Testing for General Fractional Integration in the Time Domain *

Uwe Hassler[a], Paulo M.M. Rodrigues[b†]and Antonio Rubia[c]

[a] Goethe University Frankfurt
[b] University of Algarve
[c] University of Alicante

June 2008

Abstract

In this paper we propose a family of least-squares based testing procedures that look to detect general forms of fractional integration at the long-run and/or the cyclical component of a time series, and which are asymptotically equivalent to Lagrange Multiplier tests. Our setting extends Robinson's (1994) results to allow for short memory in a regression framework and generalises the procedures in Agiakloglou and Newbold (1994), Tanaka (1999) and Breitung and Hassler (2002) by allowing for single or multiple fractional unit roots at any frequency in $[0, \pi]$. Our testing procedure can be easily implemented in practical settings and is flexible enough to account for a broad family of long- and short-memory specifications, including ARMA and/or GARCH-type dynamics among others. Furthermore, these tests have power against different types of alternative hypotheses and enable inference to be conducted under critical values drawn from a standard Chi-squared distribution, irrespective of the long-memory parameters.

Keywords: LM tests, nonstationarity, fractional integration, cyclical integration
JEL classification: C20, C22

Submission is Intended for the Paul Newbold Special Issue

---

[*] **Acknowledgements:** The authors thank Peter Phillips, Luis A. Gil-Alaña, Javier Hualde, the participants of the conference in honour of Paul Newbold held at the University of Nottingham, the 2nd Meeting of the MIFN, IESE Barcelona, and seminar attendants at the University of Navarra, University of Évora and the Bank of Portugal for their useful comments and suggestions. The authors also thank co-editor Robert Taylor and two anonymous referees for valuable comments and suggestions. Financial support from POCTI/ FEDER (grant ref. PTDC/ECO/64595/2006) and the SEJ2005-09372/ECON project is gratefully acknowledged. A previous version of this paper is available in the Working Paper series edited by *Fundacion de Cajas de Ahorros* (FUNCAS), WP-380, under the title "*Testing for the General Unit Root Hypothesis in the Time Domain*".

[†] **Corresponding Author:** Paulo M. M. Rodrigues, Faculty of Economics, University of Algarve, Campus de Gambelas, 8005-139 Faro, Portugal. E-mail: prodrig@ualg.pt.

1 Introduction

Modelling and forecasting macroeconomic and financial variables is at the forefront of the applied time-series econometric literature. These series are usually characterised by strongly persistent correlation structures over long intervals of time. In this paper, we propose several time domain test statistics to detect general forms of fractional integration. Our approach follows the Lagrange-multiplier (LM) framework studied in Robinson (1991, 1994), Agiakloglou and Newbold (1994), Tanaka (1999), Breitung and Hassler (2002) and Nielsen (2004, 2005). In particular, we propose a family of tests (which are asymptotically equivalent to standard LM tests for fractional integration) in the linear regression model $Y_t = \sum_{s=1}^{n} \phi_s X_{st}\left(Y_t\right) + u_t$, where Y_t is directly determined under the null hypothesis and the regressors $X_{st}\left(Y_t\right)$ are straightforwardly computed by linearly filtering Y_t. This approach can be easily implemented in practical settings and is flexible enough to account for a broad family of long- and short-memory specifications. Furthermore, these tests also have power against different types of alternative hypotheses, and allow inference to be conducted under critical values which are drawn from a standard Chi-squared distribution, independently of the long-memory parameters.

The tests we discuss are formally intended to detect general long memory patterns embedded in the autoregressive filter

$$\Delta_{\boldsymbol{\gamma}}\left(L;\boldsymbol{\delta}\right) \equiv \prod_{i=1}^{n} \xi_{\gamma_i}\left(L;\delta_i\right) \tag{1}$$

where $\boldsymbol{\delta} = \left(\delta_1, ..., \delta_n\right)'$, $\boldsymbol{\delta} \in \mathbb{R}^n$, $n \geq 1$, is a vector of possible non-integer values that control the extent of time dependence at any of the frequencies in $\boldsymbol{\gamma} = \left(\gamma_1, ..., \gamma_n\right)'$, with $0 \leq \gamma_1 < \gamma_2 < ... < \gamma_n \leq \pi$. These frequencies characterise the long-run and/or the seasonal (cyclical) nature of the data. The polynomials $\xi_{\gamma_i}\left(L;\delta_i\right)$ are defined as $\xi_{\gamma_1}\left(L;\delta_1\right) = \left(1-L\right)^{\delta_1}$, if $\gamma_1 = 0$; $\xi_{\gamma_n}\left(L;\delta_n\right) = \left(1+L\right)^{\delta_n}$, if $\gamma_n = \pi$; and $\xi_{\gamma_i}\left(L;\delta_i\right) = (1 - 2\cos\gamma_i L + L^2)^{\delta_i}$, if $\gamma_i \in (0, \pi)$, where the latter frequencies are also known in the literature as Gegenbauer frequencies; (see, for instance, Gray, Zhang and Woodword, 1989, p.237). Finally, L denotes the conventional back-shift operator.

The filter in (1), also considered in Robinson (1994), generates theoretical autocovariances that decay hyperbolically and sinusoidally, an empirical feature that is manifested in a number of periodic time series. Serial dependence may be present at the long-run ($\gamma_1 = 0$), and/or at any of the remaining (cyclical) frequencies involved. Hence, particular cases include the well-known fractional integration model, as well as pure cyclical and seasonal models which are routinely applied to economic and non-economic variables. For instance, cyclical models have been used to explain interest rate dynamics (Ramachandran and Beaumont, 2001), industrial production (Dalla and Hidalgo, 2005), and nominal exchange rates (Smallwood and Norrbin, 2006), among others. Recent studies focusing on non-economic variables have analysed, for instance, atmospheric levels of CO_2 (Woodward, Cheng and Gray, 1998), wind speed (Bouette *et al.*, 2006), or power demand (Soares and Souza, 2006). The extant literature on seasonal and non-seasonal models embedded in this general framework (both integrated and fractionally integrated) is overwhelming; see, for instance, Porter-Hudak (1990) for empirical applications of seasonal long-memory models on monetary aggregates, and Gil-Alana and Robinson (2001) and Gil-Alana (2005) for studies on consumption and income data, and inflation, respectively.

Our setting extends Robinson's (1994) results to allow for short memory in a regression framework and thus also generalises the procedures in Agiakloglou and Newbold (1994), Tanaka (1999) and Breitung and Hassler (2002) by allowing for single or multiple fractional integration at any frequency in $[0, \pi]$. Furthermore, we allow for different types of errors in the data generating process (DGP) which include martingale difference sequences and weakly correlated errors, thus allowing for ARMA and/or time varying volatility patterns. As in the frequency-domain case, the tests do not require formal knowledge of the true values of the long-memory vector $\boldsymbol{\delta}$. These are mainly intended for formally pretesting hypotheses about the extent of cyclical and non-cyclical persistence, and to construct confidence sets that include the true values of the long-memory coefficients with a certain asymptotic coverage level. This is valuable for descriptive inference and provides reliable values for initiating optimisation routines important for estimation procedures such as (quasi) maximum likelihood procedures.

The remaining of the paper is organised as follows. Section 2 introduces the testing procedures and discusses their asymptotic distributions, section 3 analyses the finite-sample performance of the tests by means of Monte Carlo experimentation and section 4 summarises the main conclusions. Finally, the mathematical proofs of the main statements are collected in a technical appendix.

In what follows, '$\Rightarrow$' and '$\xrightarrow{p}$' denote weak convergence and convergence in probability, respectively, as the sample size is allowed to diverge. The variable $\mathbb{I}_{(\cdot)}$ is an indicator function that takes value equal to one if the condition in the subscript is fulfilled and zero otherwise. Finally, vectors and matrices are represented in bold letters.

2 Testing procedures

In the general case considered in (1), we will say that the observable x_t is generated by a *General Fractionally Integrated (GFI)* process of order $\boldsymbol{\delta}$, denoted as $x_t \sim$GFI$(\boldsymbol{\delta})$, *i.e.*, $\Delta_{\boldsymbol{\gamma}}(L; \boldsymbol{\delta})\, x_t = \varepsilon_t$, where the properties of ε_t will be discussed below. The study of particular cases is straightforward by imposing restrictions on (1). The pure trend or zero-frequency model is obtained for $n = 1$ and $\gamma_1 = 0$. A seasonal filter arises for seasonal frequencies γ_i, see e.g., Hassler (1994). Pure cyclical models are captured for $0 < \gamma_1 < \cdots < \gamma_n < \pi$. If $n = 1$, the latter case is often said to result in a GARMA model, whereas $n > 1$ leads to so-called n-factor GARMA models; see Woodward *et al.* (1998), and Ramachandran and Beaumont (2001) for a discussion of the statistical properties of these models. The generalisations (for instance, allowing for stationary short-run dynamics) can encompass both ARMA and ARFIMA models as particular cases.

The main interest of this paper lies in testing whether $\boldsymbol{\delta} = \mathbf{d}$, with $\mathbf{d} \in \mathbb{R}^n$ being specified *a priori*, against the alternative for which the order of integration is $\mathbf{d} + \boldsymbol{\theta}$, with $\boldsymbol{\theta} \neq \mathbf{0}$. Thus, the hypothesis of interest is stated as

$$\mathrm{H}_0 : \boldsymbol{\delta} = \mathbf{d} \text{ or } \mathrm{H}_0 : \boldsymbol{\theta} = \mathbf{0},$$

against the alternative hypothesis $\mathrm{H}_1 : \boldsymbol{\delta} \neq \mathbf{d}$, or $\mathrm{H}_1 : \boldsymbol{\theta} \neq \mathbf{0}$.

We start our theoretical analysis by introducing and discussing the initial set of assumptions, as well as general notational issues and several fundamental definitions used throughout the text.

Assumptions:
$\mathcal{A}$.1) *The observable data* $\{x_t\}$ *is generated from* $\Delta_{\boldsymbol{\gamma}}(L; \mathbf{d})\, x_t = \varepsilon_t$, $t = 1, ..., T$, *with* $\Delta_{\boldsymbol{\gamma}}(L; \mathbf{d})$ *defined in* (1), *and* $\mathbf{d}$ *being a possibly non-integer vector in* $\mathbb{R}^n$, $n \geq 1$.
$\mathcal{A}$.2) *The innovation process* $\{\varepsilon_t, \mathcal{G}_t\}_{-\infty}^{\infty}$, $\mathcal{G}_t = \sigma(\varepsilon_j : j \leq t)$, *forms a martingale difference sequence (MDS) and verifies* $E(\varepsilon_t^2) = \sigma^2 < \infty$, $E(\varepsilon_t^2 | \mathcal{G}_{t-1}) > 0$ *almost surely, with one of the following restrictions holding true:*

$\mathcal{A}.2.1$) $\{\varepsilon_t\}$ *is independent and identically distributed and* $E(|\varepsilon_t^4|^{1+r})$ *uniformly bounded for some* $r > 0$.

$\mathcal{A}.2.2$) $\{\varepsilon_t\}$ *is strictly stationary and ergodic with*

$$\sum_{l_1=-\infty}^{\infty} \sum_{l_2=-\infty}^{\infty} \ldots \sum_{l_7=-\infty}^{\infty} |\kappa_\varepsilon (0, l_1, ..., l_7)| < \infty,$$

where $\kappa_\varepsilon (0, l_1, ..., l_7)$ *is the eight-order joint cumulant of* $\{\varepsilon_t\}$.

In our analysis, we consider the general case of (1) under the null hypothesis given by $x_t \sim$GFI($\mathbf{d}$). Simpler specifications (*e.g.*, pure seasonal models) arise considering restricted versions of $\Delta_\gamma (L; \mathbf{d}) x_t$, for which our conclusions extend straightforwardly. Owing to nonstationarity, it is customary in the literature related to fractional integration to assume $x_t \mathbb{I}_{(t\leq 0)} = 0$, either explicitly (*e.g.*, Tanaka, 1999; Demetrescu, Kuzin and Hassler, 2008), or indirectly (*e.g.*, Nielsen, 2004, 2005), since this restriction ensures that the observable process is well-defined in the mean-square sense regardless of the values of $\mathbf{d}$; see Marinucci and Robinson (1999) and Robinson (2005) for further details. We note that, under the null hypothesis, this restriction is not formally necessary to derive the asymptotic distribution of the Lagrange Multiplier (LM) test statistics studied in this paper, although it would conveniently simplify the theoretical characterisation of the power function under the alternative hypothesis.[1] Assumption $\mathcal{A}.2.1$ can be weakened by requiring that, conditional on the σ-field of events $\mathcal{G}_t$, moments up to the fourth-order (and suitable cross-products of elements of ε_t) equal the corresponding unconditional moments, so that essentially $\{\varepsilon_t\}$ is only required to behave as an i.i.d process up to the fourth-order moment. The main purpose of $\mathcal{A}.2.2$ is to allow for (unknown) time-varying conditional volatility patterns in $\{\varepsilon_t\}$. For instance, GARCH-type and Stochastic Volatility models are permitted, among other forms of conditional heteroskedasticity, under restrictions that limit the extent of temporal dependence. As in Gonçalves and Kilian (2007) and Demetrescu *et al.* (2008), this holds by requiring the absolute summability of the eight-order joint cumulants.

In our analysis we will further relax Assumption $\mathcal{A}.2$, by also allowing for stationary AR(p) dynamics in the generating process, which may appear jointly with time-varying volatility patterns. Therefore, we consider as an alternative to assumption $\mathcal{A}.2$ the following:

$\mathcal{A}.2'$) *The innovation process satisfies* $a(L)\varepsilon_t = v_t$, *where* $a(L) = 1 - \sum_j^p a_j L^j$, $p \geq 0$, *such that* $a(z)$ *has all its roots outside the unit circle and* $\{v_t, \mathcal{G}_t\}$, *is a strictly stationary and ergodic MDS satisfying the restrictions in either Assumption* $\mathcal{A}.2.1$ or $\mathcal{A}.2.2$.

The proofs under Assumption $\mathcal{A}.2'$ are formally discussed for the case in which the autoregressive order p is known. For practical purposes, the short-run dynamics may be characterised by a stationary and invertible linear process $\varepsilon_t = \sum_{j=0}^{\infty} b_j v_{t-j}$ such that the AR(p) model, for some large enough $p < \infty$, approaches the underlying AR representation reasonably well. Since the actual performance of this approximation, when the underlying correlation structure in the short-run component is unknown, is ultimately an empirical question we shall study in detail the effects on the finite-sample properties of the regression-based tests in the Monte Carlo section.

Next, we formally introduce the definitions of the main processes and variables that characterise the test procedures in our study.

---

[1] Other initialisations are also possible; see Tanaka (1999) and references therein. We thank an anonymous referee for valuable comments on the truncation restriction used in this paper.

Definition 2.1. *For all* $j \geq 1$ *and any* $\gamma \in [0, \pi]$, *define the non-stochastic weighting process* $\omega_j(\gamma)$ *as,*

$$\omega_j(\gamma) = \begin{cases} 1/j, & if\ \gamma = 0 \\ 2j^{-1}\cos(j\gamma), & if\ \gamma \in (0, \pi) \\ (-1)^j/j, & if\ \gamma = \pi \end{cases}.$$

More generally, given $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_n)'$, $\gamma_s \in [0, \pi]$, *define* $\boldsymbol{\omega}_j(\boldsymbol{\gamma}) = (\omega_j(\gamma_1), ..., \omega_j(\gamma_n))'$.

Definition 2.2. *Given the observable process* $\{x_t, t = 1, ..., T\}$ *as defined in* $\mathcal{A}.1$, *and a vector* $\boldsymbol{\delta} \in \mathbb{R}^n$, *define the filtered series* $\varepsilon_{\boldsymbol{\delta}t} = \Delta_{\boldsymbol{\gamma}}(L; \boldsymbol{\delta})\, x_t$, *where, if* $\boldsymbol{\delta} = \mathbf{d}$, *then* $\Delta_{\boldsymbol{\gamma}}(L; \mathbf{d})\, x_t = \varepsilon_t$ *and thus* $\varepsilon_{\mathbf{d}t} = \varepsilon_t$. *For any frequency* $\gamma_s \in [0, \pi]$, *define the following stochastic processes:*

$$\varepsilon^*_{\gamma_s, t-1} = \sum_{j=1}^{t-1} \omega_j(\gamma_s)\, \varepsilon_{\boldsymbol{\delta}, t-j} \text{ and } \varepsilon^{**}_{\gamma_s, t-1} = \sum_{j=1}^{\infty} \omega_j(\gamma_s)\, \varepsilon_{\boldsymbol{\delta}, t-j}.$$

Definition 2.3. *Given* $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_n)'$, *define the n-dimensional vectors*

$$\begin{aligned} \boldsymbol{\varepsilon}^*_{\boldsymbol{\gamma}, t-1} &= \left(\boldsymbol{\varepsilon}^*_{\gamma_1, t-1}, ..., \boldsymbol{\varepsilon}^*_{\gamma_n, t-1}\right)' = \sum_{j=1}^{t-1} \boldsymbol{\omega}_j(\boldsymbol{\gamma})\, \varepsilon_{\boldsymbol{\delta}, t-j}; \\ \boldsymbol{\varepsilon}^{**}_{\boldsymbol{\gamma}, t-1} &= \left(\boldsymbol{\varepsilon}^{**}_{\gamma_1, t-1}, ..., \boldsymbol{\varepsilon}^{**}_{\gamma_n, t-1}\right)' = \sum_{j=1}^{\infty} \boldsymbol{\omega}_j(\boldsymbol{\gamma})\, \varepsilon_{\boldsymbol{\delta}, t-j}. \end{aligned}$$

Some comments on these definitions follow. The process $\omega_j(\gamma_s)$, $1 \leq s \leq n$, in Definition 2.1 is related to the asymptotic expansions of the polynomials $\log \xi_{\gamma_s}(L; \delta_s)$ which characterise the score vector under the null hypothesis and, therefore, plays a major role in the construction of LM test statistics; see the next section for details. Similarly, Definitions 2.2. and 2.3 introduce two key variables for this context: the sample-based (or observable) vector-series $\varepsilon^*_{\boldsymbol{\gamma}, t-1}$, and its asymptotic (or theoretical) counterpart $\varepsilon^{**}_{\boldsymbol{\gamma}, t-1}$, which are determined by weighting the filtered series $\varepsilon_{\boldsymbol{\delta}t}$. As formally shown in the technical appendix, constructing the test statistics by using the observable information does not impact on the asymptotic distribution.

2.1 The Lagrange Multiplier Test

In this section, we present an LM procedure for testing fractional integration under assumptions $\mathcal{A}.1$ and $\mathcal{A}.2$, which will be useful in order to better understand the regression framework presented below. We construct a Gaussian likelihood function, as though the innovations were normally distributed, but note that this is not required in order to ensure the validity of the asymptotic results.

