models the power of the tests is increasing in β. We also find that a break located in the middle of the sample ($\tau^* = 0.5$) is more easily detected than a late break ($\tau^* = 0.75$). In the case where $\tau^* = 0.5$, for both the case where δ is known and where it is estimated, our proposed R tests, are very competitive on power with the SW test, often displaying higher power than SW, but lag behind the power of SW when $\tau^* = 0.75$.

(iv) For smoothly varying trend models our proposed R tests are overall the best performing on power, both for the case where δ is known and where it is estimated.

(v) In most cases the W test displays lower power, often significantly lower, than the other tests. The only exception is for a martingale process where it is competitive with the other tests on power, other than for the case where $\delta = 0$ and $T = 128$.

(vi) Trimming of the LW estimate in general appears to increase the power of the $R(\widehat{\delta})$ test, albeit marginally.

6 Application to U.S. CPI Inflation

We apply our proposed $R(\widehat{\delta})$ test to U.S. inflation over the period 1970–2022. Historically, this period is characterised by different monetary policy regimes and changing underlying conditions in both the financial markets and macroeconomic circumstances. The inflationary burst brought on by the oil shocks of the 1970s was eventually curbed by the more aggressive attitude to inflation control ushered in during the Volker-Greenspan era and the resulting, so-called, Great Moderation. More recently, U.S. inflation has reverted to periods of instability, most notably the 2008 financial crisis and recent international turmoil.

The inflation series is computed as the first differences of the (natural) logarithm of monthly CPI, the Consumer Price Index for all Urban Consumers (all items in U.S. city average), seasonally adjusted: series CPIAUCSL from the FRED database. The sample size is $T = 635$. The plot of the series (in log-first differences) is given in Figure 1. This is scaled by 1200 to be visually compatible with the measure of inflation that is commonly used. Also depicted in red is a non-parametric estimate of the mean of the series computed over a rolling window of width 12 months: this is at least suggestive of the presence of some time-variation in the unconditional mean of the series across the sample.

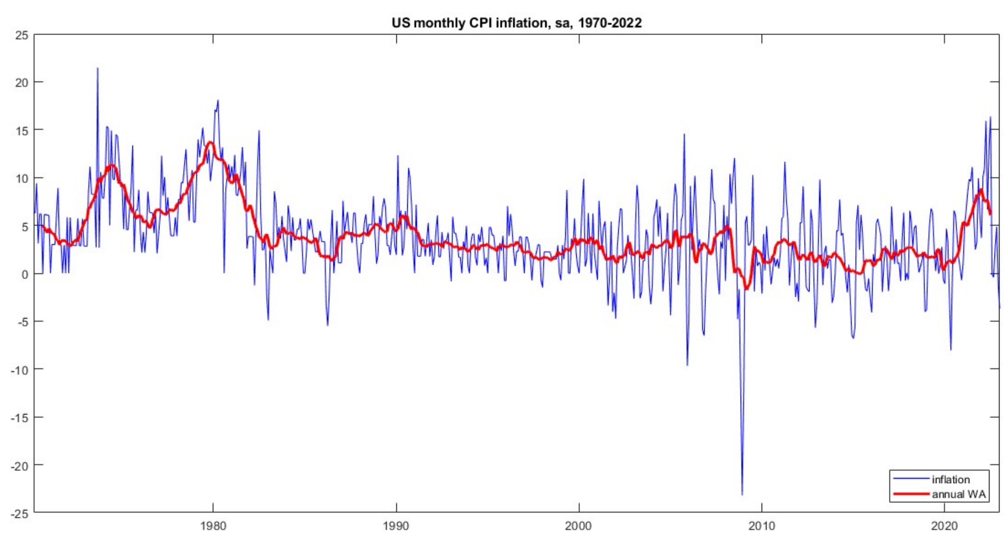

Figure 1: U.S. Monthly CPI Inflation (seasonally adjusted), 1970-2022

In Table 1 we report the outcomes of the $R(\widehat{\delta})$ statistic, for a range of values of the trimming and bandwidth parameters, l and m. The statistics were computed using either the untrimmed LW estimate, denoted $\widehat{\delta}$, or the trimmed LW estimate with $l^* = 2$, denoted $\widehat{\delta}_2$, and $l^* = 3$, denoted $\widehat{\delta}_3$. In each case we used a bandwidth of $m^* = \lfloor T^{0.65} \rfloor$. Bold entries denote cases where the outcome of the statistic exceeds the 5% bootstrap critical value (with the bootstrap critical values calculated as outlined in footnote 1 with $B = 999$ bootstrap replications). Missing entries are where we consider the range $\{l, m\}$ to be too small to deliver a reliable estimate of the long run variance.

Table 1: Application of $R(\widehat{\delta})$ Tests to Monthly Seasonally Adjusted U.S. CPI

l \ m	13 $\lfloor T^{0.4} \rfloor$	18 $\lfloor T^{0.45} \rfloor$	25 $\lfloor T^{0.5} \rfloor$	30	34 $\lfloor T^{0.55} \rfloor$	40	48 $\lfloor T^{0.6} \rfloor$	66 $\lfloor T^{0.65} \rfloor$
	Panel A: Untrimmed LW Estimate ($l^* = 1$), $\widehat{\delta} = 0.45$							
2	2.07	2.58	2.91	3.10	3.15	2.77	2.75	2.46
4	2.01	2.60	2.97	**3.16**	**3.21**	2.79	2.76	2.46
6	1.86	2.54	2.96	3.17	**3.22**	2.78	2.75	2.44
9	—	239	2.92	3.17	**3.22**	2.74	2.72	2.41
12	—	—	**5.44**	**5.15**	**4.79**	**3.44**	**3.24**	2.64
	Panel B: Trimmed LW Estimate ($l^* = 2$), $\widehat{\delta}_2 = 0.41$							
2	2.48	3.12	3.58	3.85	3.96	3.57	3.61	3.34
4	2.45	3.19	3.70	3.99	**4.10**	3.64	3.67	3.37
6	2.28	3.15	3.72	4.04	**4.14**	3.65	3.68	3.37
9	—	2.99	3.70	4.06	4.18	3.63	3.66	3.35
12	—	—	**7.13**	**6.82**	**6.40**	**4.70**	**4.47**	3.74
	Panel C: Trimmed LW Estimate ($l^* = 3$), $\widehat{\delta}_3 = 0.37$							
2	2.87	3.63	4.23	**4.59**	**4.76**	4.38	**4.48**	4.28
4	2.88	3.78	4.43	**4.82**	**4.98**	**4.52**	**4.61**	4.35
6	2.70	3.76	4.49	**4.91**	**5.08**	**4.56**	**4.65**	4.37
9	—	3.58	4.49	**4.98**	**5.16**	**4.57**	**4.66**	4.36
12	—	—	**8.89**	**8.58**	**8.11**	**6.05**	**5.80**	**4.96**

As might be anticipated, given the apparent non-constancy of the unconditional mean in

Figure 1, trimming is seen to have a significant effect on the LW estimate of δ for the inflation series. In particular, the LW estimate decreases as the amount of trimming increases, passing from $\widehat{\delta} = 0.45$ when no trimming is used, to $\widehat{\delta}_3 = 0.37$ when $l^* = 3$. Relatedly, we see that the evidence against the null hypothesis that the unconditional mean of the inflation series is constant across the sample is much higher when trimming with $l^* = 3$ is used, compared to the cases where no trimming or trimming with $l^* = 2$ is used. The outcomes of the statistics are also seen to be uniformly larger when $l^* = 3$ than for the other cases, for given values of l and m. For the case where we use the trimmed LW estimate with $l^* = 3$ we see that we can reject the null hypothesis for most of the values of l and m considered; indeed, for $l = 12$, we can reject for all the values of m considered. However, it is also worth observing that for our recommended tuning parameter settings of $m = \lfloor T^{0.55} \rfloor$ and $l = 4$, we are able to reject the null hypothesis of a constant mean, regardless of whether a trimmed or untrimmed LW estimate is used.

