A Multivariate Test Against Spurious Long Memory

Philipp Sibbertsen,[1] Christian Leschinski and Marie Holzhausen

Institute of Statistics, Faculty of Economics and Management,
Leibniz University Hannover, D-30167 Hannover, Germany

This version: March 3, 2015

Abstract

This paper provides a multivariate score-type test to distinguish between true and spurious long memory. The test is based on the weighted sum of the partial derivatives of the multivariate local Whittle likelihood function. This approach takes phase shifts in the multivariate spectrum into account. The resulting pivotal limiting distribution is independent of the dimension of the process, which makes it easy to apply in practice. We prove the consistency of our test against the alternative of random level shifts or monotonic trends. A Monte Carlo analysis shows good finite sample properties of the test in terms of size and power. Additionally, we apply our test to the log-absolute returns of the S&P 500, DAX, FTSE, and the NIKKEI. The multivariate test gives formal evidence that these series are contaminated by level shifts.

JEL-Numbers: C12, C32

Keywords: Multivariate Long Memory · Semiparametric Estimation · Spurious Long Memory · Volatility

[1]Corresponding author:
Phone: +49-511-762-3783
Fax: +49-511-762-3923
E-Mail: sibbertsen@statistik.uni-hannover.de

1 Introduction

Distinguishing between true and spurious long memory is of major importance for the empirical modeling of many macroeconomic and financial time series. Usually, fractionally integrated processes are used to model long memory. Nevertheless, several authors point out that other data generating processes can have similar autocovariance features. Examples of this literature include Granger and Ding (1996), who argue that time varying coefficient models and other non-linear models can generate spurious long memory. Diebold and Inoue (2001) derive parameter constellations for which random level shift processes, STOPBREAK models, and markov switching models can generate spurious long memory and support these arguments with extensive simulation studies. Similar results are obtained by Granger and Hyung (2004), who also report that the evidence for long memory in absolute returns of the S&P 500 vanishes if breaks are accounted for. Further contributions that give similar evidence are Lobato and Savin (1998) and Mikosch and Stărică (2004), among others.

Motivated by these findings, several tests have been proposed to distinguish true long memory from spurious long memory. Berkes et al. (2006) or Yau and Davis (2012), among others, suggest tests for the null hypothesis of spurious long memory. Here, we focus on the literature with the null hypothesis of true long memory. Dolado et al. (2005) propose a time domain test based on the testing principle of previously derived fractional Dickey Fuller tests. Shimotsu (2006) suggests two tests. The first one is based on the observation that for a true fractionally integrated process the memory parameter d must be the same across all subsamples. On the other hand, the implied d of a spurious long-memory process depends on the number of shifts and their location in the sample. Splitting the sample into subsamples, therefore, changes the implied d in every subsample. Shimotsu's second approach is to test whether the d-th difference of a process is $I(0)$ or its partial sum $I(1)$, using KPSS and Phillips-Perron tests. Another property of fractionally integrated processes is that the parameter d remains unchanged if the process is temporally aggregated. Ohanissian et al. (2008) use this property to test for true long memory by testing the equality of ds estimated from different aggregation levels of the same process. A similar test based on periodic sub-sampling, also known as skip-sampling, is presented by Davidson and Rambaccussing (2015). Their testing procedure compares the memory estimator of the skip-sampled data with the estimated d from the original data. Haldrup and Kruse (2014) propose a test based on the fact that nonlinear transformations of an $I(d)$ process will reduce the order of memory when the Hermite rank of the transformation is greater than one.

Perron and Qu (2010) derive the properties of the periodogram of processes with short memory plus level shifts. They find that for low frequencies the effect of the shifts dominates the behavior of the spectral density and the implied value of d is one. For larger frequencies, on the other hand, the short memory component is dominant and the implied d is zero. These findings explain the sensitivity of semiparametric d-estimators with respect to the bandwidth choice. Therefore, Perron and Qu (2010) propose a test statistic based on the difference between memory parameters estimated with different bandwidths. The same results on the spectral density of level shift processes are used by Qu (2011), who derives a score-type test that is based on the derivative of the local Whittle likelihood function. Simulation studies conducted by Qu (2011) and Leccadito et al. (2015) show that among the tests suggested so far, overall the Qu test has the best power against a wide range of alternatives.

All of these approaches discussed so far are univariate. There is, however, a strand of literature that has considered multivariate extensions of fractionally integrated processes. Early examples include the work of Sowell (1989) and Lobato (1997). In particular, Lobato (1999) and Shimotsu (2007) extend the local Whittle estimator to a multivariate framework.

We contribute to this literature by generalizing the approach of Qu (2011) to test for true long memory in multivariate processes. The test statistic is based on the weighted sum of the partial derivatives of the multivariate local Whittle likelihood function in the form introduced by Shimotsu (2007). In this specification the cross-spectral densities contain information on the phase and coherence of the process. We prove that our test is consistent. Due to the weighting scheme it has a pivotal distribution that does not depend on the dimension of the process and for univariate processes it reduces to the special case of the Qu test.

To our knowledge, this is the first multivariate test against spurious long memory. The idea behind the test is that under the null hypothesis the derivative of the local Whittle likelihood function evaluated at $\hat{d}$ for the first $[mr] < m$ Fourier frequencies with $r \in [\varepsilon, 1]$ is approximately equal to zero, independent of the number of frequencies used for the local Whittle estimation of d. Under the alternative the derivative diverges if it is evaluated for a lower number of Fourier frequencies than used for the estimation of d since it is based on a wrong assumption about the shape of the spectral density.

Our test statistic is derived in a multivariate long memory framework which excludes fractional cointegration. As this is a drawback for the applicability of our test we also provide a modified test statistic for the situation of fractionally cointegrated data. This modified test statistic has the same asymptotic properties as the original test statistic, including the same limiting distribution.

