Residual log-periodogram inference for long-run relationships

U. Hassler[a], F. Marmol[b], C. Velasco[b,*]

[a]*Institute of Statistics and Econometrics, Free University of Berlin, Berlin*
[b]*Departamento de Economía, Universidad Carlos III de Madrid, Calle Madrid 126, E28903 Getafe, Madrid, Spain*

Received 12 April 2004
Available online 9 April 2005

Abstract

We assume that some consistent estimator $\widehat{\beta}$ of an equilibrium relation between non-stationary series integrated of order $d \in (0.5, 1.5)$ is used to compute residuals $\hat{u}_t = y_t \quad \widehat{\beta} x_t$ (or differences thereof). We propose to apply the semiparametric log-periodogram regression to the (differenced) residuals in order to estimate or test the degree of persistence δ of the equilibrium deviation u_t. Provided $\widehat{\beta}$ converges fast enough, we describe simple semiparametric conditions around zero frequency that guarantee consistent estimation of δ. At the same time limiting normality is derived, which allows to construct approximate confidence intervals to test hypotheses on δ. This requires that $d \quad \delta > 0.5$ for superconsistent $\widehat{\beta}$, so the residuals can be good proxies of true cointegrating errors. Our assumptions allow for stationary deviations with long memory, $0 \leqslant \delta < 0.5$, as well as for non-stationary but transitory equilibrium errors, $0.5 < \delta < 1$. In particular, if x_t contains several series we consider the joint estimation of d and δ. Wald statistics to test for parameter restrictions of the system have a limiting χ^2 distribution. We also analyse the benefits of a pooled version of the estimate. The empirical applicability of our general cointegration test is investigated by means of Monte Carlo experiments and illustrated with a study of exchange rate dynamics.

JEL classification: C14; C22

Keywords: Fractional cointegration; Semiparametric inference; Limiting normality; Long memory; Non-stationarity; Exchange rates

*Corresponding author. Fax: +34 91 624 9875.
E-mail address: carlos.velasco@uc3m.es (C. Velasco).

1. Introduction

A substantial part of economic theory deals with long run equilibrium relation ships generated by market forces and behavioral rules. Granger (1981) and Engle and Granger (1987) were the first to formalize the idea of integrated variables sharing an equilibrium relation which turned out to be either stationary or have a lower degree of integration than the original series. They denoted this property by cointegration, signifying co movements among trending variables which could be exploited to test for the existence of equilibrium relationships within a fully dynamic specification framework.

The presence of, at least, a unit root in economic time series is implied in many economic models as those based on the rational use of available information or on the existence of very high adjustment costs in some markets. Interesting examples include future contracts, stock prices, yield curves, exchange rates, money velocity, hysteresis theories of unemployment and, perhaps the most popular, the implications of the permanent income hypothesis for real consumption under rational expectations. Thus, most of the cointegration literature has focused on the case where variables contain a single unit root. Moreover, in most of the occasions, the equilibrium relation turned out to be modeled as a weakly stationary or short memory $I(0)$ process. Within this $I(1)/I(0)$ set up, Engle and Granger (1987) suggested a two step estimation procedure for single equation dynamic modeling which has become very popular in applied research. First, an OLS regression is run among the levels of the series of interest. Then, Dickey Fuller type unit root tests are performed on the residual sequence to determine whether it has a unit root. Under the null hypothesis the residuals are $I(1)$, and under the alternative the residuals are $I(0)$.

Some economic applications, however, suggest that even if the data are $I(1)$, the residual term representing the potential equilibrium error might be fractionally integrated. See, e.g., Robinson (1994a), Baillie (1996) and Gil Alaña and Robinson (1997). Loosely speaking, a series u_t is said to be fractionally integrated of order δ, in short $I(\delta)$, if $\Delta^{\delta}u_t$ is $I(0)$, where δ is not an integer but a real number. The degree of integration determines the key dynamic or memory properties of the economic series. A fractionally integrated process is stationary if $\delta<0.5$ and nonstationary otherwise (cf. Granger and Joyeux, 1980; Hosking, 1981). In spite of being nonstationary, if $0.5\leqslant\delta<1$ the process is mean reverting with transitory memory, i.e. any random shock has only a temporary influence on the series, in contrast with the case when $\delta\geqslant 1$, where the process is both nonstationary and not mean reverting with permanent memory, i.e., any random shock having now a permanent effect on the future path of the series. Consequently, a wide range of dynamic behavior is ruled out a priori if δ is restricted to integer values and a much broader range of cointegration possibilities is permitted when fractional cases are considered. More importantly, now the degree of memory of the residual series, δ, is a parameter suitable, in principle, of estimation and testing by means of any of the existing methods.

In this sense, the most widespread estimation method of the memory parameter δ with observed series is the so called log periodogram estimator (Geweke and Porter Hudak, 1983; Robinson, 1995a) due to its semiparametric nature and simplicity. In this paper we provide theoretical grounds on the behavior of the log periodogram estimator when applied to the residual equilibrium series. Indeed, the residual based log periodogram regression for (fractional) cointegration testing has been applied in a series of papers recently. Cheung and Lai (1993), Masih and Masih (1995) and Soofi (1998) test the purchasing power parity hypothesis, while Booth and Tse (1995) and Masih and Masih (1998) investigate interest rate future markets and exchange rate dynamics, respectively. Their approach also relies on a two step procedure, where the log periodogram regression is applied in a second step to regression residuals obtained in a first step from a cointegrating regression.[1] Experimentally, they collected evidence that the t statistics associated with the estimator $\widehat{\delta}$ may not be approximately normally distributed, cf. also the recent Monte Carlo results by Tse et al. (1999). Their experimental evidence, however, is limited with two respects. First, they only consider bivariate regressions, second, they assume that the observed series are integrated of order one. Our analysis overcomes these drawbacks in that we allow the observed series to be integrated of order d, $0.5<d<1.5$, and moreover multiple regressions are also considered. But most important, our asymptotic treatment reveals that the residual based log periodogram regression does result in a limiting normal distribution provided the very first harmonic frequencies are neglected. This modification, which has been called trimming in the statistical literature, had not been considered in the experimental studies previously quoted.

In this paper we assume that the series of interest are a (linearly) cointegrated set of $I(d)$ processes, with the corresponding innovation being an $I(\delta)$ process such that $d>\delta$ with $0.5<d<1.5$. In case of single equation regression, given an estimator $\widehat{\beta}$ of the corresponding cointegrating coefficient, we consider the residuals $\hat{u}_t \quad y_t \quad \widehat{\beta}x_t$ and estimate δ from a log periodogram regression of the residuals, or of the differenced residuals. With the gap between d and δ being large enough, $\delta<d \quad 0.5$, we obtain sufficient conditions for the consistency of the estimators of the memory parameter δ of the cointegration error. In particular, we require trimming of the very first frequencies of the residual periodogram. Furthermore, assumptions are strengthened in order to establish limiting normality. Given a consistent and asymptotically normal estimator it is straightforward to compute at what level of significance the estimators of δ are (i) positive, (ii) less than 0.5, (iii) larger than 0.5, or (iv) less than 1. Such inference is of immediate economic interest, because the degree of integration δ measures the persistence of the deviations from long run equilibrium. Depending on our null hypothesis of interest, e.g. $\delta \quad 0$ or 1, we propose alternative procedures based on either original or differenced residuals that

[1] A multivariate approach in contrast to single equation regressions was employed by Baillie and Bollerslev (1994) and Dueker and Startz (1998). Two recent papers provide asymptotic theory for determining the cointegration rank in a fractional context: Robinson and Yajima (2002) suggest a frequency domain approach designed for stationary processes, while Breitung and Hassler (2002) consider a time domain approach valid in the nonstationary case.

lead to a consistent characterization of the long run relationship among some economic series.

The rest of the paper is organized as follows. The next section sets the scene by introducing the basic bivariate cointegrated regression model and the relevant theory for the residual log periodogram regression. The third section is reserved for the extension to multiple regressions, investigating the situation where the degree of integration of the regressors and the error are jointly estimated. Residual Wald statistics testing parameter restrictions remain asymptotically χ^2 distributed just as found by Robinson (1995a) for observed series. Section 4 considers non Gaussian series and situations ruled out in previous sections when δ can be arbitrarily close to d. In Section 5 Monte Carlo experiments are reported with respect to the empirical relevance of some of the assumptions used. We propose an empirical research strategy that is illustrated in Section 6 with a study of exchange rate dynamics. The final section contains a more detailed summary of our main findings. Technical assumptions and auxiliary results are collected in Appendix A, while proofs are relegated to Appendix B.

2. Residual log-periodogram regression

In this section, we restrict ourselves to the leading case of a simple regression between two non stationary series. Let the observable $y_t \sim I(d)$ and $x_t \sim I(d)$, $0.5 < d < 1.5$, satisfy

$$y_t \quad \beta x_t + u_t, \quad \beta \neq 0, \quad t \quad 1, \ldots, T,$$

with $u_t \sim I(\delta)$, $0 \leqslant \delta < d$, and let $\widehat{\beta}$ be a consistent estimate of β based on T observations of y_t and x_t. The interval $0.5 < d < 1.5$ covers most empirically relevant cases. Extensions to higher order integration might be possible but are not considered here. The properties of estimates of β depend on d and δ, bridging the gap between root T consistency for stationary regressions and T superconsistency for $I(1)$ regressions with $I(0)$ residuals. We will assume the following condition on $\widehat{\beta}$ distinguishing the case where the overall memory of regressors and errors, $\delta + d$, is strictly less than 1, and the case where it is equal or larger than 1.

Assumption 1. Let $d \in (0.5, 1.5)$, $\delta \in [0, d)$,
Case I : If $\delta + d \geqslant 1$ then $\widehat{\beta} \quad \beta \quad O_p(T^{\delta - d})$.
Case II : If $\delta + d < 1$ then $\widehat{\beta} \quad \beta \quad O_p(T^{1-2d})$.

This assumption holds when the $\widehat{\beta}$ are the OLS estimates for $\delta \in [0, 1.5) \quad \{0.5\}$ (see De Jong and Davidson, 2000; Robinson and Marinucci, 2001). There are several alternative estimates that try to improve the asymptotic and finite sample properties of OLS estimates. Robinson and Marinucci (2001) proposed a narrow band frequency domain LS estimate which satisfies Assumption 1 when a bandwidth is chosen appropriately but under a somewhat different definition of non stationary long memory processes than the one we use in this paper. This alternative definition implies different initial conditions for integrated processes than ours, and is also less

tractable for our purposes because it implies that the series are nonstationary for any value of $\delta \neq 0$ (though asymptotically stationary for $\delta < 0.5$). Moreover, though convergence rates for slope estimates are the same, the asymptotic theory is different for each definition (see Marinucci and Robinson, 1999). Alternatively, Kim and Phillips (2001) developed fully modified version of LS under Gaussian assumptions for Case I.

We say that a covariance stationary time series u_t is $I(\delta)$ if it has a spectral density $f_{uu}(\lambda)$, defined by $\mathrm{Cov}(u_t, u_{t+j}) \quad \int_{-\pi}^{\pi} f_{uu}(\lambda)\cos(j\lambda)\,\mathrm{d}\lambda$, satisfying for some positive constant G_u

$$f_{uu}(\lambda) \sim G_u|\lambda|^{-2\delta} \quad \text{as } \lambda \to 0, \quad \delta < 0.5. \tag{1}$$

This reflects a persistent behaviour or long memory at low frequencies when $\delta > 0$, weak dependence when $\delta \quad 0$ and negative memory when $\delta < 0$, but leaving unparameterized the rest of the spectrum. This definition covers standard fractional parametric models such as stationary ARFIMA (Granger and Joyeux, 1980; Hosking, 1981),

$$\phi(L)(1 \quad L)^{\delta} u_t \quad \theta(L)\varepsilon_t, \tag{2}$$

where ε_t is white noise, ϕ and θ are polynomials in the lag operator L with all their roots outside the unit circle and $(1 \quad L)^{\delta}$ is the fractional difference operator.

For non stationary data we adopt a parallel definition of the memory parameter δ in terms of stationary increments. Thus u_t is $I(\delta)$, $0.5 \leqslant \delta < 1.5$, if $\Delta u_t \quad (1 \quad L)u_t$ is zero mean $I(\delta \quad 1)$, encompassing the $I(0)$ and $I(1)$ terminology of the standard cointegration literature. For non stationary $u_t \sim I(\delta)$ we consider a generalized or pseudo spectral density using the difference operator transfer function $|1 \quad \mathrm{e}^{\mathrm{i}\lambda}|^2$ $(2\sin(\lambda/2))^2$ as

$$f_{uu}(\lambda) \quad (2\sin(\lambda/2))^{-2} f_{\Delta u \Delta u}(\lambda) \sim G_u|\lambda|^{-2\delta} \quad \text{as } \lambda \to 0, \quad 0.5 \leqslant \delta < 1.5,$$

which has a similar behaviour around the origin in terms of the memory parameter δ as the spectral density (1) of stationary long memory processes. These definitions allow us to directly extend standard frequency domain assumptions and analysis to non stationary data.

A variety of estimates of the long memory parameter of stationary series has been proposed. Many of them are parametric in the sense that a full parametric model is also specified for the short memory behaviour of the series as in (2). These include estimates based on different approximations to Gaussian likelihoods in frequency and time domains (e.g. Fox and Taqqu, 1986; Dahlhaus, 1989; Sowell, 1992). On the other hand, semiparametric estimates have the advantage of avoiding short run specification and have become of wide use in practice, although they are asymptotically inefficient compared to parametric competitors.

The most popular of semiparametric estimates is probably the log periodogram regression estimate proposed by Geweke and Porter Hudak (1983) because of its intuitive and computational appeal, though some competitors have been studied under more general conditions (e.g. Robinson, 1995b). As many semiparametric estimates of long memory parameters, it is based on the properties of the spectral

density for low frequencies, cf. expression (1). Define the (cross) periodogram of two sequences p_t and q_t, t $1,\ldots,T$, as

$$I_{pq}(\lambda) \quad w_p(\lambda)w_q^*(\lambda), \quad w_p(\lambda) \quad (2\pi T)^{-1/2}\sum_{t\ 1}^{T} p_t \exp(\mathrm{i}\lambda t),$$

where the star $*$ superscript denotes simultaneous transposition and complex conjugation. The periodogram $I_{uu}(\lambda)$ is the sample equivalent of the spectral density for an observed sequence u_t and constitutes the basic statistic for frequency domain inference. Robinson (1995a) showed that, as for short memory series, the periodogram of long memory series is asymptotically unbiased and uncorrelated when evaluated at the harmonic frequencies λ_j $2\pi j/T$ for j growing with sample size T (see e.g. Lemmas A.1 and A.2 in Appendix A). This is the basis to write the logarithm of (1) as

$$\log I_{uu}(\lambda_j) \approx \log G_u \quad 2\delta \log \lambda_j + \log \frac{I_{uu}(\lambda_j)}{f_{uu}(\lambda_j)}, \quad j \quad 1,\ldots,m,$$

where m is small compared to T. This is a linear regression model with the log periodogram as dependent variable, non stochastic regressor r_j $2\log\lambda_j$, slope δ and approximately homoscedastic and independent errors. The log periodogram regression estimate of δ is the least squares estimate $\widehat{\delta}(u)$.

The asymptotic properties of $\widehat{\delta}(u)$ were analysed rigorously for multiple stationary Gaussian series ($0.5<\delta<0.5$) by Robinson (1995a) for m growing with T under some smoothness conditions on $f_{uu}(\lambda)$ (cf. Theorem A.1 in Appendix A). He also considered a pooling of contributions from adjacent frequencies to achieve efficiency gains (see Section 4) and excluded the very low frequencies, following the findings of Künsch (1987).

When the equilibrium errors u_t are non stationary but not observable it is sensible to estimate δ from the increments of the observed residuals $\Delta\hat{u}_t$ Δu_t $(\widehat{\beta}$ $\beta)\Delta x_t$ which we may expect to have memory close to δ 1, so the periodogram of the residual differences is

$$I_{\Delta\hat{u}\Delta\hat{u}}(\lambda_j) \quad I_{\Delta u\Delta u}(\lambda_j) \quad (\widehat{\beta} \quad \beta)\{I_{\Delta u\Delta x}(\lambda_j) + I_{\Delta x\Delta u}(\lambda_j)\} + (\widehat{\beta} \quad \beta)^2 I_{\Delta x\Delta x}(\lambda_j). \qquad (3)$$

However, when u_t are stationary, δ could be estimated directly from the levels of the observed residuals $\hat{u}_t$ u_t $(\widehat{\beta}$ $\beta)x_t$ through

$$I_{\hat{u}\hat{u}}(\lambda_j) \quad I_{uu}(\lambda_j) \quad (\widehat{\beta} \quad \beta)\{I_{ux}(\lambda_j) + I_{xu}(\lambda_j)\} + (\widehat{\beta} \quad \beta)^2 I_{xx}(\lambda_j),$$

avoiding problems of non invertibility that may arise with differenced stationary data. For inference on δ using the residuals $\hat{u}_t$ or increments $\Delta\hat{u}_t$ the key point is that $\widehat{\beta}$ β has to be small enough in probability to make the contribution of the slope estimation negligible in the residual periodograms. We show that this is the case using only Assumption 1, where the estimates $\widehat{\beta}$ can be obtained by any method and we do not need their asymptotic distribution or moments. On the other hand, as our proofs rely on Robinson's (1995a) analysis, Gaussianity of x_t and u_t is required. We also note that Robinson (1997) considered semiparametric memory estimation from

nonparametric regression residuals using a local Gaussian likelihood (see Robinson, 1995b), but avoiding such assumption.