7 Conclusions

We have developed portmanteau tests, based on ratios of the periodogram ordinates of the series, for detecting general forms of non-constancy in the level component of a (possibly) fractionally integrated time series process. The numerator contains low frequency ordinates designed to become large when there is non-constancy in the level, our leading case being where the statistic which includes only the lowest frequency ordinate in the numerator. The denominators use higher frequency ordinates to the scale the numerator by an estimate of the long run variance of the series. For this leading case, we have shown that the periodogram ratio tests admit pivotal limiting distributions of a well known form under the null hypothesis that the level of the series is constant across the sample and have also established consistency against a wide class of time-varying mean components, including deterministic level shift models, smoothly varying trend components, power trends, and martingales. A Monte Carlo simulation study, again focussing on our leading case, showed that the test displays good finite sample size control and is very competitive on power with extant tests in the literature. An empirical application to U.S. inflation suggests the presence of statistically significant time-variation in the mean over the period 1970-1922.

We end with a suggestion for further research. Our recommended tests require the practitioner to specify values for the bandwidth and trimming parameters used in constructing the test statistic. While we have made recommendations for the values of these to use in practical applications, it would also be worth exploring if data-based choices for these tuning parameters, such as those discussed for the bandwidth in the context of estimating the long memory parameter in Henry (2001), have the potential to improve the finite sample performance of the tests.

References

Abadir, K. M., W. Distaso, and L. Giraitis. 2007. Nonstationarity-extended local whittle estimation. *Journal of Econometrics* 141, 1353–1384.

Abadir, K. M., W. Distaso, and L. Giraitis. 2009. Two estimators of the long-run variance: beyond short memory. *Journal of Econometrics* 150, 56–70.

Andrews, D.W.K. 1993. Tests for parameter instability and structural change with unknown change point. *Econometrica* 61, 821–856.

Aue, A., and L. Horváth. 2013. Structural breaks in time series. *Journal of Time Series Analysis* 34, 1–16.

Bai, J. 1999. Likelihood ratio tests for multiple structural changes. *Journal of Econometrics* 91, 299–323.

Bhattacharya, R. N., V. K. Gupta, and E. Waymire. 1983. The hurst effect under trends. *Journal of Applied Probability* 20, 649–662.

Christensen, B. J., and R. T. Varneskov. 2017. Medium band least squares estimation of fractional cointegration in the presence of low-frequency contamination. *Journal of Econometrics* 197, 218–244.

Clements, M., and D. Hendry. 1998. *Forecasting Economic Time Series* (Cambridge

University Press).

Cobb, G. W. 1978. The problem of the Nile: conditional solution to a changepoint problem. *Biometrika* 65, 243–251.

Dalla, V., L. Giraitis, and J. Hidalgo. 2006. Consistent estimation of the memory parameter for nonlinear time series. *Journal of Time Series Analysis* 27, 211–251.

Dalla, V., L. Giraitis, and P. M. Robinson. 2020. Asymptotic theory for time series with changing mean and variance. *Journal of Econometrics* 219, 281–313.

Diebold, F. X., and A. Inoue. 2001. Long memory and regime switching. *Journal of Econometrics* 105, 131–159.

Giraitis, L., P. Kokoszka, and R. Leipus. 2001. Testing for long memory in the presence of a general trend. *Journal of Applied Probability* 38, 1033–1054.

Giraitis, L., R. Leipus, and A. Philippe. 2006. A test for stationarity versus trends and unit roots for a wide class of dependent errors. *Econometric Theory* 22, 989–1029.

Granger, C. W. J., and N. Hyung. 2004. Occasional structural breaks and long memory with an application to the sp 500 absolute stock returns. *Journal of Empirical Finance* 11, 399–421.

Henry, M. 2001. Robust automatic bandwidth for long memory. *Journal of Time Series Analysis* 22, 293–316.

Horváth, L., and G. Rice. 2014. Extensions of some classical methods in change point analysis. *TEST* 23, 219–255.

Hurvich, C. M., and K. I. Beltrao. 1993. Asymptotics for the low frequency ordinates of the periodogram of a long-memory time series. *Journal of Time Series Analysis* 14, 455-472.

Iacone, F. 2010. Local whittle estimation of the memory parameter in presence of deterministic components. *Journal of Time Series Analysis* 31, 37–49.

Iacone, F., M. Ø. Nielsen, and A. M. R. Taylor. 2022. Semiparametric tests for the order of integration in the possible presence of level breaks. *Journal of Business and Economic Statistics* 40, 880–896.

Iacone, F., S. J. Leybourne, and A. M. R. Taylor. 2014. A fixed-b test for a break in level at an unknown time under fractional integration. *Journal of Time Series Analysis* 35, 40–54.

Iacone, F., S. J. Leybourne, and A.M. R. Taylor. 2017. Testing for a change in mean under fractional integration. *Journal of Time Series Econometrics*.

Künsch, H. 1986. Discrimination between monotonic trends and long-range dependence. *Journal of Applied Probability* 23, 1025–1030.

Leschinski, C., and P. Sibbertsen. 2018. The periodogram of spurious long- memory processes. Hannover Economic Papers (HEP) dp-632, Leibniz Universität Hannover, Wirtschaftswissenschaftliche Fakultät, June.

Lobato, I. N., and N. E. Savin. 1998. Real and spurious long-memory properties of stock-market data. *Journal of Business and Economic Statistics* 16, 261–268.

Lobato, I. N., and P. M. Robinson. 1998. A Nonparametric Test for I(0), *Review of Economic Studies* 65, 475–495.

Marinucci, D., and P.M. Robinson. 1999. Alternative forms of fractional brownian motion. *Journal of Statistical Planning and Inference* 80, 111–122.

McCloskey, A., and J. B. Hill. 2017. Parameter estimation robust to low-frequency contamination.*Journal of Business and Economic Statistics* 35, 598–610.

McCloskey, A., and P. Perron. 2013. Memory parameter estimation in the presence of level shifts and deterministic trends.*Econometric Theory* 29, 1196–1237.

Mikosch, T., and C. Stărică. 2004. Nonstationarities in financial time series, the long-range dependence, and the igarch effects. *The Review of Economics and Statistics* 86, 378–390.

Moschopoulos, P.G. 1985. The distribution of the sum of independent gamma random variables. *Annals of the Institute of Statistical Mathematics* 37, 541–544.

Nyblom, J. 1989. Testing for the constancy of parameters over time, *Journal of the American Statistical Association* 84, 223–230.

Perron, P. 1989. The great crash, the oil price shock, and the unit root hypothesis. *Econometrica* 57, 1361–1401.

Perron, P., and Z. Qu. 2010. Long-memory and level shifts in the volatility of stock market return indices. *Journal of Business and Economic Statistics* 28, 275–290.

Qu, Z. 2011. A test against spurious long memory. *Journal of Business and Economic Statistics* 29, 423–438.

Reeves, J., J. Chen, X. L. Wang, R. Lund, and Q. Lu. 2007. A review and comparison of changepoint detection techniques for climate data. *Journal of Applied Meteorology and Climatology* 46, 900–915.

Robinson, P. M. 1997. Large-sample inference for nonparametric regression with dependent errors. *The Annals of Statistics* 25, 2054–2083.

Robinson, P. M., and D. Marinucci. 2001. Narrow-band analysis of nonstationary processes, *Annals of Statistics* 29, 947 - 986.

Shao, X. 2011. A simple test of changes in mean in the possible presence of long-range dependence. *Journal of Time Series Analysis* 32, 598–606.

Shao, X., and W. B. Wu. 2007a. Local whittle estimation of fractional integration for nonlinear processes. *Econometric Theory* 23, 899–929.

Shao, X., and W. B. Wu. 2007b. Local asymptotic powers of nonparametric and semiparametric tests for fractional integration, *Stochastic Processes and their Applications* 117 , 251-261.

Terrin, N., and C. M. Hurvich. 1994. An asymptotic Wiener-Itô representation for the low frequency ordinates of the periodogram of a long memory time series. *Stochastic Processes and their Applications* 54, 297-307.

Teverovsky, V., and M. Taqqu. 1997. Testing for long-range dependence in the presence of shifting means or a slowly declining trend, using a variance-type estimator. *Journal of Time Series Analysis* 18, 279–304.

Wenger, K., C. Leschinski, and P. Sibbertsen. 2019. Change-in-mean tests in long-memory time series: a review of recent developments. *AStA Advances in Statistical Analysis* 103, 237–256.

Wu, W. B., and X. Shao. 2006. Invariance principles for fractionally integrated nonlinear processes. *Lecture Notes-Monograph Series* 50, 20–30.

Table 2: Empirical Size of $R(\delta^{\dagger})$. DGP1-DGP5.