In the empirical example we apply our test to the log-absolute returns of four stock market indices, the Standard & Poor 500, DAX, FTSE and NIKKEI. Even though the log-absolute values of S&P 500 returns have been studied in many of the aforementioned

contributions on the possibility of spurious long memory, the tests proposed so far often fail to reject the null hypothesis of a true long-memory process. We, therefore, reconsider this example by extending it to a multivariate framework and can clearly reject the null hypothesis, indicating that the series are indeed contaminated by level shifts.
The rest of the paper is structured as follows. After stating the model and the assumptions in Section 2, the test statistic is derived in Section 3. Some Monte Carlo simulations are given in Section 4. The empirical application is presented in Section 5 and Section 6 concludes. Proofs can be found in the appendix.

2 Model Specification and Assumptions

Point of departure is the multivariate q-dimensional long memory model

$$\begin{pmatrix} (1-L)^{d_1} & & 0 \\ & \ddots & \\ 0 & & (1-L)^{d_q} \end{pmatrix} \begin{pmatrix} X_{1t} - EX_{1t} \\ \vdots \\ X_{qt} - EX_{qt} \end{pmatrix} = \begin{pmatrix} u_{1t} \\ \vdots \\ u_{qt} \end{pmatrix},$$

with $-1/2 < d_1, \ldots, d_q < 1/2$ and $t = 1, \ldots, T$. This can alternatively be written as

$$D(d_1, ..., d_q)(X_t - EX_t) = u_t, \tag{1}$$

where X_t is a $(q \times 1)$ column vector and $u_t = (u_{1t}, u_{2t}, ..., u_{qt})'$ is a covariance stationary process with spectral density $f_u(\lambda)$ which is bounded and bounded away from zero in a matrix sense at the zero frequency, $\lambda = 0$. The operator $D(d_1, ..., d_q) = diag((1-L)^{d_1}, ..., (1-L)^{d_q})$ is a $(q \times q)$ matrix with zeros on the non-diagonal elements.
In a univariate framework a type II fractionally integrated process (e.g., Marinucci and Robinson, 1999) is defined by $(1-L)^d x_t = u_t 1(t \geq 0)$, where u_t is an $I(0)$ process having the Wold representation $u_t = \sum_{j=0}^{\infty} \theta_j \epsilon_{t-j}$ with $\sum_{j=0}^{\infty} ||\theta_j||^2 < \infty$. The innovations ϵ_t are assumed to be a martingale difference sequence satisfying $E(\epsilon_t | \mathfrak{F}_{t-1}) = 0$ and $E(\epsilon_t^2 | \mathfrak{F}_{t-1}) < \infty$ with $\mathfrak{F}_t = \sigma(\{\epsilon_s, s \leq t\})$. Furthermore, it is $u_t = 0$ for $t \leq 0$. The order of fractional integration is given by d and $(1-L)^d$ is defined by its binomial expansion

$$(1-L)^d = \sum_{j=0}^{\infty} \frac{\Gamma(j-d)}{\Gamma(-d)\Gamma(j+1)} L^j,$$

with $\Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t} dt$. L denotes the Backshift operator, i.e. $Le_t = e_{t-1}$. Details about recent developments on long-memory time series can be found in Beran et al. (2013) or Giraitis et al. (2012).

The spectral density of the multivariate long-memory process X_t as in (1), with $d = (d_1, d_2, ..., d_q)'$ being the memory vector, is local to the origin given by

$$f(\lambda_j) \sim \Lambda_j(d) G \Lambda_j^*(d), \tag{2}$$

with $\Lambda_j(d) = diag(\Lambda_{ja}(d))$ and $\Lambda_{ja}(d) = \lambda_j^{-d_a} e^{i(\pi-\lambda_j)d_a/2}$, where $\lambda_j = 2\pi j/T$ denotes the j-th Fourier frequency, $j = 1, \ldots, \lfloor T/2 \rfloor$. G is a real, positive definite, symmetric and finite matrix and the asterix A^* denotes the conjugate transpose of the matrix A. Further, the imaginary number is denoted by i and d_a is the memory parameter in dimension a.

The assumption on G excludes fractional cointegration as it stands. We first derive our test statistic under this assumption and consider the case of fractionally cointegrated data afterwards. It turns out that the asymptotic properties of the test statistic remain unchanged under slight modifications of our test statistic.

The spectral density representation in (2) accounts for phase shifts in the spectrum. Phase shifts occur as the covariance function $\gamma(h)$ of the process is no longer necessarily time-reversible in the multivariate setting, that is $\gamma(h) \neq \gamma(-h)$. Therefore, the off-diagonal elements of the spectral matrix of X_t contain complex valued elements which are not vanishing even at $\lambda = 0$ and which depend on the memory parameter d_a. These complex valued elements vanish if and only if the matrix G in (2) is diagonal or $d_a = d$ for all dimensions a.

Our model in equation (1) is a causal filter and therefore the phase parameter in the off-diagonal element $f_{ab}(\lambda_j)$ equals $\pi/2(d_a - d_b)$, as it has been pointed out by Kechagias and Pipiras (2015).

The spectral density of the process u_t in (1) is assumed to fulfill the local condition

$$f_u(\lambda) \sim G, \qquad \lambda \to 0.$$

This condition is fulfilled whenever u_t has the Wold decomposition $u_t = C(L)\varepsilon_t$, where $C(1)$ is finite and has full rank, and $C(L)$ is a polynomial in the lag operator with absolute summable weights.

Furthermore, define the periodogram of X_t evaluated at frequency λ as

$$I(\lambda) = w(\lambda) w^*(\lambda), \qquad \text{with} \qquad w(\lambda) = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} X_t e^{it\lambda}.$$

In the rest of the paper the superscript 0 denotes the true value of a parameter, for example d^0 is the true memory parameter.