Hence, consider $\boldsymbol{\delta} = \mathbf{d} + \boldsymbol{\theta}$, with i-th element $\delta_i = d_i + \theta_i$ and recall that $\varepsilon_{\boldsymbol{\delta}t} = \Delta_{\boldsymbol{\gamma}}(L; \boldsymbol{\delta})\, x_t$. The Gaussian log-likelihood function for $(\boldsymbol{\delta}', \sigma^2)'$, given $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_n)'$ and conditional on the set of observable information $\mathbf{x}_T = \{x_t, t = 1, ..., T\}$ is

$$\mathcal{L}(\boldsymbol{\delta}, \sigma^2 | \mathbf{x}_T) = -\frac{T}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}(\varepsilon_{\boldsymbol{\delta}t})^2 \qquad (2)$$

and the respective gradient of (2) evaluated under $\mathrm{H}_0 : \boldsymbol{\theta} = \mathbf{0}$ is

$$\left.\frac{\partial \mathcal{L}(\boldsymbol{\delta}, \sigma^2|\mathbf{x}_T)}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=0} = -\frac{1}{\sigma^2}\sum_{t=1}^{T} \varepsilon_t \left.\left(\frac{\partial \varepsilon_{\boldsymbol{\delta} t}}{\partial \boldsymbol{\theta}}\right)\right|_{\boldsymbol{\theta}=0}.$$

Note that if $\gamma_1 = 0$, *i.e.* the zero frequency is considered, then the partial derivative of $\varepsilon_{\boldsymbol{\delta} t}$ on θ_1 is

$$\frac{\partial \varepsilon_{\boldsymbol{\delta} t}}{\partial \theta_1} = \log\left(1-L\right)\left(1-L\right)^{\theta_1}\left(1-L\right)^{d_1}\left[\prod\nolimits_{i=2}^{n} \xi_{\gamma_i}\left(L;\delta_i\right)\right] x_t$$

which reduces to $\log\left(1-L\right)\Delta_{\boldsymbol{\gamma}}\left(L;\mathbf{d}\right)x_t = \log\left(1-L\right)\varepsilon_t$ when the score vector is evaluated at $\boldsymbol{\theta} = \mathbf{0}$. Similarly, the partial derivatives with respect to θ_s, $s = 2, ..., n-1$, for $\gamma_s \in (0, \pi)$ and θ_n for $\gamma_n = \pi$, when evaluated under the null hypothesis are given, respectively as,

$$\left.\frac{\partial \varepsilon_{\boldsymbol{\delta} t}}{\partial \theta_s}\right|_{\mathrm{H}_0:\boldsymbol{\theta}=0} = \log\left(1 - 2\cos\gamma_s L + L^2\right)\varepsilon_t, \text{ and } \left.\frac{\partial \varepsilon_{\boldsymbol{\delta} t}}{\partial \theta_n}\right|_{\mathrm{H}_0:\boldsymbol{\theta}=0} = \log\left(1+L\right)\varepsilon_t.$$

Following Chung (1996), Gradshteyn and Ryzhik (2000, sect. 1.514), and Breitung and Hassler (2002), the elements that characterise the score vector under the null hypothesis can be expanded as:

$$\log\left(\mathcal{F}_{\gamma_k}\right)\varepsilon_t = -\sum_{j=1}^{\infty}\omega_j\left(\gamma_k\right)\varepsilon_{t-j},$$

where $\mathcal{F}_{\gamma_1} = 1 - L$, $\mathcal{F}_{\gamma_l} = \xi_{\gamma_l}\left(L;1\right)$, $l = 2, ..., n-1$, $\mathcal{F}_{\gamma_n} = 1 + L$ and $\omega_j\left(\gamma_k\right)$ is as given in Definition 2.1. Using Definitions 2.2 and 2.3, it follows that,

$$\left.\frac{\partial \mathcal{L}(\boldsymbol{\delta}, \sigma^2|\mathbf{x}_T)}{\partial \boldsymbol{\theta}}\right|_{\mathrm{H}_0:\boldsymbol{\theta}=0} = \frac{1}{\sigma^2}\sum_{t=1}^{T}\varepsilon_t\left(\sum_{j=1}^{\infty}\boldsymbol{\omega}_j\varepsilon_{t-j}\right) \equiv \frac{1}{\sigma^2}\sum_{t=1}^{T}\varepsilon_t\left(\boldsymbol{\varepsilon}^{**}_{\boldsymbol{\gamma},t-1}\right)$$

for which the final sample version is,

$$\left.\frac{\partial \mathcal{L}(\boldsymbol{\delta}, \sigma^2|\mathbf{x}_T)}{\partial \boldsymbol{\theta}}\right|_{\mathrm{H}_0:\boldsymbol{\theta}=0} = \frac{1}{\sigma^2}\sum_{t=2}^{T}\varepsilon_t\left(\sum_{j=1}^{t-1}\boldsymbol{\omega}_j\varepsilon_{t-j}\right) \equiv \frac{1}{\sigma^2}\sum_{t=2}^{T}\varepsilon_t\left(\boldsymbol{\varepsilon}^{*}_{\boldsymbol{\gamma},t-1}\right).$$

Under the null hypothesis and given the restrictions provided in Assumption $\mathcal{A}.2$, ε_t is uncorrelated with $\boldsymbol{\varepsilon}^{*}_{\boldsymbol{\gamma},t-1}$ and $\boldsymbol{\varepsilon}^{*}_{\boldsymbol{\gamma},t-1}$ is (asymptotically) covariance stationary, and so is the score vector. The Fisher information matrix, estimated as the outer product of gradients, is given by the inverse of

$$\frac{1}{\sigma^4}\frac{1}{T}\sum_{t=2}^{T}\varepsilon_t^2\left(\boldsymbol{\varepsilon}^{*}_{\boldsymbol{\gamma},t-1}\boldsymbol{\varepsilon}'^{*}_{\boldsymbol{\gamma},t-1}\right)$$

which converges in probability to a finite, invertible covariance matrix under Assumptions $\mathcal{A}.1$ and $\mathcal{A}.2$. Therefore, a suitable test statistic for $\mathrm{H}_0 : \boldsymbol{\theta} = \mathbf{0}$ under the Lagrange Multiplier principle can be devised. This is formally stated next in Theorem 2.1.

Theorem 2.1. *Let $\{x_t, t = 1, ..., T\}$ be an observable process such that Assumptions $\mathcal{A}.1$ and $\mathcal{A}.2$ hold true. Given some arbitrary $\mathbf{d} \in \mathbb{R}^n$, define the test statistic*

$$LM_T = \left(\sum_{t=2}^{T}\varepsilon_{\mathbf{d}t}\boldsymbol{\varepsilon}^{*}_{\boldsymbol{\gamma},t-1}\right)'\left[\sum_{t=2}^{T}\varepsilon^2_{\mathbf{d}t}\boldsymbol{\varepsilon}^{*}_{\boldsymbol{\gamma},t-1}\boldsymbol{\varepsilon}'^{*}_{\boldsymbol{\gamma},t-1}\right]^{-1}\left(\sum_{t=2}^{T}\varepsilon_{\mathbf{d}t}\boldsymbol{\varepsilon}^{*}_{\boldsymbol{\gamma},t-1}\right) \qquad (3)$$

with $\left\{\varepsilon_{\mathbf{d}t}, \boldsymbol{\varepsilon}^{*}_{\gamma,t-1}\right\}_{t=1}^{T}$ *determined based on* $\mathbf{d}$ *according to Definitions 2.1-2.3. Then, under the null hypothesis* $\mathrm{H}_0 : \boldsymbol{\delta} = \mathbf{d}$ or *equivalently,* $\mathrm{H}_0 : \boldsymbol{\theta} = \mathbf{0}$*, it follows as* $T \to \infty$ that,

$$LM_T \Rightarrow \chi^2_{(n)},$$

where $\chi^2_{(n)}$ *stands for a Chi-squared distribution with* n *degrees of freedom.*

Proof. See Appendix B.

Theorem 2.1 generalises the LM test proposed by Tanaka (1999) considering single or multiple fractional integration at any frequency in $[0, \pi]$, and with innovations which are not necessarily independent but simply MDS. Hence, the testing procedure suggested is robust against (conditional) heteroskedasticity of unknown form provided that the regularity conditions are observed. Under the i.i.d assumption in $\mathcal{A}$.2.1, the asymptotic variance of the score vector is given by $\sigma^2\mathbf{\Gamma}_\gamma$, $\mathbf{\Gamma}_\gamma \equiv \sum_{j=1}^{\infty} \boldsymbol{\omega}_j(\boldsymbol{\gamma})\,\boldsymbol{\omega}'_j(\boldsymbol{\gamma})$, which equals $\sigma^2\pi^2/6$ for the restricted case $\gamma_1 = 0$ and $n = 1$ analysed in Tanaka (1999); see Appendix A for further details. The variance parameter σ^2 can be estimated consistently as $\widehat{\sigma}^2_T = \sum_{t=2}^{T} \varepsilon^2_{\mathbf{d}t}/T$, where the non-stochastic matrix $\mathbf{\Gamma}_\gamma$ can be determined by the close-form representations given in Appendix A, or by simple numerical approximation.

Remark 2.1: For theoretical purposes we have considered that the vector of frequencies $\boldsymbol{\gamma}$ is known. This allows us to discuss the asymptotic distribution of the LM test under fractional integration at any frequency, or combination of frequencies, in $[0, \pi]$. For empirical purposes, this restriction holds naturally for the zero-frequency case as well as for pure seasonal models, as $\boldsymbol{\gamma}$ is predetermined, but it may prove restrictive when analysing cyclical models by means of Gegenbauer polynomials. This limitation also extends to the frequency-domain test studied in Robinson (1994) and, as a result, both methods would require consistent estimates of the unknown frequencies in the most general context. Different estimation methods have been proposed in the literature; see, among others, Yajima (1996), Giriatis, Hidalgo, and Robinson (2001), Hidalgo and Soulier (2004), Dalla and Hidalgo (2005), and Hidalgo (2007). However, the formal proof for consistency is limited to the case $|d| < 1/2$ and remains to be shown for the most general case treated in this paper, which motivates an interesting topic for further research. In any case, the estimation bias that may arise when inferring $\boldsymbol{\gamma}$ in small samples, may imply further biases when using the estimated values in subsequent testing.

Remark 2.2: The LM test previously described in Theorem 2.1 may also be obtained under short-run dynamics in the errors, *i.e.*, the more general assumption $\mathcal{A}.2'$; see, for instance, Robinson (1994) and Tanaka (1999, pp. 563-565). The results in Tanaka (1999) regarding the handling of short-run dynamics, although relating to the zero frequency only, are interesting and illustrative of the difficulties involved in the correction of autocorrelation in the residuals. Tanaka (1999, p.564), in reference to ARMA errors, shows that the computation of the variance/covariance matrix necessary to robustify the test becomes more involved as the orders of the autoregressive and moving average components become larger. This characteristic of the time domain Lagrange Multiplier tests makes the regression based test procedures described next an appealing approach from an empirical point of view.

2.2 Regression-based tests for fractional integration

As an alternative to the approach discussed in the previous section, we propose a testing procedure that belongs to the linear regression context which is asymptotically equivalent to the LM_T test statistic. The regression-based approach was pioneered by Agiakloglou and Newbold (1994) for the context of fractional integration at the zero frequency and further developed in Breitung and Hassler (2002), Hassler and Breitung (2006), and Demetrescu *et al.* (2008) for the same context. Regression-based tests are particularly useful for the empirically relevant case in which data exhibit weak correlation. Hence, we discuss the testing principle and the asymptotic distribution of the relevant tests under the more general assumption $\mathcal{A}.2'$. The results for the restricted case studied in Assumption $\mathcal{A}.2$ follow straightforwardly. The following proposition states the general testing strategy in the regression framework.

Proposition 2.1. *Given* $\{x_t, t = 1, ..., T\}$ *under Assumptions* $\mathcal{A}.1$ and $\mathcal{A}.2'$, *the null hypothesis* $\mathrm{H}_0 : \boldsymbol{\theta} = \mathbf{0}$, *i.e.,* $x_t \sim$GFI$(\mathbf{d})$, $\mathbf{d} \in \mathbb{R}^n$, *can be tested against the alternative* $\mathrm{H}_1 : x_t \sim$GFI$(\mathbf{d} + \boldsymbol{\theta})$, $\boldsymbol{\theta} \neq \mathbf{0}$, *through a test for the joint significance of the regression coefficients* $\{\phi_s\}_{s=1}^n$ *(i.e.,* $\mathrm{H}_0 : \phi_1 = ... = \phi_n = 0$), *in the following augmented least-squares auxiliary regression:*

$$\varepsilon_{\mathbf{d}t} = \sum_{s=1}^{n} \phi_l \varepsilon^*_{\gamma_s,t-1} + \sum_{i=1}^{p} \zeta_i \varepsilon_{\mathbf{d},t-i} + e_{tp}, \quad t = p + \mathit{1}, ..., T \tag{4}$$

where $\left\{\varepsilon_{\mathbf{d}t}, \varepsilon^*_{\gamma_s,t-1}\right\}_{t=2}^{T}$ *is determined based on* $\mathbf{d}$ *according to Definitions 2.1-2.3, and p represents the order of augmentation considered.*

The statistical properties of the LS estimates of $\boldsymbol{\phi} = (\phi_1, ..., \phi_n)'$ under the null differ from those under the alternative hypothesis, which provides us with the basis to statistically identify the order of integration of the data. Note that, by sharp contrast to the LM_T test, the problem of short-run dynamics can easily be handled in the regression context by means of augmentation, as in the case of the well-known Dickey-Fuller (DF) test. Theorem 2.2 shows the limit results for the estimated coefficients in the auxiliary regression (4) under the set of restrictions considered, and Theorem 2.3 discusses the asymptotic distribution of a suitable test statistic.

Theorem 2.2. *Denote* $\boldsymbol{\beta} = \left(\phi_1, .., \phi_n, \zeta_1, ..., \zeta_p\right)'$ and *let* $\boldsymbol{\beta}_T$ *be the (n+p) estimated vector of parameters in the pth order augmented auxiliary regression* $\varepsilon_{\mathbf{d}t} = \boldsymbol{\beta}'\mathbf{X}^*_{tp} + e_{tp}$, *with* $\mathbf{X}^*_{tp} = \left(\boldsymbol{\varepsilon}'^*_{\boldsymbol{\gamma},t-1}, \varepsilon_{\mathbf{d},t-1}, ..., \varepsilon_{\mathbf{d},t-p}\right)'$. *Let the (n+p) vector* $\boldsymbol{\mu}_0 = (0, ..., 0, a_1, ..., a_p)'$, *with the* a_i *parameters corresponding to the autoregressive coefficients in* $\left(1 - \sum_{i=1}^{p} a_i L\right) \varepsilon_t = v_t$. *Then, under the null hypothesis and Assumptions* $\mathcal{A}.1$ and $\mathcal{A}.2'$, *as* $T \rightarrow \infty$,

$$\sqrt{T}\left(\boldsymbol{\beta}_T - \boldsymbol{\mu}_0\right) \Rightarrow \mathcal{N}\left(\mathbf{0},\ \left(\boldsymbol{\Omega}^{**}_p\right)^{-1} \boldsymbol{\Lambda}_p \left(\boldsymbol{\Omega}^{**}_p\right)^{-1}\right)$$

with $\boldsymbol{\Omega}^{**}_p \equiv E\left(\mathbf{X}^{**}_{tp}\mathbf{X}'^{**}_{tp}\right)$ *and* $\boldsymbol{\Lambda}_p \equiv E\left(v_t^2 \mathbf{X}^{**}_{tp}\mathbf{X}'^{**}_{tp}\right)$, *where* $\mathbf{X}^{**}_{tp} = \left(\boldsymbol{\varepsilon}'^{**}_{\boldsymbol{\gamma},t-1}, \varepsilon_{\mathbf{d},t-1}, ..., \varepsilon_{\mathbf{d},t-p}\right)'$.

Proof. *See Appendix B.*

Remark 2.3: Consider $\mathcal{A}.1$ and the more restrictive condition $\mathcal{A}.2$ which sets $p = 0$. Let $\boldsymbol{\phi}_T$ be the estimated vector of parameters from an auxiliary regression with no augmentation, $\varepsilon_{\mathbf{d}t} = \boldsymbol{\phi}'\boldsymbol{\varepsilon}^*_{\boldsymbol{\gamma},t-1} + e_t$, $t = 2, ..., T$. Then, under the null hypothesis $\mathrm{H}_0 : \boldsymbol{\theta} = \mathbf{0}$, and as $T \rightarrow \infty$,

$$\sqrt{T}\boldsymbol{\phi}_T \Rightarrow \mathcal{N}\left(\mathbf{0}, \mathbf{V}_{\boldsymbol{\gamma}}\right),$$

where $\mathbf{V}_{\gamma} = \left(\frac{1}{\sigma^4}\right) \mathbf{\Gamma}_{\gamma}^{-1} \mathbf{\Lambda}_{\varepsilon,\gamma} \mathbf{\Gamma}_{\gamma}^{-1}$, with $\mathbf{\Gamma}_{\gamma} \equiv \sum_{j=1}^{\infty} \boldsymbol{\omega}_j(\gamma) \boldsymbol{\omega}_j'(\gamma)$ defined previously, and $\mathbf{\Lambda}_{\varepsilon,\gamma} = E\left(\varepsilon_t^2 \boldsymbol{\varepsilon}_{\gamma,t-1}^{**} \boldsymbol{\varepsilon}_{\gamma,t-1}^{\prime **}\right)$; for details, see Hassler, Rodrigues and Rubia (2008).

Owing to asymptotic normality, and since the null hypothesis only implies linear restrictions on the parameters involved, we can easily test $\mathrm{H}_0 : \boldsymbol{\theta} = \mathbf{0}$ by means of a test statistic based on the Wald representation which tests for $\mathrm{H}_0 : \boldsymbol{\phi} = \mathbf{0}$ in the auxiliary regression. Note that, although we use the functional form of a Wald-type test, our testing procedure is an LM or score test because it builds directly on the gradient of the likelihood function. Theorem 2.3 discusses its asymptotic distribution in the general case.

Theorem 2.3. *Let* $\mathbf{R}$ *be an* $n \times (n+p)$ *matrix such* that $[\mathbf{R}]_{ij} = 1$ *for all* $i = j$ *and zero otherwise. Consider the Wald-type test statistic on the estimates of the augmented auxiliary regression, i.e.,*

$$\Upsilon_{Wp}^{(n)} = \left[\sqrt{T}\mathbf{R}\boldsymbol{\beta}_T\right]' \left[\mathbf{R}\widehat{\mathbf{V}}_{\mathbf{T}}\mathbf{R}'\right]^{-1} \left[\sqrt{T}\mathbf{R}\boldsymbol{\beta}_T\right] \tag{5}$$

with $\widehat{\mathbf{V}}_{\mathbf{T}}$ *being the sample estimation of the covariance matrix of* $\boldsymbol{\beta}_T$ *such that*

$$\widehat{\mathbf{V}}_{\mathbf{T}} = \left(\frac{1}{T}\sum_{t=p+1}^{T} \mathbf{X}_{tp}^{*}\mathbf{X}_{tp}^{*\prime}\right)^{-1} \left(\frac{1}{T}\sum_{t=p+1}^{T} \widehat{e}_{tp}^{2}\mathbf{X}_{tp}^{*}\mathbf{X}_{tp}^{*\prime}\right) \left(\frac{1}{T}\sum_{t=p+1}^{T} \mathbf{X}_{tp}^{*}\mathbf{X}_{tp}^{*\prime}\right)^{-1},$$

where $\widehat{e}_{tp}$ *denotes the estimated residuals from (4). Under the same conditions of Theorem 2.2,* $\Upsilon_{Wp}^{(n)}$ *is asymptotically equivalent to* LM_T*, i.e.,* $\Upsilon_{Wp}^{(n)} \Rightarrow \chi^2_{(n)}$.

Proof. See Appendix B.

Corollary 2.1. *Consider the restricted joint hypothesis* $\boldsymbol{\theta} = \theta\mathbf{1}_n$, for some scalar $\theta \neq 0$, *and where* $\mathbf{1}_n$ *is a vector of ones in* $\mathbb{R}^n$*. This is the case, for instance, when analysing the suitability of so-called (seasonal) rigid models, which assume homogeneity in the order of fractional integration across the set of frequencies involved; see Porter-Hudak (1990) and Hassler (1994). The auxiliary regression in this case is given as* $\varepsilon_{\mathbf{d}t} = \bar{\phi}\left(\sum_{s=1}^{n} \varepsilon_{\gamma_s,t-1}^{*}\right) + \sum_{i=1}^{p} \zeta_i \varepsilon_{\mathbf{d},t-i} + u_t$*, and the relevant test statistic, say* $\bar{\Upsilon}^{(n)}$*, analyses the significance of the* $\bar{\phi}$ *parameter. This test statistic, which is a squared t-statistic, is asymptotically distributed as* $\chi^2_{(1)}$*, since only one restriction is implied.*

Corollary 2.2. *If* $\{v_t\}$ *in assumption* $\mathcal{A}.2'$ *is i.i.d with finite fourth-order moment,* $E\left(v_t^2 \mathbf{X}_{tp}^{**}\mathbf{X}_{tp}^{**\prime}\right)$ *is proportional to* $E\left(\mathbf{X}_{tp}^{**}\mathbf{X}_{tp}^{**\prime}\right)$ *. Hence, the null hypothesis* $\mathrm{H}_0 : \boldsymbol{\phi} = \mathbf{0}$ *can easily be tested by using alternative test statistics which can be constructed under the Lagrange Multiplier and the Likelihood Ratio principles, and which are asymptotically equivalent to* LM_T*. As discussed previously, in the context of this paper all these tests are necessarily LM tests regardless of their functional form. Let* $\Upsilon_{LR,p}^{(n)} = T(\log \mathcal{S}_R - \log \mathcal{S}_u)$ *and* $\Upsilon_{LM,p}^{(n)} = T(\mathcal{S}_R - \mathcal{S}_u)/\mathcal{S}_R$*, where* $\mathcal{S}_R$ *and* $\mathcal{S}_u$ *denote the squared sum of restricted and unrestricted residuals, respectively. Then, under the null, and as* $T \rightarrow \infty$, $\Upsilon_{LR,p}^{(n)} \Rightarrow \chi^2_{(n)}$ and $\Upsilon_{LM,p}^{(n)} \Rightarrow \chi^2_{(n)}$.