	$\boldsymbol{T = 128}$				$\boldsymbol{T = 512}$		
	$m = T^{\lfloor 0.5 \rfloor}$	$m = T^{\lfloor 0.65 \rfloor}$	$m = T^{\lfloor 0.8 \rfloor}$		$m = T^{\lfloor 0.5 \rfloor}$	$m = T^{\lfloor 0.65 \rfloor}$	$m = T^{\lfloor 0.8 \rfloor}$
			DGP1, $\delta^{\dagger} = 0$				
$l = 2$	0.054	0.057	0.056	$l = 2$	0.050	0.047	0.052
$l = 4$	0.055	0.059	0.056	$l = 4$	0.048	0.050	0.049
$l = 6$	0.054	0.058	0.055	$l = 8$	0.048	0.049	0.053
			DGP2, $\delta^{\dagger} = 0$				
$l = 2$	0.083	0.159	0.314	$l = 2$	0.058	0.095	0.210
$l = 4$	0.094	0.176	0.336	$l = 4$	0.059	0.098	0.212
$l = 6$	0.098	0.195	0.360	$l = 8$	0.061	0.100	0.222
			DGP3, $\delta^{\dagger} = 0$				
$l = 4$	0.051	0.054	0.056	$l = 4$	0.052	0.054	0.061
			DGP4, $\delta^{\dagger} = 0.3$				
$l = 4$	0.053	0.055	0.055	$l = 4$	0.048	0.051	0.051
			DGP5, $\delta^{\dagger} = -0.3$				
$l = 4$	0.055	0.056	0.053	$l = 4$	0.046	0.040	0.040
			DGP4, $\delta^{\dagger} = 0$				
$l = 4$	0.362	0.486	0.592	$l = 4$	0.450	0.609	0.726
			DGP5, $\delta^{\dagger} = 0$				
$l = 4$	0.002	0.000	0.000	$l = 4$	0.000	0.000	0.000

Table 3: Empirical Size of $R(\widehat{\delta})$. DGP1-DGP5.

		$\widehat{\delta}$ (no trimming)			$\widehat{\delta}_2$ (trimming, $l^* = 2$)			$\widehat{\delta}_3$ (trimming, $l^* = 3$)		
Panel A: $T = 128$										
DGP	$l \setminus m$	$T^{\lfloor 0.5 \rfloor}$	$T^{\lfloor 0.65 \rfloor}$	$T^{\lfloor 0.8 \rfloor}$	$T^{\lfloor 0.5 \rfloor}$	$T^{\lfloor 0.65 \rfloor}$	$T^{\lfloor 0.8 \rfloor}$	$T^{\lfloor 0.5 \rfloor}$	$T^{\lfloor 0.65 \rfloor}$	$T^{\lfloor 0.8 \rfloor}$
	2	0.053	0.053	0.050	0.059	0.063	0.062	0.066	0.071	0.074
1	4	0.050	0.053	0.049	0.058	0.063	0.062	0.070	0.074	0.077
	6	0.049	0.053	0.050	0.058	0.063	0.062	0.070	0.074	0.077
	2	0.011	0.008	0.043	0.013	0.014	0.021	0.012	0.013	0.019
2	4	0.010	0.008	0.045	0.012	0.014	0.021	0.012	0.014	0.019
	6	0.011	0.009	0.047	0.013	0.014	0.022	0.012	0.013	0.019
3	4	0.050	0.049	0.047	0.056	0.058	0.058	0.066	0.073	0.076
4	4	0.046	0.043	0.047	0.053	0.055	0.051	0.057	0.057	0.054
5	4	0.051	0.054	0.057	0.072	0.081	0.095	0.081	0.094	0.110
Panel B: $T = 512$										
DGP	$l \setminus m$	$T^{\lfloor 0.5 \rfloor}$	$T^{\lfloor 0.65 \rfloor}$	$T^{\lfloor 0.8 \rfloor}$	$T^{\lfloor 0.5 \rfloor}$	$T^{\lfloor 0.65 \rfloor}$	$T^{\lfloor 0.8 \rfloor}$	$T^{\lfloor 0.5 \rfloor}$	$T^{\lfloor 0.65 \rfloor}$	$T^{\lfloor 0.8 \rfloor}$
	2	0.041	0.047	0.044	0.044	0.043	0.042	0.050	0.050	0.052
1	4	0.044	0.046	0.043	0.041	0.043	0.041	0.049	0.052	0.053
	6	0.036	0.048	0.044	0.039	0.042	0.039	0.047	0.051	0.053
	2	0.016	0.014	0.062	0.016	0.016	0.031	0.015	0.014	0.032
2	4	0.014	0.015	0.062	0.015	0.015	0.031	0.014	0.014	0.030
	6	0.013	0.014	0.065	0.017	0.015	0.032	0.016	0.014	0.029
3	4	0.050	0.047	0.047	0.048	0.044	0.041	0.052	0.053	0.050
4	4	0.044	0.046	0.041	0.046	0.047	0.046	0.049	0.053	0.054
5	4	0.042	0.048	0.050	0.046	0.051	0.055	0.049	0.060	0.067

Table 4: Empirical Power of $R(\delta^\dagger)$. DGP1-DGP5. Single Level Break.

	$\boldsymbol{T=128}$				$\boldsymbol{T=512}$		
	$m=T^{\lfloor 0.5\rfloor}$	$m=T^{\lfloor 0.65\rfloor}$	$m=T^{\lfloor 0.8\rfloor}$		$m=T^{\lfloor 0.5\rfloor}$	$m=T^{\lfloor 0.65\rfloor}$	$m=T^{\lfloor 0.8\rfloor}$
DGP1, $\delta^\dagger=0$							
$l=2$	0.515	0.568	0.593	$l=2$	0.992	0.997	0.998
$l=4$	0.509	0.573	0.595	$l=4$	0.993	0.998	0.998
$l=6$	0.486	0.571	0.597	$l=8$	0.994	0.997	0.998
DGP2, $\delta^\dagger=0$							
$l=4$	0.236	0.380	0.574	$l=4$	0.594	0.716	0.879
DGP3, $\delta^\dagger=0$							
$l=4$	0.309	0.343	0.357	$l=4$	0.875	0.891	0.904
DGP4, $\delta^\dagger=0.3$							
$l=4$	0.128	0.135	0.128	$l=4$	0.182	0.196	0.194
DGP5, $\delta^\dagger=-0.3$							
$l=4$	1.000	1.000	1.000	$l=4$	1.000	1.000	1.000
DGP4, $\delta^\dagger=0.0$							
$l=4$	0.494	0.616	0.711	$l=4$	0.697	0.811	0.877
DGP5, $\delta^\dagger=0.0$							
$l=4$	0.850	0.785	0.534	$l=4$	1.000	1.000	1.000

Table 5: Empirical Power of $R(\widehat{\delta})$. DGP1-DGP5. Single Level Break.

		$\widehat{\delta}$ (no trimming)			$\widehat{\delta}_2$ (trimming, $l^* = 2$)			$\widehat{\delta}_3$ (trimming, $l^* = 3$)		
Panel A: $T = 128$										
DGP	$l \setminus m$	$T^{\lfloor 0.5 \rfloor}$	$T^{\lfloor 0.65 \rfloor}$	$T^{\lfloor 0.8 \rfloor}$	$T^{\lfloor 0.5 \rfloor}$	$T^{\lfloor 0.65 \rfloor}$	$T^{\lfloor 0.8 \rfloor}$	$T^{\lfloor 0.5 \rfloor}$	$T^{\lfloor 0.65 \rfloor}$	$T^{\lfloor 0.8 \rfloor}$
	2	0.276	0.287	0.157	0.297	0.296	0.256	0.273	0.264	0.230
1	4	0.258	0.286	0.154	0.297	0.298	0.257	0.254	0.255	0.222
	6	0.234	0.284	0.148	0.292	0.300	0.254	0.247	0.254	0.222
2	4	0.026	0.026	0.083	0.033	0.037	0.050	0.030	0.033	0.041
3	4	0.179	0.196	0.116	0.205	0.206	0.180	0.189	0.192	0.176
4	4	0.078	0.068	0.064	0.092	0.094	0.083	0.088	0.086	0.079
5	4	0.745	0.830	0.266	0.835	0.846	0.721	0.720	0.706	0.586
Panel B: $T = 512$										
DGP	$l \setminus m$	$T^{\lfloor 0.5 \rfloor}$	$T^{\lfloor 0.65 \rfloor}$	$T^{\lfloor 0.8 \rfloor}$	$T^{\lfloor 0.5 \rfloor}$	$T^{\lfloor 0.65 \rfloor}$	$T^{\lfloor 0.8 \rfloor}$	$T^{\lfloor 0.5 \rfloor}$	$T^{\lfloor 0.65 \rfloor}$	$T^{\lfloor 0.8 \rfloor}$
	2	0.878	0.895	0.540	0.891	0.895	0.770	0.850	0.833	0.688
1	4	0.867	0.895	0.534	0.887	0.895	0.771	0.836	0.829	0.688
	6	0.845	0.886	0.526	0.884	0.885	0.767	0.824	0.829	0.681
2	4	0.170	0.189	0.393	0.200	0.217	0.309	0.170	0.173	0.252
3	4	0.663	0.688	0.442	0.682	0.695	0.577	0.641	0.634	0.517
4	4	0.127	0.131	0.113	0.138	0.141	0.170	0.158	0.163	0.160
5	4	1.000	1.000	0.621	1.000	1.000	0.998	1.000	1.000	0.987