We need to state the following assumptions which follow those in Shimotsu (2007):

Assumption 1. *For $\beta \in (0,2]$ and $a,b = 1,\dots,q$ as $\lambda \to 0+$*

$$f_{ab}(\lambda) - \exp\left(i(\pi-\lambda)\left(d_a^0 - d_b^0\right)/2\right)\lambda^{-d_a^0-d_b^0}G_{ab}^0 = O\left(\lambda^{-d_a^0-d_b^0+\beta}\right).$$

Here and in the following f_{ab} and G_{ab} are the respective elements of the matrices f and G.

Assumption 2. *It holds that*

$$X_t - EX_t = A(L)\varepsilon_t = \sum_{j=0}^{\infty} A_j \varepsilon_{t-j},$$

with $\sum_{j=0}^{\infty} \|A_j\|^2 < \infty$ and $\|\cdot\|$ denotes the supremum norm. It is assumed that $E(\varepsilon_t|\mathfrak{F}_{t-1}) = 0$, $E(\varepsilon_t\varepsilon_t^{'}|\mathfrak{F}_{t-1}) = I_q$ a.s. for $t = 0,\pm1,\pm2,\dots$ where $\mathfrak{F}_t$ denotes the σ-field generated by ε_s and I_q is an identity matrix, $s \le t$. Furthermore, there exists a scalar random variable ε such that $E\varepsilon^2 < \infty$ and for all $\tau > 0$ and some $K > 0$ it is $P(\|\varepsilon_t\|^2 > \tau) \le KP(\varepsilon^2 > \tau)$. In addition, it holds for $a,b,c,d = 1,2$, $t = 0,\pm1,\pm2,\dots$ that

$$E(\varepsilon_{at}\varepsilon_{bt}\varepsilon_{ct}|\mathfrak{F}_{t-1}) = \mu_{abc} \qquad a.s.$$

and

$$E(\varepsilon_{at}\varepsilon_{bt}\varepsilon_{ct}\varepsilon_{dt}|\mathfrak{F}_{t-1}) = \mu_{abcd} \qquad a.s.,$$

where $|\mu_{abc}| < \infty$ and $|\mu_{abcd}| < \infty$.

Assumption 3. *In a neighborhood $(0,\delta)$ of the origin, $A(\lambda) = \sum_{j=0}^{\infty} A_j e^{ij\lambda}$ is differentiable and*

$$\frac{\partial}{\partial\lambda}\, {}_aA(\lambda) = O\left(\lambda^{-1}\|{}_aA(\lambda)\|\right), \qquad \lambda \to 0+,$$

where ${}_aA(\lambda)$ is the a-th row of $A(\lambda)$.

Assumption 4. *As $T \to \infty$ it holds for any $\gamma > 0$*

$$\frac{1}{m} + \frac{m^{1+2\beta}(\log m)^2}{T^{2\beta}} + \frac{\log T}{m^{\gamma}} \to 0,$$

where m is the bandwidth parameter.

Assumption 5. *There exists a finite real matrix Q such that*

$$\Lambda_j\left(d^0\right)^{-1} A(\lambda_j) = Q + o(1), \qquad \lambda_j \to 0.$$

These assumptions are multivariate versions of the assumptions in Qu (2011). They allow for non-Gaussianity. Assumption 1 and 5 are satisfied by multivariate ARFIMA processes. Assumption 2 excludes conditional heteroscedasticity as it is, but it is possible to relax this assumption which is omitted here for the ease of presentation. Assumption 4 is slightly stronger than the assumption used in Qu (2011) for the univariate local Whittle estimator. However, this stronger assumption is necessary for the Hessian of the objective function of the local Whittle estimator to converge, which is needed in our proof. It should be mentioned that Assumption 4 gives a sharp upper bound for the number of frequencies m which can be used for the local Whittle estimator and thus for our test statistic. It is $m = o(T^{0.8})$.

3 Testing for Spurious Long Memory

In this section we consider multivariate testing for pure long memory. Our test is spectral based and uses the different properties of the periodogram of long-memory processes and processes with structural breaks or trends. Special use will be made from the fact that the slope of the spectral density of a process with structural breaks is nearly zero for $j > \sqrt{T}$.

3.1 The MLWS Statistic

To be concrete, we are interested in testing the hypothesis that the spectral density local to the origin has the shape given in equation (2):

$$H_0 : f(\lambda_j) \sim \Lambda_j(d) G \Lambda_j^*(d)$$

as $\lambda \to 0+$ with $d_a \in (-1/2, 1/2)\ \forall\ a = 1, \dots, q$. Thus, under the null hypothesis X_t is a multivariate long-memory process. The alternative is that X_t contains spurious long memory, this is X_t is contaminated either by level shifts or smooth trends. Note that this does not exclude the existence of a long memory component under the alternative. To motivate our test statistic, we first need to consider the properties of the periodogram under the alternative. We consider two alternative models. The first one is a multivariate random level shift model defined by

$$\begin{aligned} X_t &= \mu_t + \kappa_t \quad \text{with} \\ \mu_t &= (I_q - \Pi_t \Phi)\mu_{t-1} + \Pi_t e_t, \end{aligned} \tag{3}$$

where κ_t, $\Pi_t = diag(\pi_{1t}, \dots, \pi_{qt})$ and e_t are mutually independent. The Bernoulli variables