Denote by $\widehat{\delta}(\Delta\hat{u})$ the log periodogram regression estimate of δ based on the differences of the observed residuals, $\Delta\hat{u}_t$,

$$\widehat{\delta}(\Delta\hat{u}) \quad 1 + \left(\sum_{j \quad \ell+1}^{m} W_j^2 \right)^{-1} \sum_{j \quad \ell+1}^{m} W_j \log I_{\Delta\hat{u}\Delta\hat{u}}(\lambda_j) \tag{4}$$

where

$$W_j \quad r_j \quad \bar{r}_\ell, \quad r_j \qquad 2 \log \lambda_j \quad \text{and} \quad \bar{r}_\ell \quad (m \quad \ell)^{-1} \sum_{j \quad \ell+1}^{m} r_j. \tag{5}$$

When $\ell > 0$ in the above definition of $\widehat{\delta}(\Delta\hat{u})$ we allow for the trimming of the very low frequencies as in Robinson (1995a). However Hurvich et al. (1998) have shown that the log periodogram regression maintains desirable properties if all frequencies from 1 up to m are used. Nevertheless we later provide an alternative justification for the policy of removing the first ℓ frequencies when residuals are used instead of observational data, for ℓ growing with T as in the next assumption.

Assumption 2. We choose

$$m \sim AT^a, \quad \ell \sim BT^b, \quad 0 < b < a < 1, \quad 0 < A, B < \infty,$$

where $p \sim q$ means that $\lim_{T\to\infty} p/q \quad 1$.

This technical assumption restricts the bandwidth numbers ℓ and m to a particular choice in terms of powers of T to simplify the presentation of the results, but more general choices are possible, though depending on unknown parameters such as d and δ. In practice only small values of ℓ are usually chosen.

In the following theorem we summarize the properties of differenced residual log periodogram regression. We concentrate on asymptotic normality and $\log T$ consistency, for studentization purposes of statistics such as $\widehat{\beta}$ whose convergence rate depends on δ as was pointed out by Robinson (1994b, 1997). Note that only Case I of Assumption 1 is relevant for $\widehat{\delta}(\Delta\hat{u})$ when $\delta > 0.5$. Additional technical assumptions on the smoothness of the spectral densities and bandwidth choice are detailed in Appendix A.

Theorem 1. *Under Assumptions* 1, 2 *and* A.1, *for Gaussian* u_t *and* x_t, $0.5 < \delta < d$ $0.5 < 1$, *as* $T \to \infty$,

$$\log T(\widehat{\delta}(\Delta\hat{u}) \quad \delta) \to_p 0.$$

If additionally Assumption A.5 *holds then*

$$m^{1/2}(\widehat{\delta}(\Delta\hat{u}) \quad \delta) \to_{\mathrm{d}} \mathrm{N}\left(0, \frac{\pi^2}{24}\right).$$

For both consistency and asymptotic normality of $\widehat{\delta}(\Delta\hat{u})$ our proofs require the trimming of an increasing number of frequencies and that $d \quad \delta > 0.5$, to obtain

uniform convergence of the normalized residual periodogram for λ_j, $\ell \leqslant j \leqslant m$. This problem prevents us from using Hurvich et al. (1998) results to completely avoid the trimming of low frequencies, though any $b > 0$ is enough for our results. The condition $d \quad \delta > 0.5$, which implies that $\widehat{\beta}$ is superconsistent (cf. Case I in Assumption 1), can be relaxed to something close to $d > \delta$ for consistency of $\widehat{\delta}(\Delta\hat{u})$ (cf. Section 4), but it seems necessary for root m consistency. This confirms Robinson's (1995a) Remark 7 that a sufficiently fast convergence rate of the estimates of the appropriate filter should be necessary for log periodogram inference based on residuals.

When $\delta \leqslant 0.5$ the previous procedure is likely to fail because Δu_t are non invertible, so we are led to work with the original residuals. The study of the asymptotic properties of the log periodogram regression estimate of δ based on the original residuals,

$$\widehat{\delta}(\hat{u}) \quad \left(\sum_{j \quad \ell+1}^{m} W_j^2 \right)^{-1} \sum_{j \quad \ell+1}^{m} W_j \log I_{\hat{u}\hat{u}}(\lambda_j),$$

is additionally complicated because we have to distinguish the cases $d + \delta \geqslant 1$ and $d + \delta < 1$, for which the estimates of β have different convergence rates. We did not have this problem before because $0.5 < \delta < d \quad 0.5$, so we now add Assumption A.4 introduced in Appendix A.

Theorem 2. *Under Assumptions* 1, 2, A.1 *and* A.4 *for Gaussian* u_t *and* x_t, $0 \leqslant \delta < 0.5$, $\delta < d \quad 0.5 < 1$, *then as* $T \to \infty$,

$$\log T(\widehat{\delta}(\hat{u}) \quad \delta) \to_p 0.$$

If additionally Assumption A.10 *holds then*

$$m^{1/2}(\widehat{\delta}(\hat{u}) \quad \delta) \to_{\mathrm{d}} \mathrm{N}\left(0, \frac{\pi^2}{24}\right).$$

The range of values of δ in our asymptotic theory for $\widehat{\delta}(\hat{u})$ is more limited than when the u_t are observable, where any $0.5 < \delta < 1$ can be consistently estimated, see Velasco (1999a). In case of residual inference, $\widehat{\delta}(\hat{u})$ is consistent only if $0 \leqslant \delta < 0.5$, with $d \quad \delta > 0.5$, as for differenced residuals, and with m and ℓ chosen appropriately. We do not consider $\delta < 0$ because this is not likely to occur for observed undifferenced data. Tapering, as suggested in Hurvich and Ray (1995), may allow consistent estimation of situations excluded in Theorems 1 and 2, e.g. $\delta < 0.5$ using $\widehat{\delta}(\Delta\hat{u})$ and $\delta \geqslant 0.5$ using $\widehat{\delta}(\hat{u})$, as was showed for observed data in Velasco (1999a). We explore the latter possibility in Section 4.

Only when $d + \delta < 1$, strong enough trimming is essential for our analysis of $\widehat{\delta}(\hat{u})$, cf. Assumption A.4. Only in this case the choices of m and ℓ are limited by the values of d and δ, due to the slower convergence rate of $\widehat{\beta}$ in Case II of Assumption 1, leaving situations where it is not possible to find sequences m and ℓ to show the root m asymptotic normality of $\widehat{\delta}(\hat{u})$, for example when $0.5 < d < 6.5/9$. However the most relevant situation of $d \quad 1$ and $0 \leqslant \delta < 0.5$ is covered by Theorem 2 with any $b > 0$.

The Gaussianity assumption can be removed for a pooled version of the log periodogram regression for some linear processes (see Section 4 below and Velasco, 2000), but in this and the following section Gaussianity plays a decisive simplifying role for residual based inference on δ.

Remark (*linear detrending*). The deterministic regressor τ_t t has similar properties to $I(1.5)$ stochastic data, so the least squares estimate of its coefficient is $T^{1.5-\delta}$ consistent (see e.g. Robinson and Marinucci, 2000). Therefore, it can be shown that Theorems 1 and 2 hold if residuals are obtained after linear detrending, $\delta<1$.

For memory estimation some a priori knowledge on δ is necessary in order to use either $\widehat{\delta}(\Delta\hat{u})$ or $\widehat{\delta}(\hat{u})$ appropriately, though use of tapered original residuals may help in providing consistent estimates for any $\delta<1$. For hypothesis testing this information can be obtained from the maintained null hypothesis. Thus consistent procedures can be obtained from asymptotic N(0, 1) t statistics based on $\widehat{\delta}(\hat{u})$ for testing of H_0: δ 0 against H_1: $\delta>0$, or on $\widehat{\delta}(\Delta\hat{u})$ for testing of H_0: δ 1 against H_1: $\delta<1$.

3. Multiple regression and estimation

We now consider the case of multivariate regressors and joint estimation of the memory parameters of the regressors and cointegrating errors. Let the observable y_t satisfy

$$y_t \quad \sum_{i\ 1}^{k} \beta_i x_{it} + u_t,$$

for $x_{it}\sim I(d_i)$, $0.5<d_i<1.5$, $u_t\sim I(\delta)$, $0\leqslant\delta<d_{\min}$, and $y_t\sim I(d_{\max})$, where $d_{\min}$ $\min_i d_i$ and $d_{\max}$ $\max_i d_i$. Let $\widehat{\beta}$ be a consistent estimate of the vector β $(\beta_1,\ldots,\beta_k)'$ based on T observations of y_t and x_t $(x_{1t},\ldots,x_{kt})'$. We make the following assumption on $\widehat{\beta}$ distinguishing the two cases of Assumption 1 and allowing for regressors with different memory parameters.

Assumption 3. Let $x_{it}\sim I(d_i)$, $d_i\in(0.5,1.5)$, i $1,\ldots,k$, $u_t\sim I(\delta)$, $0\leqslant\delta<d_i$:
Case I: If $\delta+d_i\geqslant 1$ then $\widehat{\beta}_i$ β_i $O_p(T^{\delta-d_i})$, i $1,\ldots,k$.
Case II: If $\delta+d_i<1$ then $\widehat{\beta}_i$ β_i $O_p(T^{1-d_{\min}-d_i})$, i $1,\ldots,k$.

It would be possible to consider more general set ups with Cases I and II mixed. However Robinson and Marinucci (2001) only consider Cases I and II separately, and showed that the convergence rates of Assumption 3 hold for OLS estimates, $\delta\neq 0.5$. In any case residual based log periodogram regression asymptotics would depend on the slowest rate of convergence, given by $O_p(T^{1-d_{\min}-d_i})$. Then it is quite straightforward to show that Theorems 1 and 2 continue to hold when we use residuals from multivariate regressions, where the assumptions on the regressors x_{it} are now to be understood componentwise. Therefore the proof of the following result is omitted.

Corollary 1. *Theorems* 1 *and* 2 *hold for multiple regression residuals* $\hat{u}_t = \widehat{\beta}' x_t$ *where the* $\widehat{\beta}$ *satisfy Assumption* 3.

The previous remark on linear detrending applies to multivariate regressions when one of the regressors is t and also the results are unaffected if the regressions include an intercept or seasonal dummies, since these variables have zero variance at the relevant frequencies, so hence on we concentrate only on stochastic regressors.

Furthermore, the memory parameters of the stationary vector $(u_t, \Delta x_t')' \sim I(\delta, D_1, \ldots, D_k)$, $D_i = d_i - 1$, $0.5 < \delta < 0.5$, can be simultaneously investigated as if the u_t were observable using Robinson's (1995a) multivariate log periodogram estimate as long as sufficient smoothness conditions are assumed for the spectral density matrix. For example, the case where a set of the regressors x_t is cointegrated is excluded (cf. Assumption A.2 in Appendix A), and some trimming is incorporated in the log periodogram regression. This permits hypothesis testing on the differences $d_i - \delta$ and efficiency gains for inference on δ or $\mathbf{d} = (\delta, d_1, \ldots, d_k)'$ using generalized LS estimation under linear restrictions on the memory parameters, like $d_i = d$, $i = 1, \ldots, k$.

To this end we set the system of $k+1$ equations, $j = \ell + 1, \ldots, m$, where we allow for the trimming of the first $\ell = 0, 1, \ldots$ frequencies,

$$\log I_{\hat{u}\hat{u}}(\lambda_j) = c_u - 2\delta \log \lambda_j + v_{u,j},$$

$$\log I_{\Delta i \Delta i}(\lambda_j) = c_i - 2D_i \log \lambda_j + v_{ij}, \quad i = 1, \ldots, k,$$

and $I_{\Delta i \Delta i}(\lambda_j)$ is the periodogram of Δx_{it}, $c_i = \log G_i$ and $D_i = d_i - 1$. The vector of OLS estimates $\widehat{\mathbf{D}}(\hat{u}) = (\widehat{\delta}, \widehat{D}_1, \ldots, \widehat{D}_k)'$ and $\widehat{\mathbf{c}}(\hat{u}) = (\widehat{c}_u, \widehat{c}_1, \ldots, \widehat{c}_k)'$,

$$\begin{pmatrix} \widehat{\mathbf{c}}(\hat{u}) \\ \widehat{\mathbf{D}}(\hat{u}) \end{pmatrix} = \mathrm{vec}(V(\hat{u})' S (S'S)^{-1}),$$

is the generalization of the log periodogram estimate of the previous section, where $S = (S_{\ell+1}, \ldots, S_m)'$, $S_j = (1, r_j)'$, and $V(\hat{u}) = (V_o(\hat{u}), V_1, \ldots, V_k)$, $V_o(\hat{u}) = (\log I_{\hat{u}\hat{u}}(\lambda_{\ell+1}), \ldots, \log I_{\hat{u}\hat{u}}(\lambda_m))'$ and $V_i = (\log I_{\Delta i \Delta i}(\lambda_{\ell+1}), \ldots, \log I_{\Delta i \Delta i}(\lambda_m))'$, $i = 1, \ldots, k$. We set the estimate $\widehat{\mathbf{d}}(\hat{u}) = (\widehat{\delta}, \widehat{d}_1, \ldots, \widehat{d}_k)'$ of $\mathbf{d}$, with $\widehat{d}_i = \widehat{D}_i + 1$.

To obtain asymptotically normal estimates of $\mathbf{d}$ when the u_t are non stationary we use differenced residuals $\Delta\hat{u}_t$, substituting the first equation in the log periodogram regression by

$$\log I_{\Delta\hat{u}\Delta\hat{u}}(\lambda_j) = c_u - 2(\delta - 1) \log \lambda_j + v_{u,j},$$

and obtain the least squares estimates $\widehat{\mathbf{D}}(\Delta\hat{u}) = (\widehat{\delta - 1}, \widehat{D}_1, \ldots, \widehat{D}_k)'$,

$$\begin{pmatrix} \widehat{\mathbf{c}}(\Delta\hat{u}) \\ \widehat{\mathbf{D}}(\Delta\hat{u}) \end{pmatrix} = \mathrm{vec}(V(\Delta\hat{u})' S (S'S)^{-1}),$$

setting $\widehat{\mathbf{d}}(\Delta\hat{u}) = (\widehat{\delta}, \widehat{d}_1, \ldots, \widehat{d}_k)'$, $\widehat{\delta} = \widehat{\delta - 1} + 1$, $\widehat{d}_i = \widehat{D}_i + 1$, and $V_o(\Delta\hat{u}) = (\log I_{\Delta\hat{u}\Delta\hat{u}}(\lambda_{\ell+1}), \ldots, \log I_{\Delta\hat{u}\Delta\hat{u}}(\lambda_m))'$. The next result gives sufficient conditions described in Appendix A for the asymptotic normality of these estimates, generalizing the univariate set up of Theorem 1, cf. Assumptions 2 and A.1, and

excluding the possibility of the components of x_t from being cointegrated themselves.

Theorem 3. *Under Assumptions* 2, 3, A.2, A.6, $(u_t, x_t')'$ *jointly Gaussian*, $\delta < d_i$ $0.5 < 1$, i $1, \ldots, k$, *then if* $0.5 < \delta < 1$, *as* $T \to \infty$,

$$2m^{1/2}(\widehat{\mathbf{d}}(\Delta\hat{u}) \quad \mathbf{d}) \to_{\mathrm{d}} \mathrm{N}(0, \Omega).$$

If $0 \leqslant \delta < 0.5$, *and additionally Assumption* A.7 *holds then*

$$2m^{1/2}(\widehat{\mathbf{d}}(\hat{u}) \quad \mathbf{d}) \to_{\mathrm{d}} \mathrm{N}(0, \Omega).$$

The covariance matrix Ω has diagonal elements $\pi^2/6$ and can be estimated consistently by the sample regression residuals covariance matrix, $\widehat{\Omega}$ $(m$ $\ell)^{-1}\sum_j^m{}_{\ell+1}\hat{v}_j\hat{v}_j'$. We do not consider Robinson's (1995a) pooled version of $\widehat{\mathbf{d}}$ in this section, nor the estimation of the constants G_r, but the same results as for observed data can be shown to hold when using residuals.

We can now follow Robinson (1995a) to test the homogeneous restriction

$$\mathrm{H}_0 : P\mathbf{d} \quad \mathbf{0}, \tag{6}$$

where P is an $n \times (k+1)$ matrix of rank $n < k+1$, as in the case of equal memory among some of the non stationary series x_{it}. The test statistics is

$$\widehat{\mathbf{d}}' P'[(0, P)\{(S'S) \otimes \widehat{\Omega}^{-1}\}(0, P)']^{-1} P\widehat{\mathbf{d}},$$

where $\widehat{\mathbf{d}}$ is either $\widehat{\mathbf{d}}(\hat{u})$ or $\widehat{\mathbf{d}}(\Delta\hat{u})$ and which has asymptotic χ^2_n distribution under (6) and the appropriate conditions of Theorem 3. A typical example is the estimation under the restriction of regressors of equal memory, imposed by

$$\underset{(k \quad 1) \times (k+1)}{P} \quad \begin{pmatrix} 0 & 1 & & 1 & 0 & \cdots \\ \vdots & \ddots & \ddots & & \ddots & \ddots \\ 0 & \cdots & 0 & 1 & & 1 \end{pmatrix}.$$

We can also achieve efficiency gains if we assume that some of the k series x_{it} share a common d parameter or any other homogeneous linear restriction

$$\mathbf{D} \quad Q\boldsymbol{\theta},$$

where Q is a given $(k+1) \times q$ matrix of rank $q < k+1$ and θ is a q dimensional column vector of unrelated parameters. The GLS type vector estimate incorporating such restrictions is

$$\begin{pmatrix} \widetilde{\mathbf{c}}(\hat{u}) \\ \widetilde{\theta}(\hat{u}) \end{pmatrix} \quad \{Q_1'((S'S) \otimes \widehat{\Omega}^{-1})Q_1\}^{-1}\mathrm{vec}(\widehat{\Omega}^{-1}V(\hat{u})'S),$$

where $\widetilde{\mathbf{D}}(\hat{u}) \quad Q\widetilde{\theta}(\hat{u})$, $\widetilde{\mathbf{d}}(\hat{u}) \quad \widetilde{\mathbf{D}}(\hat{u}) + (0, 1, \ldots, 1)'$, and

$$Q_1 \quad \begin{pmatrix} I_{k+1} & 0 \\ 0 & Q \end{pmatrix}.$$

If u_t is known to be non stationary we may substitute $V(\hat{u})$ by $V(\Delta\hat{u})$, and set $\widetilde{\mathbf{d}}(\Delta\hat{u}) \quad \widetilde{D}(\Delta\hat{u}) + (1, 1, \ldots, 1)'$, $\widetilde{\mathbf{D}}(\Delta\hat{u}) \quad Q\widetilde{\theta}(\Delta\hat{u})$. Then under the appropriate assumptions of Theorem 3 it can be shown that

$$2m^{1/2}(\widetilde{\mathbf{d}} \quad \mathbf{d}) \to_{\mathrm{d}} \mathrm{N}(0, Q(Q'\Omega^{-1}Q)^{-1}Q'),$$

where $\widetilde{\mathbf{d}}$ is either $\widetilde{\mathbf{d}}(\hat{u})$ or $\widetilde{\mathbf{d}}(\Delta\hat{u})$, and the covariance matrix of the asymptotic distribution can be estimated consistently by $Q(Q'\widetilde{\Omega}^{-1}Q)^{-1}Q'$ using the GLS residuals in $\widetilde{\Omega}$.