Supplementary Appendix to “Nonparametric Detection of a Time-Varying Mean”

Fabrizio Iacone[a] and A.M. Robert Taylor[b]

[a] Università degli Studi di Milano and University of York

[b] Essex Business School, University of Essex

June 2025

Abstract

This supplementary appendix contains two sections. Section S.1 provides proofs of the technical results stated in the main paper. Section S.2 provides the additional Monte Carlo results referred to in Section 5.3 of the main paper.

S.1 Proofs

We now provide derivations of the key large sample results stated in the paper, with accompanying regularity conditions where needed. In what follows we use C as generic notation for a finite constant.

Lemma S.1. *(i) If $\mu_t = \mu + \beta DU_t(\tau^*)$, for $j = o(T)$,*

$$I_{\mu\mu}(\lambda_j) = T\frac{\beta^2}{4\pi^3} j^{-2}\left(1 - \cos\left(2\pi j \tau^*\right)\right) + O(\frac{j}{T})$$

(ii) If $\mu_t = \mu + \beta h(t/T)$, where $h(s)$ is a Lipschitz continuous function on $[0, 1]$, $h(s) \neq h(r)$ for some $s \neq r$, and $\beta \neq 0$, then $I_{\mu\mu}(\lambda_j) \leq C{\lambda_j}^{-1} j^{-1}$ for all j in $\{1,\ ...,\ \lfloor T/2 \rfloor\}$.

(iii) For μ_t as in part (ii), and such that $1/T\sum\mu_t^2 - \left(1/T\sum\mu_t\right)^2 \to \kappa$ for some $\kappa > 0$, there is a frequency λ_j, $j \in \{1,\ ...,\ \iota\}$ with $1/\iota + \iota/T \to 0$ as $T \to \infty$ such that $I_{\mu\mu}(\lambda_j) > CT^{1-\epsilon}$ for some $C > 0$.

(iv) For $\mu_t = \mu + \beta t^{\varphi - 1/2}$, with $\varphi \in (0, 1/2)$, $\beta \neq 0$, then $I_{\mu\mu}(\lambda_1) = O_e(T^{2\varphi})$.

Proof. (i) Using standard properties of trigonometric functions at frequencies λ_j we have that

$$w_\mu(\lambda_j) = \frac{1}{\sqrt{2\pi T}}\left(\sum_{t=1}^{T}\mu e^{-i\lambda_j t} + \sum_{t=\lfloor\tau^* T\rfloor+1}^{T}\beta e^{-i\lambda_j t}\right) = -\frac{1}{\sqrt{2\pi T}}\beta\left(\sum_{t=1}^{\lfloor\tau^* T\rfloor} e^{-i\lambda_j t}\right)$$

$$I_{\mu\mu}(\lambda_j) = \frac{1}{2\pi T}\beta^2\left\{\left(\sum_{t=1}^{\lfloor\tau^* T\rfloor}\cos(\lambda_j t) + i\sum_{t=1}^{\lfloor\tau^* T\rfloor}\sin(\lambda_j t)\right) + \left(\sum_{t=1}^{\lfloor\tau^* T\rfloor}\cos(\lambda_j t) - i\sum_{t=1}^{\lfloor\tau^* T\rfloor}\sin(\lambda_j t)\right)\right\}$$

and

$$\frac{2\pi T}{\beta^2} I_{\mu\mu}(\lambda_j) = \left(\sum_{t=1}^{\lfloor\tau^* T\rfloor}\cos(\lambda_j t)\right)^2 + \left(\sum_{t=1}^{\lfloor\tau^* T\rfloor}\sin(\lambda_j t)\right)^2.$$

Next, using standard trigonometric identities we have that

$$\sum_{t=1}^{r}\cos(\theta t) = \frac{\sin\left(r + \frac{1}{2}\right)\theta - \sin\left(\frac{1}{2}\theta\right)}{2\sin\left(\frac{1}{2}\theta\right)}, \quad \sum_{t=1}^{r}\sin(\theta t) = \frac{\cos\left(\frac{1}{2}\theta\right) - \cos\left(r + \frac{1}{2}\right)\theta}{2\sin\left(\frac{1}{2}\theta\right)}$$

see Robinson (1995b), page 1845, and the reference therein to Zygmund (1977). Consequently,

$$\begin{aligned}
\left(\sum_{t=1}^{r}\cos(\theta t)\right)^2 &+ \left(\sum_{t=1}^{r}\sin(\theta t)\right)^2 \\
&= \frac{\left(\sin\left(r+\frac{1}{2}\right)\theta\right)^2+\left(\sin\left(\frac{1}{2}\theta\right)\right)^2-2\left(\sin\left(r+\frac{1}{2}\right)\theta\right)\left(\sin\left(\frac{1}{2}\theta\right)\right)}{\left(2\sin\left(\frac{1}{2}\theta\right)\right)^2} \\
&+\frac{\left(\cos\left(r+\frac{1}{2}\right)\theta\right)^2+\left(\cos\left(\frac{1}{2}\theta\right)\right)^2-2\left(\cos\left(r+\frac{1}{2}\right)\theta\right)\left(\cos\left(\frac{1}{2}\theta\right)\right)}{\left(2\sin\left(\frac{1}{2}\theta\right)\right)^2} \\
&= \frac{1+1-2\cos\left(\left(r+\frac{1}{2}\right)\theta-\frac{1}{2}\theta\right)}{\left(2\sin\left(\frac{1}{2}\theta\right)\right)^2} = \frac{2-2\cos(r\theta)}{\left(2\sin\left(\frac{1}{2}\theta\right)\right)^2}
\end{aligned}$$

using the trigonometric identities, $(\cos(\omega))^2+(\sin(\omega))^2=1$ and $\cos(\alpha-\omega)=\cos(\alpha)\cos(\omega)+\sin(\alpha)\sin(\omega)$. Therefore, for $\lambda_j=\theta$, $\lfloor\tau^*T\rfloor=r$,

$$\begin{aligned}
I_{\mu\mu}(\lambda_j) &= \frac{\beta^2}{2\pi T}\frac{1-\cos(\lfloor\tau^*T\rfloor\lambda_j)}{2\left(\sin\left(\frac{1}{2}\lambda_j\right)\right)^2} \\
&= \frac{\beta^2}{2\pi T}\frac{1}{\left(\frac{1}{2}\lambda_j\right)^2}\frac{\left(\frac{1}{2}\lambda_j\right)^2}{2\left(\sin\left(\frac{1}{2}\lambda_j\right)\right)^2}\left(1-\cos(\lfloor\tau^*T\rfloor\lambda_j)\right) \\
&= T\frac{\beta^2}{4\pi^3 j^2}\frac{\left(\frac{1}{2}\lambda_j\right)^2}{\left(\sin\left(\frac{1}{2}\lambda_j\right)\right)^2}\left(1-\cos(\lfloor\tau^*T\rfloor\lambda_j)\right).
\end{aligned}$$

Finally, a standard Taylor series expansion for $\sin(\lambda)$ in a neighbourhood of 0 gives

$$\sin(\frac{1}{2}\lambda_j)=\frac{1}{2}\lambda_j+O\left((\lambda_j)^3\right)=\frac{1}{2}\lambda_j+O\left((\frac{j}{T})^3\right)$$

and

$$\cos(\lfloor\tau^*T\rfloor\lambda_j)=\cos(\tau^*T\lambda_j)+(\tau^*T\lambda_j-\lfloor\tau^*T\rfloor\lambda_j)\sin(\lfloor\tau^*T\rfloor|_{mvt}\lambda_j)=\cos(\tau^*T\lambda_j)+O(\frac{j}{T^2})$$

where $\lfloor\tau^*T\rfloor\leq\lfloor\tau^*T\rfloor|_{mvt}\leq\tau^*T$, and we used the bounds $\sin(\lfloor\tau^*T\rfloor|_{mvt}\lambda_j)=O(1)$, $(\tau^*T-\lfloor\tau^*T\rfloor)=O(\frac{1}{T})$.