π_{it} and π_{jt} for the different dimensions of the q-dimensional process X_t are correlated with correlation matrix Υ for $i, j = 1, ..., q$. We consider a shift probability that is defined by $p = p^*/T$, where p^* is the expected number of shifts. Furthermore, the magnitude of the shifts is characterized by the q-dimensional column vector e_t, with $e_t \sim N(0, \Sigma_e)$, and the noise process κ_t, $\kappa_t \sim N(0, \Sigma_\kappa)$. The pairwise correlation coefficients of π_{it} and π_{jt}, e_{it} and e_{jt}, and κ_{it} and κ_{jt} are labeled as $\rho_{\pi,ij}$, $\rho_{e,ij}$ and $\rho_{\kappa,ij}$, $\forall\ i, j = 1, ..., q$.
The q-dimensional diagonal matrix $\Phi = diag(\phi, \ldots, \phi)$, with $0 \leq \phi \leq 1$, contains autoregressive coefficients and determines the persistence of the level shifts. This allows us to consider stationary as well as non-stationary multivariate random level shift processes. This formulation of our random level shift model is a multivariate version of the autoregressive random level shift process suggested in Xu and Perron (2014).
The second model under the alternative is the smooth trend model:

$$X_t = H\left(\frac{t}{T}\right) + \kappa_t, \tag{4}$$

where all variables are q-dimensional column vectors, $H(t/T) = (h_1(t/T), \ldots, h_q(t/T))'$ and $h_i(t/T)$ is a Lipschitz continuous function on $[0, 1]$, $\forall\ i = 1, ..., q$. The noise term κ_t is defined as in equation (3).
In analogy to Perron and Qu (2010), the periodogram of X_t in (3) or (4) can be decomposed in three components by

$$\begin{aligned} I_X(\lambda_j) &= \frac{1}{2\pi T}\left(\sum_{t=1}^{T}\kappa_t \exp(i\lambda_j t)\right)\left(\sum_{t=1}^{T}\kappa_t \exp(i\lambda_j t)\right)^* \\ &+ \frac{1}{2\pi T}\left(\sum_{t=1}^{T}\mu_t \exp(i\lambda_j t)\right)\left(\sum_{t=1}^{T}\mu_t \exp(i\lambda_j t)\right)^* \\ &+ \frac{2}{2\pi T}\sum_{t=1}^{T}\sum_{s=1}^{T}\kappa_t \mu_t' \cos(\lambda_j(t-s)), \end{aligned}$$

where again the asterix denotes the complex conjugate. By similar arguments as in Proposition 3 of Perron and Qu (2010) for $\lambda_j = o(1)$ the first summand is of order $O_P(1)$, the second is of order $O_P(T^{-1}\lambda_j^{-2})$, and the third term is of order $O_P(T^{-1/2}\lambda_j^{-1})$. Therefore, for each component in X_t the level shifts affect the periodogram only up to $j = O(T^{1/2})$. The stochastic orders are exact in the case of level shifts as in equation (3) and approximate for slowly varying trends in (4).
This decomposition of the periodogram can now be used to construct a multivariate local Whittle score-type test (MLWS test). It is based on the difference between the spectral

density of a fractionally integrated process and the periodogram of a series contaminated by mean shifts or smooth trends that is almost flat for frequencies $m > \sqrt{T}$. This property also explains why the bias of the estimate $\hat{d}$ of the memory parameter depends heavily on the bandwidth choice if a local semiparametric estimator is used.

The test statistic is based on the derivative of the local Whittle likelihood function evaluated at $\hat{d}$, where $\hat{d}$ is the local Whittle estimate obtained using the first m Fourier frequencies. Qu (2011) now evaluates the derivative of the local Whittle likelihood function at the first $[mr]$ Fourier frequencies, where $r \in [\varepsilon, 1]$ with $\varepsilon > 0$. For $r = 1$ the derivative is exactly zero and for smaller r the derivative should be close to zero as long as the estimate of d remains stable when the bandwidth is decreased. This is the case under the null hypothesis. If the alternative is true, the non-uniform behavior of the spectral density leads to a divergence of the derivative. The test statistic is obtained by taking the supremum of the derivative over all r.

Our test statistic extends this idea to the multivariate case. It is based on the weighted sum of the partial derivatives of the multivariate local Whittle likelihood as defined in Shimotsu (2007), which is given in (2).

As the Gaussian log-likelihood of X_t and G are real, the local Whittle likelihood localized to the origin can be written as

$$Q_m(G,d) = \frac{1}{m}\sum_{j=1}^{m}\left\{\log\det\Lambda_j(d)G\Lambda_j^*(d) + tr\left[G^{-1}Re\left[\Lambda_j(d)^{-1}I(\lambda_j)\Lambda_j^*(d)^{-1}\right]\right]\right\}. \tag{5}$$

The first order condition with respect to G gives

$$G = \frac{1}{m}\sum_{j=1}^{m}Re\left[\Lambda_j(d)^{-1}I(\lambda_j)\Lambda_j^*(d)^{-1}\right].$$

Substituting this into $Q_m(G,d)$ and

$$\begin{aligned}\log\det\Lambda_j(d) + \log\det\Lambda_j^*(d) &= \log\det\Lambda_j(d)\Lambda_j^*(d)\\ &= \log\det\left(diag\left(\lambda_j^{-2d_a}\right)\right)\\ &= -2\sum_{a=1}^{q}d_a\log\lambda_j\end{aligned}$$

gives the objective function of the multivariate Gaussian semiparametric estimate (GSE) of Shimotsu (2007):

$$R(d) = \log\det\hat{G}(d) - 2\sum_{a=1}^{q}d_a\frac{1}{m}\sum_{j=1}^{m}\log\lambda_j \tag{6}$$

with

$$\hat{G}(d) = \frac{1}{m}\sum_{j=1}^{m} Re\left[\Lambda_j(d)^{-1} I(\lambda_j)\Lambda_j^*(d)^{-1}\right].$$

To state our test statistic we need to introduce some algebra on the first derivative of the objective function $R(d)$ which is condensed in Lemma 1 below. Denote by $\eta = (\eta_1,\ldots,\eta_q)'$ a $(q\times 1)$ vector of real numbers and $\nu_j = log\lambda_j - 1/m\sum_{j=1}^{m} log\lambda_j$. Furthermore, set ${}_aG^{-1}$ to be the a-th row of G^{-1} and set i_a to be the $(q\times q)$ matrix with a one on the a-th diagonal element and zeros elsewhere. Additionally, M_a denotes the a-th column of the matrix M. Then, we can write:

Lemma 1. *Under Assumptions 1 to 5 we have*

$$\begin{aligned}\sum_{a=1}^{q}\eta_a\sqrt{m}\frac{\partial R(d)}{\partial d_a} &= \frac{2}{\sqrt{m}}\sum_{a=1}^{q}\eta_a\sum_{j=1}^{m}\nu_j\left({}_aG^{-1}Re\left[\Lambda_j(d)^{-1}I(\lambda_j)\Lambda_j^*(d)^{-1}\right]_a - 1\right)\\ &+\frac{1}{\sqrt{m}}\sum_{a=1}^{q}\eta_a\sum_{j=1}^{m}(\lambda_j-\pi)_aG^{-1}Im\left[\Lambda_j(d)^{-1}I(\lambda_j)\Lambda_j^*(d)^{-1}\right]_a + o_P(1)\end{aligned}$$

The right hand side of Lemma 1 is the main ingredient of our test statistic which is asymptotically equivalent to the weighted sum of the components of the gradient vector. The test statistic is given by:

$$\begin{aligned}MLWS &= \frac{1}{2}\sup_{r\in[\varepsilon,1]}\Bigg\| \frac{2}{\sqrt{\sum_{j=1}^{m}\nu_j^2}}\sum_{a=1}^{q}\eta_a\sum_{j=1}^{[mr]}\nu_j\left({}_aG^{-1}(\hat{d})\left[Re\left[\Lambda_j(\hat{d})^{-1}I(\lambda_j)\Lambda_j^*(\hat{d})^{-1}\right]\right]_a - 1\right) \qquad (7)\\ &+\frac{1}{\sqrt{\sum_{j=1}^{m}\nu_j^2}}\sum_{a=1}^{q}\eta_a\left({}_aG^{-1}(\hat{d})\right)\sum_{j=1}^{[mr]}(\lambda_j-\pi)Im\left[\Lambda_j(\hat{d})^{-1}I(\lambda_j)\Lambda_j^*(\hat{d})^{-1}\right]_a\Bigg\|.\end{aligned}$$

Remark 1: The factor 1/2 is added in order to obtain comparability with the univariate case.

Remark 2: As usual, a small sample correction is applied by replacing $m^{-1/2}$ with $(\sum_{j=1}^{m}\nu_j^2)^{-1/2}$ which improves the size of the test and is asymptotically equivalent.

In the univariate case our test reduces exactly to that of Qu (2011). The imaginary part in our test statistic accounts for the phase shifts in the multivariate spectrum.

By combining the results of Shimotsu (2007) with those of Qu (2011) we are able to derive the limiting distribution of the test statistic (7). It is stated in the following theorem, where B(s) denotes standard one-dimensional Brownian motion:

Theorem 1. *Under Assumptions 1 to 5 we have for $T \to \infty$*

$$
\begin{aligned}
MLWS \quad \Rightarrow \quad & \frac{1}{2} \sup_{r \in [\epsilon,1]} \left\| \int_0^r \left[(1+\log s)^2 \left(2\eta'\eta + 2\eta' \left(G^0 \odot \left(G^0\right)^{-1} \right) \eta \right) \right. \right. \\
+ \quad & \left. \frac{\pi^2}{4} \left(2\eta' \left(G^0 \odot \left(G^0\right)^{-1} \right) \eta - 2\eta'\eta \right) \right]^{1/2} dB(s) \\
- \quad & 2\eta' B(1) \int_0^r (1+\log s) ds \eta \\
- \quad & \left. 2\eta' F(r) \eta \int_0^1 (1+\log s) dB(s) \right\|,
\end{aligned}
$$

where

$$
F(r) = \int_0^r (1+\log s)^2 ds \left(G^0 \odot \left(G^0\right)^{-1} + I_q \right) + \frac{\pi^2}{4} \left(G^0 \odot \left(G^0\right)^{-1} - I_q \right).
$$

The test statistic as it stands and its limiting distribution in Theorem 1 hold for any choice of the weight vector η. However, the test statistic is not pivotal as the limiting distribution depends on G^0 and thus on the unknown memory parameter d^0. Furthermore, the limiting distribution depends on the dimension q.

To overcome this problem, we fix the weighting scheme η to $\eta_a = 1/\sqrt{q}$, $\forall a = 1, \ldots, q$, to obtain a pivotal test independent of the unknown parameter d^0. Further, this choice guarantees that for every dimension q the limiting distribution is exactly the same as in Qu (2011). This is stated in the following lemma:

Lemma 2. *Under Assumptions 1 to 5 and setting $\eta_1 = \ldots = \eta_q = 1/\sqrt{q}$ we have for $T \to \infty$*

$$
\begin{aligned}
MLWS \quad \Rightarrow \quad & \sup_{r \in [\epsilon,1]} \left\| \int_0^r (1+\log s) dB(s) \right. \\
- \quad & B(1) \int_0^r (1+\log s) ds \\
- \quad & \left. F(r) \int_0^1 (1+\log s) dB(s) \right\|,
\end{aligned}
$$

where

$$
F(r) = \int_0^r (1+\log s)^2 ds.
$$

Remark 3: For $\varepsilon = 0.02$, the asymptotic critical values of the MLWS test are given by 1.118, 1.252, 1.374, and 1.517 for a 10%, 5%, 2.5%, and 1% significance level respectively. The corresponding critical values for a larger trimming parameter, $\varepsilon = 0.05$, equal 1.022, 1.155, 1.277, and 1.426, as shown by Qu (2011).