Proof: For proof of corollaries 2.1 and 2.2, see Appendix B.

Remark 2.4. The regression-based tests discussed (either with augmentation under short-run dynamics, or no-augmentation under the MDS assumption) are asymptotically equivalent to the time-domain LM test in Section 2.1, to the frequency-domain test in Robinson (1994), and to the general maximum likelihood-based tests in Nielsen (2004). The LM regression-based test in Breitung and Hassler (2002), focusing on the fractionally integrated model, $\Delta_{\boldsymbol{\gamma}}(L;\mathbf{d}) = (1-L)^d$, arises for $n = 1$ at frequency zero as a particular case in our context. It is worth mentioning that, as remarked in Nielsen (2004), the experimental simulations in Tanaka (1999), and Breitung and Hassler (2002), show that in finite samples the time domain test procedures tend to be superior to the frequency domain tests, both in size and power behaviour, hence a similar performance is likely to be observed in a more general setting as well.

Remark 2.5. The tests presented above are robust against conditional heteroskedasticity of unknown form. This is achieved by using a consistent estimate of the asymptotic covariance matrix $\mathbf{V}_{\boldsymbol{\gamma}}$ based on a version of the Eicker-White estimator. If the data are believed to be generated under assumptions $\mathcal{A}.1$ and $\mathcal{A}.2.1$ then $\mathbf{V}_{\boldsymbol{\gamma}} = \boldsymbol{\Gamma}_{\boldsymbol{\gamma}}^{-1}$, and subsequently used directly.

Remark 2.6. As discussed in Breitung and Hassler (2002), the auxiliary regression centered on the zero-frequency, $\varepsilon_{\mathbf{d}t} = \phi_1 \varepsilon^*_{0,t-1} + e_t$, is reminiscent of the Dickey-Fuller regression and the Wald-test in Dolado, Gonzalo and Mayoral (2002). Nevertheless, meaningful differences arise since in the DF test the regressor is I(0) under the alternative, whereas $\varepsilon^*_{0,t-1}$ is FI$(d+\theta)$ owing to the different types of weights used in constructing these variables. Similarly, for pure seasonal models, the general auxiliary regression in Proposition 2.1 is evocative of the Hylleberg, Engle, Granger and Yoo (1990) [HEGY] test regression, in the sense that the regressors $\varepsilon^*_{\gamma_s,t-1}$ are weighted linear combinations of lags of $\varepsilon_{\mathbf{d}t}$ related to a specific (seasonal) frequency. Further differences arise in this case, because regressors in the HEGY context are ensured to be asymptotically orthogonal by construction, whereas the LM-based regressors are not. This feature advises against testing partial hypothesis (*i.e.*, subsets of m parameters) based on the estimates of the general model (*i.e.*, after estimating a regression with $n > m$ parameters), as the covariance matrix is not (block) diagonal.

Remark 2.7. GFI models are particularly difficult to estimate in practical settings owing to their strong non-linear nature. Proposition 2.1 provides a valuable tool to construct confidence sets that include the true value, $\mathbf{d} \in \mathbb{R}^n$, with $(1-\alpha)\,\%$ asymptotic nominal probability. These sets could be used to obtain reliable starting values for optimisation routines aiming to estimate $\mathbf{d}$, such as the (quasi)-maximum likelihood methods discussed in Chung (1996) and Nielsen (2004). Confidence sets can be obtained from a grid-search on Θ, a compact subset of $\mathbb{R}^n$, by using the results in Proposition 2.1. For instance, denote $\Upsilon^{(n)}_{W,\boldsymbol{\delta}}$ as the value of the test statistic in Theorem 2.3 when evaluated at any $\boldsymbol{\delta} \in \Theta$, and let $\mathcal{D}_{T\alpha} = \left\{\boldsymbol{\delta} : \Pr\left[\chi^2_{(n)} \leq \Upsilon^{(n)}_{W,\boldsymbol{\delta}}\right] \leq 1-\alpha\right\}$, *i.e.*, the subset of Θ containing all the vectors for which the null hypothesis cannot be rejected at the $(1-\alpha)\,\%$ asymptotic nominal confidence level. If $\mathcal{D}_{T\alpha}$ is in the interior of Θ, then the probability of $\mathbf{d}$ being in the closure of $\mathcal{D}_{T\alpha}$ is at least $(1-\alpha)\,\%$. The grid-search process is computationally feasible because the dimension parameter n is not large in empirical models, and because the order of integration in observable data usually assumes values in a small range. For rigid models, a confidence interval of the form $\left[d^{\alpha}_{T,l}, d^{\alpha}_{T,u}\right]$ can easily be constructed from Corollary 2.1, given $\bar{\mathcal{D}}_{T\alpha} = \left\{\delta : \Pr\left[\chi^2_{(1)} \leq \bar{\Upsilon}^{(n)}_{\boldsymbol{\delta}}\right] \leq 1-\alpha\right\}$, by setting $d^{\alpha}_{T,l} = \inf \bar{\mathcal{D}}_{T\alpha}$ and $d^{\alpha}_{T,u} = \sup \bar{\mathcal{D}}_{T\alpha}$.

Remark 2.8 Demetrescu *et al.* (2008) analyse the performance of several procedures to determine the order of augmentation, p, of the test regression in finite samples and conclude in favour of the rule of thumb proposed by Schwert (1989) which shows relatively good performance in finite-samples. This rule sets $p = \left[c(T/100)^{1/4}\right]$, where c is a positive constant and $[\cdot]$ denotes the integer value of the argument.

Remark 2.9. Throughout our analysis, we have focused on the model $\Delta_{\gamma}(L; \mathbf{d})(x_t - \mu_t) = \varepsilon_t$, by allowing for different dynamics in ε_t, and restricting $\mu_t = 0$ for simplicity of analysis. As commented in Breitung and Hassler (2002), the simplest way to deal with non-zero deterministic patterns, $\mu_t \neq 0$, is to detrend x_t prior to computing the relevant test statistics. This does not affect the limit distribution of the relevant statistics; see the discussion in Robinson (1994).

Remark 2.10. The theoretical derivation of the local power functions under the alternative is a nontrivial problem due to the multiple hypothesis context. For restricted cases, it becomes more tractable, and it can be shown, following for instance Tanaka (1999) and Demetrescu *et al.* (2008) under the additional restriction $x_t \mathbb{I}_{(t \leq 0)} = 0$, that the test procedures will converge to a noncentral Chi-squared distribution under local alternatives for which $\theta_i = O\left(T^{-1/2}\right)$. Since for applied purposes the behaviour of the power function in finite-samples is particularly relevant, we shall address this issue carefully next in the Monte Carlo section.

3 Finite-sample analysis

In this section, we address the empirical properties of the regression-based test statistic in finite samples. The zero-frequency fractionally integrated process, $\Delta_{\gamma}(L; \mathbf{d}) = (1 - L)^d$, has received considerable attention in the literature; see for instance, Breitung and Hassler (2002), and Nielsen (2004), among others. These show the good finite-sample performance of LM tests, both in absolute terms and in relation to alternative frequency domain-based procedures. We therefore analyse cyclical and seasonal models aiming to contribute to better understand the properties of LM tests in the general context.

The applied literature on cyclical or seasonal fractionally integrated models has focused on both economic and non-economic variables. Empirical datasets are characterised by quite different features. The number of observations available for financial and many geophysical variables is relatively large, and often includes several thousand observations, whereas the length of macroeconomic variables is much more limited.[2] Data recorded on a high-frequency basis typically exhibit persistent short-run dynamics, whereas aggregated data tend to display considerably weaker forms of serial dependence. We consider the possibility of different types of short-run dynamics as well as different sample sizes to analyse the empirical size and power. In particular, we focus on samples of length $T \in \{100, 250, 500\}$. For data sets involving a large number of observations, as some of those analysed in the literature, the asymptotic theory is expected to provide a good approximation.

In the first experiment we consider a simple pure cyclical model,

$$(1 - 2\cos\gamma_s L + L^2)^{d+\theta} x_t = \varepsilon_t$$

in order to analyse the empirical size and power properties of $\Upsilon_W^{(1)}$, which is asymptotically distributed as $\chi^2_{(1)}$, when testing $\mathrm{H}_0 : d = 1$ with true values given by $d = 1$ and $\theta \in [-0.3, 0.3]$.

---

[2] The dataset in Bouette *et al.* (2006), relating to hourly average wind speeds measured between 1951 and 2003, includes over 16,000 observations, Soares and Souza (2006) consider two years of hourly electricity demand and Gil-Alana (2005) studies US monthly inflation in a dateset with more than 1000 observations.

We consider 5000 replications and $\varepsilon_t \sim iid\mathcal{N}(0,1)$. Since the Gegenbauer frequency γ_s is a 'free' parameter, we set $\gamma_s = s\pi/10$, with $s = 1, ..., 9$. The rejection frequencies for a nominal significance level of 5% and sample sizes of $T = 100$ and $T = 250$ are shown in Table 1.

[Insert Table 1 around here]

The test shows approximately correct size and good power performance even in small samples. For $\theta = 0$, only minor differences in the empirical size of the tests, following no particular pattern, arise across the γ_s frequencies considered. For non-zero values of θ, we observe several interesting features in the empirical power functions. First, given γ_s and T, power tends to exhibit a symmetric U-shaped figure around the $\pi/2$ frequency, which is more evident for small values of $|\theta|$. This suggests that, the larger the difference $|\gamma_s - \pi/2|$ with $\gamma_s \in (0, \pi)$, the more powerful the testing procedure becomes. The dependence of power on the particular frequency the test is related to is not surprising, since the variance of the regressor (and hence, the signal-to-noise ratio and, ultimately, the power of the test) depends on the specific frequency, γ_s, considered and, more generally, on $\boldsymbol{\gamma}$; see appendix A for further technical details. Furthermore, if we compare these results to those in Breitung and Hassler (2002, Table 1, p.176) for the zero-frequency case, the power observed at the long-run frequency is approximately of the same order as that for $\gamma_s = \pi/2$. This suggests that, everything else equal, fractionally integrated dynamics is generally easier to detect at the cyclical than at the zero-frequency. A similar feature appears when dealing with $\gamma_s = \pi$ (not reported here) for which power is similar to that of $\gamma_s = \pi/2$.[3] Dealing with the non-zero frequency also has other benefits in terms of power. For fixed T and γ_s, the power functions tend to be symmetric around $\theta = 0$, since only the size of $\theta - 0$, and not its sign, seems to drive the probability of rejection. This does not seem to be the case for the zero-frequency case analysed in Breitung and Hassler (2002), where the LM test is likely to reject more easily if $\theta < 0$. Finally, power is largely enhanced even for a small sample of $T = 250$, and virtually reaches 100% for all tests when $T = 500$, thus showing the consistency of the testing procedure in cases of small samples.

As a second experiment, we consider a more general two-factor cyclical model given by,

$$(1 - 2\cos\gamma_1 L + L^2)^{d_1+\theta_1}(1 - 2\cos\gamma_2 L + L^2)^{d_2+\theta_2} x_t = \varepsilon_t.$$

We address the ability of the unrestricted joint test $\Upsilon_W^{(2)}$, asymptotically distributed as $\chi^2_{(2)}$, as well as that of the restricted joint test $\bar{\Upsilon}^{(2)}$ discussed in Corollary 2.1 and asymptotically distributed as $\chi^2_{(1)}$, to detect fractionally integrated dynamics. As before, we set $d_1 = d_2 = 1$, and $\theta_1, \theta_2 \in [-0.3, 0.3]$, considering 5000 replications and $\varepsilon_t \sim iid\mathcal{N}(0,1)$. The joint test $\Upsilon_W^{(2)}$ is expected to reject the null hypothesis if fractional integration is present in, at least, one of the frequencies involved. The restricted joint test $\bar{\Upsilon}^{(2)}$ should be more efficient than $\Upsilon_W^{(2)}$ when the restriction $\theta_1 = \theta_2$ is true, but it is expected to exhibit less comparative power to reject the false null otherwise.

In view of the previous experiment, we expect the power function to depend on the value of $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)'$. We set $\gamma_1 = 0.15 \approx \pi/20$, based on the estimated frequency of the business cycle by the NBER, and consider what seems to be the most unfavourable frequency for the tests when dealing with frequencies in $(0, \pi)$, given by $\gamma_2 = \pi/2$, which also corresponds to one

---

[3] Note that the asymptotic variance is proportional to $\psi(\gamma)$, see appendix A. This is a positive, symmetric and non-continuous function in $[0, \pi]$ that takes the minimum value $\psi(\gamma) = \pi^2/6$ for $\gamma \in \{0, \pi/2, \pi\}$, and the maximum value given by $\lim_{\gamma \to 0^+} \psi(\gamma) = \lim_{\gamma \to \pi^-} \psi(\gamma) = 2\pi^2/3$. We can therefore expect a discontinuity in the power function for the case $\gamma = 0 + \epsilon$ or $\gamma = \pi - \epsilon$ even for an arbitrarily small $\epsilon > 0$.

of the harmonics of quarterly and monthly seasonality. For frequencies $\gamma \in (0, \pi)$ away from $\pi/2$, further simulations (not reported here) show much better statistical performance both in terms of size and power. The rejection frequencies for a nominal significance level of 5% and sample length $T = 100$ are shown in Table 2.

[Insert Table 2 around here]

Several interesting features emerge from this experiment in relation to the joint test statistics $\bar{\Upsilon}^{(2)}$ and $\Upsilon_W^{(2)}$. We observe that the restricted test is more powerful than the latter when the restriction $\theta_1 = \theta_2$ is true, but is also considerably less efficient in the general context $\theta_1 \neq \theta_2$, particularly for small values of $|\theta|$. Both tests tend to reject the (false) null more easily when fractional integration is present at frequency 0.15, *i.e.*, at the frequency for which the magnitude $|\gamma_s - \pi/2\ |$ is larger. For instance, if $d_1 = 1 - 0.1$ and $d_2 = 1$, the power of $\bar{\Upsilon}^{(2)}$ and $\Upsilon_W^{(2)}$ is approximately 39.8% and 48.7%, respectively. In contrast, for $d_1 = 1$ and $d_2 = 1 - 0.1$, the power is only 8.2% and 16.1%. When both θ_1 and θ_2 move away from the origin, the power of the joint tests, particularly that of $\Upsilon_W^{(2)}$, increases significantly. We note that the power of $\Upsilon_W^{(2)}$ seems to be symmetric for the set of frequencies considered, whereas the restricted joint test $\bar{\Upsilon}^{(2)}$ tends to reject more easily when $\theta_1 > 0$ and $\theta_2 < 0$ when compared to the converse case. For instance, the power of $\bar{\Upsilon}^{(2)}$ for $\theta_1 = 0.3$ and $\theta_2 = -0.3$ is almost 100%, and around 25% for $\theta_1 = -0.3$ and $\theta_2 = 0.3$. By contrast, the power of the unrestricted test $\Upsilon_W^{(2)}$ in any of these cases is almost 100%. As in the case of the one-factor model, considering larger samples, $T \in \{250, 500\}$ leads to considerable improvement of the statistical properties of all the tests. We do not present these results to save space but can be provided upon request.

Finally, the last set of experiments also considers the two-factor filter $\Delta_\gamma(L; \boldsymbol{\delta}) = (1 - 2\cos\gamma_1 L + L^2)^{d_1+\theta_1}(1 - 2\cos\gamma_2 L + L^2)^{d_2+\theta_2}$, but now allowing for stationary and invertible ARMA patterns in the error term, *i.e.*, we analyse the performance of the augmentation-based test statistics when the DGP is,

$$\Delta_\gamma(L; \boldsymbol{\delta})\, x_t = \varepsilon_t \quad \text{and} \quad (1 - aL)\,\varepsilon_t = (1 - bL)\, v_t,$$

under the restriction $|a| < 1$ and $|b| < 1$. We first focus on ARMA(1,1) dynamics and, as in Demetrescu *et al.* (2008), set $a = 0.5$ and $b = -0.5$. The ARMA(1,1) model is particularly relevant because short-run dynamics in empirical applications are usually characterised parsimoniously through this specification. Additionally, we analyse in more detail the effects of persistence through an AR(1) with parameter $a \in \{0.5, 0.75, 0.9\}$ and $b = 0$ in the above specification. Since for empirical purposes the underlying structure of the short-run component is typically unknown, we explore the effects on the tests when the number of lags to be included in the auxiliary regression are determined according to Schwert's rule, $p = \left[4(T/100)^{1/4}\right]$, as this showed the best performance in the empirical analysis in Demetrescu *et al.* (2008). The rejection frequencies for the joint tests given ARMA(1,1) patterns for $T \in \{100, 500\}$ are shown in Table 3, whereas Tables 4 and 5 report the respective empirical results for AR(1) errors for the given values of the autoregressive coefficient a.

[Insert Table 3 around here]

We first discuss the results for the ARMA(1,1) dynamics. The general conclusions that arise for the weakly-dependent case are similar to those observed for the *i.i.d* case, although we observe several quantitative changes. Augmentation enables correction of the empirical size for all tests, and only small undersizing effects are observed in our simulations. However, and as shown in previous literature, ensuring correct empirical size against general ARMA dynamics through augmentation in small samples, such as $T = 100$, occurs usually at the cost of potentially large power reductions when compared to the i.i.d. case. This pervasive effect has been widely documented in the unit root literature, where the augmented Dickey-Fuller regression is perhaps the most widely used in applied settings. In fact, the power of the joint tests shows figures similar in magnitude to those observed in Demetrescu *et al.* (2008) for the zero frequency fractionally integrated case. By contrast to the unit root case, importantly, power improves considerably faster at frequencies away from zero. For instance, for the ARMA model considered, the power of $\Upsilon_{Wp}^{(2)}$ is not larger than 39% in the range $\boldsymbol{\theta} = (-0, 3, 0.3)'$ when only 100 observations are available. For a larger sample of $T = 500$, all else equal, power increases up to 98%. Similarly, the joint restricted test $\bar{\Upsilon}_{p}^{(2)}$ has a peak of approximately 30% for $T = 100$ when $\theta_1 = \theta_2 = -0.3$, which increases significantly to 99% when $T = 500$.

[**Insert Tables 4 and 5 around here**]

Similar results can be observed when analysing the effects of persistence in residuals. Although the empirical size is approximately correct in all cases, as the autoregressive root approaches one in a small sample with 100 observations, power reductions with respect to the i.i.d. case are far more evident. For small values of $|\boldsymbol{\theta}|$ it becomes difficult to reject the false null, and even for some configurations which include relatively large values of $\boldsymbol{\theta}$ when $a = 0.9$. As in the previous case, the power of the tests considerably improves as the number of observations increases. Therefore, for the test $\Upsilon_{Wp}^{(2)}$, given the samples typically available for many empirical applications, augmenting the regression proves a valid tool to ensure empirical sizes close to the asymptotic nominal level and good power properties.