(ii) This bound result is only stated to be local in Lemma 1 of Qu (2011), but using the same arguments and the fact that $(\pi/\lambda_j+1)<3\pi/\lambda_j$ is sufficient to establish the bound in general.

(iii) From

$$\frac{1}{T}\sum_{t}\left(\mu_t - (1/T\sum_{t}\mu_t)\right)^2 = 2\frac{1}{T}2\pi\sum_{j=1}^{T/2} I_{\mu\mu}(\lambda_j)$$

when T is even, and

$$\frac{1}{T}2\pi\sum_{j=1}^{\iota} I_{\mu\mu}(\lambda_j) = \frac{1}{T}2\pi\sum_{j=1}^{T/2} I_{\mu\mu}(\lambda_j) - \frac{1}{T}2\pi\sum_{j=\iota+1}^{T/2} I_{\mu\mu}(\lambda_j)$$

then

$$\frac{1}{T}2\pi\sum_{j=1}^{\iota} I_{\mu\mu}(\lambda_j) = 1/2\kappa + o(1) + O(\iota^{-1}). \tag{S.1}$$

Notice that this implies that $I(\lambda_j)$ is at least diverging with rate $T/(\iota^{1+\epsilon})$. To see this, suppose that it is not true, and $\max I_{\mu\mu}(\lambda_j) < CT/(\iota^{1+\epsilon})$ This would imply that $\frac{1}{T}2\pi\sum_{j=1}^{\iota} I_{\mu\mu}(\lambda_j) = o(1)$, which would be in contrast with (S.1) so there must be a $I_{\mu\mu}(\lambda_j)$ which is at least diverging with order $T/(\iota^{1+\epsilon})$. The conclusion follows on taking $\iota = \lfloor T^{\epsilon/2}\rfloor$.

(iv) For $j = 1$, we can rewrite the periodogram as

$$\begin{aligned} I_{\mu\mu}(\lambda_1) &= \frac{1}{2\pi}T^{2\varphi}\left\{\left(\frac{1}{T}\sum_{t=1}^{T}(t/T)^{\varphi-1/2}\cos(2\pi t/T)\right)^2 + \left(\frac{1}{T}\sum_{t=1}^{T}(t/T)^{\varphi-1/2}\sin(2\pi t/T)\right)^2\right\} \\ &= \frac{1}{2\pi}T^{2\varphi}\left\{\left(\int_0^1 s^{\varphi-1/2}\cos(2\pi s)ds\right)^2 + \left(\int_0^1 s^{\varphi-1/2}\sin(2\pi s)ds\right)^2\right\} + o(T^{2\varphi}) \end{aligned}$$

using integral approximation. □

Note: the exact bound for the periodogram of an abrupt change in the mean has previously been presented in Leschinski and Sibbertsen (2018), using a different method of proof.

Proof of Theorem 1.

(i) Let $W_{\delta+1}(r)$ denote a Type I fractional Brownian motion, as defined in Mandelbrot and

Van Ness (1968) and Marinucci and Robinson (1999), and define

$$\begin{aligned}
\widehat{W}_{\delta+1}(r) &:= W_{\delta+1}(r) - rW_{\delta+1}(1) \\
\kappa(\delta)^2 &:= \Gamma\left(1-2\delta\right) / \left[\left(1+2\delta\right)\Gamma\left(1+\delta\right)\Gamma\left(1-\delta\right)\right] \\
Q_\delta\left(j\right) &:= \left\{\left(2\pi j \int_0^1 \sin\left(2\pi jr\right)\widehat{W}_{\delta+1}\left(r\right)dr\right)^2 + \left(2\pi j \int_0^1 \cos\left(2\pi jr\right)\widehat{W}_{\delta+1}\left(r\right)dr\right)^2\right\}.
\end{aligned}$$

We will prove the result that

$$R(\delta) \xrightarrow{d} (2\pi)^{2\delta}\kappa(\delta)Q_\delta(1) \tag{S.2}$$

We establish this limiting result under the regularity conditions stated in Assumption A1, but we note that this includes the possibility that ξ_t is a linear process, or indeed linear and Gaussian, and so the expression also follows when ξ_t is a Gaussian linear process. Thus, the expression in (S.2) is equivalent to the limit given in Hurvich and Beltrao (1993).

We discuss the limits for the numerator and the denominator in turn. First, for the numerator, we note that, under Assumption A1, the FCLT

$$\frac{1}{T^{1/2+\delta}}\sum_{t=1}^{\lfloor rT\rfloor}\xi_t \Rightarrow \left(2\pi f_{\eta\eta}(0)\right)^{(1/2)}\kappa(\delta)W_{\delta+1}(r)$$

holds, where $f_{\eta\eta}(\lambda)$ is the spectral density of η_t; see Wu and Shao (2006). Notice that, strictly speaking, Wu and Shao (2006) do not include the case where $\delta = 0$ in the statement of their Theorem 2.1; however the result for $\delta = 0$ follows from Theorem 3 coupled with assumption (ii), both of Wu (2007).

Next, the convergence result

$$2\pi T^{-2\delta}I_{\xi\xi}(\lambda_j) \Rightarrow 2\pi f_{\eta\eta}(0)\kappa(\delta)^2 Q_\delta(j) \tag{S.3}$$

follows from Lemma 1 of Hualde and Iacone (2017). Noting that $I_{xx}(\lambda_j) = I_{\xi\xi}(\lambda_j)$, under the null hypothesis that μ_t is constant, we readily obtain from (S.3) that

$$(\lambda_1)^{2\delta} I_{xx}(\lambda_1) \Rightarrow (2\pi)^{2\delta} f_{\eta\eta}(0) Q_\delta(1)\kappa(\delta)^2.$$

Turning to the denominator, and recalling that $I_{xx}(\lambda_j) = I_{\xi\xi}(\lambda_j)$ under the null hypothesis, we then establish that

$$\frac{1}{m-l+1}\sum_{j=l}^{m}(\lambda_j)^{2\delta} I_{\xi\xi}(\lambda_j) \xrightarrow{p} f_{\eta\eta}(0) \tag{S.4}$$

as $l/m+m/T \to 0$. For the linear case, under Assumption A2 the convergence in (S.4) follows using the bounds in Theorem 2 of Robinson (1995a) and a law of large number argument as in the proof of Theorem 1 of Robinson (1995b), pages 1636=1638. For the nonlinear case, we note that Assumption A.1 in Lemma A.1 and Lemma A.2 of Shao and Wu (2007a) corresponds to part (ii) of our Assumption A2. The other assumption made for Lemma A.2 in Shao and Wu (2007a) is that $\sum_{k\in\mathbf{Z}}|\gamma_\eta(k)| < \infty$. Using $\gamma_\eta(k) = \sum_{j\in\mathbf{Z}} E(\mathcal{P}_j\eta_0\mathcal{P}_j\eta_k)$, $\sum_{k\in\mathbf{Z}}|\gamma_\eta(k)| \le \sum_{k\in\mathbf{Z}}\sum_{j\in\mathbf{Z}}\|\mathcal{P}_j\eta_0\|_2\|\mathcal{P}_j\eta_k\|_2$ by the Chauchy-Schwarz inequality, as in Lemma A.1 of Shao and Wu (2007a), then $\sum_{k\in\mathbf{Z}}|\gamma_\eta(k)| < \infty$ is implied by Assumption A1. Thus, both the assumptions for Lemma A.2 of Shao and Wu (2007a) are met. The bounds in Lemma A.2 of Shao and Wu (2007a) can be used in the same way as those of Theorem 2 of Robinson (1995a), to obtain (S.4).