After deriving the limiting distribution of the test, we have to prove its consistency under the alternatives (3) and (4). This is done in the following theorem:

Theorem 2. *Suppose that the process X_t is generated by (3) or (4). Furthermore, assume that $m/T^{1/2} \to \infty$, $P(\hat{d}_j - d_j^0 > \epsilon) \to 1$, $P(m^{-1}\sum_{j=1}^{m} I_j(\lambda)\lambda_j^{2\hat{d}} > \epsilon) \to 1$ for each j as $T \to \infty$ with ϵ being some arbitrary small constant, and assume that Assumptions 1 to 5 hold. Then, $MLWS \xrightarrow{P} \infty$ as $T \to \infty$.*

3.2 MLWS Test for Fractionally Cointegrated Series

So far, fractional cointegration has been ruled out by our assumptions on the matrix G, which has reduced rank if components of X_t are cointegrated. However, our test can be robustified against fractional cointegration without altering the limiting distribution. In order to do this, we adopt a stationary fractional cointegration setup with exactly one cointegrating relationship. As before, the observed series X_t follows $X_t = D(d_1,...,d_q)^{-1}u_t$, as in (1). However, now the components of X_t obey the cointegrating relation

$$X_{1t} = \beta' X_{(-1)t} + w_t, \tag{8}$$

where $X_{(-1)t}$ denotes the vector X_t but excluding the first element, $\beta = (\beta_2,...,\beta_q)'$ and $w_t \sim I(d_w)$. We assume with no loss in generality that X_{1t} is the dependent variable since the ordering in X_t is arbitrary. Since fractional cointegration requires an equal degree of integration in the relevant component series, it is required that $d_1 = \max_{2\leq a\leq q} d_a$, where the max only considers those a, where $\beta_a \neq 0$. The component series are cointegrated if the residual series w_t has reduced memory so that $0 \leq d_w < d_1 < 1/2$. This is a standard setup for stationary fractional cointegration, similar to that of Nielsen (2007), but allowing for phase shifts in the spectrum.

If the cointegrating vector β is known, the MLWS test can be carried out on the transformed series $\mathbf{B}X_t = (w_t, X_{2t},...,X_{qt})'$. The matrix $\mathbf{B}$ is $(q\times q)$ with ones on the diagonal elements and the off-diagonal elements in the first row of $\mathbf{B}$ correspond to the elements of the cointegrating vector β, that is $B_{1a} = \beta_a$, for all $a = 2,...,q$. All other elements of $\mathbf{B}$ are zeros. This means if fractional cointegration is present, one of the cointegrated variables is removed from X_t and replaced by the cointegration residuals w_t.

The transformed series $\mathbf{B}X_t$ has the spectral density

$$f(\lambda) \sim \Lambda_j(d)\tilde{G}\Lambda_j^*(d),$$

with $\Lambda_j(d)$ as in (2) but with $d_1 = d_w$. Importantly, after this transformation $\tilde{G}$ is a real, positive definite, symmetric and finite matrix, which is obviously different from the one in (2).

The cointegration robust test statistic is therefore given by

$$\begin{aligned}\widetilde{MLWS} \quad = \quad & \frac{1}{2}\sup_{r\in[\varepsilon,1]}\left\| \frac{2}{\sqrt{\sum_{j=1}^{m}\nu_j^2}}\sum_{a=1}^{q}\eta_a\sum_{j=1}^{[mr]}\nu_j\left({}_a\tilde{G}^{-1}(\hat{d})\left[Re\left[\Lambda_j(\hat{d})^{-1}I(\lambda_j)\Lambda_j^*(\hat{d})^{-1}\right]\right]_a - 1\right)\right. \\ & \left. + \frac{1}{\sqrt{\sum_{j=1}^{m}\nu_j^2}}\sum_{a=1}^{q}\eta_a\left({}_a\tilde{G}^{-1}(\hat{d})\right)\sum_{j=1}^{[mr]}(\lambda_j-\pi)Im\left[\Lambda_j(\hat{d})^{-1}I(\lambda_j)\Lambda_j^*(\hat{d})^{-1}\right]_a\right\|.\end{aligned}$$

Exactly like the original MLWS test, the modified statistic $\widetilde{MLWS}$ has the limit distribution given in Lemma 2.

So far, we assumed that the cointegrating vector β in (8) is known, since estimating the cointegrating relationship is beyond the scope of this paper. In empirical applications β is often implied by economic theory. If this is not the case, the procedure can be made feasible if a consistent estimator for β is applied that converges with a faster rate than $\sqrt{m}$ which is the bandwidth used for the MLWS statistic. A possible choice is the fully modified narrow-band least squares (FMNBLS) estimator proposed by Nielsen and Frederiksen (2011), for which $\beta_a - \beta_a^0 = O_p\left(T^{\delta_{FMNBLS}/2+(\delta_{FMNBLS}-1)(d_w-d_a)}\right)$, if it is applied with bandwidth $m = \lfloor T^{\delta_{FMNBLS}}\rfloor$, where $0 < \delta_{FMNBLS} < 1$.[1] Depending on the strength $(d_w - d_1)$ of the cointegrating relationship, this estimator can achieve convergence rates close to parametric ones. It is always possible to achieve a convergence rate faster than $\sqrt{m}$, if

$$\delta_{FMNBLS} > \frac{(d_w-d_1)+\delta/2}{(d_w-d_1)+1/2} \tag{9}$$

is selected as the bandwidth parameter for the FMNBLS estimator. Since $0.5 < \delta < 0.8$ is required for the bandwidth parameter δ used to determine m for the MLWS test, the right hand side of (9) is always less than one, so that there is a δ_{FMNBLS} that fulfills the condition. Simulation studies, not reported here, show that this is sufficient to leave the distribution of the MLWS statistic applied to the estimated series $\hat{\mathbf{B}}X_t$ unaffected - even

[1] cf. Nielsen and Frederiksen (2011) for further conditions on m.

in relatively small samples.[2] Therefore, the assumption of known β does not restrict the applicability of our approach in practice.

3.3 Pre-Whitening

Although the MLWS statistic in (7) is asymptotically independent of short memory dynamics, we need to apply a pre-whitening procedure to avoid negative effects on the size and power properties of the test in finite samples if short memory dynamics are present. Similar to Qu (2011), we do so by approximating the short memory dynamics of u_t in (1) with a low order VARMA process that is given by

$$u_t = \mathcal{A}(L)^{-1}\mathcal{M}(L)v_t,$$

where $\mathcal{A}(L) = I_q - \mathcal{A}_1 L - ... - \mathcal{A}_{p_A} L^{p_A}$ and $\mathcal{M}(L) = I_q - \mathcal{M}_1 L - ... - \mathcal{M}_{q_M} L^{q_M}$ are matrix lag polynomials and v_t is a $(q \times 1)$ Gaussian white noise with zero mean and variance-covariance matrix Σ_v.