4 Conclusion

In this paper, we analyse time domain regression-based tests that allow testing for fractionally integrated patterns against integer or fractional integration in general models. The tests involving single or multiple parameters can be computed from simple least-squares regressions, and are asymptotically equivalent to the frequency-domain LM tests of Robinson (1994) and the likelihood-based tests in Nielsen (2004), for which the relevant critical values are obtained from a Chi-square distribution with as many degrees of freedom as the number of restrictions being tested, and independent of the order of integration. Augmented versions of these tests are asymptotically robust against weakly-dependent errors following unknown patterns under quite general conditions, and exhibit good statistical performance in samples of moderate size. This makes the general regression-based LM testing strategy discussed in this paper a valuable tool when addressing preliminary data analysis in which parsimonious yet potentially restrictive hypothesis related to the order of integration of the data is formally validated or refuted.

Appendix A: Asymptotic covariance matrix in the i.i.d case

In this appendix we present the limit expressions which characterise the asymptotic variances and covariances of the score vector under i.i.d observations.

Definition A.1. *For any* $\gamma \in [0, \pi]$, *let* $\psi(\gamma) = \lim_{T\to\infty} \sum_{j=1}^{T} \omega_j^2(\gamma)$. *Following Gradshteyn and Ryzhik (2000, sect. 1.443), it follows that* $\psi(\gamma) = \pi^2/6$, if $\gamma \in \{0, \pi\}$, and $\psi(\gamma) = 2\left(\pi^2/3 - \pi\gamma + \gamma^2\right)$, *otherwise. Similarly, given* $\gamma_k, \gamma_m \in [0, \pi]$, $\gamma_k \neq \gamma_m$, *let* $\psi(\gamma_k, \gamma_m) = \lim_{T\to\infty} \sum_{j=1}^{T} \omega_j(\gamma_k)\,\omega_j(\gamma_m)$. *Note that* $|\psi(\gamma_k, \gamma_m)| < \infty$, *in particular, we have*

$$\psi(\gamma_k, \gamma_m) = \begin{cases} -\psi(\gamma_m)/2 & if\ \gamma_k = 0, \gamma_m = \pi \\ \left(\psi(\gamma_m) - \gamma_m^2\right)/2 & if\ \gamma_k = 0, \gamma_m \in (0, \pi) \\ -\psi(\gamma_m)/4 - \gamma_m(\pi/2 - \gamma_m) & if\ \gamma_k = \pi, \gamma_m \in (0, \pi) \end{cases},$$

and, if γ_k,$\gamma_m \in (0, \pi)$, *then*

$$\begin{aligned} \psi(\gamma_k, \gamma_m) &= \frac{2\pi^2}{3} - \pi\left(\gamma_k + \gamma_m + |\gamma_k - \gamma_m|\right) + \frac{\left[(\gamma_k + \gamma_m)^2 + (|\gamma_k - \gamma_m|)^2\right]}{2} \\ &= \psi(0, |\gamma_k - \gamma_m|) + \psi(0, \gamma_k + \gamma_m). \end{aligned}$$

Definition A.2. *Given* $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_n)'$, *with* $0 \leq \gamma_1 < \gamma_2 < ... < \gamma_n \leq \pi$, *denote* $\boldsymbol{\Gamma}_{\boldsymbol{\gamma}} = \lim_{T\to\infty} \sum_{j=1}^{T} \boldsymbol{\omega}_j(\boldsymbol{\gamma})\,\boldsymbol{\omega}_j(\boldsymbol{\gamma})'$, i.e.,

$$\boldsymbol{\Gamma}_{\boldsymbol{\gamma}} = \begin{pmatrix} \psi(\gamma_1) & \psi(\gamma_1, \gamma_2) & \cdots & \psi(\gamma_1, \gamma_n) \\ \psi(\gamma_2, \gamma_1) & \psi(\gamma_2) & \cdots & \psi(\gamma_2, \gamma_n) \\ \vdots & \vdots & \ddots & \vdots \\ \psi(\gamma_n, \gamma_1) & \psi(\gamma_n, \gamma_2) & \cdots & \psi(\gamma_n) \end{pmatrix},$$

with $\boldsymbol{\Gamma}_{\boldsymbol{\gamma}} < \infty$ *being a symmetric positive definite matrix. Under the i.i.d restriction, the asymptotic variance of the score vector is proportional to* $\boldsymbol{\Gamma}_{\boldsymbol{\gamma}}$; *see Appendix B for further details.*

Appendix B: Technical Proofs

Before proceeding, consider the following additional notation. For an $(n \times 1)$ vector A, $||A||$ denotes the Euclidean vector norm, such that $||A||^2 = A'A$. For an $(n \times m)$ matrix A, $||A||$ denotes the Euclidean matrix norm, $||A||^2 = tr(A'A)$. The constant K is used throughout the proofs to refer to some generic strictly positive constant which does not depend on the sample size. The notation, $\Rightarrow$, $\xrightarrow{p}$, $\xrightarrow{ms}$, $\rightarrow$ denotes weak convergence, convergence in probability, mean square convergence and convergence of a series of real numbers, respectively. The conventional notation $o(1)$ $(o_p(1))$ is used to represent a series of numbers (random numbers) converging to zero (in probability), while $O(1)$ $(O_p(1))$ denotes a series of numbers (random numbers) that are bounded (in probability). As in the main text, $\mathbb{I}_{(.)}$ is an indicator function, and vectors and matrices are denoted through bold letters. Finally, since $\boldsymbol{\gamma}$ is used to refer to the vector of frequencies that characterise the filter $\Delta_{\boldsymbol{\gamma}}(L; \boldsymbol{\delta})$, we shall use the short-hand notation $\boldsymbol{\omega}_j \equiv \boldsymbol{\omega}_j(\boldsymbol{\gamma})$ as there is no risk of confusion.

Next, we provide some preliminary Lemmae necessary for the proofs of the theorems presented in the text.

Lemma B.1. Consider assumptions A.1 and A.2, and *let* $\varepsilon_t = \Delta_{\boldsymbol{\gamma}}(L; \mathbf{d})\,x_t$ *and* $\boldsymbol{\gamma} \equiv (\gamma_1, .., \gamma_n)'$. *Consider the random vectors,* $\boldsymbol{\varepsilon}_{\boldsymbol{\gamma},t-1}^{*}$ *and* $\boldsymbol{\varepsilon}_{\boldsymbol{\gamma},t-1}^{**}$ *as given in Definition 2.3,* $\boldsymbol{\varepsilon}_{\boldsymbol{\gamma},t-1}^{**} - \boldsymbol{\varepsilon}_{\boldsymbol{\gamma},t-1}^{*} = \boldsymbol{\vartheta}_{\boldsymbol{\gamma},t}$,

$\boldsymbol{\Omega}_t^{**} = \boldsymbol{\varepsilon}^{**}_{\boldsymbol{\gamma},t-1}\boldsymbol{\varepsilon}'^{**}_{\boldsymbol{\gamma},t-1}$, and $\boldsymbol{\Omega}^{*}_t = \boldsymbol{\varepsilon}^{*}_{\boldsymbol{\gamma},t-1}\boldsymbol{\varepsilon}'^{*}_{\boldsymbol{\gamma},t-1}$. *Then, for any arbitrary constants* $\alpha > 0$, $\beta > 1/2$, *it follows as* $T \to \infty$ that,

i) $\boldsymbol{\vartheta}_{\boldsymbol{\gamma},t} = O_p\left(t^{-1/2}\right)$, and $E||\varepsilon_t \boldsymbol{\vartheta}_{\boldsymbol{\gamma},t}||^2 = O\left(t^{-1}\right) + o\left(t^{-2}\right)$,
ii) $||T^{-\alpha}\sum_{t=2}^{T}\varepsilon_t \boldsymbol{\vartheta}_{\boldsymbol{\gamma},t}|| = o_p(1)$ and $||T^{-\alpha}\sum_{t=2}^{T}\boldsymbol{\vartheta}_{\boldsymbol{\gamma},t}|| = o_p(1)$,
iii) $||T^{-\beta}\sum_{t=2}^{T}\left(\boldsymbol{\Omega}_t^{**} - \boldsymbol{\Omega}_t^{*}\right)|| = o_p(1)$,
iv) $||T^{-\beta}\sum_{t=2}^{T}\varepsilon_t^2\left(\boldsymbol{\Omega}_t^{**} - \boldsymbol{\Omega}_t^{*}\right)|| = o_p(1)$.

Proof of Lemma B.1.
For part *i*), let $\gamma \in [0, \pi]$ and denote $\vartheta_{\gamma,t} = \sum_{j=t}^{\infty}\omega_j(\gamma)\varepsilon_{t-j}$. Since $\omega_j(\gamma) = O(1/j)$, it follows that $E\left[(\vartheta_{\gamma,t})^2\right] = O\left(\sum_{j=t}^{\infty} 1/j^2\right) = O\left(t^{-1}\right)$ and, therefore, $\sqrt{t}\vartheta_{\gamma,t} = O_p(1)$. Hence, $\boldsymbol{\varepsilon}^{**}_{\boldsymbol{\gamma},t-1} - \boldsymbol{\varepsilon}^{*}_{\boldsymbol{\gamma},t-1} \equiv \boldsymbol{\vartheta}_{\boldsymbol{\gamma},t} = O_p\left(t^{-1/2}\right)$. Also, $E||\varepsilon_t\boldsymbol{\vartheta}_{\boldsymbol{\gamma},t}||^2 = \sum\limits_{s=1}^{n}\sum\limits_{j=t,l=t}^{\infty}\omega_j(\gamma_s)\,\omega_l(\gamma_s)\,E\left(\varepsilon_t^2\varepsilon_{t-j}\varepsilon_{t-l}\right)$, where, from stationarity, $E\left(\varepsilon_t^2\varepsilon_{t-j}\varepsilon_{t-l}\right) = \kappa_\varepsilon(0,j,l,0) + \sigma^4\mathbb{I}_{(j=l)}$ and, since $\kappa_\varepsilon(0,j,l,0) = o\left(\frac{1}{|j|\;|l|}\right)$ necessarily under the assumption of absolute summability, then

$$\begin{aligned}\sum_{j=t,l=t}^{\infty}\omega_j(\gamma_s)\,\omega_l(\gamma_s)\,E\left(\varepsilon_t^2\varepsilon_{t-j}\varepsilon_{t-l}\right) &= \sigma^4\sum_{j=t}^{\infty}\omega_j^2(\gamma_s) + o\left(\sum_{j=t,l=t}^{\infty}\frac{1}{j^2l^2}\right)\\ &= O\left(\sum_{j=t}^{\infty}1/j^2\right) + o\left(\sum_{j=t}^{\infty}\frac{1}{j^2}\right)o\left(\sum_{l=t}^{\infty}\frac{1}{l^2}\right)\end{aligned}$$

and therefore $E||\varepsilon_t\boldsymbol{\vartheta}_{\boldsymbol{\gamma},t}||^2 = O\left(t^{-1}\right) + o\left(t^{-2}\right)$ as required. Note that, under assumption $\mathcal{A}.2.1$ and $\kappa_\varepsilon(0,j,l,0) = 0$ the required result simplifies trivially to $E||\varepsilon_t\boldsymbol{\vartheta}_{\boldsymbol{\gamma},t}||^2 = O\left(t^{-1}\right)$. For part *ii*), since $E\left(\varepsilon_t\varepsilon_s\varepsilon_{t-j}\varepsilon_{s-l}\right) = 0$ for all $t \neq s$ owing to the MDS property of ε_t, we have

$$\begin{aligned}E\left\|\frac{1}{T^\alpha}\sum_{t=2}^{T}\varepsilon_t\boldsymbol{\vartheta}_{\boldsymbol{\gamma},t}\right\|^2 &\leq \frac{1}{T^{2\alpha}}\sum_{t=2}^{T}E||\varepsilon_t\boldsymbol{\vartheta}_{\boldsymbol{\gamma},t}||^2 + o(1)\\ &= \frac{1}{T^{2\alpha}}\left(\sum_{t=2}^{T}\left[O\left(t^{-1}\right) + o\left(t^{-2}\right)\right]\right) + o(1)\\ &= O\left(\frac{\log T}{T^{2\alpha}}\right) + o\left(T^{-2\alpha}\right) + o(1) = o(1)\end{aligned}$$

for any $\alpha > 0$ under Assumptions $\mathcal{A}.1$ and $\mathcal{A}.2$, by using (i). From Markov's inequality,

$$\left\|\frac{1}{T^\alpha}\sum_{t=2}^{T}\varepsilon_t\left(\boldsymbol{\varepsilon}^{**}_{\boldsymbol{\gamma},t-1} - \boldsymbol{\varepsilon}^{*}_{\boldsymbol{\gamma},t-1}\right)\right\| = O_p\left(\sqrt{\log T}/T^\alpha\right) = o_p(1).$$

Similarly,

$$\begin{aligned}E\left\|\frac{1}{T^\alpha}\sum_{t=2}^{T}\boldsymbol{\vartheta}_{\boldsymbol{\gamma},t}\right\|^2 &\leq \frac{1}{T^{2\alpha}}\sum_{s=1}^{n}\sum_{t=2}^{T}\sum_{j=t,l=t}^{\infty}\omega_j(\gamma_s)\,\omega_l(\gamma_s)\,E\left(\varepsilon_{t-j}\varepsilon_{t-l}\right) + o(1)\\ &= \sum_{s=1}^{n}\left(\frac{1}{T^{2\alpha}}\sum_{t=2}^{T}\sum_{j=t}^{\infty}\omega_j^2(\gamma_s)\,E\left(\varepsilon_{t-j}^2\right)\right) + o(1)\\ &= O\left(\frac{\log T}{T^{2\alpha}}\right).\end{aligned}$$

For part *iii*), first note that

$$\begin{aligned}
\boldsymbol{\Omega}_t^{**} - \boldsymbol{\Omega}_t^{*} &= \left(\sum_{j,l=t}^{\infty} \boldsymbol{\omega}_j \boldsymbol{\omega}_l' \varepsilon_{t-j}\varepsilon_{t-l}\right) + \left(\sum_{j=1}^{t-1}\sum_{l=t}^{\infty} \boldsymbol{\omega}_j \boldsymbol{\omega}_l' \varepsilon_{t-j}\varepsilon_{t-l}\right) + \left(\sum_{j=t}^{\infty}\sum_{l=1}^{t-1} \boldsymbol{\omega}_j \boldsymbol{\omega}_l' \varepsilon_{t-j}\varepsilon_{t-l}\right) \\
&= \mathbf{D}_{1t} + \mathbf{D}_{2t} + \mathbf{D}_{3t},
\end{aligned}$$

where these terms have been defined implicitly. For the first component, note that $\mathbf{D}_{1t} = \boldsymbol{\vartheta}_{\boldsymbol{\gamma},t}\boldsymbol{\vartheta}_{\boldsymbol{\gamma},t}'$. Then, from the triangle and Cauchy-Schwarz matrix inequalities and the MDS property of ε_t it follows that

$$\begin{aligned}
E\left\|\frac{1}{T^{\alpha}}\sum_{t=2}^{T}\boldsymbol{\vartheta}_{\boldsymbol{\gamma},t}\boldsymbol{\vartheta}_{\boldsymbol{\gamma},t}'\right\| &\leq \frac{1}{T^{\alpha}}\sum_{t=2}^{T} E\left\|\boldsymbol{\vartheta}_{\boldsymbol{\gamma},t}\boldsymbol{\vartheta}_{\boldsymbol{\gamma},t}'\right\| \leq \frac{1}{T^{\alpha}}\sum_{t=2}^{T} E\left\|\boldsymbol{\vartheta}_{\boldsymbol{\gamma},t}\right\|^2 \\
&\leq \sum_{i=1}^{n}\left(\frac{1}{T^{\alpha}}\sum_{t=2}^{T}\sum_{j=t}^{\infty}\omega_j^2\left(\gamma_i\right)E\left(\varepsilon_{t-j}^2\right)\right) + o\left(1\right) \\
&= O\left(\frac{\log T}{T^{\alpha}}\right),
\end{aligned}$$

and, hence, $||T^{-\alpha}\sum_{t=2}^{T}\mathbf{D}_{1t}|| = o_p\left(1\right)$ for any $\alpha > 0$. Similarly, $\mathbf{D}_{2t} = \sum_{j=1}^{t-1}\omega_j\varepsilon_{t-j}\left(\sum_{l=t}^{\infty}\boldsymbol{\omega}_l\varepsilon_{t-l}\right)' = \varepsilon_{\gamma,t-1}^{*}\boldsymbol{\vartheta}_{\gamma,t}'$. Therefore, for any $\beta > 1/2$, it follows by triangle and Cauchy-Schwarz inequalities joint with the properties of the matrix norm that

$$\begin{aligned}
E\left\|\frac{1}{T^{\beta}}\sum_{t=2}^{T}\mathbf{D}_{2t}\right\| &\leq \frac{1}{T^{\beta}}\sum_{t=2}^{T}E\left\|\varepsilon_{\boldsymbol{\gamma},t-1}^{*}\boldsymbol{\vartheta}_{\boldsymbol{\gamma},t}'\right\| \leq \frac{1}{T^{\beta}}\sum_{t=2}^{T}\sqrt{E\left\|\varepsilon_{\boldsymbol{\gamma},t-1}^{*}\right\|^2}\sqrt{E\left\|\boldsymbol{\vartheta}_{\boldsymbol{\gamma},t}\right\|^2} \\
&= O\left(\frac{T^{1/2}}{T^{\beta}}\right) = o_p\left(1\right)
\end{aligned}$$

because $E\left\|\varepsilon_{\boldsymbol{\gamma},t-1}^{*}\right\|^2 \leq E\left\|\varepsilon_{\boldsymbol{\gamma},t-1}^{**}\right\|^2 = O\left(1\right)$ and $E\left\|\boldsymbol{\vartheta}_{\boldsymbol{\gamma},t}\right\|^2 = O\left(1/t\right)$, as discussed in (i) above. Finally, $\mathbf{D}_{3t} = \left(\sum_{j=t}^{\infty}\boldsymbol{\omega}_j\varepsilon_{t-j}\right)\left(\sum_{l=1}^{t-1}\boldsymbol{\omega}_l\varepsilon_{t-l}\right)' = \mathbf{D}_{2t}' = \boldsymbol{\vartheta}_{\gamma,t}\varepsilon_{\boldsymbol{\gamma},t-1}^{\prime *}$, and consequently $\left\|T^{-\beta}\sum_{t=2}^{T}\mathbf{D}_{3t}\right\| = O_p\left(\frac{T^{1/2}}{T^{\beta}}\right)$, which renders the required result. For part *iv*), note that $\varepsilon_t^2\left(\boldsymbol{\Omega}_t^{**} - \boldsymbol{\Omega}_t^{*}\right) = \varepsilon_t^2\left(\mathbf{D}_{1t} + \mathbf{D}_{2t} + \mathbf{D}_{2t}'\right)$, and the required result then holds as in previous lemmae. First, $\varepsilon_t^2\mathbf{D}_{1t} = \left(\varepsilon_t\boldsymbol{\vartheta}_{\boldsymbol{\gamma},t}\right)\left(\boldsymbol{\vartheta}_{\boldsymbol{\gamma},t}\varepsilon_t\right)'$, and hence, by the triangle and Cauchy-Schwarz inequalities $E\left\|\frac{1}{T^{\alpha}}\sum_{t=2}^{T}\varepsilon_t^2\ \boldsymbol{\vartheta}_{\boldsymbol{\gamma},t}\boldsymbol{\vartheta}_{\boldsymbol{\gamma},t}'\right\| \leq \frac{1}{T}\sum_{t=2}^{T}E\left\|\varepsilon_t\boldsymbol{\vartheta}_{\boldsymbol{\gamma},t}\right\|^2 = o\left(1\right)$ for any $\alpha > 0$ from (i). Also, $\varepsilon_t^2\mathbf{D}_{2t} = \left(\varepsilon_t\varepsilon_{\boldsymbol{\gamma},t-1}^{*}\right)\left(\boldsymbol{\vartheta}_{\boldsymbol{\gamma},t}\varepsilon_t\right)'$, so for any $\beta > 1/2$ we have

$$\begin{aligned}
E\left\|\frac{1}{T^{\beta}}\sum_{t=2}^{T}\varepsilon_t^2\ \mathbf{D}_{2t}\right\| &\leq \frac{1}{T^{\beta}}\sum_{t=2}^{T}\sqrt{E\left\|\varepsilon_t\varepsilon_{\boldsymbol{\gamma},t-1}^{*}\right\|^2}\sqrt{E\left\|\varepsilon_t\boldsymbol{\vartheta}_{\boldsymbol{\gamma},t}\right\|^2} \\
&\leq \frac{1}{T^{\beta}}\sum_{t=2}^{T}\sqrt{E\left\|\varepsilon_t\varepsilon_{\boldsymbol{\gamma},t-1}^{**}\right\|^2}\sqrt{E\left\|\varepsilon_t\boldsymbol{\vartheta}_{\boldsymbol{\gamma},t}\right\|^2} \\
&= O\left(\frac{T^{1/2}}{T^{\beta}}\right) = o\left(1\right).
\end{aligned}$$

Since obviously $\left\|T^{-\beta}\sum_{t=2}^{T}\varepsilon_t^2\mathbf{D}_{2t}'\right\| = O_p\left(\frac{T^{1/2}}{T^{\beta}}\right) = o_p\left(1\right)$, this completes the proof.■

Lemma B.2. *Let* $\boldsymbol{\Lambda}_{\varepsilon\gamma} = E\left(\varepsilon_t^2 \boldsymbol{\varepsilon}^{**}_{\boldsymbol{\gamma},t-1}\boldsymbol{\varepsilon}'^{**}_{\boldsymbol{\gamma},t-1}\right)$. *Then, under the null hypothesis and Assumptions* $\mathcal{A}.1$ and $\mathcal{A}.2$, *as* $T \to \infty$, $\frac{1}{\sqrt{T}}\left(\sum_{t=2}^{T} \varepsilon_{\mathbf{d}t}\boldsymbol{\varepsilon}^{*}_{\boldsymbol{\gamma},t-1}\right) \Rightarrow \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Lambda}_{\varepsilon\gamma}\right)$, *with* $\boldsymbol{\Lambda}_{\varepsilon\gamma} = \sigma^4\boldsymbol{\Gamma}_{\boldsymbol{\gamma}} + \sum_{j,l\geq 1} \boldsymbol{\omega}_j\boldsymbol{\omega}_l' \, \kappa_\varepsilon\left(0,j,l,0\right)$, *and* $\boldsymbol{\varepsilon}^{*}_{\boldsymbol{\gamma},t-1}$ *generated from* $\varepsilon_{\mathbf{d}t} = \Delta_{\boldsymbol{\gamma}}\left(L;\mathbf{d}\right)x_t$.