4. Residual log-periodogram for non-Gaussian data

The previous results have three main limitations. First, they rely on Gaussianity, employed for reference to Robinson (1995a) and to show the negligible effect of residual based estimates compared to original data. Second, we always have required $d \quad \delta > 0.5$ for root m consistency and asymptotic normality, but such condition is likely to be too stringent for consistency of semiparametric estimates of δ. And third, we have to avoid non stationary residuals ($\delta \geqslant 0.5$) when analysing $\widehat{\delta}(\hat{u})$.

Recently Velasco (2000) has moved in the direction of relaxing Gaussianity for the consistency of the log periodogram regression estimate. The two main devices used for this are a fixed pooling of periodogram ordinates in the regression, as originally proposed by Robinson (1995a), and tapering. We analyse in this section the consistency of a version of the residual log periodogram regression for linear processes with well behaved independent and identically distributed (i.i.d.) innovations. Pooling also permits to relax the condition $d \quad \delta > 0.5$, allowing a trade off between the cointegration degree $d \quad \delta$ and the pooling employed, while non stationary residuals ($\delta \geqslant 0.5$) can be treated consistently by tapering. Assuming enough moments for the innovations of the observed data, we could also analyse the asymptotic distribution of the estimates, see Velasco (2000) and Fay and Soulier (2001) for details.

Tapering downweights the observations at both extremes of the observed stretch of data, using a smooth function that leaves mainly unchanged the central part of the sample. We use the full cosine window

$$h_t \quad \frac{1}{2}\left\{1 \quad \cos\left(\frac{2\pi t}{T}\right)\right\},$$

so the tapered periodogram is

$$I^h_{uu}(\lambda) \quad |w^h_u(\lambda)|^2, \quad w^h_u(\lambda) \quad \left(2\pi \sum_{t \ 1}^{T} h_t^2\right)^{-1/2} \sum_{t \ 1}^{T} h_t u_t \exp(\mathrm{i}\lambda t),$$

and define for $J \quad 1, 2, \ldots,$ fixed with T, the pooled tapered periodogram

$$I^{(J)}_{uu}(\lambda_j) \quad \sum_{r \ 1}^{J} I^h_{uu}(\lambda_{j+r-J}), \quad j \quad \ell + J + 1, \ell + 2J + 1, \ldots, m,$$

suppressing reference to tapering and assuming that $(m \quad \ell \quad 1)/J$ is integer. Note that even for $\ell \quad 0$ we suppress the first tapered periodogram ordinate $I^h_{uu}(\lambda_1)$ to avoid zero frequency leakage (see Velasco, 1999a). The pooled log periodogram estimate of the memory parameter δ considered in Robinson (1995a) using the mentioned frequencies is

$$\widehat{\delta}^{(J)}(u) \quad \left(\sum_{j \quad \ell(J)}^{m} W_j^2\right)^{-1} \sum_{j \quad \ell(J)}^{m} W_j \log I^{(J)}_{uu}(\lambda_j),$$

where it is shown that letting $J>1$, fixed with T, improves the efficiency of $\widehat{\delta}^{(J)}(u)$. Note that at the same time, since the tapered periodograms at Fourier frequencies λ_j are not asymptotically uncorrelated, there is now serial correlation among the $\log I^{(J)}_{uu}(\lambda_j)$, increasing the asymptotic variance of the tapered $\widehat{\delta}^{(J)}(u)$.

We adapt the set up of Velasco (2000) to investigate the consistency of $\widehat{\delta}^{(J)}(\hat{u})$ for non Gaussian data as follows. Instead of Gaussianity we introduce a fourth order stationary linear process condition, with filter coefficients compatible with (1). Let $\lfloor x \rfloor$ denote the largest integer equal or less than x.

Assumption 4. Any $z_t \in \{\Delta^{\bar{\delta}} u_t, \Delta x_{1t}, \ldots, \Delta x_{kt}\}$, $\bar{\delta} \quad \lfloor \delta + 0.5 \rfloor$, satisfies

$$z_t \quad \sum_{j \quad 0}^{\infty} \alpha_j^{(z)} \varepsilon_{t-j}^{(z)}, \quad \sum_{j \quad 0}^{\infty} (\alpha_j^{(z)})^2 < \infty,$$

where the $\varepsilon_t^{(z)}$ are i.i.d. with $\mathrm{E}[\varepsilon_t^{(z)}] \quad 0$, $\mathrm{E}[(\varepsilon_t^{(z)})^2] \quad 1$ and $\mathrm{E}[(\varepsilon_t^{(z)})^4] < \infty$, and in a neighbourhood $(0, \varepsilon)$ of the origin, $\alpha_z(\lambda) \quad \sum_{j \quad 0}^{\infty} \alpha_j^{(z)} \exp(ij\lambda)$ is differentiable with $|\mathrm{d}/\mathrm{d}\lambda \alpha_z(\lambda)| \quad \mathrm{O}(|\lambda|^{-1}|\alpha_z(\lambda)|)$ as $\lambda \to 0$.

Assumption 4 was used in Robinson (1995b) with martingale difference innovations; four bounded moments are enough for all our consistency results. We next introduce a further assumption following Chen and Hannan (1980):

Assumption 5. $\varepsilon_t^{(z)}$ has characteristic function $\psi(\theta) \quad \mathrm{E}[\exp(i\theta\varepsilon_t^{(z)})]$ satisfying

$$\sup_{|\theta| \geq \theta_0} |\psi(\theta)| \quad \delta(\theta_0) < 1, \quad \forall \theta_0 > 0, \quad \text{and}$$

$$\int_{-\infty}^{\infty} |\psi(\theta)|^p \, \mathrm{d}\theta < \infty \quad \text{for some integer } p > 1.$$

The first part of Assumption 5 is a Cramér condition, satisfied by distributions with a non zero absolute continuous component, while the second part implies that ε_t has a probability density function. We need this condition to use an asymptotic approximation for the probability density of a finite length vector of discrete Fourier transforms of the innovations ε_t (see Velasco, 2000). It holds for Gaussian series, as the first part of Assumption 4, but also for all usual continuous distributions.

In the next theorem we consider residual based estimates of δ using choices of bandwidths ℓ, m which are powers of T as in previous sections. Furthermore the pooling parameter has to satisfy certain conditions in order to control bias, see

Assumptions A.8 and A.9 in Appendix A. We only analyse consistency in multiple regressions.

Theorem 4. *Under Assumptions* 1, 2, 4, 5 *for* u_t *and* x_{jt}, A.3 *and* A.8 *then as* $T \to \infty$, $0.5 < \delta < d < 1.5$,

$$\log T(\widehat{\delta}^{(J)}(\Delta\hat{u}) \quad \delta) \to_p 0.$$

If additionally Assumption A.9 *holds when* $d + \delta < 1$, $0 \leqslant \delta < d < 1.5$, *then*

$$\log T(\widehat{\delta}^{(J)}(\hat{u}) \quad \delta) \to_p 0.$$

When using original residuals we are now able to deal with values $0.5 \leqslant \delta < d < 1.5$ because the tapered periodogram of the non stationary data x_t remains asympto tically unbiased for $f_{xx}(\lambda)$ when $d \geqslant 1$, unlike for untapered data. The consistency of $\widehat{\delta}^{(J)}(\Delta\hat{u})$ when $0 \leqslant \delta \leqslant 0.5$ could be analysed following the methods of Theorems 8 and 9 of Velasco (1999a).

We find, as with Gaussian data, that the conditions on the bandwidths m and ℓ imposed by Theorem 4 are more restrictive in Case II, $d + \delta < 1$. These imply no further restrictions on the values of d and δ because sufficiently large values of J guarantee that feasible choices of m and ℓ exist to construct consistent estimates of δ, even for arbitrarily small values of $d \quad \delta > 0$, as can be deduced from Assumptions A.8 and A.9.

5. Monte Carlo evidence

In this section we investigate the residual based log periodogram regression according to Theorems 1 and 2 and Corollary 1 experimentally. Let x_t' $(x_{1t}, \ldots, x_{kt})$ consist of $ARFIMA(0, d_i, 0)$ series,

$$(1 \quad L)^{d_i} x_{it} \quad \varepsilon_{it}, \quad i \quad 1, 2, \ldots, k.$$

The true regression model is

$$y_t \quad \sum_{i\ 1}^{k} x_{it} + u_t, \quad t \quad 1, 2, \ldots, T,$$

with $(1 \quad L)^{\delta} u_t \quad \varepsilon_{0t}$, where ε_{it} are i.i.d.$(0, \sigma^2)$ processes independent of each other, $i \quad 0, 1, \ldots, k$. Stationary fractionally integrated series are simulated without approximation using the algorithm by Hosking (1984), and non stationary series are obtained by integration. The ε_{it} are $\mathcal{N}(0, 1)$, or drawn from a t distribution with 3 degrees of freedom or from a χ^2 distribution with 5 degrees of freedom. In the latter two cases the variates have been standardized. This t distribution has only finite second but not higher moments, while the χ^2 is skewed to the right.

The regression model is estimated by OLS,

$$y_t \quad \widehat{\alpha} + \widehat{\beta}' x_t + \hat{u}_t, \quad t \quad 1, 2, \ldots, T.$$

Next, the periodogram is computed from the differenced or original residuals. The corresponding log periodogram regressions are

$$\log(I_{\hat{u}\hat{u}}(\lambda_j)) \quad \widehat{c} + \widehat{\delta}(\hat{u})R_j + \hat{v}_j, \quad j \quad \ell + 1, \ell + 2, \ldots, m, \quad \lambda_j \quad 2\pi j/T,$$

$$\log(I_{\Delta\hat{u}\Delta\hat{u}}(\lambda_j)) \quad \widehat{c} + \widehat{\delta\,\,1}(\Delta\hat{u})R_j + \hat{v}_j, \quad j \quad \ell + 1, \ell + 2, \ldots, m,$$
$$\lambda_j \quad 2\pi j/(T \quad 1),$$

with $R_j \quad \log(4\sin^2(\lambda_j/2)) \approx \quad 2\log(\lambda_j) \quad r_j$. Three different standard errors can be considered. The usual empirical standard error is given by

$$\sqrt{\frac{1}{m \quad \ell}\sum_{j\ \ell+1}^{m}\hat{v}_j^2\left(\sum_{j\ \ell+1}^{m}(R_j \quad \bar{R}_\ell)^2\right)^{-1}}, \quad \bar{R}_\ell \quad \frac{1}{m \quad \ell}\sum_{j\ \ell+1}^{m}R_j.$$

A theoretical modification of the empirical standard errors has been motivated already by Geweke and Porter Hudak (1983):

$$s.e. \quad \sqrt{\frac{\pi^2}{6}\left(\sum_{j\ \ell+1}^{m}(R_j \quad \bar{R}_\ell)^2\right)^{-1}}, \quad \bar{R}_\ell \quad \frac{1}{m \quad \ell}\sum_{j\ \ell+1}^{m}R_j. \tag{7}$$

Finally, the asymptotic standard error due to Robinson (1995a) is $\pi/\sqrt{24m}$. Throughout all experiments we found that the theoretical modification given in (7) outperforms the empirical and the asymptotic standard errors in terms of coverage probabilities. Therefore, only the outcome of t statistics relying on (7) is reported. The test statistics considered hence are

$$t_\delta \quad \frac{\widehat{\delta}(\hat{u}) \quad \delta}{s.e.}, \quad \tau_\delta \quad \frac{\widehat{\delta\,\,1}(\Delta\hat{u}) + 1 \quad \delta}{s.e.}.$$

In our experiments the t statistics are compared with standard normal percentiles. Two sided tests at the 1%, 5% and 10% level are applied. We only report results for $m \quad T^{0.5}$, although a more elaborate choice of optimal m has recently been suggested by Hurvich and Deo (1999) and other deterministic choices such as $m \quad T^{0.4}, T^{0.7}$ have been tried. These produced similar results, as expected, since given fractionally integrated noise models the choice of m should not matter to our main interest, the size of the test (as long as m is big enough) though power increases with m (and does not vary with T). The trimming parameter ℓ is varied very slowly.

Simulations not reported here, in agreement with previous analysis, indicate that the normal approximation is valid for true errors irrespective of any trimming ($\ell \geqslant 0$), use of nonstationary levels ($\delta > 0.5$), or leptokurtic t or skewed χ^2 distributions. Our Monte Carlo design tries to address the validity of these points for residuals and we start investigating how the experimental level of residual based cointegration tests depends on some of the assumptions that we found sufficient to establish limiting normality.

Tables 1 and 2 report percentages of rejection from 2000 replications when testing for the true value of δ using (differences of) residuals from bivariate regressions with Gaussian variables. We observe:

(a) Without trimming, $\ell = 0$, the normal approximation is not valid, at least with the original residuals without differencing.
(b) Trimming only the first frequency, $\ell = 1$, provides a satisfactory normal approximation for $T = 250$ and $T = 1000$ (and also $T = 500$ not reported here).
(c) Even if the gap between δ and d is not as big as it should be according to the theory, i.e. $\delta < d = 0.5$ does not hold, the normal distribution in Tables 1 and 2 still yields a useful approximation in case of trimming the first frequency.

Table 1
Experimental level in case of differenced residuals (τ_δ), $k = 1$

d	α (%)	$\delta = 1.0$	$\delta = 0.9$	$\delta = 0.8$	$\delta = 0.7$	$\delta = 0.6$	ℓ
$T = 250$, $m = 16$							
	1	1.10	1.15	1.45	1.65	2.10	
1.4	5	4.70	5.55	5.10	6.00	6.20	0
	10	9.75	11.20	10.20	9.85	11.20	
	1	1.45	1.30	1.15	1.10	1.45	
1.4	5	5.65	4.90	5.25	4.95	5.05	1
	10	10.50	10.50	10.05	9.85	9.10	
	1	1.90	1.75	1.60	1.55	1.85	
1.0	5	6.50	6.25	5.40	5.60	5.20	0
	10	12.15	12.00	10.70	11.15	10.85	
	1	1.60	2.05	1.30	1.40	1.35	
1.0	5	5.15	5.60	5.00	5.10	5.65	1
	10	10.00	9.35	8.60	9.25	10.15	
$T = 1000$, $m = 32$							
	1	1.80	1.55	1.25	0.85	1.45	
1.4	5	6.20	5.75	6.10	5.45	5.30	0
	10	10.60	10.80	10.90	10.75	10.35	
	1	1.10	1.30	1.20	1.30	0.90	
1.4	5	4.80	4.90	4.90	4.30	4.70	1
	10	10.30	9.30	9.85	8.15	10.45	
	1	1.50	1.25	1.60	1.85	1.35	
1.0	5	5.95	5.50	5.70	6.60	6.55	0
	10	10.75	10.55	10.45	11.45	11.75	
	1	1.20	1.25	1.60	1.10	1.35	
1.0	5	5.00	4.80	5.05	4.00	5.60	1
	10	9.75	9.50	9.85	9.40	10.00	

The variables are integrated of order d while the errors are $I(\delta)$. All series are constructed from innovations that follow a $\mathcal{N}(0, 1)$ distribution. Tests for the true value of δ built upon differenced residuals from bivariate regressions. Percentage of rejections of two-sided tests at the 1%, 5% and 10% level based on a standard normal approximation of τ_δ. The trimming number is ℓ.

Table 2
Experimental level in case of original residuals (t_δ), $k = 1$

d	α (%)	$\delta = 1.0$	$\delta = 0.8$	$\delta = 0.6$	$\delta = 0.4$	$\delta = 0.2$	ℓ
$T = 250$, $m = 16$							
	1	2.85	2.25	2.40	2.25	2.45	
1.4	5	8.30	8.10	8.00	7.00	7.25	0
	10	15.15	13.70	14.10	11.95	12.10	
	1	1.80	1.45	1.55	1.40	1.50	
1.4	5	5.85	5.55	5.25	5.40	4.90	1
	10	9.80	10.05	9.15	9.25	9.25	
	1	3.15	3.00	2.65	2.20	2.50	
1.0	5	8.65	8.35	8.15	7.20	6.85	0
	10	14.40	14.15	13.75	12.15	12.25	
	1	1.55	1.25	1.35	1.20	1.35	
1.0	5	4.85	5.90	5.70	5.05	5.60	1
	10	9.15	9.30	10.15	9.25	9.30	
$T = 1000$, $m = 32$							
	1	2.55	2.60	2.60	2.70	2.65	
1.4	5	7.80	7.80	7.05	8.30	7.30	0
	10	12.70	13.90	11.60	14.20	12.60	
	1	1.20	1.40	1.05	1.35	1.45	
1.4	5	4.65	5.55	4.80	5.35	5.50	1
	10	9.45	10.75	9.40	9.90	10.40	
	1	2.60	2.00	2.25	2.20	2.00	
1.0	5	7.00	7.45	7.40	6.95	6.20	0
	10	12.25	12.30	12.25	11.80	10.70	
	1	1.35	1.55	1.25	1.45	1.40	
1.0	5	4.35	6.05	5.65	5.55	5.35	1
	10	9.05	11.60	10.65	9.65	10.20	

The variables are integrated of order d while the errors are $I(\delta)$. All series are constructed from innovations that follow a $\mathcal{N}(0, 1)$ distribution. Tests for the true value of δ built upon original residuals from bivariate regressions. Percentage of rejections of two-sided tests at the 1%, 5% and 10% level based on a standard normal approximation of t_δ. The trimming number is ℓ.