Since $\mathcal{A}(L)$ and $D(d_1,\ldots,d_q)$ are not commutative in the multivariate case, there are two ways to generalize the univariate ARFIMA process to a vector process. This has been pointed out by Lobato (1997). For the specification used here, Sela and Hurvich (2009) coined the acronym FIVARMA because the process $X_t = D(d_1,\ldots,d_q)^{-1}\mathcal{A}(L)^{-1}\mathcal{M}(L)v_t$ can be interpreted as a fractionally integrated VARMA process. The process $X_t = \mathcal{A}(L)^{-1}D(d_1,\ldots,d_q)^{-1}\mathcal{M}(L)v_t$ is a vector autoregression of a fractionally integrated vector moving average process and is referred to as VARFIMA. The FIVARMA model has recently been applied in Chiriac and Voev (2011) and has first been studied by Sowell (1989). As Lütkepohl (2007) points out, the parameters of the unrestricted VARMA model are not identified. Thus, we follow Chiriac and Voev (2011) and estimate the model in its final equation form, where $\mathcal{A}(L)$ is a scalar operator. The estimation is carried out using the approximate maximum likelihood estimator of Beran (1995) which is based on an approximation of the AR(∞) representation of the FIVARMA process.[3]

Since the short memory dynamics can be approximated well with a low order model and the estimation of FIVARMA models is computationally very demanding, we restrict the model order to be $p_A = q_M = 1$. In analogy to Qu (2011), we then apply the filter $\hat{\mathcal{A}}(L)^{-1}\hat{\mathcal{M}}(L)X_t = \tilde{X}_t$ to the original series X_t. To test for spurious long memory we subsequently apply the MLWS test in (7) to the filtered series $\tilde{X}_t$.

[2] Monte Carlo results are available from the authors upon request.

[3] For the sample sizes considered here, this approach turns out to be faster than the method of Sela and Hurvich (2009). For larger samples the computing time can be further reduced by conducting the ML estimation on subsamples and using the average of these subsample estimators, as suggested by Beran and Terrin (1994).

Note that we do not assume that the short memory dynamics follow a VARMA(1,1)-process. We use it as a reasonable approximation to the true short memory dynamics in small samples. Asymptotically the test is unaffected by any form of short memory dependence because we only use the periodogram ordinates at Fourier frequencies local to the pole. The short memory dynamics have no influence on the shape of the pole. This is also why the pre-whitening procedure leaves the limiting distribution of the test unaffected.

4 Monte Carlo Study

To analyze the finite sample properties of the MLWS test, we conduct a Monte Carlo analysis that consists of three parts. In the first part, we consider a bivariate setup and conduct experiments to determine the influence of the bandwidth choice, $m = \lfloor T^{\delta} \rfloor$, and the choice of the trimming parameter ε on the size and the power of the test. Then, we turn to higher dimensional applications to analyze how the size and power depend on the dimension q of the multivariate process. Finally, in the third part we analyze the performance of the test if short memory dynamics exist. To disentangle the performance of the multivariate test from that of the pre-whitening procedure, the latter is only applied in the last part when short memory dynamics are present.
The simulation studies of Qu (2011) and Leccadito et al. (2015) show that the Qu test has good power against a wide range of different alternatives, such as non-stationary random level shifts, smooth trends, markov switching models, or the STOPBREAK proces of Engle and Smith (1999). Therefore, we focus on analyzing the properties that are specific to the multivariate case and use a stationary random level shift process for all power DGPs. Further simulation studies covering other forms of deterministic trends and conditional heteroscedasticity are included in a supplementary appendix, available online. All results presented hereafter are based on $M = 5000$ Monte Carlo replications and all tests are carried out with a nominal significance level of $\alpha = 0.05$.

4.1 Size and Power Comparison in a Bivariate Setup

The size study for the bivariate case is based on the multivariate fractionally integrated process from equation (1), where the short memory component $u_t = v_t$ with $v_t \sim N(0, \Sigma_v)$ is specified to be a bivariate white noise

$$D(d_1, d_2)X_t = v_t.$$

In this setup we want to investigate two aspects. First, we evaluate whether the size

			$\rho_v =$ **.0**				$\rho_v =$ **.8**			
T	d_1/d_2		.0		.4		.0		.4	
		δ/ε	**.02**	**.05**	**.02**	**.05**	**.02**	**.05**	**.02**	**.05**
250	.0	**.60**	.008	.011	.009	.010	.011	.015	.007	.010
		.65	.011	.016	.010	.015	.015	.023	.012	.020
		.70	.016	.019	.015	.018	.016	.021	.015	.022
		.75	.014	.022	.016	.022	.020	.028	.018	.023
	.4	**.60**			.007	.009			.006	.009
		.65			.009	.014			.012	.017
		.70			.013	.019			.014	.019
		.75			.017	.025			.020	.024
1000	.0	**.60**	.015	.022	.017	.021	.017	.025	.018	.023
		.65	.019	.028	.020	.024	.024	.030	.021	.028
		.70	.023	.031	.024	.031	.028	.037	.027	.031
		.75	.026	.037	.029	.039	.037	.049	.032	.038
	.4	**.60**			.015	.020			.015	.021
		.65			.021	.031			.022	.025
		.70			.025	.032			.025	.031
		.75			.029	.035			.030	.038

Table 1: Size of MLWS test for FIVARMA (0,d,0): $D(d_1,d_2)X_t = v_t$ with $v_t \sim N(0,\Sigma_v)$ and $\sigma_v^2 = 1$. The bandwidth m is determined by $m = \lfloor T^{\delta} \rfloor$.

depends on the correlation ρ_v between the components of the innovation vector v_t, or whether it depends on the (possibly different) degrees of memory d_1 and d_2 in the two series. Second, we want to determine the effect of the bandwidth m and the trimming parameter ε. Since the trimming parameter ε can be chosen discretionary, we follow Qu (2011) and conduct our simulations for $\varepsilon \in \{0.02, 0.05\}$.