(ii) The replacement of δ by $\widehat{\delta}$ is accounted for by using a mean value theorem expansion. For the numerator, to establish $(\lambda_1)^{2\widehat{\delta}} I_{\xi\xi}(\lambda_1) - (\lambda_1)^{2\delta} I_{\xi\xi}(\lambda_1) = o_p(1)$ it is sufficient to use a mean value expansion $T^{2\widehat{\delta}}$ around $T^{2\delta}$ up to the n-th term, $n > 2/\epsilon$. The denominator can be treated in a similar way.

Proof of Theorem 2.

Denoting $w^*_\xi(\lambda)$ as the complex conjugate of $w_\xi(\lambda)$ and $I\mu\xi(\lambda) = w_\mu(\lambda)w^*_\xi(\lambda)$ as the cross periodogram, we can write

$$I_{xx}(\lambda) = I_{\mu\mu}(\lambda) + I_{\mu\xi}(\lambda) + I_{\xi\mu}(\lambda) + I_{\xi\xi}(\lambda).$$

From the assumption on the order of $I(\lambda_1)$, the properties of the periodogram, and the Cauchy Schwarz inequality, we have that

$$I_{\mu\mu}(\lambda_1) = O_e(T^{2\phi}), \quad I_{\xi\xi}(\lambda_1) = O_e(T^{2\delta}), \quad I_{\mu\xi}(\lambda_1) = O_p(T^{(\phi+\delta)})$$

and so

$$\lambda_1^{2\delta} I_{xx}(\lambda_1) = \lambda_1^{2\delta} I_{\mu\mu}(\lambda_1) + o_p(T^{2(\phi-\delta)}) \text{ if } \phi > \delta$$
$$\lambda_1^{2\delta} I_{xx}(\lambda_1) = \lambda_1^{2\delta} I_{\xi\xi}(\lambda_1) + o_p(1) \text{ if } \phi < \delta$$

Similarly, $\frac{1}{m-l+1}\sum_{j=l}^{m} \lambda^{2\delta} I_{\xi\xi}(\lambda_j) = G + o_p(1)$, and

$$\frac{1}{m-l+1}\sum_{j=l}^{m} \lambda_j^{2\delta} I_{\mu\mu}(\lambda_j) = O_e(\frac{1}{m} T^{2(\phi-\delta)} \sum_{j=l}^{m} j^{2(\delta-\phi)-1}). \tag{S.5}$$

The latter is $O((m/T)^{2(\delta-\phi)} m^{-1}) = o(1)$ if $\delta > \phi$. Otherwise, the bound in (S.5) is $O_e(T^{2(\phi-\delta)}\ m^{-1}\ l^{2(\delta-\phi)})$. In both cases, bounds for $\frac{1}{m-l+1}\sum_{j=l}^{m} \lambda_j^{2\delta} I_{\mu\xi}(\lambda_j)$ follow from the Cauchy Schwarz inequality. The result follows combining these bounds. For example, in the $\phi = 0.5$, $\delta = 0$ case,

$$I_{\mu\mu}(\lambda_1) = O_e(T), \quad I_{\xi\xi}(\lambda_1) = O_e(1), \quad I_{\mu\xi}(\lambda_1) = O_p(T^{\frac{1}{2}})$$
$$\frac{1}{m-l+1}\sum_{j=l}^{m} I_{\mu\mu}(\lambda_j) = O(\frac{1}{m} T \sum_{j=l}^{m} j^{-2}) = O(\frac{T}{lm})$$
$$\frac{1}{m-l+1}\sum_{j=l}^{m} I_{\xi\xi}(\lambda_j) = O_p(1)$$
$$\frac{1}{m-l+1}\sum_{j=l}^{m} I_{\mu\xi}(\lambda_j) = O_p((\frac{T}{lm})^{\frac{1}{2}})$$

and so $I_{\mu\mu}(\lambda_1) = O_e(T)$, and hence

$$\text{if } \frac{T}{lm} \to \infty, \quad \frac{1}{m-l+1}\sum_{j=l}^{m} I_{xx}(\lambda_j) = O_e(\frac{T}{lm})$$
$$\text{if } \frac{T}{lm} \to 0, \quad \frac{1}{m-l+1}\sum_{j=l}^{m} I_{xx}(\lambda_j) = O_e(1).$$

from which the result in Corollary 1 follows.

S.2 Additional Monte Carlo Results - Comparisons with Extant Tests

In the last part of our Monte Carlo exercise, we have run a comparative study of the finite sample size and power properties of the $R(\delta)$ and $R(\widehat{\delta})$ tests against a set of benchmark tests from the extant literature. In doing so, for $R(\delta)$ and $R(\widehat{\delta})$ we follow the recommended settings for the tuning parameters given at the end of Section 5.2, setting $m = \lfloor T^{0.55} \rfloor$.

The two most natural comparator tests are the W test of Qu (2011), discussed in Remark 3, and the $T_n(\widehat{\delta})$ test of Giraitis *et al.* (2006). The latter is a modification of the V/S test of Giraitis *et al.* (2001) to allow for long memory in x_t. Like the tests we propose, both of these tests are designed to detect general forms of non-constancy in μ_t and allow for long memory in ξ_t. A further test we compare against is the SW test of Iacone *et al.* (2014), which is based on fixed-b asymptotics. The SW test, in contrast to our proposed tests and the tests of Qu (2011) and Giraitis *et al.* (2006), is specifically designed to detect a single deterministic level break. In a large Monte Carlo exercise, Wenger *et al.* (2019) find that the SW test generally outperforms other available tests for a level shift in long memory series. It therefore provides a useful benchmark for power in the single level shift case.

In what follows, we will refer to these three benchmark tests, with an obvious notation, as W, VS, and SW. To distinguish between cases where δ is estimated and where it is not, when evaluated at a user-chosen value of the long memory parameter, $\delta^{\dagger}$, we denote these tests as $W(\delta^{\dagger})$, $VS(\delta^{\dagger})$ and $SW(\delta^{\dagger})$, and we denote the corresponding tests based on an estimate of δ as $W(\widehat{\delta})$, $VS(\widehat{\delta})$ and $SW(\widehat{\delta})$. All of these tests require choices to be made of tuning parameters: in all cases, we follow the recommended settings given by the authors of the tests. In particular: for the W test the trimming parameter ε is set to 0.05 and the highest frequency m, is set to $\lfloor T^{0.7} \rfloor$, respectively; for the VS test the bandwidth parameter is set as $q = \lfloor T^{1/3} \rfloor$; finally for the SW test the scale parameter for the bandwidth is set as $b = 0.1$.

All of the tests under consideration depend on δ: the SW and the VS for the selection of the critical value (since the limiting null distribution depends on δ), and, in common with our proposed tests, the VS and W tests for the computation of the test statistic. With the exception of the W test, all of the tests can validly be computed at either a user specified value $\delta^{\dagger}$ or at an estimate of δ. In the case of the W statistic of Qu (2011), the use of the (untrimmed) LW estimate is crucial to the derivation of its limiting null distribution. Where an estimate of δ is used, the SW and VS tests, like our $R(\widehat{\delta})$ test, only require a consistent (subject to minimum consistency rates) estimate of δ. In all of the cases where δ is estimated we use the LW estimate of the form in (20) with $m^* = \lfloor T^{0.65} \rfloor$, except that for the W test, for which, following the recommendation in Qu (2011), $m^* = \lfloor T^{0.7} \rfloor$ is used. As in the experiments reported in Sections 5.1 and 5.2, we bound the support for the estimation of δ to $[-0.49, 0.49]$, except that for the VS test where we follow the authors' recommendation of $[-0.45, 0.45]$.

In Table S.1, we first report the finite sample size properties of the tests against DGP1-DGP5 from Section 5.1. The results for the W test under known δ are where we replace the LW estimate $\widehat{\delta}$ in the W statistic by $\delta^{\dagger}$, noting that this test will not have the same limiting null distribution as for the case where it is evaluated at the LW estimate.

From the results in Table S.1 we see that the SW test has good size properties, both for the known and estimated δ cases. The only exceptions occurs, as expected, in cases where the user specifies a value for $\delta^{\dagger}$ which is different from the true value of δ. Even in such cases, however, the size distortion is the smallest of the three tests, suggesting that the distorting effect due to an imprecise estimate of δ is lowest for this test. This is confirmed by the performances of the VS and W tests: the VS test is subject to some potentially large size distortion even in the larger sample size, when δ is assumed known, at least when the spectral density of η_t is subject to some curvature, as in the AR(1) case or when $\delta = 0.3$. The size performance of the VS test is improved if the LW estimate is used, but is still significantly over-sized for DGP4 where $\delta = 0.3$. Finally, the results verify the invalidity of the W when based on an assumed value of δ, even where the correct value of δ is assumed.