(d) Even without cointegration, e.g. $\delta \quad d \quad 1$ in Tables 1 and 2, the normal approximation seems to provide a reasonable guideline as long as trimming is applied.

Table 3 considers the power of residual based tests from bivariate regressions and can be summarized as follows:

(e) As the trimming parameter grows, power decreases.
(f) The difference in power between the log periodogram regression of differences or levels of residuals when testing for $\delta_0 \quad 1$ is negligible.
(g) From the levels of residuals one may test for $\delta_0 \quad 0$, while tests for this hypothesis from differences (not reported here) suffer from gross size distortion.

Table 3
5% power in case of residuals, $k = 1$

T	$\delta_0 = 1.0$	$\delta_1 = 0.9$	$\delta_1 = 0.8$	$\delta_1 = 0.7$	$\delta_1 = 0.6$	ℓ
Differenced residuals ($\tau_{\delta=\delta_0}$)						
250	5.65	6.40	9.50	18.05	26.95	1
($m = 16$)	5.35	6.30	8.35	12.80	19.50	2
1000	5.15	8.25	19.70	44.35	67.40	1
($m = 32$)	5.55	7.60	15.95	32.25	56.05	2
Original residuals ($t_{\delta=\delta_0}$)						
250	5.85	6.35	10.35	17.90	29.35	1
($m = 16$)	5.85	5.90	8.30	13.55	20.20	2
1000	4.85	8.30	19.05	43.85	71.05	1
($m = 32$)	4.30	7.45	15.40	32.10	57.45	2
Original residuals ($t_{\delta=\delta_0}$)						
T	$\delta_0 = 0.0$	$\delta_1 = 0.1$	$\delta_1 = 0.2$	$\delta_1 = 0.3$	$\delta_1 = 0.4$	ℓ
250	5.75	6.25	8.85	18.25	28.45	1
($m = 16$)	5.00	6.00	9.00	14.45	21.15	2
1000	5.75	8.25	23.90	48.05	71.60	1
($m = 32$)	4.35	8.15	17.95	35.05	57.85	2

The regressor is integrated of order 1.4 while the errors are $I(\delta_0)$ under the respective null hypotheses and $I(\delta_1)$ under the alternatives. All series are constructed from innovations that follow a $\mathcal{N}(0, 1)$ distribution. Tests for δ_0 built upon differenced or original residuals from bivariate regressions. Percentage of rejections of two-sided tests at the 5% level based on a standard normal approximation of τ_δ or t_δ, respectively. The trimming number is ℓ.

Tables 4 and 5 are constructed from residuals from bivariate regressions where regressors and errors rely on either t or χ^2 distributions (similar results not reported here arise for t distributed regressors and χ^2 distributed residuals and the other way round). We observe:

(h) The statements (a) (d) continue to hold in case of the considered leptokurtic and skewed distributions.

Next, we investigated the log periodogram regression (of differences) of residuals from trivariate regressions, k 2, where all variables are constructed from Gaussian variates. Again, without trimming the normal approximation is clearly not useful. Furthermore, the following findings arise from Tables 6 and 7.

(i) If d_1 d_2 1.4, trimming of only the first harmonic frequency, ℓ 1, results in a fairly reliable normal approximation. This is also true for the log periodogram regression of residuals without differencing even if $\delta > 0.5$. Moreover, it seems to hold in case of a cointegration gap smaller than 0.5, e.g. for δ 1.

Table 4
Experimental level in case of differenced residuals (τ_δ), $k = 1$, $T = 250$

d	α (%)	$\delta = 1.0$	$\delta = 0.9$	$\delta = 0.8$	$\delta = 0.7$	$\delta = 0.6$	ℓ
t_3							
	1	2.40	2.00	1.35	2.10	1.70	
1.4	5	6.60	6.25	5.35	7.35	7.30	0
	10	11.30	11.15	10.20	12.35	12.65	
	1	0.85	1.15	1.40	0.95	0.85	
1.4	5	5.55	5.10	5.30	5.00	4.10	1
	10	10.65	9.75	10.45	9.55	9.00	
	1	2.75	2.30	1.85	2.10	1.65	
1.0	5	6.75	7.05	6.15	5.80	5.55	0
	10	11.85	11.35	11.35	10.30	10.70	
	1	1.30	1.25	1.15	1.44	1.40	
1.0	5	5.05	5.30	5.05	4.55	4.35	1
	10	10.10	9.25	9.45	9.10	9.95	
$\chi^2(5)$							
	1	1.85	1.55	2.40	1.95	1.75	
1.4	5	7.25	5.95	6.05	6.25	5.95	0
	10	12.20	11.50	11.10	10.60	11.55	
	1	1.00	0.90	1.40	1.30	1.65	
1.4	5	4.25	5.20	5.70	5.25	5.65	1
	10	10.05	10.20	10.80	9.80	11.05	
	1	2.50	2.50	1.95	2.80	1.90	
1.0	5	8.00	6.50	6.55	7.40	6.65	0
	10	14.05	12.35	11.55	12.85	11.45	
	1	1.95	1.25	1.35	1.55	1.15	
1.0	5	5.00	5.90	5.10	5.30	4.65	1
	10	10.40	10.55	10.00	9.85	10.20	

The variables are integrated of order d while the errors are $I(\delta)$. All series are constructed from innovations that follow either t_3 or $\chi^2(5)$ distributions. Tests for the true value of δ built upon differenced residuals from bivariate regressions. Percentage of rejections of two-sided tests at the 1%, 5% and 10% level based on a standard normal approximation of τ_δ. The trimming number is ℓ and $m = 16$.

(j) If d_1 1.4 and d_2 0.6, cases I and II are mixed when δ 0.2, which violates Assumption 3. Nevertheless, the normal approximation provides a valid guideline in case of trimming the first frequency (original residuals). Surprisingly, this also seems to hold for T 250 even if $\delta > d_2$ 0.6, where Assumption 3 is again violated. For T 1000 observations slightly different results emerge: in case that $\delta > d_2$ 0.6 trimming only one harmonic frequency is not sufficient for a normal approximation, so trimming may need to grow with sample size.

We have also investigated in Tables 8 and 9 the effects of pooling a small number J of periodogram ordinates. In this case the asymptotic variance of the log periodogram estimate is reduced and we replace in the expression for the standard

Table 5
Experimental level in case of original residuals (t_δ), $k = 1$, $T = 250$

d	α (%)	$\delta = 1.0$	$\delta = 0.8$	$\delta = 0.6$	$\delta = 0.4$	$\delta = 0.2$	ℓ
t_3							
	1	3.35	3.65	3.30	3.25	2.75	
1.4	5	9.50	9.05	9.55	8.50	8.30	0
	10	15.30	14.80	16.20	14.15	14.50	
	1	1.45	1.20	1.20	1.70	1.65	
1.4	5	5.25	5.10	4.85	4.80	6.25	1
	10	10.20	9.75	9.50	9.00	10.30	
	1	3.00	2.60	2.25	2.35	2.20	
1.0	5	8.35	7.25	7.90	7.10	7.20	0
	10	14.10	12.35	14.00	12.80	12.55	
	1	1.55	0.90	1.60	1.35	1.60	
1.0	5	5.80	4.70	4.75	5.35	5.90	1
	10	10.05	10.25	10.15	9.90	10.85	
$\chi^2(5)$							
	1	3.45	3.40	2.75	3.30	2.95	
1.4	5	10.20	8.25	9.35	10.70	7.45	0
	10	15.30	13.95	15.25	17.15	13.10	
	1	1.25	1.50	1.10	1.15	1.65	
1.4	5	4.80	5.05	5.05	5.05	5.80	1
	10	9.15	9.25	10.85	10.45	9.60	
	1	3.55	3.40	2.95	2.65	1.75	
1.0	5	9.10	8.65	7.85	7.40	6.50	0
	10	14.40	13.70	12.85	12.90	11.15	
	1	1.30	1.20	1.40	1.10	1.35	
1.0	5	4.90	4.75	4.70	4.90	4.90	1
	10	9.05	9.95	9.10	9.20	8.85	

The variables are integrated of order d while the errors are $I(\delta)$. All series are constructed from innovations that follow either t_3 or $\chi^2(5)$ distributions. Tests for the true value of δ built upon the levels of residuals from bivariate regressions. Percentage of rejections of two-sided tests at the 1%, 5% and 10% level based on a standard normal approximation of t_δ. The trimming number is ℓ and $m = 16$.

errors $\pi^2/6$ by the general expression $\psi'(J)/J$ for J 1, 2, ..., where ψ is the digamma function (cf. Robinson, 1995a). Further to the previous findings, we can state that for Gaussian and other distributions (not reported here),

(k) The larger J, the larger the power with ℓ 1 when testing δ_0 1 with differenced residuals or δ_0 0 with original ones, keeping good size properties.
(l) Use of original residuals with $J > 1$ when testing δ_0 1 requires ℓ 2 to maintain size, resulting in a noticeable power loss compared to testing based on errors (Table 8).

This Monte Carlo evidence can be summarized as follows as a rule of thumb for empirical work with bivariate and multiple regressions: If the log periodogram

Table 6
Experimental level in case of differenced residuals (τ_δ), $k = 2$

d_1	d_2	α (%)	$\delta = 1.0$	$\delta = 0.9$	$\delta = 0.8$	$\delta = 0.7$	$\delta = 0.6$	ℓ
$T = 250$, $m = 16$								
		1	3.30	3.20	3.45	2.65	1.70	
1.4	1.4	5	8.90	8.00	8.65	8.70	6.40	0
		10	14.70	13.75	14.10	14.65	11.95	
		1	1.15	1.45	1.50	1.55	0.95	
1.4	1.4	5	5.65	5.25	4.50	6.35	4.60	1
		10	10.60	9.85	9.75	10.75	9.45	
		1	2.25	1.85	1.70	1.90	1.80	
1.4	0.6	5	7.40	5.95	6.00	6.90	5.45	0
		10	12.65	12.00	11.20	11.15	11.55	
		1	1.40	1.40	1.65	1.55	1.55	
1.4	0.6	5	5.10	5.20	6.15	5.55	5.10	1
		10	9.25	9.80	11.35	10.75	10.20	
$T = 1000$, $m = 32$								
		1	2.60	2.40	2.35	2.40	2.60	
1.4	1.4	5	7.40	7.60	7.20	6.85	7.35	0
		10	12.40	13.65	12.10	12.00	12.70	
		1	0.90	1.00	1.15	0.95	1.35	
1.4	1.4	5	4.45	5.40	4.60	4.65	6.20	1
		10	10.35	10.45	10.20	10.15	11.00	
		1	3.00	2.35	1.90	1.50	2.10	
1.4	0.6	5	9.50	7.30	7.25	6.00	6.45	0
		10	14.60	12.20	12.30	11.00	11.80	
		1	3.00	1.85	1.35	1.30	1.35	
1.4	0.6	5	8.65	7.20	5.70	5.20	5.20	1
		10	15.25	12.85	10.65	9.80	9.05	

The regressors are integrated of order d_1 and d_2 while the errors are $I(\delta)$. All series are constructed from innovations that follow a $\mathcal{N}(0, 1)$ distribution. Tests for the true value of δ built upon differenced residuals from trivariate regressions. Percentage of rejections of two-sided tests at the 1%, 5% and 10% level based on a standard normal approximation of τ_δ. The trimming number is ℓ.

regression is applied to the level of OLS residuals with trimming of the first harmonic frequency only, then the normal approximation of the t statistic t_δ with theoretical standard errors *s.e.* should yield reliable inference when we test for any δ between 0 and 1. This is valid even for not Gaussian data and even if $\delta < d_i$ 0.5 does not hold for all i, except of the extreme case where $d_i < \delta$ for some i. The same seems to hold true for the t statistic τ_δ from differences for any δ between 0.5 and 1. If sample size is large enough, pooling increases power with ℓ 1.

Finally, we want to propose an empirical research strategy as an overall summary of our results. In most economic applications there are two null hypotheses (with corresponding alternatives) of major interest:

$$H_0^{(1)} : \delta \quad 1 \text{ vs. } \delta < 1, \quad H_0^{(0)} : \delta \quad 0 \text{ vs. } \delta > 0.$$

Table 7
Experimental level in case of original residuals (t_δ), $k = 2$

d_1	d_2	α (%)	$\delta = 1.0$	$\delta = 0.8$	$\delta = 0.6$	$\delta = 0.4$	$\delta = 0.2$	ℓ
$T = 250$, $m = 16$								
		1	9.35	7.65	5.85	6.50	5.75	
1.4	1.4	5	18.80	17.40	16.00	14.90	14.15	0
		10	26.80	25.50	23.25	21.60	20.30	
		1	1.10	1.70	1.60	1.60	1.50	
1.4	1.4	5	5.75	5.30	5.05	5.80	5.65	1
		10	9.85	10.15	9.70	10.15	9.90	
		1	4.90	4.35	3.75	5.10	2.65	
1.4	0.6	5	12.70	10.70	10.40	11.85	8.40	0
		10	19.10	17.35	17.45	16.85	13.85	
		1	1.60	1.55	2.05	1.90	1.20	
1.4	0.6	5	5.90	6.10	6.25	6.40	4.80	1
		10	10.50	11.10	10.80	10.80	9.60	
$T = 1000$, $m = 32$								
		1	7.45	5.50	5.80	5.30	5.65	
1.4	1.4	5	14.05	13.50	13.55	13.45	13.10	0
		10	21.95	20.75	20.55	20.85	18.90	
		1	1.35	1.65	1.95	1.95	0.90	
1.4	1.4	5	4.85	5.70	6.80	6.45	4.15	1
		10	9.70	11.75	11.55	11.45	9.35	
		1	5.05	3.65	3.80	3.50	3.05	
1.4	0.6	5	13.60	10.10	10.25	8.55	7.55	0
		10	20.50	16.15	15.05	14.20	13.35	
		1	2.85	1.35	1.05	1.60	1.60	
1.4	0.6	5	8.70	5.20	4.25	6.30	5.50	1
		10	15.10	8.90	8.80	10.65	11.20	

The regressors are integrated of order d_1 and d_2 while the errors are $I(\delta)$. All series are constructed from innovations that follow a $\mathcal{N}(0,1)$ distribution. Tests for the true value of δ built upon original residuals from trivariate regressions. Percentage of rejections of two-sided tests at the 1%, 5% and 10% level based on a standard normal approximation of t_δ. The trimming number is ℓ.

We suggest to test $H_0^{(1)}$ from the differences of residuals, while clearly $H_0^{(0)}$ should be tested from levels. If, first, both hypotheses are rejected, there is fractional cointegration, i.e. we have long memory but transitory equilibrium deviations. The degree of persistence δ should then be estimated from the levels of the residuals; approximate confidence intervals allow to test whether the estimate is significantly different from 0.5, the borderline of non stationarity. If, second, $H_0^{(0)}$ is not rejected while $H_0^{(1)}$ is, we have the strong cointegration result that the errors may be considered as $I(0)$. If, third, $H_0^{(0)}$ is rejected while $H_0^{(1)}$ must be accepted, the error should be considered as $I(1)$, i.e. persistent, and there is no long run equilibrium. If, finally, none of these hypotheses can be rejected, more data should be used to increase power.

Table 8
5% power in case of errors with pooling, $k = 1$

Pooling	$\delta_0 = 1.0$	$\delta_1 = 0.9$	$\delta_1 = 0.8$	$\delta_1 = 0.7$	$\delta_1 = 0.6$	ℓ
Differenced errors ($\tau_{\delta\ \delta_0}$)						
	5.10	8.50	17.35	32.15	50.00	0
$J = 2$	5.05	6.45	11.10	18.55	29.35	1
	5.25	6.25	9.10	14.90	23.20	2
	5.20	7.60	16.50	29.25	45.40	0
$J = 3$	5.45	7.25	12.55	21.70	35.55	1
	4.70	6.05	8.55	12.10	17.75	2
Original errors ($t_{\delta\ \delta_0}$)						
Pooling	$\delta_0 = 1.0$	$\delta_1 = 0.9$	$\delta_1 = 0.8$	$\delta_1 = 0.7$	$\delta_1 = 0.6$	ℓ
	10.85	7.05	6.85	12.40	24.00	0
$J = 2$	6.35	5.60	6.95	11.35	20.35	1
	4.30	5.15	6.95	10.35	16.55	2
	31.45	19.25	9.85	7.75	12.00	0
$J = 3$	11.80	8.30	7.10	8.95	15.35	1
	6.00	5.70	5.50	8.05	10.85	2
Original errors ($t_{\delta\ \delta_0}$)						
Pooling	$\delta_0 = 0.0$	$\delta_1 = 0.1$	$\delta_1 = 0.2$	$\delta_1 = 0.3$	$\delta_1 = 0.4$	ℓ
	5.10	7.30	18.20	37.50	59.65	0
$J = 2$	5.05	7.20	12.00	22.00	35.25	1
	5.25	6.00	9.45	16.00	26.50	2
	5.20	6.95	17.75	35.95	59.45	0
$J = 3$	5.45	6.75	13.30	25.15	39.75	1
	4.70	6.00	8.75	13.50	21.70	2

Errors are $I(\delta_0)$ under the respective null hypothesis and $I(\delta_1)$ under the alternatives. All series are constructed from innovations that follow a $\mathcal{N}(0, 1)$ distribution. Percentage of rejections of two-sided tests at the 5% level based on a standard normal approximation of τ_δ or t_δ, respectively. Sample size is $T = 250$. The bandwidth is $m = 16$. The trimming number is ℓ. The pooling number is J.