Proof of Lemma B.2.
Under the null hypothesis, $\varepsilon_{\mathbf{d}t} = \varepsilon_t$, from which $\sum_{t=2}^{T} \varepsilon_{\mathbf{d}t}\boldsymbol{\varepsilon}^{**}_{\boldsymbol{\gamma},t-1} = \sum_{t=2}^{T}\left(\sum_{j=1}^{\infty}\boldsymbol{\omega}_j\varepsilon_{t-j}\varepsilon_t\right) = \sum_{t=2}^{T}\mathbf{Z}_t$, say, where $E\left(\mathbf{Z}_t|\mathcal{G}_{t-1}\right) = \mathbf{0}$, so $\{\mathbf{Z}_t, \mathcal{G}_t\}$ is a vector MDS with unconditional and conditional covariance matrices

$$
\begin{aligned}
E\left(\mathbf{Z}_t\mathbf{Z}_t'\right) &= \sum_{j=1}^{\infty}\sum_{l=1}^{\infty}\boldsymbol{\omega}_j\boldsymbol{\omega}_l' \, E\left(\varepsilon_t^2\varepsilon_{t-j}\varepsilon_{t-l}\right) \equiv \boldsymbol{\Lambda}_{\varepsilon\gamma}, \\
E\left(\mathbf{Z}_t\mathbf{Z}_t'|\mathcal{G}_{t-1}\right) &= \sum_{j,l=1}^{\infty}\boldsymbol{\omega}_j\boldsymbol{\omega}_l'\varepsilon_{t-j}\varepsilon_{t-l}E\left(\varepsilon_t^2|\mathcal{G}_{t-1}\right).
\end{aligned}
$$

It is interesting to briefly comment the conditions upon which $\boldsymbol{\Lambda}_{\varepsilon\gamma}$ is well-defined. Owing to stationarity, $E\left(\varepsilon_t^2\varepsilon_{t-j}\varepsilon_{t-l}\right) = \kappa_\varepsilon\left(0,j,l,0\right) + \sigma^4\mathbb{I}_{(j=l)}$, and thus $\boldsymbol{\Lambda}_{\varepsilon\gamma} = \sigma^4\boldsymbol{\Gamma}_{\boldsymbol{\gamma}} + \sum_{j,l\geq 1}\boldsymbol{\omega}_j\boldsymbol{\omega}_l' \, \kappa_\varepsilon\left(0,j,l,0\right)$. The first component is bounded and positive definite, as discussed in Appendix A. Since $\boldsymbol{\omega}_j$ is not absolute summable, the second component requires additional summability conditions making $\kappa_\varepsilon\left(0,j,l,0\right)$ negligible as $j,l \to \infty$. Under i.i.d errors, $\kappa_\varepsilon\left(0,j,l,0\right) = 0$ for all l,j, and hence $\boldsymbol{\Lambda}_{\varepsilon\gamma} = \sigma^4\boldsymbol{\Gamma}_{\boldsymbol{\gamma}}$ is bounded and bounded away from zero. Under the more general MDS assumption, the absolute summability of the fourth-order cumulants ensures $\boldsymbol{\Lambda}_{\varepsilon\gamma} < \infty$, and as a result the asymptotic covariance matrix is characterized by the pattern of conditional heteroskedasticity. Since $\boldsymbol{\Lambda}_{\varepsilon\gamma} - \sigma^4\boldsymbol{\Gamma}_{\boldsymbol{\gamma}}$ is obviously semipositive definite, $\boldsymbol{\Lambda}_{\varepsilon\gamma}$ is bounded and bounded away from zero.

We now prove the required result by using the central limit theory for vector MDS. For any $\boldsymbol{\lambda} \in \mathbb{R}^n$ such that $\boldsymbol{\lambda}'\boldsymbol{\lambda} = 1$ define $z_t = \boldsymbol{\lambda}'\mathbf{Z}_t$. Then, we require $(\mathcal{C}1)$ $T^{-1}\sum_{t=2}^{T} z_t^2 - E\left(z_t^2\right) \xrightarrow{p} 0$, and $(\mathcal{C}2)$ $\max_{2\leq t\leq T}|z_t/\sqrt{T}| \xrightarrow{p} 0$, (cf. Davidson, 1994, Thm 24.3). Note that $T^{-1}\sum_{t=2}^{T}\left(z_t^2 - E\left(z_t^2\right)\right) = \boldsymbol{\lambda}'\mathbf{S}_T\boldsymbol{\lambda}$, where $\mathbf{S}_T = T^{-1}\sum_{t=2}^{T}\left(\mathbf{Z}_t\mathbf{Z}_t' - \boldsymbol{\Lambda}_{\varepsilon\gamma}\right)$ owing to the MDS property of $\mathbf{Z}_t$, and then $(\mathcal{C}1)$ is verified if $\mathbf{S}_T = o_p\left(1\right)$ by Slutsky's theorem. It is worth noting that $\sum_{l=0}^{\infty}|\omega_l\left(\gamma_i\right)\omega_l\left(\gamma_j\right)| < \infty$ for any $\gamma_i, \gamma_j \in [0,\pi]$ by Cauchy-Schwarz inequality, so $\boldsymbol{\varepsilon}^{**}_{\boldsymbol{\gamma},\mathbf{t-1}}$ is defined through a $\mathcal{G}_t$-measurable transformation of a strictly stationary and ergodic process under Assumptions $\mathcal{A}.1$ and $\mathcal{A}.2$. Therefore, $\mathbf{Z}_t$ is a strictly stationary and ergodic MDS (cf. White, 2001, Thm. 3.35) and so is z_t.

Under Assumption $\mathcal{A}.2.1$, $T^{-1}\sum_{t=2}^{T}\left[E\left(\mathbf{Z}_t\mathbf{Z}_t'\right) - E\left(\mathbf{Z}_t\mathbf{Z}_t'|\mathcal{G}_{t-1}\right)\right] \xrightarrow{p} 0$, because $\{\mathbf{Z}_t, \mathcal{G}_t\}$ is a stationary and ergodic MDS. Furthermore, since $E\left(|\varepsilon_t|^4\right) < K < \infty$ for all t, and $E\left(\mathbf{Z}_t\mathbf{Z}_t'\right) = \sigma^4\boldsymbol{\Gamma}_{\boldsymbol{\gamma}}$, then $T^{-1}\sum_{t=2}^{T}E\left(\mathbf{Z}_t\mathbf{Z}_t'\right) \xrightarrow{p} \sigma^4\boldsymbol{\Gamma}_{\boldsymbol{\gamma}}$ from stationarity. Alternatively, under Assumption $\mathcal{A}.2.2$, for any $\gamma_i, \gamma_j \in [0,\pi]$, and the set of indices $l_h \geq 1$, $h = 1, ..., 4$, define $\varsigma_{ij}\left(l_1,l_2,l_3,l_4\right) = \omega_{l_1}\left(\gamma_i\right)\omega_{l_2}\left(\gamma_i\right)\omega_{l_3}\left(\gamma_j\right)\omega_{l_4}\left(\gamma_j\right)$ and let $E||\mathbf{S}_T - \boldsymbol{\Lambda}_{\varepsilon,\gamma}||^2 = \sum_{i,j=1}^{n}\mathcal{E}_{ij,T}$, whose characteristic element is given by

$$
\begin{aligned}
\mathcal{E}_{ij,T} &= E\left(\frac{1}{T}\sum_{t=2}^{T}\varepsilon_t^2\varepsilon^{**}_{\gamma_i,t-1}\varepsilon^{**}_{\gamma_j,t-1} - \left[\boldsymbol{\Lambda}_{\varepsilon\gamma}\right]_{ij}\right)^2 \\
&= T^{-1}\sum_{l_1,...,l_4=1}^{\infty}\varsigma_{ij}\left(l_1,...,l_4\right)\left\{T^{-1}\sum_{t=2}^{T}\sum_{s=2}^{T}Cov\left(\varepsilon_{t-l_1}\varepsilon_{t-l_2}\varepsilon_t^2,\varepsilon_{s-l_3}\varepsilon_{s-l_4}\varepsilon_s^2\right)\right\} + o\left(1\right).
\end{aligned}
$$

The covariances on the right-hand side do not depend on any of the elements of $\boldsymbol{\gamma}$. Furthermore, under the assumption of stationarity, they can be written as the sum of products of cumulants of ε_t of order eight and lower (cf. Brillinger, 1981, Thm. 2.3.2), which eventually rule the asymptotic behavior of $\mathcal{E}_{ij,T}$. First, we examine the case $i = j = 1$, for which we can assume $\gamma_1 = 0$ with no loss of generality for the discussion that follows. Under the restriction of absolute summability, $T|\mathcal{E}_{11,T}|$ is uniformly bounded by

$$\sum_{\tau=-\infty}^{\infty} \sum_{l_1,...,l_4=1}^{\infty} |\varsigma_{11}(l_1, ..., l_4)| \; |\kappa_\varepsilon(0, l_1 - l_4, l_1, l_1, \tau - l_3 + l_1, \tau - l_4 + l_1, \tau + l_1, \tau + l_1)|$$

with $\tau \equiv t - s$; see Gonçalves and Kilian (2007) and Proposition 2 in Demetrescu *et al.* (2008). By Lemma 10 in the latter paper, and noting that $\varsigma_{11}(l_1, ..., l_4) = O\left(\frac{1}{l_1 \times ... \times l_4}\right)$, this term can be shown to be uniformly bounded as well. Then, for the generic term $\mathcal{E}_{ij,T}$, $i, j \geq 1$, and noting that $|\omega_j(\gamma)|$ is uniformly bounded in $[0, \pi]$ by $2/j$, it follows for any pair $\gamma_i, \gamma_j \in [0, \pi]$ that $|\varsigma_{ij}(l_1, ..., l_4)| \leq \prod_{h=1}^{4} |2l_h^{-1}| \leq 8|\varsigma_{11}(l_1, ..., l_4)|$, from which obviously $T\;|\mathcal{E}_{ij,T}| \leq 8T\;|\mathcal{E}_{11,T}| < K < \infty$, independently of T or the particular frequencies involved. Consequently, $E||\mathbf{S}_T - \mathbf{\Lambda}_{\varepsilon,\gamma}||^2 = O(T^{-1}) = o(1)$ and $T^{-1}\sum_{t=2}^{T} \varepsilon_t^2 \boldsymbol{\varepsilon}_{\boldsymbol{\gamma},\mathbf{t-1}}^{**} \boldsymbol{\varepsilon}'^{**}_{\boldsymbol{\gamma},\mathbf{t-1}} \overset{ms}{\to} \mathbf{\Lambda}_{\varepsilon\gamma}$. Since mean-square convergence implies convergence in probability, $(\mathcal{C}1)$ holds under Assumptions $\mathcal{A}.1$ and $\mathcal{A}.2$ as required. At this point it is worth recalling that $\left\|T^{-1}\sum_{t=2}^{T} \varepsilon_t^2 \left(\boldsymbol{\varepsilon}_{\boldsymbol{\gamma},t-1}^{**} - \boldsymbol{\varepsilon}_{\boldsymbol{\gamma},t-1}^{*}\right)\right\| = o_p(1)$ from Lemma B.1 *iii*), so it follows by the Asymptotic Equivalence Lemma [AEL] (cf. White, 2001, Lemma 4.7) and under the null hypothesis that $T^{-1}\sum_{t=2}^{T} \varepsilon_{\mathbf{d}t}^2 \boldsymbol{\varepsilon}_{\boldsymbol{\gamma},\mathbf{t-1}}^{*} \boldsymbol{\varepsilon}'^{*}_{\boldsymbol{\gamma},\mathbf{t-1}} \overset{p}{\to} \mathbf{\Lambda}_{\varepsilon\gamma}$.

To address $(\mathcal{C}2)$ recall that under Assumptions $\mathcal{A}.1$ and $\mathcal{A}.2$, $\{\mathbf{Z}_t, z_t\}$ is strictly stationary and ergodic, and uniformly bounded and bounded away from zero under the L_2-norms, so the Lindeberg condition in $(\mathcal{C}2)$ is trivially satisfied (cf. Davidson, 2000, Thm. 6.2.3). Therefore, the Central Limit Theorem (CLT) for MDS jointly with the Cramér-Wold device (cf. Davidson, 1994, Thm. 25.6) allows us to conclude under the null hypothesis and as $T \to \infty$ that $T^{-1/2}\sum_{t=2}^{T} \varepsilon_t^2 \boldsymbol{\varepsilon}_{\boldsymbol{\gamma},t-1}^{**} \Rightarrow \mathcal{N}(\mathbf{0}, \mathbf{\Lambda}_{\varepsilon\boldsymbol{\gamma}})$. To complete the proof, recall from Lemma B.1 *ii*) that $\left\|T^{-1/2}\sum_{t=2}^{T} \varepsilon_t^2 \boldsymbol{\vartheta}_{\boldsymbol{\gamma},t} \boldsymbol{\vartheta}'_{\boldsymbol{\gamma},t}\right\| = o_p(1)$, so by the AEL it follows that, $T^{-1/2}\sum_{t=2}^{T} \varepsilon_t^2 \boldsymbol{\varepsilon}_{\boldsymbol{\gamma},t-1}^{*} \Rightarrow \mathcal{N}(\mathbf{0}, \mathbf{\Lambda}_{\varepsilon\boldsymbol{\gamma}})$ as required.■

Lemma B.3. *Define the k-th order autocovariance* $E\left(\boldsymbol{\varepsilon}_{\boldsymbol{\gamma},t-1}^{**} \boldsymbol{\varepsilon}'^{**}_{\boldsymbol{\gamma},t-1-k}\right) = \mathbf{\Lambda}_{\varepsilon\gamma}(k)$, $k > 0$, *and let* $\widehat{e}_t$ *be the estimated residuals from an auxiliary regression as in (4) with no augmentation. Then, under Assumptions* $\mathcal{A}.1$ *and* $\mathcal{A}.2$, *the null hypothesis, and as* $T \to \infty$ *it follows that:*
i) $\sum_{k=0}^{\infty} \mathbf{\Lambda}_{\varepsilon\gamma}^{\tau}(k) < \infty$, *for* $\tau \geq 1$;
ii) $T^{-1}\sum_{t=2}^{T} \boldsymbol{\varepsilon}_{\boldsymbol{\gamma},t-1}^{*} \boldsymbol{\varepsilon}'^{*}_{\boldsymbol{\gamma},t-1} \overset{p}{\to} \sigma^2 \mathbf{\Gamma}_{\gamma}$;
iii) $T^{-1}\sum_{t=2}^{T} \widehat{e}_t^2 \boldsymbol{\varepsilon}_{\boldsymbol{\gamma},t-1}^{*} \boldsymbol{\varepsilon}'^{*}_{\boldsymbol{\gamma},t-1} \overset{p}{\to} \mathbf{\Lambda}_{\varepsilon\gamma}$, *with* $\mathbf{\Lambda}_{\varepsilon\gamma} \equiv E\left(\varepsilon_t^2 \boldsymbol{\varepsilon}_{\boldsymbol{\gamma},t-1}^{**} \boldsymbol{\varepsilon}'^{**}_{\boldsymbol{\gamma},t-1}\right)$.