Next we look at the finite sample power of these tests against a variety of time-varying μ_t DGPs. We exclude $W(\delta^\dagger)$ from this analysis given its lack of size control. To do so, we simulate data according to (1)-(2) with $\eta_t \sim IID\, N(0,1)$, for both $\delta = 0$ and $\delta = 0.3$. For the time-varying mean component, μ_t, we considered a range of models, as follows. For the case where $\delta = 0$ we consider:

- DGP-P0. $\mu_t = 0$, $t = 1, ..., T$;
- DGP-P1. $\mu_t = \beta DU_t(\tau^*)$, with $\beta = 0.5$ and $\tau^* = 0.5$;
- DGP-P2. $\mu_t = \beta DU_t(\tau^*)$, with $\beta = 0.25$ and $\tau^* = 0.5$;
- DGP-P3. $\mu_t = \beta DU_t(\tau^*)$, with $\beta = 0.5$ and $\tau^* = 0.75$;
- DGP-P4. $\mu_t = \beta \times \phi(a_t)$ for $a_t = t/T - 1/2 \in [-1/2, 1/2]$, where $\phi(a)$ is the pdf of a standard normal, with $\beta = 3$;
- DGP-P5. $\mu_t = -\beta_1(a_t) + \beta_2 \times \Phi(a_t)$ for $a_t \in [-5, 5]$, where $\Phi(a)$ is the cdf of a standard normal, with $\beta_1 = 1$ and $\beta_2 = 0.25$;
- DGP-P6. $\mu_t = \beta\Phi(a_t)$ for $a_t \in [-4, 0]$, with $\beta = 1$
- DGP-P7. $\mu_t = \frac{1}{\sqrt{5}} \sum_{s=1}^{t} \beta_s \epsilon_s$ where β_t IID $B(1, 6.10/T)$, $\epsilon_t \sim IID\, N(0,1)$, β_t and ϵ_t mutually independent, and where $B(1, p)$ denotes a Bernoulli random variable that takes the value 1 with probability p, and 0 with probability $1 - p$.

For $\delta = 0.3$ the same DGPs were used but the parameters β, β_1 and β_2 were each multiplied by three, reflective of the fact that it is more difficult to detect non-constancy in μ_t for $\delta = 0.3$ *vis-à-vis* $\delta = 0$. We label these DGPs, DGP-P0*,...,DGP-P7* in the tables.

Plots of six of the non-constant specifications of μ_t, DGP-P1-DGP-P6, are graphed in Figure S.1 (where the horizontal time axis is scaled to [0,1]), with the corresponding periodograms of the μ_t component, for $T = 128$, graphed in Figure S.2.

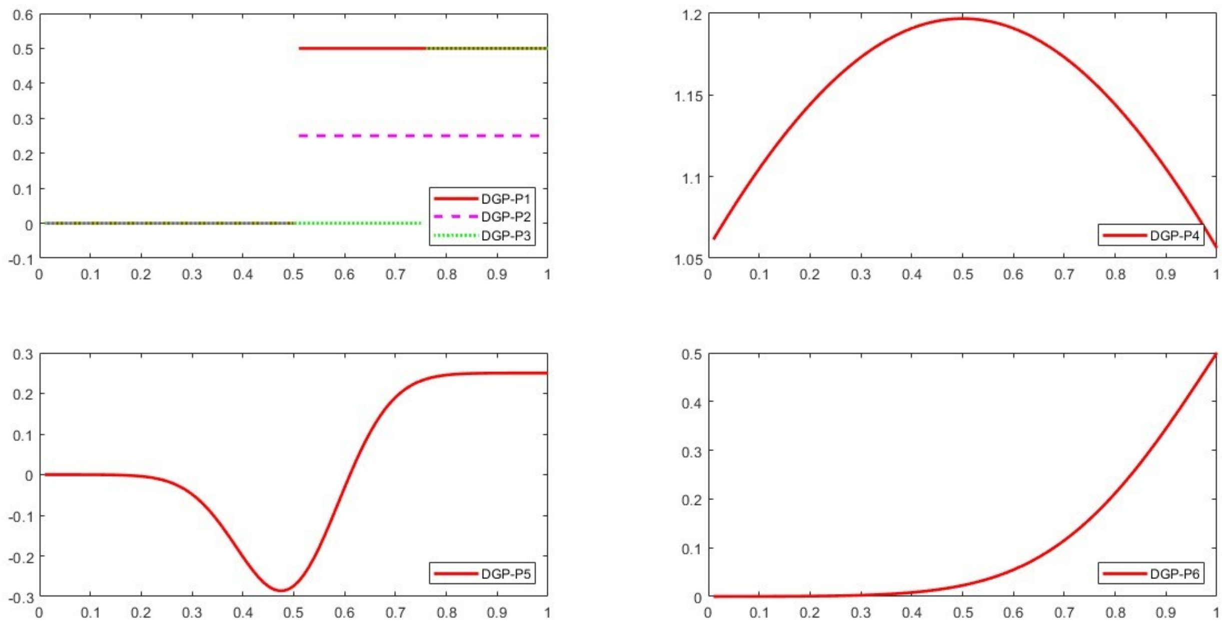

Figure S.1: Plots of μ_t for DGP-P1-DGP-P6.

DGP-P0 sets μ_t to be a zero constant, and provides a size benchmark when comparing the powers of the various tests. DGP-P1-DGP-P3 are single abrupt (deterministic) level shift models, with varying break location and break magnitude parameters. DGP-P4-DGP-P6 are models where μ_t displays a smoothly varying trend, of varying shape, over the sample. In DGP-P7, μ_t follows a martingale process, specified such that when $\delta = 0$ the signal-to-noise ratio associated with μ_t coincides with that of model 1 of Qu (2011, p.430). We note that the SW test is specifically designed to detect single abrupt level shifts and so would be expected to be most competitive for DGP-P1-DGP-P3.

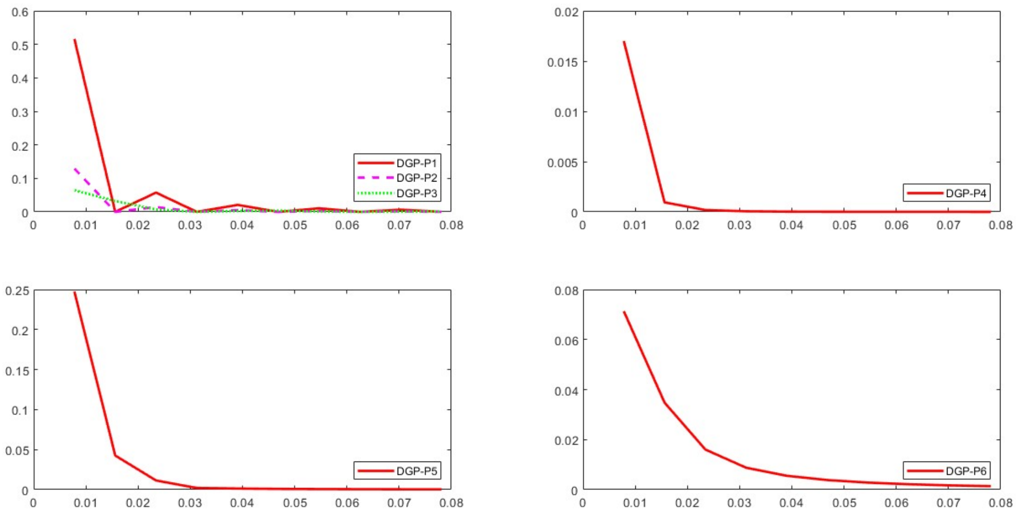

Figure S.2: Plots of the periodograms of μ_t, λ_1,...,$\lambda_{\lfloor T^{1/2} \rfloor}$, $T = 128$, DGP-P1-DGP-P6

Results relating to the case where the practitioner correctly sets $\delta^\dagger = \delta$ are reported in Table S.2, while Table S.3 reports the corresponding results for the case where δ is estimated. In the case of the $R(\widehat{\delta})$ test we report results based on the LW estimate with either no trimming ($l^* = 1$) denoted $R(\widehat{\delta})$, trimming with $l^* = 2$ denoted $R(\widehat{\delta}_2)$, and trimming with $l^* = 3$ denoted $R(\widehat{\delta}_3)$. A summary of the findings from these results is provided in section 5.3 of the main text.