6. Exchange rate dynamics

In a cointegration study with integer orders of integration, Baillie and Bollerslev (1989) argued that seven different nominal spot exchange rates, namely, Germany, the United Kingdom, Japan, Canada, France, Italy and Switzerland, all relative to the US Dollar and observed daily from 1980 to 1985, do contain unit roots in their univariate time series representations, giving also evidence in support of the existence of a single cointegrating vector between this set of nominal exchange rates. Such a cointegration relation has been questioned and found to be fragile by Sephton and Larsen (1991) and Diebold et al. (1994) even though both used the same data set. Diebold et al. note that the lack of cointegration is reinforced when using data covering the post 1973 floating exchange rate regime. Subsequently, Baillie and Bollerslev (1994) collected more reliable evidence in a fractional set up, generalizing

Table 9
5% power in case of residuals with pooling, $k = 1$

Pooling	$\delta_0 = 1.0$	$\delta_1 = 0.9$	$\delta_1 = 0.8$	$\delta_1 = 0.7$	$\delta_1 = 0.6$	ℓ
Differenced residuals ($\tau_{\delta\ \delta_0}$)						
	5.20	8.50	16.90	32.05	50.35	0
$J = 2$	5.55	6.20	10.15	17.75	29.15	1
	5.20	6.15	9.10	13.85	23.10	2
	5.20	7.25	15.80	29.10	45.20	0
$J = 3$	5.55	7.05	12.00	21.40	35.10	1
	5.50	6.05	8.20	12.30	18.05	2
Original residuals ($t_{\delta\ \delta_0}$)						
Pooling	$\delta_0 = 1.0$	$\delta_1 = 0.9$	$\delta_1 = 0.8$	$\delta_1 = 0.7$	$\delta_1 = 0.6$	ℓ
	9.65	8.55	13.95	25.80	42.05	0
$J = 2$	6.70	5.80	7.75	12.50	21.55	1
	5.60	5.50	6.45	10.80	17.35	2
	21.30	11.80	8.25	13.15	23.10	0
$J = 3$	13.15	7.55	6.65	9.80	16.65	1
	7.05	5.65	5.60	7.45	10.50	2
Original residuals ($t_{\delta\ \delta_0}$)						
Pooling	$\delta_0 = 0.0$	$\delta_1 = 0.1$	$\delta_1 = 0.2$	$\delta_1 = 0.3$	$\delta_1 = 0.4$	ℓ
	6.90	6.25	12.20	25.10	45.35	0
$J = 2$	5.10	6.55	11.40	19.80	31.10	1
	5.15	5.40	9.20	15.10	25.00	2
	6.80	6.50	12.25	24.90	44.40	0
$J = 3$	5.65	6.20	11.85	22.25	37.30	1
	4.75	5.70	8.05	13.00	21.60	2

The regressors are integrated of order 1.4 while the errors are $I(\delta_0)$ under the respective null hypothesis and $I(\delta_1)$ under the alternatives. All series are constructed from innovations that follow a $\mathcal{N}(0,1)$ distribution. Tests for δ_0 built upon original residuals from bivariate regressions. Percentage of rejections of two-sided tests at the 5% level based on a standard normal approximation of τ_δ or t_δ, respectively. Sample size is $T = 250$. The bandwidth is $m = 16$. The trimming number is ℓ. The pooling number is J.

the error correction formulation to allow for possible fractional cointegration. They find evidence that a linear combination of the same spot exchange rates contains long range dependence. In particular, they estimate an error correction term with memory 0.89 in a fractional white noise model, with an (asymptotic) standard error of 0.02.

In this section we confirm their results for the same seven currencies. We use monthly data taken from Citibase and run from 1974.1 until 1997.12, which leaves us with T 288 observations. Following Baillie and Bollerslev (1994), the logarithms of the data are analysed. The use of monthly observations may help to control changing conditional variances and should not affect the analysis of long run properties compared to higher frequency data.

Table 10
Correlogram of CAN

	Levels	First differences	Levels	First differences
Lag	*ACF*		*PACF*	
1	0.988	0.204	0.988	0.204
2	0.975	0.019	0.041	0.063
3	0.963	0.015	0.024	0.034
4	0.951	0.056	0.023	0.046
5	0.937	0.047	0.074	0.071
6	0.923	0.024	0.019	0.056
7	0.910	0.010	0.027	0.015
8	0.895	0.193	0.017	0.207
9	0.879	0.107	0.081	0.028
10	0.863	0.229	0.025	0.233

Correlogram and partial correlogram for the Canada exchange rate. The asymptotic standard error is 0.117 under the null of no correlation.

On application of the well known *ADF* test to our data set, we obtain p values greater than 0.05. Moreover, in some cases we cannot reject the presence of a unit root at any conventional significance level. For example, in the Canadian case, the value obtained of the *ADF* test is 0.63, whereas the 10% critical value is 2.87. To further confirm this claim, in Table 10 we present the *ACF* and *PACF* of the levels and first differences of the Canada exchange rate series. It can be observed that the autocorrelations exhibit the typical very slow decline associated with a nonstationary process, and that the autocorrelations of the *change*, i.e. the autocorrelations of the approximate rate of return, are all them small.

Nonetheless, an alternative potential explanation for the high persistence of the exchange rate is the possibility that the memory parameters of these series may be fractional, since it is well known that standard integer order unit root tests have low power against fractional alternatives (cf., e.g., Hassler and Walter, 1994; Dolado and Marmol, 1997).

In order to confirm this possibility, we start with determining the memory of the individual series by applying the log periodogram regression without trimming, ℓ 0, to the differences of the original data. The regression range was chosen as m 18, 20, 22. This choice provided fairly stable estimates and avoids the first seasonal frequency, which given our monthly data is $\lambda_{T/12}$ λ_{24}. However these bandwidths are far from mean square optimal choices, $T^{4/5} \approx 93$, which would lead to serious bias in semiparametric estimates and distortions in our statistical inference.

With the standard errors *s.e.* from (7), the estimates presented in Table 11 for Germany, UK, Switzerland and Japan are not significantly different from 1, while France, Italy and Canada have significantly larger values. Consequently, if we test according to Robinson (1995a) that all seven estimates are equal, the p value of the

Table 11
Individual memory, 1974.1 1997.12

		GER	UK	SWI	JAP	FRA	ITA	CAN
$m = 18$	$\widehat{d}(\Delta u)$	1.17	1.02	0.86	1.24	1.41	1.34	1.40
	$\tau_{d=1}$	0.88	0.10	0.72	1.23	2.11	1.75	2.06
$m = 20$	$\widehat{d}(\Delta u)$	1.15	1.10	0.84	1.23	1.33	1.34	1.44
	$\tau_{d=1}$	0.81	0.53	0.88	1.29	1.84	1.86	2.42
$m = 22$	$\widehat{d}(\Delta u)$	1.18	1.15	0.84	1.16	1.31	1.32	1.35
	$\tau_{d=1}$	1.07	0.88	0.93	0.94	1.80	1.90	2.05

Log-periodogram regression of differences of logarithms with $\ell = 0$. The t statistics built on the standard errors $s.e. = 0.194, 0.181, 0.170$ for $m = 18, 20, 22$, respectively.

Table 12
Residual analysis for separate regressions

m	17	18	19	20	21	22
s.e.	0.262	0.250	0.240	0.230	0.221	0.213
FRA on ITA, CAN						
$\widehat{\delta}(\Delta\hat{u})$	1.37	1.23	1.18	1.14	1.05	0.99
$\tau_{\delta=1}$	1.41	0.92	0.75	0.61	0.23	0.05
GER on UK, SWI, JAP						
$\widehat{\delta}(\Delta\hat{u})$	0.79	0.74	0.88	0.89	0.84	0.92
$\tau_{\delta=1}$	0.80	1.04	0.50	0.48	0.72	0.38

Log-periodogram regression of differenced residuals with $\ell = 1$. The t statistics built on the standard error *s.e.* from (7).

Wald statistics are always less than 0.001. Note that this multivariate inference is only valid under no cointegration (cf. Assumption A.2). The null hypothesis that France, Italy and Canada have the same memory parameter, however, is clearly not rejected (p value >0.964), while the hypothesis of a common d of Germany, UK, Switzerland and Japan is not rejected at the 5% level for small m. We conclude that there are two groups of data: Germany, UK, Switzerland and Japan may be considered as $I(1)$, while the order of integration of France, Italy and Canada is roughly 1.4.

We hence start with separate cointegrating regressions and apply the log periodogram regression to differenced residuals. First, France is regressed on Italy and Canada, see the upper panel in Table 12. With trimming the first frequency, ℓ 1, and varying m we clearly cannot reject that the residuals are integrated of

Table 13
Final residual analysis
GER on UK, SWI, JAP and RES(FRA on ITA, CAN)

m	17	18	19	20	21	22
From differences, $H_0 : \delta = 1$						
$\widehat{\delta}(\Delta\hat{u})$	0.46	0.49	0.62	0.67	0.65	0.66
$\tau_{\delta\ 1}$	2.06	2.04	1.58	1.44	1.58	1.60
p-val.	0.020	0.021	0.057	0.075	0.057	0.055
From levels, $H_0 : \delta = 0$						
$\widehat{\delta}(\hat{u})$	0.63	0.61	0.67	0.69	0.66	0.68
$t_{\delta\ 0}$	2.41	2.44	2.79	3.00	2.99	3.19
p-val.	0.008	0.007	0.003	0.001	0.001	0.001

Log-periodogram regression of differenced and original residuals with $\ell = 1$. The t statistics built on the standard error *s.e.* from Table 12.

order one. Hence, we have three $I(1.4)$ series that cointegrate to $I(1)$ residuals. In the lower panel of Table 12 it is reported that the null hypothesis that the $I(1)$ series from Germany, UK, Switzerland and Japan do not cointegrate (δ 1), cannot be rejected.

Finally, we regress the German data on UK, Switzerland, Japan and the $I(1)$ residual RES from the regression of France on Italy and Canada. The results with trimming one frequency are presented in Table 13. From differences we first test the null of no cointegration, δ 1. For all m from 17 to 22 it is rejected at least at the 10% level, and most of the times the p values are close or below the 5% level. At the same time, the log periodogram regression of the original residuals clearly rejects the null hypothesis δ 0. We conclude that it is fractional cointegration that links the considered exchange rates. The memory parameter δ of the equilibrium deviations is estimated as approximately 0.65 from levels as well as from differences. It is never significantly different from 0.5, i.e. we cannot not reject that the error term is non stationary, although we have found that it is not persistent ($\delta < 1$).

We also did the analysis from Table 13 without trimming, ℓ 0. The resulting t statistics not reported here are very similar to those from Table 13, because the standard errors are smaller without trimming and the estimates are closer to the null. From levels one roughly estimates $\widehat{\delta}(\hat{u})$ 0.5, while the log periodogram regression from differences yields approximately $\widehat{\delta}(\Delta\hat{u})$ 0.7. The findings with trimming from Table 13 where $\widehat{\delta}(\hat{u}) \approx \widehat{\delta}(\Delta\hat{u})$ seem to be more reliable.

7. Concluding remarks

In this paper we followed the route opened by Robinson (1995a, Remark 7) for sound statistical inference on memory properties of fractional models. He suggested that given a sufficiently fast rate of convergence of the regression estimator the

residual based log periodogram regression should result in asymptotic normality just as with observed series (confer the application in Robinson and Marinucci, 2001). Indeed, we found that given the gap between the orders of integration of regressors and error is big enough, the log periodogram regression of residuals gives rise to limiting normality. This result essentially relies on trimming the very first few frequencies of the periodogram, a policy that was not employed by the empirical and experimental papers reviewed in the introduction. We hence obtained simple conditions for consistent estimation of the degree of persistence in the deviations from the long run equilibrium which are more general than most parametric models used in common practice. Given asymptotically normal estimators this allows for statistical inference of immediate economic interest. We are now able to discriminate on sound asymptotic grounds between short memory errors, stationary long memory innovations, non stationary but transitory equilibrium deviations, and finally non stationary and persistent errors.

Our results also cover the integer cointegration case of $I(1)$ regressors with $I(0)$ errors. But contrasting the residual based work by Phillips and Ouliaris (1990), Shin (1994) or more recently Xiao (1999) the asymptotic theory we suggest is standard and moreover does not depend on the number of regressors. What is more, a system approach of joint estimation of the orders of integration of regressors and disturbance term is possible, and a pooled version was shown to be robust to departures from Gaussianity and from *strongly* cointegrated systems with $d - \delta > 0.5$. We evaluated the asymptotic results by means of Monte Carlo experiments where it turned out that trimming only one frequency should be enough for practical purposes with usual sample sizes.

To illustrate these points we applied the log periodogram regression to a set of seven nominal exchange rates, collecting evidence that exchange rates are linked by a fractional cointegration relation. In this respect, with our semiparametric set up we conclude that there could be two clusters of currencies. On the one hand, Germany, UK, Switzerland and Japan, that may be considered as $I(1)$ processes. On the other hand, France, Italy and Canada, with an order of integration about 1.4. We fail to find evidence of cointegration among the first group of exchange rates, whereas we cannot reject that the residuals from a regression of France on Italy and Canada are $I(1)$. However, we find polynomial cointegration when regressing the German data on UK, Switzerland, Japan and the residuals from the regression of France on Italy and Canada, so that they do not drift apart in the long run. The memory parameter of the equilibrium deviations of this extended regression is about 0.65, i.e. the error correction term is non stationary but not persistent.

Acknowledgements

We are grateful to P.M. Robinson and two referees, and to the participants at the 8th World Congress of the Econometric Society, Seattle, 2000, and at the Workshop on New Approaches to the Study of Economic Fluctuations by CEPR, Hydra, 2000, for very helpful comments on an earlier version. This paper was completed while the

first author was visiting Universidad Carlos III de Madrid. Financial support through the European Commission programme *Training and Mobility of Researchers* and CICYT Refs. No. SEC 2001 08090 and BEC 2001 1270 and SEJ 2004 04583/ECON is gratefully acknowledged.

Appendix A. Assumptions and auxiliary results

For our asymptotic theory we will need to impose the following regularity assumption (cf. Assumptions 1 and 2 of Robinson, 1995a) which applies either to the spectral density (of stationary processes) or to the pseudospectral density (of non stationary processes), imposing the rate in (1).

Assumption A.1. The (pseudo) spectral density $f_{zz}(\lambda)$ of z_t, $z \in \{x, u\}$ (d_x d, d_u δ) satisfies, $0 < \gamma \leqslant 2$, $0 < G_z < \infty$,

$$f_{zz}(\lambda) \quad G_z \lambda^{-2d_z}(1 + \mathrm{O}(|\lambda|^{\gamma})) \quad \text{as } \lambda \to 0,$$

and is differentiable in a neighbourhood $(0, \varepsilon)$ of the origin with

$$\left| \frac{\mathrm{d}}{\mathrm{d}\lambda} f_{zz}(\lambda) \right| \quad \mathrm{O}(|\lambda|^{-1-2d_z}) \quad \text{as } \lambda \to 0.$$

This assumption holds for standard *ARFIMA* series with γ 2 and for any fractional model with $f(\lambda)$ $(2\sin(\lambda/2))^{-2d_z} f^*(\lambda)$, if in an interval of the origin either the $I(0)$ short memory spectral density $f^*(\lambda)$ is Lipschitz(γ), $0 < \gamma \leqslant 1$, or its derivative is Lipschitz (γ 1), $1 < \gamma \leqslant 2$. The following assumption is a multivariate generalization of this set up for z_t containing possibly both stationary and non stationary elements (cf. Robinson, 1995a).

Assumption A.2. The (pseudo) spectral density matrix $\mathbf{f}(\lambda)$ $(f_{ij}(\lambda))$ of z_t $(\Delta^{\bar{\delta}} u_t, \Delta x_t')'$ satisfies, $0 < \gamma \leqslant 2$, $0 < G_i < \infty$, i, j $0, 1, \ldots, k$,

$$f_{ii}(\lambda) \quad G_i \lambda^{-2D_i}(1 + \mathrm{O}(|\lambda|^{\gamma})) \quad \text{as } \lambda \to 0,$$

where $\bar{\delta}$ $\lfloor \delta + 0.5 \rfloor$ and D_o δ $\bar{\delta}$, and is differentiable in a neighbourhood $(0, \varepsilon)$ of the origin with

$$\left| \frac{\mathrm{d}}{\mathrm{d}\lambda} f_{ij}(\lambda) \right| \quad \mathrm{O}(|\lambda|^{-1-D_i-D_j}) \quad \text{as } \lambda \to 0.$$

Set the coherence matrix $R(\lambda)$ of $(\Delta^{\bar{\delta}} u_t, \Delta x_t')'$, with typical element $R_{ij}(\lambda)$ $f_{ij}(\lambda)/(f_{ii}(\lambda) f_{jj}(\lambda))^{1/2}$, the coherence between z_{it} and z_{jt}. Then $R(0)$ is not singular and for some $\alpha \in (0, 2]$,

$$|R_{ij}(\lambda) \quad R_{ij}(0)| \quad \mathrm{O}(|\lambda|^{\alpha}) \quad \text{as } \lambda \to 0.$$

For tapered periodograms we impose the following assumption strengthening Assumption A.1, and which holds, for e.g. *ARFIMA* models with γ 2, and relaxes conditions such as $f_{zz}(\lambda)|\lambda|^{2d^2}$ $G_z + \mathrm{E}_{\gamma}|\lambda|^{\gamma} + \mathrm{o}(|\lambda|^{\gamma})$, as $\lambda \to 0$, $0 < \mathrm{E}_{\gamma} < \infty$ used in Velasco (2000).