Proof of Lemma B.3.
In *i*), the asymptotic k-th order autocovariance matrix, $k \geq 0$, is given by

$$E\left(\boldsymbol{\varepsilon}_{\boldsymbol{\gamma},t-1}^{**} \boldsymbol{\varepsilon}'^{**}_{\boldsymbol{\gamma},t-1-k},\right) = \sum_{j,l=1}^{\infty} \boldsymbol{\omega}_j \boldsymbol{\omega}'_l E\left(\varepsilon_{t-j}\varepsilon_{t-k-l}\right) = \sigma^2 \sum_{j=1}^{\infty} \boldsymbol{\omega}_j \boldsymbol{\omega}'_{j+k} \equiv \mathbf{\Lambda}_{\varepsilon\gamma}(k) < \infty$$

with $\mathbf{\Lambda}_{\varepsilon\gamma}(0) = \mathbf{\Lambda}_{\varepsilon\boldsymbol{\gamma}}$. More specifically,

$$\mathbf{\Lambda}_{\varepsilon\gamma}(k) = o\left(\sum_{j=1}^{\infty} \frac{1}{j(j+k)}\right) = o\left(\frac{1}{k}\left(\sum_{j=1}^{\infty} \frac{1}{j} - \frac{1}{j+k}\right)\right) = o\left(\frac{\log k}{k}\right),$$

and, as a result, $\left\{\mathbf{\Lambda}^{\tau}_{\varepsilon\gamma}(k)\right\}_{k=0}^{\infty}$ is summable for any $\tau \geq 1$. For part *ii*), let again $\mathbf{\Omega}_t^{**} = \boldsymbol{\varepsilon}^{**}_{\gamma,t-1}\boldsymbol{\varepsilon}'^{**}_{\gamma,t-1}$ and $\mathbf{\Omega}^{*}_t = \boldsymbol{\varepsilon}^{*}_{\gamma,t-1}\boldsymbol{\varepsilon}'^{*}_{\gamma,t-1}$, with $\overline{\mathbf{\Omega}}_T^{**}$ and $\overline{\mathbf{\Omega}}_T^{*}$ being their respective sample means. Clearly, $E\left(\overline{\mathbf{\Omega}}_T^{**}\right) = \sigma^2\mathbf{\Gamma}_{\boldsymbol{\gamma}}$, whereas $\overline{\mathbf{\Omega}}_T^{*}$ is asymptotically unbiased, since

$$
\begin{aligned}
E\left(\overline{\mathbf{\Omega}}_T^{*}\right) &= \sigma^2\sum_{j=1}^{T}\boldsymbol{\omega}_j\boldsymbol{\omega}_j' - \sigma^2T^{-1}\sum_{j=2}^{T}j\left[\boldsymbol{\omega}_j\boldsymbol{\omega}_j'\right] + \sigma^2T^{-1}\sum_{j=2}^{T}\boldsymbol{\omega}_j\boldsymbol{\omega}_j' \\
&= \sigma^2\sum_{j=1}^{T}\boldsymbol{\omega}_j\boldsymbol{\omega}_j' - o(1) + O\left(T^{-1}\right) \to \sigma^2\mathbf{\Gamma}_{\boldsymbol{\gamma}}.
\end{aligned}
$$

We can show that $\overline{\mathbf{\Omega}}_T^{**} \xrightarrow{ms} \sigma^2\mathbf{\Gamma}_{\boldsymbol{\gamma}}$ using a similar approach as in Lemma B.2., from which $\overline{\mathbf{\Omega}}_T^{*} \xrightarrow{p} \sigma^2\mathbf{\Gamma}_{\boldsymbol{\gamma}}$ by Lemma B.1 *iii*) and the AEL. In particular, note that we can write

$$
\begin{aligned}
TE\left(\left[\frac{1}{T}\sum_{t=2}^{T}\mathbf{\Omega}_t^{**} - \sigma^2\mathbf{\Gamma}_{\gamma}\right]_{ij}\right)^2 &= \sum_{l_1,\ldots,l_4=1}^{\infty}\varsigma_{ij}\left(l_1,\ldots,l_4\right) \\
&\quad\times\frac{1}{T}\sum_{t=2}^{T}\sum_{s=2}^{T}Cov\left(\left[\varepsilon_{t-l_1}\varepsilon_{t-l_2}\right],\left[\varepsilon_{s-l_3}\varepsilon_{s-l_4}\right]\right) + o(1).
\end{aligned}
$$

Following Lemma A.2 in Gonçalves and Kilian (2004) and Lemma 8 in Demetrescu *et al.* (2008), this term is uniformly bounded by $\mathfrak{B}_{ij} + 2\sum_{k=-\infty}^{\infty}\left[\mathbf{\Lambda}^2_{\varepsilon\gamma}(k)\right]_{ij}$, with $\mathbf{\Lambda}_{\varepsilon\gamma}(k)$ defined in (i) and

$$
\mathfrak{B}_{ij} = \sum_{t=-\infty}^{\infty}\sum_{l_1,\ldots,l_4=0}^{\infty}\left|\varsigma_{ij}\left(l_1,\ldots,l_4\right)\right|\,\left|\kappa_{\varepsilon}\left(0,l_2-l_1,t+l_3-l_1,t+l_4-l_1\right)\right|.
$$

By considering again $\gamma_1 = 0$ with no loss of generality, we note that $\left|\varsigma_{ij}\left(l_1,\ldots,l_4\right)\right| \leq 8\left|\varsigma_{11}\left(l_1,\ldots,l_4\right)\right|$ and, since $\sum_{k=-\infty}^{\infty}\mathbf{\Lambda}^2_{\varepsilon\gamma}(k) < \infty$ from stationarity and from (i)of this Lemma, then for any pair $\gamma_i,\gamma_j \in [0,\pi]$, $\mathfrak{B}_{ij} + 2\sum_{k=-\infty}^{\infty}\left[\mathbf{\Lambda}^2_{\varepsilon\gamma}(k)\right]_{ij} < \infty$ as a corollary of Lemma 8 in Demetrescu *et al.* (2008). Hence, $E||\overline{\mathbf{\Omega}}_T^{**} - \sigma^2\mathbf{\Gamma}_{\boldsymbol{\gamma}}||^2 = O\left(T^{-1}\right)$ and therefore $\overline{\mathbf{\Omega}}_T^{**} \xrightarrow{ms} \sigma^2\mathbf{\Gamma}_{\boldsymbol{\gamma}}$. But since from Lemma B.1 *iv*) $||T^{-1}\sum_{t=2}^{T}\left(\mathbf{\Omega}_t^{**} - \mathbf{\Omega}_t^{*}\right)|| = o_p(1)$, the AEL allows us to conclude that $T^{-1}\sum_{t=2}^{T}\boldsymbol{\varepsilon}^{*}_{\gamma,t-1}\boldsymbol{\varepsilon}'^{*}_{\gamma,t-1} \xrightarrow{p} \sigma^2\mathbf{\Gamma}_{\boldsymbol{\gamma}}$, as required, with convergence in probability being implied by the stronger convergence in the mean square sense.
In *iii*), the null hypothesis implies $\boldsymbol{\phi} = \mathbf{0}$ and $e_t = \varepsilon_t$ in the auxiliary regression, thereby $\widehat{e}_t^2 - \varepsilon_t^2 = \left(\sum_{s=1}^{n}\phi_{s,T}\varepsilon^{**}_{\gamma_s,t-1}\right)^2 = \left(\boldsymbol{\varepsilon}'^{**}_{\gamma,t-1}\boldsymbol{\phi}_T\right)\left(\boldsymbol{\phi}_T'\boldsymbol{\varepsilon}^{**}_{\gamma,t-1}\right)$. Hence,

$$
\begin{aligned}
\left\|\frac{1}{T}\sum_{t=2}^{T}\boldsymbol{\varepsilon}^{**}_{\gamma,t-1}\left(\widehat{e}_t^2 - \varepsilon_t^2\right)\boldsymbol{\varepsilon}'^{**}_{\gamma,t-1}\right\| &= \left\|\frac{1}{T}\sum_{t=2}^{T}\mathbf{\Omega}_t^{**}\boldsymbol{\phi}_T\boldsymbol{\phi}_T'\mathbf{\Omega}_t^{**}\right\| \leq \frac{1}{T}\sum_{t=2}^{T}\|\mathbf{\Omega}_t^{**}\boldsymbol{\phi}_T\boldsymbol{\phi}_T'\mathbf{\Omega}_t^{**}\| \\
&\leq \frac{1}{T}\sum_{t=2}^{T}||\mathbf{\Omega}_t^{**}||\;||\boldsymbol{\phi}_T\boldsymbol{\phi}_T'||\;||\mathbf{\Omega}_t^{**}||
\end{aligned}
$$

by the triangle inequality first and finally by the Cauchy-Schwarz inequality. The estimated parameter vector $\boldsymbol{\phi}_T$ is $\sqrt{T}$-consistent (see proof of Theorem 2.2 below), so $\|\boldsymbol{\phi}_T\boldsymbol{\phi}_T'\| = O_p\left(T^{-1}\right)$. Since from Assumptions $\mathcal{A}.1$ and $\mathcal{A}.2$

$$
E\left\|\mathbf{\Omega}_t^{**}\right\|^2 \leq \sum_{i,j=1}^{n}\sum_{l_1,\ldots,l_4=1}^{\infty}\left|\varsigma_{ij}\left(l_1,\ldots,l_4\right)\right|\left|E\left(\varepsilon_{t-l_1}\varepsilon_{t-l_2}\varepsilon_{t-l_3}\varepsilon_{t-l_4}\right)\right|
$$

is uniformly bounded, it follows that

$$\left\| \frac{1}{T} \sum_{t=2}^{T} \left(\boldsymbol{\varepsilon}_{\boldsymbol{\gamma},t-1}^{**} \boldsymbol{\varepsilon}_{\boldsymbol{\gamma},t-1}^{\prime **} \right) \left(\widehat{e}_t^2 - \varepsilon_t^2 \right) \right\| = O_p \left(\frac{1}{T} \sum_{t=2}^{T} O \left(T^{-1} \right) \right) = O_p \left(T^{-1} \right)$$

as $T \rightarrow \infty$. Finally, as for Lemma B.1 *iv*), we can readily show that $\left\| T^{-1} \sum_{t=2}^{T} \left(\boldsymbol{\vartheta}_{\boldsymbol{\gamma},t-1} \boldsymbol{\vartheta}_{\boldsymbol{\gamma},t-1}^{\prime} \right) \left(\widehat{e}_t^2 - \varepsilon_t^2 \right) \right\| = o_p(1)$, so the AEL renders the required result. ■

Proof of Theorem 2.1.
The proof of Theorem 2.1 is now obvious in view of the results in Lemmas B.1-B.3, and holds straightforwardly by the CMT. In particular,

$$\begin{aligned} LM_T &= \left(\frac{1}{\sqrt{T}} \sum_{t=2}^{T} \varepsilon_{\mathbf{d}t} \boldsymbol{\varepsilon}_{\boldsymbol{\gamma},t-1}^{*} \right)^{\prime} \left[\frac{1}{T} \sum_{t=2}^{T} \varepsilon_{\mathbf{d}t}^2 \boldsymbol{\varepsilon}_{\boldsymbol{\gamma},t-1}^{*} \boldsymbol{\varepsilon}_{\boldsymbol{\gamma},t-1}^{\prime *} \right]^{-1} \left(\frac{1}{\sqrt{T}} \sum_{t=2}^{T} \varepsilon_{\mathbf{d}t} \boldsymbol{\varepsilon}_{\boldsymbol{\gamma},t-1}^{*} \right) \\ &= \mathbf{A}_T^{\prime} \left[\mathbf{B}_T^{-1} \right] \mathbf{A}_T, \text{ say.} \end{aligned}$$

Under the null hypothesis, $\varepsilon_{\mathbf{d}t} = \varepsilon_t$, so under Assumptions $\mathcal{A}.1$ and $\mathcal{A}.2$, as $T \rightarrow \infty$, $\mathbf{A}_T \Rightarrow \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Lambda}_{\varepsilon\boldsymbol{\gamma}}\right)$ and $\mathbf{B}_T \xrightarrow{p} \boldsymbol{\Lambda}_{\varepsilon\boldsymbol{\gamma}}$ according to Lemma B.1 *i*) and *iv*), Lemmas B.2, B.3 and the AEL. The required convergence then follows by the CMT from which $LM_T \Rightarrow \mathbf{N}_n^{\prime}\mathbf{N}_n$, where $\mathbf{N}_n$ is a n-dimensional standard normal distribution and, hence, $LM_T \Rightarrow \chi_{(n)}^2$. ■

Corollaries.

For proof of Corollary 2.1, notice that the score of the log-likelihood function when $\boldsymbol{\theta} = \theta \mathbf{1}_n$, with $\mathbf{1}_n$ being a vector of ones in $\mathbb{R}^n$, is given by

$$\begin{aligned} \left. \frac{\partial \mathcal{L}(\boldsymbol{\delta}, \sigma^2 | \mathbf{x}_T)}{\partial \theta} \right|_{\mathrm{H}_0 : \theta = 0} &= -\frac{1}{\sigma^2} \sum_{t=1}^{T} \varepsilon_t \left(\log\left[1 - L\right] + \sum_{i=2}^{n-1} \log\left[\xi_{\gamma_i}\left(L; 1\right) \right] + \log\left[1 + L\right] \right) \varepsilon_t \\ &= \frac{1}{\sigma^2} \sum_{t=1}^{T} \varepsilon_t \sum_{s=1}^{n} \left(\sum_{j=1}^{\infty} \omega_j\left(\gamma_s\right) \varepsilon_{t-j} \right) \equiv \frac{1}{\sigma^2} \sum_{t=1}^{T} \varepsilon_t \left(\sum_{s=1}^{n} \varepsilon_{\gamma_s,t-1}^{**} \right) \end{aligned}$$

which suggests that $\mathrm{H}_0 : \boldsymbol{\theta} = \mathbf{0}$ can be tested by analyzing the statistical significance of the $\bar{\phi}$ parameter in the auxiliary regression $\varepsilon_{\mathbf{d}t} = \bar{\phi} \left(\sum_{s=1}^{n} \varepsilon_{\gamma_s,t-1}^{*} \right) + u_t$. Since $\sum_{s=1}^{n} \varepsilon_{\gamma_s,t-1}^{*} = \mathbf{1}_n^{\prime} \boldsymbol{\varepsilon}_{\boldsymbol{\gamma},t-1}^{*}$ is a linear transformation of the regressors in the basic auxiliary regression, we have that $\bar{\phi}_T = \left(\mathbf{1}_n^{\prime} \bar{\boldsymbol{\Omega}}_T^{*} \mathbf{1}_n \right)^{-1} \left(\mathbf{1}_n^{\prime} \left[\varepsilon_t \boldsymbol{\varepsilon}_{\boldsymbol{\gamma},t-1}^{*} \right] \right)$ and, hence, it follows from Theorem 2.1 and the CMT that $\sqrt{T}\bar{\phi}_T \Rightarrow \mathcal{N}\left(\mathbf{0}, \mathbf{1}_n^{\prime} \mathbf{V}_{\boldsymbol{\gamma}} \mathbf{1}_n\right)$ as $T \rightarrow \infty$. ■

Corollary 2.2 holds from the asymptotic normality in Theorem 2.2 owing to the fact that $\boldsymbol{\Lambda}_p$ is proportional to $\boldsymbol{\Omega}_p^{**}$ under the restrictions considered; see Theorems 4.32 and 4.37, and comments in White (2001).

Lemma B.4. *Let* $\{b_j\}_{j \geq 0}$ *be the coefficients in the Wold representation,* $\varepsilon_t = \sum_{j=0}^{\infty} b_j v_{t-j}$ *under Assumption* $\mathcal{A}.2^{\prime}$*. Let* $\varphi_j(\gamma)$ *be the j-th element in the serial convolution of* $\{\omega_{j+1}(\gamma)\}_{j \geq 0}$ *and* $\{b_j\}_{j \geq 0}$ *for any* $\gamma \in [0, \pi]$*. Then,* $\varphi_j(\gamma) = \omega_1(\gamma)$*, if j=0, and* $\varphi_j(\gamma)$ *is* $O(\omega_j(\gamma))$ *otherwise.*

Proof of Lemma B.4.
Recall that, for all $\gamma \in [0, \pi]$, $|\omega_j(\gamma)| \leq 2/j$, and hence $\omega_j(\gamma) = O(1/j)$. The serial convolution of $\{b_j\}_{j\geq 0}$ and $\{\omega_{j+1}\}_{j\geq 0}$ determines coefficients as a function of the γ frequency which are given by $\varphi_j(\gamma) = \sum_{k=0}^{j} b_k \omega_{j-k+1}(\gamma)$, where $\varphi_j(\gamma) \leq |\varphi_j(\gamma)| \leq 2\sum_{k=0}^{j} \frac{j}{j-k+1}|b_k|$, with $\left(\frac{j}{j-k+1}\right) \leq k$ for all $1 \leq k \leq j$, so $\varphi_j(\gamma) \leq |b_0| \left(\frac{2j}{j+1}\right) + 2\sum_{j=1}^{k} j|b_k|$. Since for any stationary AR(p) model $\sum_{j=1}^{k} j|b_k| < \infty$, the coefficient $|\varphi_j(\gamma)|$ is bounded by a constant as $j \to \infty$, and hence $\varphi_j(\gamma) = O(1/j)$, which leads us to the desired result.

As a result, $\{\varphi_j(\gamma)\}$ belongs to the same space of square-summable coefficient series as $\{\omega_j(\gamma)\}$ does, so the results discussed under MDS errors follow under Assumption $\mathcal{A}.2'$ in most cases by simply modifying the limit variances. Also, note that since γ is taken from $[0, \pi]$, this lemma trivially generalizes the results in Demetrescu *et al.* (2008), discussed for $\gamma = 0$, to any other frequency. ■

Lemma B.5. *Considering Assumption $\mathcal{A}.2'$, the asymptotic and observable processes under the null hypothesis are now given by $\boldsymbol{\varepsilon}_{\boldsymbol{\gamma},t-1}^{**} = \sum_{j=0}^{\infty} \boldsymbol{\varphi}_j v_{t-j-1}$, $\boldsymbol{\varepsilon}_{\boldsymbol{\gamma},t-1}^{*} = \sum_{j=0}^{t-1} \boldsymbol{\varphi}_j v_{t-j-1}$, respectively, with $\boldsymbol{\varphi}_j \equiv (\varphi_j(\gamma_1), ..., \varphi_j(\gamma_n))'$, and $\{\varphi_j(\cdot)\}_{j\geq 0}$ given in Lemma B.4. Then, as T is allowed to diverge, Lemma B.1 still holds under Assumption $\mathcal{A}.2'$ with trivial modifications, i.e.,: i)* $\boldsymbol{\vartheta}_{\boldsymbol{\gamma},t} = O_p(t^{-1/2})$ and $E||v_t \boldsymbol{\vartheta}_{\boldsymbol{\gamma},t}|| = O(t^{-1}) + o(t^{-2})$, *ii)* $||T^{-\alpha} \sum_{t=p+1}^{T} v_t \boldsymbol{\vartheta}_{\boldsymbol{\gamma},t}|| = o_p(1)$ and $||T^{-\alpha} \sum_{t=p+1}^{T} \boldsymbol{\vartheta}_{\boldsymbol{\gamma},t}|| = o_p(1)$, *iii)* $||T^{-\beta} \sum_{t=p+1}^{T} \left(\boldsymbol{\varepsilon}_{\boldsymbol{\gamma},t-1}^{**} \boldsymbol{\varepsilon}_{\boldsymbol{\gamma},t-1}^{\prime **} - \boldsymbol{\varepsilon}_{\boldsymbol{\gamma},t-1}^{*} \boldsymbol{\varepsilon}_{\boldsymbol{\gamma},t-1}^{\prime *}\right)|| = o_p(1)$, *iv)* $||T^{-\beta} \sum_{t=p+1}^{T} v_t^2 \left(\boldsymbol{\varepsilon}_{\boldsymbol{\gamma},t-1}^{**} \boldsymbol{\varepsilon}_{\boldsymbol{\gamma},t-1}^{\prime **} - \boldsymbol{\varepsilon}_{\boldsymbol{\gamma},t-1}^{*} \boldsymbol{\varepsilon}_{\boldsymbol{\gamma},t-1}^{\prime *}\right)|| = o_p(1)$, *for any* $\alpha > 0$, $\beta > 1/2$.

Proof of Lemma B.5. This Lemma holds directly from Lemma B.1 and Lemma B.4.

Lemma B.6. *Let* $\mathbf{X}_{tp} = (\varepsilon_{\mathbf{d},t-1}, ..., \varepsilon_{\mathbf{d},t-p})'$ *be the p-dimensional vector of lagged values of the dependent variable, and define the $n+p$ dimensional vectors* $\mathbf{X}_{tp}^{*} = \left(\boldsymbol{\varepsilon}_{\boldsymbol{\gamma},t-1}^{*\prime}, \mathbf{X}_{tp}'\right)'$, $\mathbf{X}_{tp}^{**} = \left(\boldsymbol{\varepsilon}_{\boldsymbol{\gamma},t-1}^{**\prime}, \mathbf{X}_{tp}'\right)'$. *Define* $\boldsymbol{\Omega}_p^{**} = E\left(\mathbf{X}_{tp}^{**} \mathbf{X}_{tp}^{\prime **}\right)$, *and let* $\overline{\boldsymbol{\Omega}}_p^{*} = T^{-1} \sum_{t=2}^{T} \mathbf{X}_{tp}^{*} \mathbf{X}_{tp}^{\prime *}$. *Then, i)* $\boldsymbol{\Omega}_p^{**}$ *is bounded and bounded away from zero, and ii)* $||\overline{\boldsymbol{\Omega}}_p^{*} - \boldsymbol{\Omega}_p^{**}|| = o_p(1)$.