Table S.1: Empirical Sizes of SW, W and VS Tests.

Panel A: Tests evaluated at $\delta^\dagger$

	$T=128$			$T=512$		
	$SW(\delta^\dagger)$	$W(\delta^\dagger)$	$VS(\delta^\dagger)$	$SW(\delta^\dagger)$	$W(\delta^\dagger)$	$VS(\delta^\dagger)$
DGP1, $\delta^\dagger=0$	0.032	0.339	0.031	0.045	0.385	0.042
DGP2, $\delta^\dagger=0$	0.051	0.826	0.098	0.046	0.857	0.086
DGP3, $\delta^\dagger=0$	0.032	0.372	0.026	0.040	0.399	0.042
DGP4, $\delta^\dagger=0.3$	0.041	0.348	0.054	0.048	0.391	0.049
DGP5, $\delta^\dagger=-0.3$	0.030	0.352	0.088	0.044	0.401	0.105
DGP4, $\delta^\dagger=0$	0.193	0.916	0.367	0.201	0.999	0.642
DGP5, $\delta^\dagger=0$	0.001	0.926	0.000	0.001	1.000	0.000

Panel B: Tests based on (untrimmed) LW estimate $\widehat{\delta}$

	$T=128$			$T=512$		
	$SW(\widehat{\delta})$	$W(\widehat{\delta})$	$VS(\widehat{\delta})$	$SW(\widehat{\delta})$	$W(\widehat{\delta})$	$VS(\widehat{\delta})$
DGP1	0.032	0.009	0.013	0.043	0.024	0.022
DGP2	0.011	0.020	0.013	0.023	0.108	0.006
DGP3	0.033	0.009	0.014	0.043	0.028	0.020
DGP4	0.039	0.010	0.138	0.046	0.026	0.053
DGP5	0.021	0.017	0.028	0.038	0.026	0.048

Table S.2: Empirical Power of $R(\delta)$, $SW(\delta)$ and $VS(\delta)$ Tests.

Panel A: $\delta = 0$

	$T = 128$			$T = 512$		
	$R(\delta)$	$SW(\delta)$	$VS(\delta)$	$R(\delta)$	$SW(\delta)$	$VS(\delta)$
DGP-P0	0.057	0.032	0.031	0.050	0.045	0.042
DGP-P1	0.537	0.495	0.485	0.996	0.988	0.994
DGP-P2	0.170	0.137	0.128	0.584	0.519	0.575
DGP-P3	0.277	0.365	0.251	0.880	0.947	0.915
DGP-P4	0.071	0.037	0.043	0.119	0.053	0.101
DGP-P5	0.283	0.160	0.243	0.871	0.512	0.882
DGP-P6	0.235	0.170	0.114	0.764	0.621	0.526
DGP-P7	0.555	0.499	0.550	0.834	0.672	0.852

Panel B: $\delta = 0.3$

	$T = 128$			$T = 512$		
	$R(\delta)$	$SW(\delta)$	$VS(\delta)$	$R(\delta)$	$SW(\delta)$	$VS(\delta)$
DGP-P0	0.055	0.041	0.054	0.047	0.048	0.049
DGP-P1*	0.632	0.607	0.625	0.923	0.846	0.900
DGP-P2*	0.218	0.184	0.211	0.371	0.307	0.362
DGP-P3*	0.314	0.451	0.298	0.643	0.703	0.582
DGP-P4*	0.078	0.003	0.078	0.087	0.007	0.086
DGP-P5*	0.364	0.185	0.376	0.666	0.263	0.638
DGP-P6*	0.304	0.204	0.164	0.532	0.326	0.290
DPG-P7*	0.540	0.447	0.511	0.732	0.488	0.684

Table S.3: Empirical Powers of $R(\widehat{\delta})$, $R(\widehat{\delta}_2)$, $R(\widehat{\delta}_3)$, $SW(\widehat{\delta})$, $W(\widehat{\delta})$ and $VS(\widehat{\delta})$ Tests.

Panel A: $\delta = 0$

	T=128						T=512					
	$R(\widehat{\delta})$	$R(\widehat{\delta}_2)$	$R(\widehat{\delta}_3)$	$SW(\widehat{\delta})$	$W(\widehat{\delta})$	$VS(\widehat{\delta})$	$R(\widehat{\delta})$	$R(\widehat{\delta}_2)$	$R(\widehat{\delta}_3)$	$SW(\widehat{\delta})$	$W(\widehat{\delta})$	$VS(\widehat{\delta})$
DGP-P0	0.051	0.059	0.067	0.032	0.009	0.013	0.051	0.043	0.053	0.043	0.024	0.022
DGP-P1	0.276	0.298	0.274	0.267	0.042	0.110	0.877	0.900	0.844	0.925	0.530	0.730
DGP-P2	0.115	0.124	0.129	0.089	0.015	0.029	0.391	0.401	0.361	0.399	0.126	0.228
DGP-P3	0.130	0.123	0.167	0.204	0.028	0.041	0.475	0.429	0.562	0.834	0.373	0.293
DGP-P4	0.062	0.070	0.078	0.034	0.008	0.015	0.105	0.101	0.106	0.047	0.033	0.043
DGP-P5	0.156	0.164	0.181	0.083	0.026	0.009	0.596	0.595	0.632	0.324	0.314	0.420
DGP-P6	0.142	0.152	0.162	0.110	0.014	0.024	0.543	0.546	0.529	0.477	0.131	0.153
DGP-P7	0.232	0.234	0.260	0.273	0.087	0.267	0.582	0.564	0.614	0.472	0.585	0.519

Panel B: $\delta = 0.3$

	T=128						T=512					
	$R(\widehat{\delta})$	$R(\widehat{\delta}_2)$	$R(\widehat{\delta}_3)$	$SW(\widehat{\delta})$	$W(\widehat{\delta})$	$VS(\widehat{\delta})$	$R(\widehat{\delta})$	$R(\widehat{\delta}_2)$	$R(\widehat{\delta}_3)$	$SW(\widehat{\delta})$	$W(\widehat{\delta})$	$VS(\widehat{\delta})$
DGP-P0	0.047	0.055	0.055	0.039	0.010	0.138	0.049	0.046	0.053	0.046	0.026	0.053
DGP-P1*	0.271	0.293	0.244	0.405	0.193	0.825	0.538	0.582	0.511	0.696	0.324	0.911
DGP-P2*	0.114	0.134	0.126	0.118	0.035	0.383	0.224	0.251	0.225	0.230	0.083	0.366
DGP-P3*	0.092	0.087	0.136	0.276	0.134	0.659	0.204	0.202	0.303	0.528	0.229	0.683
DGP-P4*	0.059	0.068	0.070	0.036	0.012	0.176	0.077	0.080	0.080	0.054	0.034	0.091
DGP-P5*	0.155	0.171	0.192	0.105	0.090	0.611	0.344	0.370	0.405	0.169	0.209	0.649
DGP-P6*	0.152	0.178	0.179	0.129	0.035	0.373	0.325	0.357	0.344	0.233	0.081	0.316
DGP-P7*	0.290	0.178	0.174	0.305	0.503	0.764	0.463	0.411	0.423	0.348	0.607	0.764

Note: $R(\widehat{\delta})$, $R(\widehat{\delta}_2)$, and $R(\widehat{\delta}_3)$ denote the $R(\cdot)$ test evaluated using the LW estimate with no trimming, trimming with $l^* = 2$, and trimming with $l^* = 3$, respectively.

Additional References

Hualde, J., and F. Iacone. 2017. Fixed bandwidth asymptotics for the studentized mean of fractionally integrated processes. *Economics Letters* 150, 39–43.

Mandelbrot, B.B., and Van Ness, J.W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Review* 10, 422–437.

Robinson, P. M. 1995a. Log-periodogram regression of time series with long range dependence. *The Annals of Statistics* 23, 1048 – 1072.

Robinson, P. M. 1995b. Gaussian Semiparametric Estimation of Long Range Dependence. *The Annals of Statistics* 23, 1630 – 1661.

Wu, W. B. 2007. Strong invariance principles for dependent random variables. *The Annals of Probability* 35, 2294 – 2320.

Zygmund, A. 1977. *Trigonometric Series* (Cambridge University Press).