Assumption A.3. Let u_t possess a (pseudo) spectral density $f_{uu}(\lambda)$ satisfying Assumption A.1 such that for $|\omega| \leqslant \lambda/2$ and some $1 < \gamma \leqslant 2$

$$f_{uu}(\lambda \quad \omega) \quad f_{uu}(\lambda) \quad \omega f'_{uu}(\lambda) + \mathrm{O}(|\lambda|^{-2\delta-\gamma}|\omega|^{\gamma}) \quad \text{as } \lambda \to 0.$$

The following are restrictions on the bandwidths defined in Assumption 2.

Assumption A.4. $\max\{0, (1 \quad d \quad \delta)/(d \quad \delta \quad 0.5)\} < b < a < 1, \; d \quad \delta > 0.5.$

Assumption A.5. $0 < b < a < 2\gamma/(1 + 2\gamma)$.

Note that depending on the values of d, δ and γ, these two assumptions may not hold simultaneously. Thus, for example, if $\gamma \quad 2$, like for *ARFIMA* processes, we need $9d + \delta > 7$, because of $(1 \quad d \quad \delta)/(d \quad \delta \quad 0.5) < 4/5$, which holds for any $\delta \geqslant 0$ if $d > \frac{7}{9}$. However, we always require $d > 0.75$ for Assumption A.4 to hold, because of $(1 \quad d \quad \delta)/(d \quad \delta \quad 0.5) < 1$.

Assumption A.6. $0 < b < a < 2\min\{\alpha, \gamma\}/(1 + 2\min\{\alpha, \gamma\})$.

Assumption A.7. $\max\{0, (1 \quad d_i \quad \delta)/(d_i \quad \delta \quad 0.5)\} < b < a < 1, \; \min d_i \quad \delta > 0.5.$

Assumption A.8. $0 < b < a < 2Jb(d \quad \delta), \; \gamma J/(J + 2) > 1, \; J \geqslant 3.$

Assumption A.9. When $d + \delta < 1$, $0 < b < a < 2J\{b(d \quad \delta) \quad (1 \quad d \quad \delta)\}$.

The following theorem is the main result on log periodogram regressions with observed data.

Theorem A.1. *Under Assumption* A.1, *for Gaussian* $u_t \sim I(\delta)$, $0.5 < \delta < 0.5$, $\ell \quad 0$ *and*

$$m^{-1}(\log T)^2 + T^{-2\gamma} m^{1+2\gamma} \to 0 \quad \text{as } T \to \infty, \tag{8}$$

then

$$m^{1/2}(\widehat{\delta}(u) \quad \delta) \to_{\mathrm{d}} \mathrm{N}\left(0, \frac{\pi^2}{24}\right).$$

Proof of Theorem A.1. It follows from Robinson (1995a), using Hurvich et al. (1998) techniques to show that trimming of very low frequencies is not necessary for the asymptotic normality of $\widehat{\delta}$. Though Hurvich et al. (1998) only consider fractional processes with $I(0)$ innovations which possess a spectral density $f^*(\lambda)$ with three bounded derivatives around $\lambda \quad 0$, their results are easily generalized to our set up with $0 < \gamma \leqslant 2$ in Assumption A.1. Note that they also used the asymptotically equivalent regressor $\log(4\sin^2(\lambda_j/2))$ proposed by Geweke and Porter Hudak (1983) which arises naturally for fractional processes. □

The condition $T^{-2\gamma} m^{1+2\gamma} \to 0$ as $T \to \infty$ in (8) reflects the fact that when the semiparametric model $G_u \lambda^{-2\delta}$ is not very appropriate for high frequencies, i.e. γ is small in Assumption A.1, m must not grow very fast to avoid higher frequency biases in the local regression. The $\log T$ consistency holds under weaker conditions on bandwidth numbers, as is shown by estimating the mean square error of $\widehat{\delta}(u)$ as in Hurvich et al. (1998).

Theorem A.2. *Under Assumption* A.1, *for Gaussian* u_t, $0.5<\delta<0.5$, ℓ 0 *and*

$$(m^{-1}+(T^{-1}m)^{2\gamma})(\log T)^2+T^{-1}m\log m\to 0 \quad \text{as } T\to\infty, \tag{9}$$

then $\log T(\widehat{\delta}(u)$ $\delta)\to_p 0$.

In many cases we may wish to exclude the first $\ell>0$ frequencies in the regression and both asymptotic normality and $\log T$ consistency hold as shown originally by Robinson (1995a):

Corollary A.1. *Theorems* A.1 *and* A.2 *hold if* $m^{-1}\ell(\log T)^2\to 0$ *as* $T\to\infty$.

Now follows the general result for multivariate log periodogram regressions.

Theorem A.3. *Under Assumption* A.2 *for Gaussian* $(u_t,\Delta x_t')'$, $0.5<\delta, d$ $1<0.5$, *and*

$$m^{-1}(\log T)^2+T^{-2\min\{\alpha,\gamma\}}m^{1+2\min\{\alpha,\gamma\}}\to 0 \quad \text{as } T\to\infty, \tag{10}$$

we obtain that

$$2m^{1/2}(\widehat{\mathbf{d}}(u) \quad \mathbf{d})\to_{\mathrm{d}}\mathrm{N}(0,\Omega).$$

This holds if trimming is introduced as long as $\ell m^{-1}(\log T)^2\to 0$ *as* $T\to\infty$. *The covariance matrix* Ω *can be estimated consistently by the sample regression residuals covariance matrix*, $\widehat{\Omega}$ $(m$ $\ell)^{-1}\sum_j^m{}_{\ell+1}\hat{e}_j\hat{e}_j'$.

Proof of Theorem 7. This follows from Robinson (1995a), extending to a multivariate set up the results by Hurvich et al. (1998) to avoid trimming, ℓ 0. □

The following theorem is the basic result for non Gaussian log periodogram regressions.

Theorem A.4. *Under Assumptions* 4, 5, A.1, A.3, $0.5\leqslant\delta<1.5$, $\gamma J/(J+2)>1$, $J\geqslant 3$, *and*

$$\ell^{-1}+m^{-1}\ell(\log T)^2+T^{-1}m\to 0 \quad \text{as } T\to\infty, \tag{11}$$

then $\widehat{\delta}^{(J)}(u)\to_p\delta$.

Proof of Theorem A.4. This follows directly using the methods of Velasco (2000) for $0<\delta<0.5$. The extension to $0.5<\delta<0$ and $0.5\leqslant\delta<1.5$ being immediate (cf. Velasco, 2000, Lemma 3, 1999a, Theorems 4 and A.1). □

We collect in two lemmas several results repeatedly used in our proofs further down.

Lemma A.1. *Under Assumption* A.1, $0.5<\delta<d<1.5$, $\ell^{-1}+mT^{-1}\to 0$ *as* $T\to\infty$, $z\in\{x,u\}$ $(d_x$ d, d_u $\delta)$, j $\ell+1,\ldots,m$,

$$\mathrm{E}[I_{\Delta z\Delta z}(\lambda_j)] \quad f_{\Delta z\Delta z}(\lambda_j)(1+\mathrm{O}(j^{-1}\log(j+1))) \quad \mathrm{O}(\lambda_j^{2-2d_z}), \tag{12}$$

and if z_t is Gaussian,

$$\max_{\ell+1\leqslant j\leqslant m} \mathrm{E}[(I_{\Delta z\Delta z}(\lambda_j)/f_{\Delta z\Delta z}(\lambda_j))^a]<\infty, \quad a> \quad 1, \tag{13}$$

$$\max_{\ell+1\leqslant j\leqslant m} I_{\Delta z\Delta z}(\lambda_j)/f_{\Delta z\Delta z}(\lambda_j) \quad \mathrm{O_p}(\log T). \tag{14}$$

Proof of Lemma A.1. Eq. (12) follows from Robinson (1995a), Theorem 2. For $a>0$, (13) follows from Gaussianity and (12). For $1<a<0$, (13) follows from Gaussianity and the proof of Lemma 5 of Hurvich et al. (1998). They used a 1/4. Finally (14) can be proved using Gaussianity, (13) and adapting the proofs of Theorems 4.5.1. and 5.3.2. of Brillinger (1975). □

Lemma A.2. *Under Assumption* A.1, $0\leqslant\delta<d<1.5$, $\ell^{-1}+m^{-1}T\to 0$ *as* $T\to\infty$, $z\in\{x,u\}$ (d_x d, d_u δ) j $\ell+1,\ldots,m$, *some* $K<\infty$,

$$\mathrm{E}[I_{zz}(\lambda_j)/f_{zz}(\lambda_j)]\begin{cases} 1+\mathrm{O}(j^{-1}\log j+j^{2(d_z-1)}\log(j+1)), & d_z<1;\\ \leqslant Kj^{2(d_z-1)}, & 1\leqslant d_z<1.5,\end{cases}$$

and if z_t is Gaussian,

$$\max_{\ell+1\leqslant j\leqslant m} \mathrm{E}[(I_{zz}(\lambda_j)/f_{zz}(\lambda_j))^a]<\infty, \quad a> \quad 1, \quad d_z<1;$$

$$\max_{\ell+1\leqslant j\leqslant m} I_{zz}(\lambda_j)/f_{zz}(\lambda_j) \quad \mathrm{O_p}(\log T), \quad d_z<1;$$

$$\max_{\ell+1\leqslant j\leqslant m} I_{zz}(\lambda_j)j^{2(1-d_z)}/f_{zz}(\lambda_j) \quad \mathrm{O_p}(\log T), \quad 1\leqslant d_z<1.5. \tag{15}$$

Proof of Lemma A.2. It can be proved along the same lines as Lemma A.1, using now Velasco (1999a) and Hurvich and Ray (1995, Theorem 1) for non stationary series to bound $\mathrm{E}[I_{zz}(\lambda_j)/f_{zz}(\lambda_j)]$. The remaining results follow from Gaussianity. □

Appendix B. Proofs

Proof of Theorem 1. All sums run for j $\ell+1,\ldots,m$ if not indicated otherwise. First we obtain from (4) that

$$\widehat{\delta}(\Delta\hat{u}) \quad \widehat{\delta}(\Delta u) \quad \left(\sum_j W_j^2\right)^{-1}\sum_{j\ \ell+1}^{m} W_j \log\frac{I_{\Delta u\Delta u}(\lambda_j)}{I_{\Delta\hat{u}\Delta\hat{u}}(\lambda_j)}, \tag{16}$$

where W_j is defined in (5). Note that (8), (9) and the condition on the trimming hold under the conditions of the theorem with the definition of m and ℓ, so it only remains to show that the effects of the residual approximation are negligible to deduce the asymptotic properties of $\widehat{\delta}(\Delta\hat{u})$ from those of $\widehat{\delta}(\Delta u)$.

We can write from Eq. (3) that

$$\log \frac{I_{\Delta\hat{u}\Delta\hat{u}}(\lambda_j)}{I_{\Delta u\Delta u}(\lambda_j)} \quad \log\left\{1 \quad (\widehat{\beta} \quad \beta)\frac{w_{\Delta x}(\lambda_j)}{w_{\Delta u}(\lambda_j)}\right\} + \log\left\{1 \quad (\widehat{\beta} \quad \beta)\frac{w_{\Delta x}(\quad \lambda_j)}{w_{\Delta u}(\quad \lambda_j)}\right\}, \tag{17}$$

so

$$\max_{\ell+1\leqslant j\leqslant m}\left|(\widehat{\beta} \quad \beta)\frac{w_{\Delta x}(\lambda_j)}{w_{\Delta u}(\lambda_j)}\right| \leqslant |\widehat{\beta} \quad \beta| \max_{\ell+1\leqslant j\leqslant m}\left|\frac{w_{\Delta x}(\lambda_j)}{\sqrt{f_{\Delta x\Delta x}(\lambda_j)}}\right| \max_{\ell+1\leqslant j\leqslant m}\left|\frac{\sqrt{f_{\Delta x\Delta x}(\lambda_j)}}{w_{\Delta u}(\lambda_j)}\right|.$$

Since $d+\delta \geqslant 1$ (Case I), from Lemma A.1 above,

$$|\widehat{\beta} \quad \beta| \max_{\ell+1\leqslant j\leqslant m}\left|\frac{w_{\Delta x}(\lambda_j)}{\sqrt{f_{\Delta x\Delta x}(\lambda_j)}}\right| \quad \mathrm{O_p}(T^{\delta-d}\sqrt{\log T}), \tag{18}$$

and for any $c>0$ fixed and $0<\varepsilon<1$, using Bonferroni's and Markov's inequalities and (13) in Lemma A.1 with $a \quad (\varepsilon/2) \quad 1> \quad 1$ and $z \quad x$, and Assumption A.1,

$$\begin{aligned}
&P\left\{T^{\delta-d}\log^2 T \max_{\ell+1\leqslant j\leqslant m}\left|\frac{\sqrt{f_{\Delta x\Delta x}(\lambda_j)}}{w_{\Delta u}(\lambda_j)}\right| > c\right\} \\
&\quad \leqslant \sum_{j \quad \ell+1}^{m} P\left\{T^{\delta-d}\log^2 T\left|\frac{\sqrt{f_{\Delta x\Delta x}(\lambda_j)}}{w_{\Delta u}(\lambda_j)}\right| > c\right\} \\
&\quad\quad \mathrm{O}\left(\sum_{j \quad \ell+1}^{m} \mathrm{E}\left[T^{\delta-d}\log^2 T\sqrt{\frac{f_{\Delta x\Delta x}(\lambda_j)}{I_{\Delta u\Delta u}(\lambda_j)}}\right]^{2-\varepsilon}\right) \\
&\quad\quad \mathrm{O}\left(\log^4 T \sum_{j \quad \ell+1}^{m} j^{-(d-\delta)(2-\varepsilon)}\right) \\
&\quad\quad \mathrm{O}(\ell^{1-(d-\delta)(2-\varepsilon)}\log^4 T) \quad \mathrm{o}(1),
\end{aligned}$$

using Assumption 2, since $d \quad \delta > a > \frac{1}{2}$ for some a and choosing $\varepsilon \quad 2 \quad (1/a) > 0$ so

$$\max_{\ell+1\leqslant j\leqslant m}\left|\frac{\sqrt{f_{\Delta x\Delta x}(\lambda_j)}}{w_{\Delta u}(\lambda_j)}\right| \quad \mathrm{o_p}(T^{d-\delta}\log^{-2} T). \tag{19}$$

Therefore from (18) and (19)

$$\max_{\ell+1\leqslant j\leqslant m}\left|(\widehat{\beta} \quad \beta)\frac{w_{\Delta x}(\lambda_j)}{w_{\Delta u}(\lambda_j)}\right| \quad \mathrm{o_p}(\log^{-1} T). \tag{20}$$

Then, using (17) and $|\log(1+x)| \leqslant 2|x|$ for $|x| \leqslant \frac{1}{2}$, and the indicator function $\mathbb{I}$,

$$\begin{aligned}
&\left| \sum_{j}^{m}{}_{\ell+1} W_j \log \frac{I_{\Delta\hat{u}\Delta\hat{u}}(\lambda_j)}{I_{\Delta u\Delta u}(\lambda_j)} \right| \\
&\quad \leqslant \sum_{j}^{m}{}_{\ell+1} |W_j| \left| \log \frac{I_{\Delta\hat{u}\Delta\hat{u}}(\lambda_j)}{I_{\Delta u\Delta u}(\lambda_j)} \right| \mathbb{I}\left[\left| (\widehat{\beta} \quad \beta) \frac{w_{\Delta x}(\lambda_j)}{w_{\Delta u}(\lambda_j)} \right| \leqslant \frac{1}{2} \right] \\
&\qquad + \sum_{j}^{m}{}_{\ell+1} |W_j| \left| \log \frac{I_{\Delta\hat{u}\Delta\hat{u}}(\lambda_j)}{I_{\Delta u\Delta u}(\lambda_j)} \right| \mathbb{I}\left[\left| (\widehat{\beta} \quad \beta) \frac{w_{\Delta x}(\lambda_j)}{w_{\Delta u}(\lambda_j)} \right| > \frac{1}{2} \right] \\
&\quad \leqslant 4|\widehat{\beta} \quad \beta| \max_{\ell+1 \leqslant j \leqslant m} |W_j| \sum_{j}^{m}{}_{\ell+1} \left| \frac{w_{\Delta x}(\lambda_j)}{w_{\Delta u}(\lambda_j)} \right| \\
&\qquad + \max_{\ell+1 \leqslant j \leqslant m} \left| \log \frac{I_{\Delta\hat{u}\Delta\hat{u}}(\lambda_j)}{I_{\Delta u\Delta u}(\lambda_j)} \right| \max_{\ell+1 \leqslant j \leqslant m} |W_j| \sum_{j}^{m}{}_{\ell+1} \mathbb{I}\left[\left| (\widehat{\beta} \quad \beta) \frac{w_{\Delta x}(\lambda_j)}{w_{\Delta u}(\lambda_j)} \right| > \frac{1}{2} \right]
\end{aligned}$$

Now, $d \quad \delta > 0.5$,

$$|\widehat{\beta} \quad \beta| \max_{\ell+1 \leqslant j \leqslant m} |W_j| \sum_{j}^{m}{}_{\ell+1} \left| \frac{w_{\Delta x}(\lambda_j)}{w_{\Delta u}(\lambda_j)} \right| \quad \mathrm{O_p}\left(\log T \sum_{j}^{m}{}_{\ell+1} j^{\delta-d} \right) \quad \mathrm{o_p}(m^{1/2}),$$

because $|\widehat{\beta} \quad \beta| \quad \mathrm{O_p}(T^{\delta-d})$, (Case I), $\max_{\ell+1 \leqslant j \leqslant m} |W_j| \quad \mathrm{O}(\log T)$ (noting definition (5) of W_t), Assumption 2, so using (12) of Lemma A.1 and applying Hölder's inequality with $1 < p < 2$, $1/q + 1/p \quad 1$),

$$\mathrm{E}\left| \frac{w_{\Delta x}(\lambda_j)}{w_{\Delta u}(\lambda_j)} \right| \leqslant C\lambda_j^{\delta-d} \left(\mathrm{E}\left| \frac{w_{\Delta x}(\lambda_j)}{\sqrt{f_{\Delta x\Delta x}(\lambda_j)}} \right|^q \right)^{1/q} \left(\mathrm{E}\left| \frac{\sqrt{f_{\Delta u\Delta u}(\lambda_j)}}{w_{\Delta u}(\lambda_j)} \right|^p \right)^{1/p} \quad \mathrm{O}\left(\lambda_j^{\delta-d} \right). \tag{21}$$

On the other hand, we obtain that

$$\max_{\ell+1 \leqslant j \leqslant m} \left| \log \frac{I_{\Delta\hat{u}\Delta\hat{u}}(\lambda_j)}{I_{\Delta u\Delta u}(\lambda_j)} \right| \quad \mathrm{o_p}(\log^{-1} T),$$

using (17) and because (20) implies that for any $c > 0$,

$$\begin{aligned}
&P\left\{ \log T \max_j \left| \log\left(1 \quad (\widehat{\beta} \quad \beta) \frac{w_{\Delta x}(\lambda_j)}{w_{\Delta u}(\lambda_j)} \right) \right| > c \right\} \\
&\quad \leqslant P\left\{ \log T \max_j \left| (\widehat{\beta} \quad \beta) \frac{w_{\Delta x}(\lambda_j)}{w_{\Delta u}(\lambda_j)} \right| > c/2 \right\} \quad \mathrm{o}(1),
\end{aligned}$$

since $|\log(1+x)| > c/\log T$ implies $|x| > c/(2\log T)$ for all $c > 0$ and T large enough.