Proof of Lemma B.6.
For part i), note that $\boldsymbol{\Omega}_p^{**}$ can be partitioned as

$$\boldsymbol{\Omega}_p^{**} \equiv \begin{pmatrix} [\boldsymbol{\Sigma}_{\varepsilon\gamma}]_{\mathrm{n}\times\mathrm{n}} & [\boldsymbol{\Sigma}_{\varepsilon X}']_{\mathrm{n}\times\mathrm{p}} \\ [\boldsymbol{\Sigma}_{\varepsilon X}]_{\mathrm{p}\times\mathrm{n}} & [\boldsymbol{\Sigma}_{X}]_{\mathrm{p}\times\mathrm{p}} \end{pmatrix},$$

where $\boldsymbol{\Sigma}_{\varepsilon\gamma} = \sigma^4 \sum_{j=1}^{\infty} \boldsymbol{\varphi}_j \boldsymbol{\varphi}_j'$ is positive definite and bounded because $\{\varphi_j(\gamma)\}$ is square-summable. Similarly, $\boldsymbol{\Sigma}_X = \sigma^2 \sum_{j=1}^{\infty} \mathbf{b}_j \mathbf{b}_j'$, with $\mathbf{b}_j = (b_{j-1}, ..., b_{j-p})'$ and $b_l = 0$ for all $l < 0$, is finite and positive definite owing to absolute summability of the coefficients in the Wold's representation of any stationary AR(p) process. From the Cauchy-Schwarz inequality, $||\boldsymbol{\Sigma}_{\varepsilon X}|| \leq ||\boldsymbol{\Sigma}_{\varepsilon\gamma}||^{1/2} \, ||\boldsymbol{\Sigma}_X||^{1/2} < \infty$, from which $||\boldsymbol{\Omega}_p^{**}|| < \infty$. Finally, $\boldsymbol{\Omega}_p^{**}$ is singular if and only if the elements of $\mathbf{X}_{tp}^{**}$ are linearly dependent, which obviously is not the case, so $\det(\boldsymbol{\Omega}_p^{**}) > \delta > 0$. Part ii) holds if (a) $||\bar{\boldsymbol{\Sigma}}_{\varepsilon\gamma}^{*} - \boldsymbol{\Sigma}_{\varepsilon,\gamma}|| = o_p(1)$, (b) $||\bar{\boldsymbol{\Sigma}}_{X}^{*} - \boldsymbol{\Sigma}_X|| = o_p(1)$, and (c) $||\bar{\boldsymbol{\Sigma}}_{\varepsilon X}^{*} - \boldsymbol{\Sigma}_{\varepsilon X}|| = o_p(1)$, given the respective sample estimators, *e.g.*, $\bar{\boldsymbol{\Sigma}}_{\varepsilon\gamma}^{*} = (T-p)^{-1} \sum_{t=p+1}^{T} \boldsymbol{\varepsilon}_{\boldsymbol{\gamma},t-1}^{*} \boldsymbol{\varepsilon}_{\boldsymbol{\gamma},t-1}^{*\prime}$. The proof of (a) follows from Lemmas B.4 and B.5 and identically as in Lemma B.3. The proof of

(b) follows as in Theorem 2.2 in Gonçalves and Kilian (2004). Finally, for part (c), consider

$$\begin{aligned} A_{1tT} &= \sum_{j=0}^{\infty} b_j \varphi_{j+i-1}(\gamma_k) \left[v_{t-j-i}^2 - \sigma^2 \right], \\ A_{2tT} &= \sum_{l=0}^{T} \sum_{\substack{j=0 \\ j \neq l-i+1}}^{T} b_l \varphi_j(\gamma_k) v_{t-j-1} v_{t-l-i} \end{aligned}$$

and let $E||\bar{\boldsymbol{\Sigma}}_{\varepsilon X}^{**} - \boldsymbol{\Sigma}_{\varepsilon X}||^2 = \sum_i^p \sum_k^n \sum_{t,s=p+1}^T E\left[(A_{1tT} + A_{2tT})(A_{1sT} + A_{2sT}) \right]$. Notice that

$$\begin{aligned} T^{-2} \sum_{t=p+1}^{T} \sum_{s=p+1}^{T} E(A_{1tT} A_{1sT}) &= \frac{1}{T} \sum_{j=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} b_j \varphi_j(\gamma) b_l \varphi_l(\gamma_k) \times \\ & \left\{ \frac{1}{T} \sum_{t=p+1}^{T} \sum_{s=p+1}^{T} Cov\left(v_{t-j-i-1}^2, v_{s-l-i-1}^2 \right) \right\} \end{aligned}$$

by setting $b_j = \varphi_l(\gamma_k) = 0$ for all $j, l < 0$. Under the restriction of stationarity and absolutely summable cumulants, the term in curly brackets is uniformly bounded in t, s and T for any $1 \leq i \leq p < \infty$. Hence, given some constant $K < \infty$, it follows by the Cauchy-Schwarz inequality that

$$T^{-2} \sum_{t=p+1}^{T} \sum_{s=p+1}^{T} E(A_{1t} A_{1s}) \leq \frac{K}{T} \left(\sum_{j=0}^{\infty} b_j \varphi_j(\gamma) \right)^2 \leq \frac{K}{T} \left(\sum_{j=0}^{\infty} b_j^2 \right) \left(\sum_{j=0}^{\infty} \varphi_{j+i}^2(\gamma) \right) = O\left(T^{-1}\right).$$

Similarly, under Assumption $\mathcal{A}.2'$, we can show that the remaining term,

$$T^{-2} \sum_{t=p+1}^{T} \sum_{s=p+1}^{T} E(A_{2tT} A_{2sT}) + T^{-2} \sum_{t=p+1}^{T} \sum_{s=p+1}^{T} E(A_{1sT} A_{2tT} + A_{1tT} A_{2sT}) = O\left(T^{-1}\right)$$

from which $||\bar{\boldsymbol{\Sigma}}_{\varepsilon X}^{**} - \boldsymbol{\Sigma}_{\varepsilon X}|| = O_p\left(T^{-1/2}\right) = o_p(1)$ by Markov's inequality. Finally, as in Lemma B.2, we can show

$$||\bar{\boldsymbol{\Sigma}}_{\varepsilon X}^{**} - \bar{\boldsymbol{\Sigma}}_{\varepsilon X}^{*}|| = O_p\left(T^{-1} \sum_{t=p+1}^{T} \left(\boldsymbol{\Omega}_t^{**} - \boldsymbol{\Omega}_t^{*} \right) \right) = O_p\left(T^{-1/2}\right) = o_p(1),$$

and then the AEL renders the required result.■

Lemma B.7. *Let* $\boldsymbol{\Lambda}_p = E\left(v_t^2 \mathbf{X}_{tp}^{**} \mathbf{X}_{tp}^{\prime **} \right)$ *be defined through the partition*

$$\begin{pmatrix} \left[\boldsymbol{\Lambda}_{\varepsilon\gamma}^b\right]_{\mathrm{n\times n}} & \left[\boldsymbol{\Lambda}_{\varepsilon X}^{\prime}\right]_{\mathrm{n\times p}} \\ \left[\boldsymbol{\Lambda}_{\varepsilon X}\right]_{\mathrm{p\times n}} & \left[\boldsymbol{\Lambda}_{X}\right]_{\mathrm{p\times p}} \end{pmatrix}$$

and let e_{tp} *and* $\widehat{e}_{tp}$ *be the residuals, and the estimated residuals, respectively, from the augmented auxiliary regression* (4)*. Then, under then null hypothesis and Assumption* $\mathcal{A}.2'$*, as* $T \rightarrow \infty$ *:*
i) $\boldsymbol{\Lambda}_p < \infty$*, and* $det(\boldsymbol{\Lambda}_p) > \delta > 0$*;*
ii) $T^{-1/2} \sum_{t=p+1}^{T} e_{tp} \mathbf{X}_{tp}^{*} \Rightarrow \mathcal{N}(0, \boldsymbol{\Lambda}_p)$*;*

iii) $T^{-1}\sum_{t=p+1}^{T}\widehat{e}_{tp}^{2}\left(\mathbf{X}_{tp}^{*}\mathbf{X}_{tp}^{\prime *}\right)\xrightarrow{p}\mathbf{\Lambda}_{p}$.

Proof of Lemma B.7.

For part *i*), $\mathbf{\Lambda}_{\varepsilon\gamma}^{b}=\mathbf{\Sigma}_{\varepsilon\gamma}+\sum_{j,l\geq 1}\left[\boldsymbol{\varphi}_{j}\boldsymbol{\varphi}_{l}^{\prime}\right]\kappa_{v}\left(0,j,l,0\right),$ with $\mathbf{\Sigma}_{\varepsilon\gamma}=\sigma^{4}\sum_{j=1}^{\infty}\boldsymbol{\varphi}_{j}\boldsymbol{\varphi}_{j}^{\prime}$ defined in Lemma B.6. From Lemma B.4 and Assumption $\mathcal{A}.2^{\prime}$, the same statistical considerations as in Lemma B.2 apply on $\mathbf{\Lambda}_{\varepsilon\gamma}^{b}$, and as a result this is a finite, positive definite covariance matrix. Similarly, we can show as in Theorem 2.2 in Gonçalves and Kilian (2007) that $\mathbf{\Lambda}_{X}<\infty$, whereas from the Cauchy-Schwarz inequality $\mathbf{\Lambda}_{\varepsilon X}<\infty$, from which $\mathbf{\Lambda}_{p}<\infty$. As in Lemma B.6, $\mathbf{\Lambda}_{p}$ is invertible, and so $\det(\mathbf{\Lambda}_{p})>\delta>0$. For part *ii*), under the null hypothesis $e_{tp}=v_{t}$, and $\mathbf{v}_{t-1}=\left(v_{t-1},...,v_{t-p}\right)^{\prime}$, we have

$$\sum_{t=p+1}^{T}e_{tp}\left(\begin{array}{c}\boldsymbol{\varepsilon}_{\gamma,t-1}^{**}\\ \mathbf{X}_{tp}\end{array}\right)=\sum_{t=p+1}^{T}\sum_{j=0}^{\infty}\left(\begin{array}{c}\boldsymbol{\varphi}_{j}v_{t-j-1}v_{t}\\ b_{j}\mathbf{v}_{t-j-1}v_{t}\end{array}\right)=\sum_{t=p+1}^{T}\left(\begin{array}{c}\mathbf{Z}_{\varepsilon t}\\ \mathbf{Z}_{Xt}\end{array}\right),\text{ say.}$$

Clearly, $\left\{\mathbf{Z}_{\varepsilon t},\mathcal{G}_{t}\right\}$ and $\left\{\mathbf{Z}_{Xt},\mathcal{G}_{t}\right\},$ $\mathcal{G}_{t}=\sigma\left(v_{j}:j\leq t\right),$ are square-integrable MDS under Assumption $\mathcal{A}.2^{\prime}$, with $E\left(\mathbf{Z}_{\varepsilon t}\mathbf{Z}_{\varepsilon t}^{\prime}\right)=\mathbf{\Lambda}_{\varepsilon\gamma}^{b}$, $E\left(\mathbf{Z}_{Xt}\mathbf{Z}_{Xt}^{\prime}\right)=\mathbf{\Lambda}_{X}$, and $E\left(\mathbf{Z}_{\varepsilon t}\mathbf{Z}_{Xt}^{\prime}\right)=\mathbf{\Lambda}_{\varepsilon X}^{\prime}$. We can use the CLT for MDS as in Lemma B.2 to show asymptotic normality of the normalized sums of $\left(\mathbf{Z}_{\varepsilon t}^{\prime},\mathbf{Z}_{Xt}^{\prime}\right)^{\prime}$. In particular, note that $(\mathcal{C}1)$ holds if *a*) $||\bar{\mathbf{\Lambda}}_{\varepsilon\gamma,T}^{b}-\mathbf{\Lambda}_{\varepsilon\gamma}^{b}||=o_{p}\left(1\right),$ *b*) $||\bar{\mathbf{\Lambda}}_{X,T}-\mathbf{\Lambda}_{X}||=o_{p}\left(1\right),$ and *c*) $||\bar{\mathbf{\Lambda}}_{\varepsilon X,T}-\mathbf{\Lambda}_{\varepsilon X}||=o_{p}\left(1\right),$ where again the first terms denote the sample estimates based on the filtered process. The proof of *a*) follows along the same lines as in Lemma B.2 owing to Lemma B.4, and the proof of *b*) follows as in Theorem 3.1 in Gonçalves and Killian (2004). To check *c*), note that for $1\leq i\leq p$, and $1\leq k\leq n$, the characteristic element of $TE||\bar{\mathbf{\Lambda}}_{\varepsilon X,T}-\mathbf{\Lambda}_{\varepsilon X}||^{2}$ can be written as $T^{-1}\sum_{t=p+1}^{T}\sum_{s=p+1}^{T}Cov\left(\varepsilon_{\mathbf{d},t-i}\varepsilon_{\gamma_{k},t-1}v_{t}^{2},\varepsilon_{\mathbf{d},s-i}\varepsilon_{\gamma_{k},s-1}v_{s}^{2}\right),$ *i.e.*,

$$T^{-1}\sum_{l_{1},...,l_{4}=-\infty}^{\infty}b_{l_{1}}b_{l_{3}}\varphi_{l_{2}}\left(\gamma_{k}\right)\varphi_{l_{4}}\left(\gamma_{k}\right)\sum_{t=p+1}^{T}\sum_{s=p+1}^{T}Cov\left(v_{t-i-l_{1}-1}v_{t-l_{2}-1}v_{t}^{2},v_{s-i-l_{3}-1}v_{s-l_{4}-1}v_{s}^{2}\right)$$

with $b_{l}=\varphi_{l}\left(\gamma_{k}\right)=0$ for all $l<0$. First, consider the case related to the smallest frequency, corresponding to $k=1$, for which we can assume again $\gamma_{1}=0$ with no loss of generality. As discussed in Proposition 2 in Demetrescu *et al.* (2008), this term is uniformly bounded by a constant that does not depend on t,s,T or i. Then, for any $1\leq k\leq n$ and all $1\leq i\leq p$, note that $\left|b_{l_{1}}b_{l_{3}}\varphi_{l_{2}}\left(\gamma_{k}\right)\varphi_{l_{4}}\left(\gamma_{k}\right)\right|\ \leq 4\ \left|b_{l_{1}}b_{l_{3}}\varphi_{l_{2}}\left(0\right)\varphi_{l_{4}}\left(0\right)\right|$ and as a result it follows that $E||\bar{\mathbf{\Lambda}}_{\varepsilon X,T}-\mathbf{\Lambda}_{\varepsilon X}||^{2}=O\left(T^{-1}\right)=o_{p}\left(1\right),$ thus implying

$$\frac{1}{T-p}\sum_{t=p+1}^{T}v_{t}^{2}\left(\mathbf{X}_{tp}^{**}\mathbf{X}_{tp}^{\prime **}\right)\overset{ms}{\rightarrow}\mathbf{\Lambda}_{p}$$

as required. Finally note that, from Lemma B.4, $\left(\mathbf{Z}_{\varepsilon t}^{\prime},\mathbf{Z}_{Xt}^{\prime}\right)^{\prime}$ is defined by an $\mathcal{G}_{t}$-measurable function on $\left\{v_{t}\right\},$ so it is a strictly stationary and ergodic MDS (cf. White, 2001, Thm. 3.35). Furthermore, from (i) in this lemma, the process is bounded and bounded away from zero under the L_{2}-norms, thus $(\mathcal{C}2)$ holds trivially. Hence, under the null hypothesis and Assumption $\mathcal{A}.2^{\prime}$, as $T\rightarrow\infty$, $T^{-1/2}\sum_{t=p+1}^{T}e_{tp}\mathbf{X}_{tp}^{**}\Rightarrow\mathcal{N}\left(\mathbf{0},\mathbf{\Lambda}_{p}\right).$ Finally, since $||T^{-1/2}\sum_{t=p+1}^{T}v_{t}\left(\mathbf{X}_{tp}^{**}-\mathbf{X}_{tp}^{*}\right)||=o_{p}\left(1\right)$ from Lemma B.5, it follows by the AEL that $T^{-1/2}\sum_{t=p+1}^{T}e_{tp}\mathbf{X}_{tp}^{*}\Rightarrow\mathcal{N}\left(\mathbf{0},\mathbf{\Lambda}_{p}\right)$ as required. For part *iii*), consider $a_{i,T}$ the LS estimate of the i-th autoregressive coefficient. Then, $v_{t}-\widehat{e}_{tp}=\sum_{i=1}^{p}\left(a_{i,T}-a_{i}\right)\varepsilon_{\mathbf{d},t-i}+\sum_{k=1}^{n}\phi_{k,T}\varepsilon_{\gamma_{k},t-1}^{*}=O_{p}\left(T^{-1/2}\right)$ owing to $\sqrt{T}$-consistency (see Theorem

2.3 below). Therefore, $v_t^2 - \widehat{e}_{tp}^2 = (v_t - \widehat{e}_{tp})(v_t + \widehat{e}_{tp}) = O_p\left(T^{-1/2}\right) + O_p\left(T^{-1}\right)$, and hence

$$\left\| T^{-1} \sum_{t=p+1}^{T} \left(v_t^2 - \widehat{e}_{tp}^2\right) \mathbf{X}_{tp}^{**}\mathbf{X}_{tp}^{\prime **} \right\| \leq \frac{1}{T} \sum_{t=p+1}^{T} \left|v_t^2 - \widehat{e}_{tp}^2\right| \; \left\|\mathbf{X}_{tp}^{**}\mathbf{X}_{tp}^{\prime **}\right\| = O_p\left(T^{-1/2}\right) = o_p(1)$$

which together with (ii) above implies that $\frac{1}{T-p}\sum_{t=p+1}^{T} \widehat{e}_{tp}^2 \left(\mathbf{X}_{tp}^{**}\mathbf{X}_{tp}^{\prime **}\right) \overset{ms}{\to} \boldsymbol{\Lambda}_p$ by the AEL. But since

$$\begin{aligned} \left\| T^{-1} \sum_{t=p+1}^{T} \left(v_t^2 - \widehat{e}_{tp}^2\right) \left(\mathbf{X}_{tp}^{**}\mathbf{X}_{tp}^{\prime **} - \mathbf{X}_{tp}^{*}\mathbf{X}_{tp}^{\prime *}\right) \right\| &= O_p\left(T^{-1} \sum_{t=p+1}^{T} \left(v_t^2 - \widehat{e}_{tp}^2\right) \left(\boldsymbol{\Omega}_t^{**} - \boldsymbol{\Omega}_t^{*}\right) \right) \\ &= O_p\left(T^{-1} \sum_{t=p+1}^{T} O_p\left(T^{-1/2}\right) O_p\left(1/\sqrt{t}\right) \right) \\ &= O_p\left(T^{-1/2}\right) = o_p(1) \end{aligned}$$

by using Cauchy-Schwarz inequality, it follows from the AEL that

$$\frac{1}{T-p} \sum_{t=p+1}^{T} \widehat{e}_{tp}^2 \mathbf{X}_{tp}^{*}\mathbf{X}_{tp}^{\prime *} = \boldsymbol{\Lambda}_p + o_p(1)$$

as $T \to \infty$.■

Proof of Theorem 2.2.

The proof of Theorem 2.2 is immediate in view of the previous results. Let $\boldsymbol{\beta}_T^{**}$ and $\boldsymbol{\beta}_T$ be the OLS estimates in the corresponding augmented auxiliary regressions $\varepsilon_{\mathbf{d}t} = \mathbf{X}_{tp}^{\prime **}\boldsymbol{\beta}^{**} + e_{tp}$, and $\varepsilon_{\mathbf{d}t} = \mathbf{X}_{tp}^{\prime *}\boldsymbol{\beta} + e_{tp}$, respectively. Since,

$$\sqrt{T}\left(\boldsymbol{\beta}_T - \boldsymbol{\mu}_0\right) = \left(\frac{1}{T} \sum_{t=p+1}^{T} \mathbf{X}_{tp}^{*}\mathbf{X}_{tp}^{*\prime} \right)^{-1} \left(\frac{1}{\sqrt{T}} \sum_{t=p+1}^{T} e_{tp}\mathbf{X}_{tp}^{*} \right)$$

then according to lemmae B.4-B.7 and the CMT, it follows under the null hypothesis and Assumption $\mathcal{A}.2'$, that as $T \to \infty$, $\sqrt{T}\left(\boldsymbol{\beta}_T^{**} - \boldsymbol{\mu}_0\right)$ and $\sqrt{T}\left(\boldsymbol{\beta}_T - \boldsymbol{\mu}_0\right)$ are asymptotically equivalent, with $\sqrt{T}\left(\boldsymbol{\beta}_T - \boldsymbol{\mu}_0\right) \Rightarrow \mathcal{N}\left(\mathbf{0},\; \left(\boldsymbol{\Omega}_p^{**}\right)^{-1} \boldsymbol{\Lambda}_p \left(\boldsymbol{\Omega}_p^{**}\right)^{-1}\right)$. ■

Proof of Theorem 2.3.