Finally,

$$\sum_{j=\ell+1}^{m} \mathbb{I}\left[\left|(\widehat{\beta} \quad \beta)\frac{w_{\Delta x}(\lambda_j)}{w_{\Delta u}(\lambda_j)}\right| > \frac{1}{2}\right] \leqslant 2|\widehat{\beta} \quad \beta| \sum_{j=\ell+1}^{m}\left|\frac{w_{\Delta x}(\lambda_j)}{w_{\Delta u}(\lambda_j)}\right|$$

$$\mathrm{O_p}\left(\sum_{j=\ell+1}^{m} j^{\delta-d}\right) \quad \mathrm{o_p}(m^{1/2}\log^{-1} T),$$

using (21), Assumption 2 and $d \quad \delta > 0.5$. The theorem follows from (16), $4\sum_{j=\ell+1}^{m} W_j^2 \sim m$ and because under Assumption A.5, (8) holds and therefore Theorem A.1 gives the result since the residual contribution is $\mathrm{o_p}(m^{-1/2})$. □

Proof of Theorem 2. First we write

$$\widehat{\delta}(\hat{u}) \quad \widehat{\delta}(u) \quad \left(\sum_j W_j^2\right)^{-1} \sum_j W_j \log\frac{I_{\hat{u}\hat{u}}(\lambda_j)}{I_{uu}(\lambda_j)}. \tag{22}$$

Next, from Robinson (1995a), and adapting the argument of Hurvich et al. (1998), since (9) holds, $(\widehat{\delta}(u) \quad \delta)\log T \to_p 0$. Also (8) holds when $a < 2\gamma/(1+2\gamma)$, $m^{1/2}(\widehat{\delta}(u) \quad \delta) \to_{\mathrm{d}} \mathrm{N}(0, \pi^2/24)$. Then it only remains to be shown that the right hand side of (22) is negligible.

Case I: When $d + \delta \geqslant 1$ and $d < 1$, we can proceed exactly as in Theorem 1, using Lemma A.2 instead of Lemma A.1, so

$$\left(\sum_j W_j^2\right)^{-1} \sum_{j=\ell+1}^{m} W_j \log\frac{I_{\hat{u}\hat{u}}(\lambda_j)}{I_{uu}(\lambda_j)} \quad \mathrm{O_p}\left(m^{-1}\log T \sum_{j=\ell+1}^{m} j^{\delta-d}\right) \quad \mathrm{o_p}(m^{-1/2}),$$

because $d \quad \delta > 0.5$.

When $d \geqslant 1$, $\mathrm{E}[I_{xx}(\lambda_j)/f_{xx}(\lambda_j)]$ is no longer bounded as $T \to \infty$ and we have to proceed differently. Let us show that

$$\max_{\ell+1\leqslant j\leqslant m}\left|(\widehat{\beta} \quad \beta)\frac{w_x(\lambda_j)}{w_u(\lambda_j)}\right| \leqslant |\widehat{\beta} \quad \beta| \max_{\ell+1\leqslant j\leqslant m}\left|\frac{w_x(\lambda_j)}{\sqrt{f_{xx}(\lambda_j)}} j^{1-d}\right| \max_{\ell+1\leqslant j\leqslant m}\left|\frac{\sqrt{f_{xx}(\lambda_j)}}{w_u(\lambda_j)} j^{d-1}\right|$$

$$\mathrm{o_p}(\log^{-1} T).$$

Using Lemma A.2, $d \geqslant 1$,

$$|\widehat{\beta} \quad \beta| \max_{\ell+1\leqslant j\leqslant m}\left|\frac{w_x(\lambda_j)}{\sqrt{f_{xx}(\lambda_j)}} j^{1-d}\right| \quad \mathrm{O_p}(T^{\delta-d}\sqrt{\log T}),$$

and for any $c > 0$ fixed and $0 < \varepsilon < 1$, using Bonferroni's and Markov's inequalities and Lemma A.2, as in the proof of Theorem 1,

$$P\left\{T^{\delta-d}\log^2 T \max_{\ell+1\leqslant j\leqslant m}\left|\frac{\sqrt{f_{xx}(\lambda_j)}}{w_u(\lambda_j)} j^{d-1}\right| > c\right\}$$

$$\leqslant \sum_{j=\ell+1}^{m} P\left\{T^{\delta-d} j^{d-1}\log^2 T\left|\frac{\sqrt{f_{xx}(\lambda_j)}}{w_u(\lambda_j)}\right| > c\right\}$$

$$
\begin{aligned}
&\mathrm{O}\left(\sum_{j\ \ell+1}^{m}\ \mathrm{E}\left[T^{\delta-d}j^{d-1}\log^2 T\sqrt{\frac{f_{xx}(\lambda_j)}{I_{uu}(\lambda_j)}}\right]^{2-\varepsilon}\right)\\
&\mathrm{O}\left(\log^4 T\sum_{j\ \ell+1}^{m}j^{-(1-\delta)(2-\varepsilon)}\right)\\
&\mathrm{O}(\ell^{1-(1-\delta)(2-\varepsilon)}\log^4 T)\quad \mathrm{o}(1),
\end{aligned}
$$

using Assumption 2 since $\delta<0.5$ implies that for some a, 1 $\delta>a>0.5$ and choosing ε 2 $(1/a)>0$, so, $d\geqslant 1$,

$$
\max_{\ell+1\leqslant j\leqslant m}\left|\frac{\sqrt{f_{xx}(\lambda_j)}}{w_u(\lambda_j)}j^{d-1}\right|\quad \mathrm{o_p}(T^{d-\delta}\log^{-2}T).
$$

Now, $\delta<0.5$,

$$
|\widehat{\beta}\quad\beta|\max_{\ell+1\leqslant j\leqslant m}|W_j|\sum_{j\ \ell+1}^{m}\left|\frac{w_x(\lambda_j)}{w_u(\lambda_j)}\right|\quad \mathrm{O_p}\left(\log T\sum_{j\ \ell+1}^{m}j^{\delta-1}\right)\quad \mathrm{o_p}(m^{1/2}),
$$

because $|\widehat{\beta}\quad\beta|\quad \mathrm{O_p}(T^{\delta-d})$, $\max_{\ell+1\leqslant j\leqslant m}|W_j|\quad \mathrm{O}(\log T)$, and because using Lem ma A.2, $(1<p<2,\ q^{-1}+p^{-1}\quad 1)$, we obtain now for $d\geqslant 1$

$$
\mathrm{E}\left|\frac{w_x(\lambda_j)}{w_u(\lambda_j)}\right|\leqslant C\lambda_j^{\delta-d}j^{d-1}\left(\mathrm{E}\left|\frac{w_x(\lambda_j)}{\sqrt{f_{xx}(\lambda_j)}}j^{1-d}\right|^q\right)^{1/q}\left(\mathrm{E}\left|\frac{\sqrt{f_{uu}(\lambda_j)}}{w_u(\lambda_j)}\right|^p\right)^{1/p}\quad \mathrm{O}(\lambda_j^{\delta-d}j^{d-1}).
$$

On the other hand, $\max_{\ell+1\leqslant j\leqslant m}|\log I_{\hat{u}\hat{u}}(\lambda_j)/I_{uu}(\lambda_j)|\quad \mathrm{o_p}(\log^{-1}T)$ as in Theorem 1, so it only remains to bound

$$
\sum_{j\ \ell+1}^{m}\mathbb{I}\left[\left|(\widehat{\beta}\quad\beta)\frac{w_x(\lambda_j)}{w_u(\lambda_j)}\right|>\frac{1}{2}\right]\leqslant 2|\widehat{\beta}\quad\beta|\sum_{j\ \ell+1}^{m}\left|\frac{w_x(\lambda_j)}{w_u(\lambda_j)}\right|=\mathrm{O_p}\left(\sum_{j\ \ell+1}^{m}j^{\delta\ \ 1}\right)=\mathrm{o_p}(m^{1/2}\log\ ^{1}T),
$$

using the definition of m and $\delta<0.5$.

Case II : If $d+\delta<1$ then $d<1$, and we obtain first that,

$$
\max_{\ell+1\leqslant j\leqslant m}\left|(\widehat{\beta}\quad\beta)\frac{w_x(\lambda_j)}{w_u(\lambda_j)}\right|\quad \mathrm{o_p}(\log^{-1}T),
$$

because $|\widehat{\beta}\quad\beta|\quad \mathrm{O_p}(T^{1-2d})$ and for any $c>0$, using Lemma A.2,

$$
\begin{aligned}
&P\left\{T^{1-2d}\log^2 T\max_{\ell+1\leqslant j\leqslant m}\left|\frac{\sqrt{f_{xx}(\lambda_j)}}{w_u(\lambda_j)}\right|>c\right\}\\
&\qquad\mathrm{O}\left(\sum_{j\ \ell+1}^{m}\ \mathrm{E}\left[T^{1-2d}\log^2 T\sqrt{\frac{f_{xx}(\lambda_j)}{I_{uu}(\lambda_j)}}\right]^{2-\varepsilon}\right)
\end{aligned}
$$

$$\mathrm{O}\left(\log^4 T \sum_{j}^{m} [T^{1-d-\delta} j^{-(d-\delta)}]^{2-\varepsilon}\right)$$
$$\mathrm{O}(T^{(1-d-\delta)(2-\varepsilon)} \ell^{1-(d-\delta)(2-\varepsilon)} \log^4 T) \quad \mathrm{o}(1),$$

if $d \quad \delta > \frac{1}{2}$ and $b > 2(1 \quad d \quad \delta)/(2(d \quad \delta) \quad 1)$ for $\varepsilon > 0$ small enough.

Thus, as in the proof of Theorem 1,

$$\left(\sum_j W_j^2\right)^{-1} \left| \sum_{j}^{m} W_j \log \frac{I_{\hat{u}\hat{u}}(\lambda_j)}{I_{uu}(\lambda_j)} \right| \quad \mathrm{O_p}(T^{1-d-\delta}[m^{-1} \log m + m^{\delta-d}] \log T), \tag{23}$$

because

$$\left(\sum_j W_j^2\right)^{-1} |\widehat{\beta} \quad \beta| \max_{\ell+1 \leqslant j \leqslant m} |W_j| \sum_{j}^{m} \left| \frac{w_x(\lambda_j)}{w_u(\lambda_j)} \right|$$
$$\mathrm{O_p}\left(m^{-1} T^{1-d-\delta} \log T \sum_{j}^{m} j^{\delta-d}\right)$$
$$\mathrm{O_p}(T^{1-d-\delta}[m^{-1} \log m + m^{\delta-d}] \log T),$$

using $|\widehat{\beta} \quad \beta| \quad \mathrm{O_p}(T^{1-2d})$ and $\mathrm{E}|w_x(\lambda_j)/w_u(\lambda_j)| \quad \mathrm{O}(\lambda_j^{\delta-d})$, and because $\max_{\ell+1 \leqslant j \leqslant m} |\log I_{\hat{u}\hat{u}}(\lambda_j) \quad \log I_{uu}(\lambda_j)| \quad \mathrm{o_p}(\log^{-1} T)$, $\max_{\ell+1 \leqslant j \leqslant m} |W_j| \quad \mathrm{O}(\log T)$ and

$$\sum_j \mathbb{I}\left[\left|(\widehat{\beta} \quad \beta) \frac{w_x(\lambda_j)}{w_u(\lambda_j)}\right| > \frac{1}{2}\right] \leqslant 2 \sum_j \left|(\widehat{\beta} \quad \beta) \frac{w_x(\lambda_j)}{w_u(\lambda_j)}\right| \quad \mathrm{O_p}(T^{1-d-\delta}[\log m + m^{\delta-d+1}]).$$

Then (23) is $\mathrm{o_p}(\log^{-1} T)$, and also $\mathrm{o_p}(m^{-1/2})$ on choosing $\max\{(1 \quad d \quad \delta)/(d \quad \delta \quad 0.5), 2(1 \quad d \quad \delta)\} < a$, but this condition is not additionally restrictive because $d + \delta < 1$ (so $d \quad \delta \quad 0.5 < 0.5$) implies that $2(1 \quad d \quad \delta) < (1 \quad d \quad \delta)/(d \quad \delta \quad 0.5) < b < a < 1$. Then the theorem follows. □

Proof of Theorem 3. We only need to consider the effect of residual estimation in a univariate regression, since the multivariate case implies no additional difficulty because only one equation in the log periodogram system is affected by the residual approximation. Working with e.g. the increments of the residuals we have to show that

$$m^{-1} \sum_j W_j \log \frac{I_{\Delta\hat{u}\Delta\hat{u}}(\lambda_j)}{I_{\Delta u \Delta u}(\lambda_j)} \quad \mathrm{o_p}(m^{-1/2}), \tag{24}$$

as in the univariate case, and additionally that

$$m^{-1} \sum_j \log \frac{I_{\Delta\hat{u}\Delta\hat{u}}(\lambda_j)}{I_{\Delta u \Delta u}(\lambda_j)} \quad \mathrm{o_p}(m^{-1/2} \log T), \tag{25}$$

$$m^{-1} \sum_j \left(\log \frac{I_{\Delta\hat{u}\Delta\hat{u}}(\lambda_j)}{I_{\Delta u \Delta u}(\lambda_j)}\right)^2 \quad \mathrm{o_p}(\log^{-2} T), \tag{26}$$

from (5.6), (5.8) and (5.13) in Robinson (1995a) for consistent covariance matrix estimation. Now to show (24) we can write now that

$$\log \frac{I_{\Delta\hat{u}\Delta\hat{u}}(\lambda_j)}{I_{\Delta u\Delta u}(\lambda_j)} \quad \log\left\{1 \quad \sum_{i\ 1}^{k} \frac{(\widehat{\beta}_i \quad \beta_i) w_{\Delta x_i}(\lambda_j)}{w_{\Delta u}(\lambda_j)}\right\} + \log\left\{1 \quad \sum_{i\ 1}^{k} \frac{(\widehat{\beta}_i \quad \beta_i) w_{\Delta x_i}(\quad \lambda_j)}{w_{\Delta u}(\quad \lambda_j)}\right\}.$$

Next,

$$\max_{\ell+1 \leqslant j \leqslant m} \left| \sum_{i\ 1}^{k} \frac{(\widehat{\beta}_i \quad \beta_i) w_{\Delta x_i}(\lambda_j)}{w_{\Delta u}(\lambda_j)} \right| \leqslant \sum_{i\ 1}^{k} |\widehat{\beta}_i \quad \beta_i| \max_j \left| \frac{w_{\Delta x_i}(\lambda_j)}{\sqrt{f_{\Delta x_i \Delta x_i}(\lambda_j)}} \right| \max_j \left| \frac{\sqrt{f_{\Delta x_i \Delta x_i}(\lambda_j)}}{w_{\Delta u}(\lambda_j)} \right| \quad o_p(\log^{-1} T), \tag{27}$$

since k is finite, bounding the above expression for each $i \quad 1, \ldots, k$ as in the univariate case. Then (25) follows in a similar but simpler way, while (26) is implied directly by (27). □

Proof of Theorem 4. For simplicity, we suppress in the notation for $I_{uu}(\lambda_j)$, etc. reference to J or tapering and only consider the case of a univariate regressor x_t. We consider successively the cases of differenced and original residuals.

Proof for differenced residuals. We obtain that $\widehat{\delta\ 1}^{(J)}(\Delta\hat{u}) \quad \widehat{\delta\ 1}^{(J)}(\Delta u)$ $o_p(\log^{-1} T)$ if we show that

$$\max_{\ell+1 \leqslant j \leqslant m} |W_j| \log \frac{I_{\Delta\hat{u}\Delta\hat{u}}(\lambda_j)}{I_{\Delta u\Delta u}(\lambda_j)} \quad o_p(\log^{-1} T),$$

which in turn follows if

$$\max_{\ell+1 \leqslant j \leqslant m} \left| (\widehat{\beta} \quad \beta) \frac{w_{\Delta x}(\lambda_j)}{w_{\Delta u}(\lambda_j)} \right| \leqslant |\widehat{\beta} \quad \beta| \max_{\ell+1 \leqslant j \leqslant m} \left| \frac{w_{\Delta x}(\lambda_j)}{\sqrt{f_{\Delta x \Delta x}(\lambda_j)}} \right| \max_{\ell+1 \leqslant j \leqslant m} \left| \frac{\sqrt{f_{\Delta x \Delta x}(\lambda_j)}}{w_{\Delta u}(\lambda_j)} \right| \quad o_p(\log^{-2} T). \tag{28}$$

We first show that

$$|\widehat{\beta} \quad \beta| \max_{\ell+1 \leqslant j \leqslant m} \left| \frac{w_{\Delta x}(\lambda_j)}{\sqrt{f_{\Delta x \Delta x}(\lambda_j)}} \right| \quad O_p(T^{\delta - d} \sqrt{\log T}). \tag{29}$$

We have $|\widehat{\beta} \quad \beta| \quad O_p(T^{\delta-d})$ while, writing $I^{\varepsilon}_{\Delta x \Delta x}(\lambda_j)$ for the periodogram of the i.i.d. innovations $\varepsilon_t^{(\Delta x)}$ of $z_t \quad \Delta x_t$ in Assumption 4, and $H_{\Delta x \Delta x}(\lambda_j)$

$I_{\Delta x\Delta x}(\lambda_j)f^{-1}_{\Delta x\Delta x}(\lambda_j) \quad 2\pi I^{\varepsilon}_{\Delta x\Delta x}(\lambda_j)$,

$$\left(\max_{\ell+1\leqslant j\leqslant m}\left|\frac{w_{\Delta x}(\lambda_j)}{\sqrt{f_{\Delta x\Delta x}(\lambda_j)}}\right|\right)^2 \leqslant \max_{\ell+1\leqslant j\leqslant m} 2\pi I^{\varepsilon}_{\Delta x\Delta x}(\lambda_j) + \max_{\ell+1\leqslant j\leqslant m}|H_{\Delta x\Delta x}(\lambda_j)|$$
$$\leqslant \max_{\ell+1\leqslant j\leqslant m} 2\pi I^{\varepsilon}_{\Delta x\Delta x}(\lambda_j) + \left(\sum_{j}^{m}{}_{\ell+1}|H_{\Delta x\Delta x}(\lambda_j)|^2\right)^{1/2}.$$

Now, $\max_{\ell+1\leqslant j\leqslant m} 2\pi I^{\varepsilon}_{\Delta x\Delta x}(\lambda_j) \quad O_p(\log T)$, see An et al. (1983), while from Velasco (2000), $1<\gamma\leqslant 2$,

$$E\sum_{j}^{m}{}_{\ell+1}\left|\frac{I_{\Delta x\Delta x}(\lambda_j)}{f_{\Delta x\Delta x}(\lambda_j)} \quad 2\pi I^{\varepsilon}_{\Delta x\Delta x}(\lambda_j)\right|^2 \quad O\left(\sum_{j}^{m}{}_{\ell+1} j^{-\gamma}\right) \quad O(\ell^{1-\gamma}),$$

so (29) follows. Then (28) follows if we show that

$$\max_{\ell+1\leqslant j\leqslant m}\left|\frac{\sqrt{f_{\Delta x\Delta x}(\lambda_j)}}{w_{\Delta u}(\lambda_j)}\right| \quad O_p(T^{d-\delta}\log^{-3}T).$$

For any sequence $M_T\to 0$ as $T\to\infty$, and with $C>0$, $0<\tau<1$ fixed, $0.5<\delta<d<1.5$, $H_{\Delta u\Delta u}(\lambda_j) \quad I_{\Delta u\Delta u}(\lambda_j)f^{-1}_{\Delta u\Delta u}(\lambda_j) \quad 2\pi I^{\varepsilon}_{\Delta u\Delta u}(\lambda_j)$, now using the innovations of Δu_t, we have that

$$\begin{aligned}
&P\left\{M_T\frac{f_{\Delta u\Delta u}(\lambda_j)}{I_{\Delta u\Delta u}(\lambda_j)}>C\right\}\\
&\quad P\left\{2\pi I^{\varepsilon}_{\Delta u\Delta u}(\lambda_j)\left(1+\frac{H_{\Delta u\Delta u}(\lambda_j)}{2\pi I^{\varepsilon}_{\Delta u\Delta u}(\lambda_j)}\right)<C^{-1}M_T\right\}\\
&\leqslant P\left\{2\pi I^{\varepsilon}_{\Delta u\Delta u}(\lambda_j)<(C^{-1}M_T)^{\tau}\right\}+P\left\{\left(1+\frac{H_{\Delta u\Delta u}(\lambda_j)}{2\pi I^{\varepsilon}_{\Delta u\Delta u}(\lambda_j)}\right)<(C^{-1}M_T)^{1-\tau}\right\}\\
&\leqslant P\left\{2\pi I^{\varepsilon}_{\Delta u\Delta u}(\lambda_j)<(C^{-1}M_T)^{\tau}\right\}+P\left\{\left|\frac{H_{\Delta u\Delta u}(\lambda_j)}{2\pi I^{\varepsilon}_{\Delta u\Delta u}(\lambda_j)}\right|>\frac{1}{2}\right\}\\
&\quad O(M_T^{J-v}+j^{-a-1}),
\end{aligned}\tag{30}$$

for any $v>0$, on choosing $\tau<1$ large enough, and some a such that $0<a<(\gamma J/(J+2) \quad 1)/2$, following Eqs. (A.8) and (A.13) in Velasco (2000).

Set $M_T \quad \ell^{2(\delta-d)}\log^3 T \quad o(1)$ as $T\to\infty$ by Assumption A.8.Then we can write for any $C>0$ and some $v, C'>0$, using (30),

$$P\left\{T^{\delta-d}\log^2 T\max_j\left|\frac{\sqrt{f_{\Delta x\Delta x}(\lambda_j)}}{w_{\Delta u}(\lambda_j)}\right|>C\right\}\leqslant P\left\{M_T\max_j\frac{f_{\Delta u\Delta u}(\lambda_j)}{I_{\Delta u\Delta u}(\lambda_j)}>C'\right\}$$

$$\leqslant \sum_j P\left\{ M_T \frac{f_{\Delta u \Delta u}(\lambda_j)}{I_{\Delta u \Delta u}(\lambda_j)} > C' \right\}$$
$$\mathrm{O}\left(\sum_j [M_T^{J-v} + j^{-a-1}] \right)$$
$$\mathrm{O}(m\ell^{2(\delta-d)(J-v)} \log^{3J} + \ell^{-a}) \quad \mathrm{o}(1)$$

choosing $v > 0$ small enough and using Assumption A.8.

Proof for original residuals. We want to show that $\widehat{\delta}^{(J)}(\hat{u}) \quad \widehat{\delta}^{(J)}(u) \quad \mathrm{o_p}(\log^{-1} T)$.

Case I: When $d + \delta \geqslant 1$ and $d < 1.5$, we can proceed exactly as for $\Delta\hat{u}$, but now we obtain that

$$|\widehat{\beta} \quad \beta| \max_{\ell+1 \leqslant j \leqslant m} \left| \frac{w_x(\lambda_j)}{\sqrt{f_{xx}(\lambda_j)}} \right| \quad \mathrm{O_p}(T^{\delta-d} \sqrt{\log T}),$$

because from Velasco (1999b) (see also Lemmas 1 and 3 of Lobato and Velasco, 2000) and for tapered data, $0.5 < d < 1.5$, $1 < \gamma \leqslant 2$,

$$\mathrm{E} \sum_{j \ell+1}^{m} \left| \frac{I_{xx}(\lambda_j)}{f_{xx}(\lambda_j)} \quad 2\pi I^{\varepsilon}_{\Delta x \Delta x}(\lambda_j) \right|^2 \quad \mathrm{O}\left(\sum_{j \ell+1}^{m} j^{-\gamma} \right) \quad \mathrm{O}(\ell^{1-\gamma}).$$

Then we can write as before, for any $C, v > 0$, $M_T \quad \ell^{2(\delta-d)} \log^3 T \quad \mathrm{o}(1)$ as $T \to \infty$, and using now the innovations of u_t if $\delta < 0.5$, or those of Δu_t if $\delta \geqslant 0.5$,

$$P\left\{ T^{\delta-d} \log^3 T \max_j \left| \frac{\sqrt{f_{xx}(\lambda_j)}}{w_u(\lambda_j)} \right| > C \right\} \quad \mathrm{O}\left(\sum_j \max_j [M_T^{J-v} + j^{-a-1}] \right) + \mathrm{o}(1)$$
$$\mathrm{O}(m\ell^{2(\delta-d)(J-v)} \log^{3J} T + \ell^{-a}) + \mathrm{o}(1)$$
$$\mathrm{o}(1),$$

with Assumption A.8, so finally

$$\max_j \left| \frac{\sqrt{f_{xx}(\lambda_j)}}{w_u(\lambda_j)} \right| \quad \mathrm{o_p}(T^{d-\delta}).$$

Case II: If $d + \delta < 1$ then $d < 1$, $\delta < 0.5$, and we proceed as before to show that

$$\max_{\ell+1 \leqslant j \leqslant m} \left| (\widehat{\beta} \quad \beta) \frac{w_x(\lambda_j)}{w_u(\lambda_j)} \right| \quad \mathrm{o_p}(\log^{-2} T).$$

Using Assumption 1 we first obtain that

$$|\widehat{\beta} \quad \beta| \max_{\ell+1 \leqslant j \leqslant m} \left| \frac{w_x(\lambda_j)}{\sqrt{f_{xx}(\lambda_j)}} \right| \quad \mathrm{O_p}(T^{1-2d} \log^{1/2} T).$$

Then we show that

$$\max_{\ell+1\leqslant j\leqslant m}\left|\frac{\sqrt{f_{xx}}(\lambda_j)}{w_u(\lambda_j)}\right| \quad o_p(T^{2d-1}\log^{-3}T),$$

because with the same notation, M_T $T^{2(1-d-\delta)}\ell^{2(\delta-d)}\log^3 T$ o(1) by Assumption A.9,

$$P\left\{T^{1-2d}\log^2 T\ \max_j\left|\frac{\sqrt{f_{xx}}(\lambda_j)}{w_u(\lambda_j)}\right|>C\right\}\leqslant P\left\{M_T\max_j\frac{f_{uu}(\lambda_j)}{I_{uu}(\lambda_j)}>C'\right\}$$
$$O(mM_T^{J-v}+\ell^{-a})+o(1),$$

which is $O(m\{T^{2(1-d-\delta)}\ell^{2(\delta-d)}\log^6 T\}^{J-v})+o(1)$ o(1), using Assumptions A.8 and A.9. □

References

An, H.-Z., Chen, Z.-G., Hannan, E.J., 1983. The maximum of the periodogram. Journal of Multivariate Analysis 13, 383 400.

Baillie, R.T., 1996. Long memory processes and fractional integration in econometrics. Journal of Econometrics 73, 5 59.

Baillie, R.T., Bollerslev, T., 1989. Common stochastic trends in a system of exchange rates. The Journal of Finance 44, 167 181.

Baillie, R.T., Bollerslev, T., 1994. Cointegration, fractional cointegration, and exchange rate dynamics. The Journal of Finance 49, 737 745.

Booth, G.G., Tse, Y., 1995. Long memory in interest rate future markets: a fractional cointegration analysis. The Journal of Future Markets 15, 573 584.

Breitung, J., Hassler, U., 2002. Inference on the cointegration rank in fractionally integrated processes. Journal of Econometrics 110, 167 185.

Brillinger, D.R., 1975. Time Series. Data Analysis and Theory. Holden Day, San Francisco.

Chen, Z.-G., Hannan, E.J., 1980. The distribution of periodogram ordinates. Journal of Time Series Analysis 1, 73 82.

Cheung, Y.-E., Lai, K.S., 1993. A fractional cointegration analysis of purchasing power parity. Journal of Business and Economic Statistics 11, 103 122.

Dahlhaus, R., 1989. Efficient parameter estimation for self-similar processes. Annals of Statistics 14, 517 532.

De Jong, R.M., Davidson, J., 2000. The functional central limit theorem and convergence to stochastic integrals II: the fractionally integrated case. Econometric Theory 16, 643 666.

Diebold, F.X., Gardeazabal, J., Yilmaz, K., 1994. On cointegration and exchange rate dynamics. The Journal of Finance 49, 727 735.

Dolado, J.J., Marmol, F., 1997. On the properties of the Dickey Pantula test against fractional alternatives. Economics Letters 57, 11 16.

Dueker, M., Startz, R., 1998. Maximum-likelihood estimation of fractional cointegration with an application to U.S. and Canadian bond rates. The Review of Economics and Statistics 80, 420 426.

Engle, R.F., Granger, C.W.J., 1987. Co-integration and error correction: representation, estimation and testing. Econometrica 55, 251 276.

Fay, G., Soulier, P., 2001. The periodogram of an i.i.d. sequence. Stochastic Processes and Their Applications 92, 315 343.

Fox, R., Taqqu, M.S., 1986. Large-sample properties of parameter estimates for strongly dependent stationary Gaussian times series. Annals of Statistics 14, 517 532.
Geweke, J., Porter-Hudak, S., 1983. The estimation and application of long memory time series models. Journal of Time Series Analysis 4, 221 238.
Gil-Alaña, L.A., Robinson, P.M., 1997. Testing of unit root and other non-stationary hypotheses in macroeconomic time series. Journal of Econometrics 80, 241 268.
Granger, C.W.J., 1981. Some properties of time series data and their in econometric model specification. Journal of Econometrics 16, 121 130.
Granger, C.W.J., Joyeux, R., 1980. An introduction to long memory time series models and fractional differencing. Journal of Time Series Analysis 1, 15 39.
Hassler, U., Walter, J., 1994. On the power of unit root tests against fractional alternatives. Economics Letters 45, 1 5.
Hosking, J.R.M., 1981. Fractional differencing. Biometrika 68, 165 176.
Hosking, J.R.M., 1984. Modeling persistence in hydrological time series using fractional differencing. Water Resources Research 20, 1898 1908.
Hurvich, C.M., Deo, R., 1999. Plug-in selection of the number of frequencies in regression estimates of the memory parameter of a long-memory time series. Journal of Time Series Analysis 20, 331 341.
Hurvich, C.M., Ray, B.K., 1995. Estimation of the memory parameter for nonstationary or noninvertible fractionally integrated processes. Journal of Time Series Analysis 16, 17 42.
Hurvich, C.M., Deo, R., Brodsky, J., 1998. The mean square error of Geweke and Porter Hudak's estimator of the memory parameter of a long memory time series. Journal of Time Series Analysis 19, 19 46.
Kim, C., Phillips, P.C.B., 2001. Fully modified estimation of fractional cointegration models. Preprint, Yale University.
Künsch, H.R., 1987. Statistical aspects of self-similar processes. Proceedings 1st World Congress of the Bernoulli Society, VNU Science Press, pp. 67 74.
Lobato, I.N., Velasco, C., 2000. Long memory in stock market trading volume. Journal of Economic and Business Statistics 18, 410 427.
Marinucci, D., Robinson, P.M., 1999. Alternative forms of fractional Brownian motion. Journal of Statistical Planning and Inference 80, 111 122.
Masih, R., Masih, A.M.M., 1995. A fractional cointegration approach to empirical tests of PPP: new evidence and methodological implications from an application to the Taiwan/US dollar relationship. Review of World Economics 131, 673 694.
Masih, A.M.M., Masih, R., 1998. A fractional cointegration approach to testing mean reversion between spot and forward exchange rates: a case of high frequency data with low frequency dynamics. Journal of Business Finance and Accounting 25, 987 1003.
Phillips, P.C.B., Ouliaris, S., 1990. Asymptotic properties of residual based tests for cointegration. Econometrica 58, 165 193.
Robinson, P.M., 1994a. Time series with strong dependence. In: Sims, C.A. (Ed.), Advances in Econometrics. Sixth World Congress, vol. I. Cambridge University Press, Cambridge, pp. 47 95.
Robinson, P.M., 1994b. Semiparametric analysis of long-memory time series. Annals of Statistics 22, 515 539.
Robinson, P.M., 1995a. Log-periodogram regression of time series with long range dependence. Annals of Statistics 23, 1048 1072.
Robinson, P.M., 1995b. Gaussian semiparametric estimation of long range dependence. Annals of Statistics 23, 1630 1661.
Robinson, P.M., 1997. Large sample inference for nonparametric regression with dependent errors. Annals of Statistics 25, 2054 2083.
Robinson, P.M., Marinucci, D., 2000. The averaged periodogram for nonstationary vector time series. Statistical Inference for Stochastic Processes 3, 149 160.
Robinson, P.M., Marinucci, D., 2001. Narrow-band analysis of nonstationary processes. Annals of Statistics 29, 947 986.

Robinson, P.M., Yajima, Y., 2002. Determination of cointegration rank in fractional systems. Journal of Econometrics 106, 217–241.
Sephton, P.S., Larsen, H.K., 1991. Tests of exchange market efficiency: fragile evidence from cointegration tests. Journal of International Money and Finance 10, 561–570.
Shin, Y., 1994. A residual-based test of the null of cointegration against the alternative of no cointegration. Econometric Theory 10, 91–115.
Soofi, A.S., 1998. A fractional cointegration test of purchasing power parity: the case of selected members of OPEC. Applied Financial Economics 8, 559–566.
Sowell, F.B., 1992. Maximum likelihood estimation of stationary univariate fractionally integrated time series. Journal of Econometrics 53, 165–188.
Tse, Y.K., Anh, V.V., Tieng, A., 1999. No-cointegration test based on fractional differencing: Some Monte Carlo results. Journal of Statistical Planning and Inference 80, 257–267.
Velasco, C., 1999a. Non-stationary log-periodogram regression. Journal of Econometrics 91, 325–371.
Velasco, C., 1999b. Gaussian semiparametric estimation of non-stationary time series. Journal of Time Series Analysis 20, 87–127.
Velasco, C., 2000. Non-Gaussian log-periodogram regression. Econometric Theory 16, 44–79.
Xiao, Z., 1999. A residual based test for the null hypothesis of cointegration. Economics Letters 64, 133–141.