Given normality of the estimated coefficients, Theorem 2.3 holds as a corollary of Theorem 2.2. Let $\mathbf{R}$ be an $n \times (n+p)$ matrix such $[\mathbf{R}]_{ij} = 1$ for all $i = j$ and zero otherwise. Consider the regression-based test statistic computed from the augmented auxiliary regression, *i.e.*, $\Upsilon_{Wp}^{(n)} = \left[\sqrt{T}\mathbf{R}\boldsymbol{\beta}_T\right]' \left[\mathbf{R}\widehat{\mathbf{V}}_{\mathbf{T}}\mathbf{R}'\right]^{-1} \left[\sqrt{T}\mathbf{R}\boldsymbol{\beta}_T\right]$ where $\widehat{\mathbf{V}}_{\mathbf{T}}$ is the sample counterpart of the asymptotic covariance matrix of $\boldsymbol{\beta}_T$, *i.e.*,

$$\widehat{\mathbf{V}}_{\mathbf{T}} = \left(\frac{1}{T} \sum_{t=p+1}^{T} \mathbf{X}_{tp}^{*}\mathbf{X}_{tp}^{*\prime} \right)^{-1} \left(\frac{1}{T} \sum_{t=p+1}^{T} \widehat{e}_{tp}^2 \mathbf{X}_{tp}^{*}\mathbf{X}_{tp}^{*\prime} \right) \left(\frac{1}{T} \sum_{t=p+1}^{T} \mathbf{X}_{tp}^{*}\mathbf{X}_{tp}^{*\prime} \right)^{-1}$$

where the inclusion of the squared estimated residuals, $\widehat{e}_{tp}^2$, is intended to provide robustness against (conditional) heteroskedastic patterns of unknown form. Given the previous lemmae and the CMT, it follows readily under the null hypothesis and as $T \to \infty$ that,

$$\sqrt{T}\left(\mathbf{R}\boldsymbol{\beta}_T\right) = \sqrt{T}\boldsymbol{\phi}_T \Rightarrow \mathcal{N}\left(\mathbf{0}, \mathbf{R}\left[\left(\boldsymbol{\Omega}_{\mathbf{p}}^{**}\right)^{-\mathbf{1}} \boldsymbol{\Lambda}_{\mathbf{p}} \left(\boldsymbol{\Omega}_{\mathbf{p}}^{**}\right)^{-\mathbf{1}}\right]\mathbf{R}'\right).$$

Hence, $\Upsilon_{Wp}^{(n)}$ converges to the distribution of a Gaussian quadratic form and, therefore, $\Upsilon_{Wp}^{(n)} \Rightarrow \chi^2_{(n)}$. ■

References

[1] Agiakloglou, C. and P. Newbold (1994) Lagrange multiplier tests for fractional difference. *Journal of Time Series Analysis* 14, 253–262.

[2] Bouette, J.C, J.F. Chassagneux, D. Sibai, R. Terron and A. Charpentier (2006) Wind in Ireland: long memory or seasonal effect? *Stochastic Environmental Research and Risk Assessment* 20, 141-151.

[3] Breitung, J. and U. Hassler (2002) Inference on the cointegration rank in fractionally integrated processes. *Journal of Econometrics* 110(2), 167-185.

[4] Brillinger, D. R. (1981) *Time Series - Data Analysis and Theory.* San Francisco: Holden Day.

[5] Chung, C.F. (1996) Estimating a generalized long memory process. *Journal of Econometrics* 73, 237-259.

[6] Dalla, V. and J. Hidalgo (2005) A parametric bootstrap test for cycles. *Journal of Econometrics* 129, 219-261.

[7] Davidson, J. (1994) *Stochastic Limit Theory*, Oxford University Press.

[8] Davidson, J. (2000) *Econometric Theory*, Blackwell Publishers.

[9] Demetrescu, M., V. Kuzin, U. Hassler (2008) Long memory testing in the time domain. *Econometric Theory*, 24, 176-215.

[10] Dolado, J.J., J. Gonzalo and L. Mayoral (2002) A fractional Dickey-Fuller test for unit roots, *Econometrica* 70, 1963-2006.

[11] Gil-Alana, L.A. (2005) Modelling US monthly inflation in terms of a jointly seasonal and nonseasonal long memory process. *Applied Stochastic Models in Business and Industry* 21, 83-94.

[12] Gil-Alana, L.A. and P.M. Robinson (2001) Testing of seasonal fractional integration in UK and Japanese consumption and income. *Journal of Applied Econometrics* 16, 95-114.

[13] Giriatis, L., J. Hidalgo and P.M. Robinson (2001) Gaussian estimation of parametric spectral density with unknown pole. *Annals of Statistics* 29, 987-1023.

[14] Gonçalves, S. and L. Kilian (2004) Bootstrapping autoregressions with conditional heteroskedasticity of unknown form. *Journal of Econometrics* 123, 89-120.

[15] Gonçalves, S. and L. Kilian (2007) Asymptotic and bootstrap inference for AR(∞) processes with conditional heteroskedasticity. *Econometric Reviews* 26, 609-641.

[16] Gradshteyn, I.S. and I.M. Ryzhik (2000) *Table of Integrals, Series, and Products*, Academic Press.

[17] Gray, H.L., N.F. Zhang and W. Woodward (1989) On generalized fractional processes. *Journal of Time Series Analysis* 10, 233-57.

[18] Hassler, U. (1994) (Mis)specification of long memory in seasonal time series. *Journal of Time Series Analysis* 15, 19–30.

[19] Hassler, U., J. Breitung (2006), A residual-based LM type test against fractional cointegration. *Econometric Theory* 22, 1091-1111.

[20] Hassler, U., P.M.M. Rodrigues and A. Rubia (2008) Testing for the General Unit Root Hypothesis in the Time Domain. Working Paper 380, Fundacion de Cajas de Ahorros (FUNCAS).

[21] Hidalgo, J. (2005) Semiparametric estimation for stationary processes whose spectra have an unknown pole. *Annals of Statistics* 33, 1843-1889.

[22] Hidalgo, J. (2007) A nonparametric test for weak dependence against strong cycles and its bootstrap analougue. *Journal of Time Series Analysis* 28, 307-349.

[23] Hidalgo, J., and P. Soulier (2004) Estimation of the location and the exponent of the spectral singularity of a long memory process. *Journal of Time Series Analysis* 25, 55-81.

[24] Hylleberg, S., R.F. Engle, C.W.J. Granger and B.S. Yoo (1990) Seasonal integration and cointegration. *Journal of Econometrics* 44, 215–238.

[25] Marinucci, D. and P.M. Robinson (1999). Alternative forms of fractional Brownian Motion. *Journal of Statistical Planning and Inference* 80, 111–122.

[26] Nielsen, M.Ø. (2004) Efficient likelihood inference in nonstationary univariate models. *Econometric Theory* 20, 116-146.

[27] Nielsen, M.Ø. (2005) Multivariate lagrange multiplier tests for fractional integration. *Journal of Financial Econometrics* 3, 372-398.

[28] Porter-Hudak, S. (1990) An application of the seasonal fractionally differenced model to the monetary aggregates. *Journal of the American Statistical Association* 85*,* 338-344*.*

[29] Ramachandran, R. and P. Beaumont (2001) Robust estimation of GARMA model parameters with application to cointegration among interest rates of industrialized countries. *Computational Economics* 17, 179 - 201.

[30] Robinson, P.M. (1991). Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. *Journal of Econometrics*, 47, 67-84.

[31] Robinson, P.M. (1994). Efficient tests of nonstationary hypotheses. *Journal of the American Statistical Association* 89, 1420-1437.

[32] Robinson, P.M. (2005). The distance between rival nonstationary fractional processes. *Journal of Econometrics* 128*,* 283-300.

[33] Schwert, G. W. (1989) Tests for unit roots: A Monte Carlo investigation. *Journal of Business and Economic Statistics* 7, 147-59.

[34] Smallwood, A.D. and S.C. Norrbin (2006) Generalized long memory processes, failure of cointegration tests and exchange rate dynamics. *Journal of Applied Econometrics* 21, 409 - 417.

[35] Soares, L.J. and L.R. Souza, (2006) Forecasting electricity demand using generalized long memory. *International Journal of Forecasting* 22, 17-28.

[36] Tanaka, K. (1999) The nonstationary fractional unit root. *Econometric Theory* 15, 549 - 582.

[37] Yajima, Y. (1996). Estimation of the frequency of the unbounded spectral density, in *Proceedings of the Business and Economic Statistical Section.* American Statistical Association.

[38] White, H. (2001). *Asymptotic Theory for Econometricians*, Academic Press.

[39] Woodward, W. A., Q. C. Cheng, and H. L. Gray (1998). A k-factor GARMA long-memory model. *Journal of Time Series Analysis* 19, 485–504.

Tables and Figures

Table 1: Empirical rejection frequencies when the DGP is the simple GARMA model $(1-2\cos\gamma_s L+L^2)^{1+\theta}x_t=\varepsilon_t, \quad \varepsilon_t\sim iid\mathcal{N}(0,1)$.

				θ			
γ_s	-.3	-.2	-.1	0	.1	.2	.3
				T=100			
$\frac{\pi}{10}$	.999	.984	.540	**.052**	.584	.981	.999
$\frac{2\pi}{10}$	.999	.933	.401	**.054**	.445	.927	.998
$\frac{3\pi}{10}$	.988	.810	.302	**.056**	.329	.832	.982
$\frac{4\pi}{10}$	.946	.689	.232	**.049**	.267	.721	.946
$\frac{5\pi}{10}$	.929	.630	.210	**.050**	.248	.686	.932
$\frac{6\pi}{10}$	.955	.683	.236	**.051**	.269	.730	.947
$\frac{7\pi}{10}$	.985	.826	.311	**.045**	.331	.836	.985
$\frac{8\pi}{10}$	.998	.929	.425	**.051**	.452	.933	.998
$\frac{9\pi}{10}$	.999	.982	.536	**.050**	.585	.984	.999
				T=250			
$\frac{\pi}{10}$	.999	.999	.924	**.043**	.921	.999	.999
$\frac{2\pi}{10}$	.999	.999	.818	**.057**	.814	.999	.999
$\frac{3\pi}{10}$	.999	.997	.653	**.050**	.686	.995	.999
$\frac{4\pi}{10}$	.999	.979	.516	**.052**	.563	.980	.999
$\frac{5\pi}{10}$	.999	.971	.468	**.051**	.545	.968	.999
$\frac{6\pi}{10}$	.999	.980	.520	**.051**	.571	.978	.999
$\frac{7\pi}{10}$	.999	.998	.664	**.045**	.682	.994	.999
$\frac{8\pi}{10}$	.999	1.00	.811	**.050**	.816	.999	.999
$\frac{9\pi}{10}$	.999	.999	.918	**.045**	.913	.999	.999

Note: Empirical size is in bold.

Table 2: Empirical rejection frequencies when the DGP is the 2-factor GARMA model $(1-2\cos(0.15)L+L^2)^{1+\theta_1}(1-2\cos(\frac{\pi}{2})L+L^2)^{1+\theta_2}x_t=\varepsilon_t,\ \varepsilon_t \sim iid\mathcal{N}(0,1)$ and T=100

Joint Restricted Test

	θ_2						
θ_1	-.3	-.2	-.1	0	.1	.2	.3
-.3	.999	.999	0.997	.959	.741	.362	.247
-.2	.996	.992	.963	.834	.512	.220	.237
-.1	.793	.731	.611	.398	.179	.098	.290
.0	.126	.102	.082	**.047**	.067	.205	.480
.1	.631	.590	.583	.574	.625	.730	.853
.2	.987	.985	.982	.981	.982	.988	.993
.3	.999	.999	.999	.999	.999	.999	.999

Joint Unrestricted Test

	θ_2						
θ_1	-.3	-.2	-.1	0	.1	.2	.3
-.3	.999	.999	.999	.999	.999	.999	.999
-.2	.994	.978	.974	.977	.990	.999	.999
-.1	.892	.684	.502	.487	.693	.911	.985
.0	.857	.510	.161	**.049**	.205	.592	.893
.1	.988	.913	.741	.556	.535	.718	.898
.2	.999	.999	.992	.980	.974	.981	.991
.3	.999	.999	.999	.999	.999	.999	.999

Note: Empirical size is in bold.

Table 3: Empirical rejection frequencies when the DGP is the 2-factor GARMA model with ARMA errors:

$$(1-2\cos(0.15)L+L^2)^{1+\theta_1}(1-2\cos(\tfrac{\pi}{2})L+L^2)^{1+\theta_2}x_t=\varepsilon_t,\ (1-0.5L)\varepsilon_t=(1+0.5L)v_t,$$
$$v_t \sim iid\mathcal{N}(0,1)$$

T=100

Joint Restricted Test

	θ_2						
θ_1	-.3	-.2	-.1	0	.1	.2	.3
-.3	.300	.233	.148	.088	.067	.099	.142
-.2	.131	.120	.089	.059	.051	.072	.127
-.1	.063	.056	.055	.045	.041	.057	.096
.0	.047	.043	.046	**.043**	.049	.062	.080
.1	.065	.059	.063	.060	.061	.075	.086
.2	.093	.087	.092	.094	.092	.104	.113
.3	.126	.127	.123	.136	.127	.130	.139

Joint Unrestricted Test

	θ_2						
θ_1	-.3	-.2	-.1	0	.1	.2	.3
-.3	.204	.142	.122	.141	.202	.315	.381
-.2	.156	.097	.058	.069	.115	.160	.228
-.1	.137	.075	.046	.039	.058	.094	.138
.0	.121	.076	.046	**.037**	.044	.063	.090
.1	.113	.079	.058	.053	.053	.062	.075
.2	.103	.077	.073	.061	.068	.075	.085
.3	.105	.094	.085	.096	.091	.100	.105

T=500

Joint Restricted Test

	θ_2						
θ_1	-.3	-.2	-.1	.0	.1	.2	.3
-.3	.992	.955	.691	.225	.082	.316	.626
-.2	.897	.794	.525	.190	.071	.228	.534
-.1	.492	.389	.230	.093	.049	.179	.424
.0	.150	.113	.073	**.048**	.067	.175	.388
.1	.087	.090	.089	.115	.159	.258	.405
.2	.239	.255	.272	.294	.345	.401	.475
.3	.437	.448	.471	.493	.530	.543	.578

Joint Unrestricted Test

	θ_2						
θ_1	-.3	-.2	-.1	.0	.1	.2	.3
-.3	.981	.926	.834	.802	.862	.949	.979
-.2	.871	.680	.463	.386	.480	.653	.815
-.1	.570	.354	.170	.117	.177	.333	.518
.0	.264	.128	.064	**.053**	.092	.206	.360
.1	.126	.095	.075	.092	.134	.222	.338
.2	.192	.205	.215	.227	.272	.341	.405
.3	.371	.367	.394	.411	.446	.475	.511

Note: Empirical size is in bold. All tests are augmented using Schwert's rule.

Table 4: Empirical rejection frequencies when the DGP is the 2-factor GARMA model with AR errors: $(1-2\cos(0.15)L+L^2)^{1+\theta_1}(1-2\cos(\frac{\pi}{2})L+L^2)^{1+\theta_2}x_t=\varepsilon_t,\ (1-0.5L)\varepsilon_t=v_t,$ $v_t \sim iid\mathcal{N}(0,1)$

T=100

Joint Restricted Test

				θ_2			
θ_1	-.3	-.2	-.1	0	0.1	.2	.3
-.3	.479	.336	.202	.108	.077	.104	.179
-.2	.250	.201	.139	.078	.061	.094	.160
-.1	.103	.087	.076	.049	.059	.093	.145
0	.056	.047	.050	**.041**	.050	.079	.132
.1	.062	.056	.054	.057	.062	.083	.122
.2	.104	.088	.083	.082	.088	.097	.135
.3	.139	.130	.123	.124	.126	.137	.151

Joint Unrestricted Test

				θ_2			
θ_1	-.3	-.2	-.1	0	.1	.2	.3
-.3	.334	.234	.191	.194	.255	.352	.467
-.2	.229	.138	.092	.089	.140	.232	.319
-.1	.224	.108	.055	.045	.082	.156	.249
0	.232	.101	.052	**.038**	.058	.127	.190
.1	.226	.118	.059	.046	.061	.115	.187
.2	.210	.124	.071	.057	.076	.128	.203
.3	.181	.118	.094	.083	.110	.164	.219

T=500

Joint Restricted Test

				θ_2			
θ_1	-.3	-.2	-.1	0	.1	.2	.3
-.3	.999	.966	.631	.153	.142	.527	.814
-.2	.967	.889	.564	.158	.108	.458	.766
-.1	.641	.538	.298	.093	.088	.409	.731
0	.205	.155	.088	**.044**	.118	.394	.706
.1	.094	.087	.090	.123	.225	.460	.693
.2	.232	.251	.269	.295	.398	.548	.710
.3	.438	.463	.481	.521	.574	.654	.750

Joint Unrestricted Test

				θ_2			
θ_1	-.3	-.2	-.1	0	.1	.2	.3
-.3	.995	.945	.833	.829	.934	.984	.998
-.2	.987	.812	.487	.397	.601	.853	.954
-.1	.947	.656	.236	.106	.268	.604	.834
0	.874	.513	.156	**.044**	.154	.464	.712
.1	.761	.437	.179	.101	.200	.453	.672
.2	.643	.431	.295	.247	.342	.527	.677
.3	.582	.496	.441	.441	.505	.600	.712

Note: Empirical size is in bold. All tests are augmented using Schwert's rule.

Table 5: Empirical rejection frequencies when the DGP is the 2-factor GARMA model with AR errors: $(1-2\cos(0.15)L+L^2)^{1+\theta_1}(1-2\cos(\frac{\pi}{2})L+L^2)^{1+\theta_2}x_t=\varepsilon_t,\ (1-0.9L)\varepsilon_t=v_t,$ $v_t \sim iid\mathcal{N}(0,1)$

T=100

Joint Restricted Test

	θ_2						
θ_1	-.3	-.2	-.1	0	.1	.2	.3
-.3	.085	.080	.062	.043	.040	.067	.114
-.2	.044	.042	.046	.037	.032	.056	.100
-.1	.037	.036	.040	.040	.041	.051	.078
0	.047	.042	.046	**.051**	.049	.065	.079
.1	.062	.068	.070	.072	.068	.082	.097
.2	.076	.075	.073	.084	.095	.087	.100
.3	.073	.077	.078	.078	.078	.086	.087

Joint Unrestricted Test

	θ_2						
θ_1	-.3	-.2	-.1	0	.1	.2	.3
-.3	.290	.143	.064	.034	.056	.105	.168
-.2	.297	.149	.070	.035	.047	.094	.167
-.1	.272	.125	.062	.039	.050	.107	.173
0	.230	.107	.056	**.043**	.065	.123	.208
.1	.157	.097	.058	.056	.081	.142	.226
.2	.120	.080	.055	.069	.094	.149	.219
.3	.081	.068	.056	.059	.078	.118	.170

T=500

Joint Restricted Test

	θ_2						
θ_1	-.3	-.2	-.1	0	.1	.2	.3
-.3	.913	.840	.667	.295	.083	.247	.587
-.2	.547	.474	.348	.184	.063	.173	.477
-.1	.172	.164	.118	.072	.0400	.111	.351
0	.063	.062	.055	**.051**	.053	.106	.248
.1	.088	.103	.080	.085	.100	.130	.197
.2	.133	.126	.123	.123	.119	.121	.161
.3	.105	.099	.092	.090	.081	.085	.088

Joint Unrestricted Test

	θ_2						
θ_1	-.3	-.2	-.1	0	.1	.2	.3
-.3	.995	.886	.549	.287	.391	.679	.866
-.2	.964	.738	.340	.148	.233	.524	.764
-.1	.843	.511	.185	.062	.138	.405	.656
0	.620	.319	.111	**.051**	.123	.320	.549
.1	.408	.217	.103	.066	.122	.258	.432
.2	.232	.150	.111	.090	.117	.184	.286
.3	.113	.091	.077	.076	.075	.102	.144

Note: Empirical size is in bold. All tests are augmented using Schwert's rule.