Table 1 shows the results. We find that the test is generally conservative in finite samples - a feature which it shares with its univariate version. For all parameter constellations, the size is better with $\varepsilon = 0.05$ than with $\varepsilon = 0.02$ and it is increasing in m. The results also improve as the sample size increases. With a sample size of $T = 1000$, $m = \lfloor T^{0.75} \rfloor$ and $\varepsilon = 0.05$ for example, we find that the size is between 3.5 and 4.9 percent for all combinations of ρ_v, d_1, and d_2. Thus, in larger samples the MLWS test achieves good size properties with the right choice of m and ε.

With regard to the correlation ρ_v between the innovations, the size tends to improve as the correlation increases since the MLWS test makes use of the coherence information. Overall, even though the test is quite conservative in small samples, the size is good

T	$\rho = \delta$ / δ/ε	.0		.5		1	
		.02	**.05**	**.02**	**.05**	**.02**	**.05**
250	**.60**	.134	.160	.147	.171	.244	.254
	.65	.300	.357	.316	.379	.421	.449
	.70	.510	.542	.537	.575	.587	.588
	.75	.739	.749	.746	.760	.735	.731
1000	**.60**	.724	.759	.730	.764	.776	.791
	.65	.934	.942	.934	.945	.914	.918
	.70	.987	.989	.989	.988	.965	.964
	.75	.996	.995	.998	.997	.988	.987

Table 2: Power of MLWS test against stationary random level shifts: $Y_t = \mu_t + v_t$ with $v_t \sim N(0, \Sigma_v)$ and $\mu_t = (I_q - \Pi_t)\mu_{t-1} + \Pi_t e_t$. The bandwidth m is determined by $m = \lfloor T^\delta \rfloor$.

in larger samples and it is remarkably stable for different degrees of memory in the components of the series and correlations among the innovation sequences.

We will now turn to the effect of m and ε on the power of the test. In contrast to the true long-memory processes under the null hypothesis, that we denote by X_t, the DGP in the power study will be denoted by Y_t. It is the sum of the white noise sequence v_t and the multivariate stationary random level shift process μ_t from equation (3) with $\Phi = I_q$:

$$\begin{aligned} Y_t &= \mu_t + v_t \\ \mu_t &= (I_q - \Pi_t)\mu_{t-1} + \Pi_t e_t. \end{aligned} \tag{10}$$

The shift probability is always kept at $p = 5/T$, so that in expectation there are five shifts in every sample and the standard deviation of the shifts is $\sigma_e = 1$. Since a different behavior of the breaks can imply different phase and coherence information, we consider different values for the correlation between the occurrence of shifts ρ_π and the correlation of the shift sizes ρ_e. For simplicity, we always set $\rho_\pi = \rho_e$. If $\rho_\pi = \rho_e = 0$ shifts occur independently in each of the components of the series, whereas shifts always coincide in timing and size if $\rho_\pi = \rho_e = 1$.[4]

The results of this experiment are shown in Table 2. We find that the power is always increasing in the bandwidth. For small sample sizes with weakly correlated shifts the

[4]Since the presence of spurious long memory depends on the location of the shifts in the sample, we discard all samples for which a test, for $H_0: d = 0$ based on the local Whittle estimate $\hat{d}_{LW}$, is not rejected for all components.

		Size							Power					
		MLWS test			Qu test				MLWS test			Qu test		
T	**q/ρ_v**	**.0**	**.4**	**.8**	**.0**	**.4**	**.8**	**q/$\rho_\pi = \rho_e$**	**.0**	**.5**	**1**	**.0**	**.5**	**1**
100	**1**	.012	.015	.012	.012	.015	.012	**1**	.124	.132	.125	.124	.132	.125
	2	.014	.016	.014	.007	.012	.012	**2**	.256	.255	.314	.157	.148	.160
	3	.015	.013	.014	.006	.008	.006	**3**	.369	.365	.453	.164	.152	.167
	4	.013	.013	.018	.007	.008	.009	**4**	.455	.460	.520	.168	.177	.180
	5	.015	.015	.011	.006	.006	.007	**5**	.548	.522	.547	.186	.175	.195
250	**1**	.021	.021	.018	.021	.021	.018	**1**	.453	.468	.453	.453	.468	.453
	2	.020	.021	.024	.017	.016	.017	**2**	.743	.736	.726	.602	.593	.549
	3	.023	.022	.020	.016	.014	.013	**3**	.886	.882	.831	.678	.669	.601
	4	.021	.021	.021	.015	.015	.014	**4**	.950	.943	.871	.743	.730	.641
	5	.023	.024	.022	.015	.014	.010	**5**	.981	.977	.885	.795	.764	.657
500	**1**	.023	.032	.030	.023	.032	.030	**1**	.791	.801	.799	.791	.801	.799
	2	.025	.033	.030	.026	.029	.024	**2**	.961	.958	.937	.929	.928	.880
	3	.029	.027	.030	.026	.020	.021	**3**	.993	.991	.964	.975	.970	.894
	4	.026	.028	.033	.020	.020	.024	**4**	.999	.999	.975	.990	.987	.917
	5	.025	.028	.027	.021	.019	.023	**5**	1.000	1.000	.982	.997	.993	.927

Table 3: Size and power of MLWS test and repeated Qu test with Simes correction for increasing dimensions q. **Left panel**: Size for FIVARMA (0,d,0): $D(d_1,\ldots,d_q)X_t = v_t$. **Right panel**: Power for $Y_t = \mu_t + v_t$ with $v_t \sim N(0,\Sigma_v)$.

test has better power with $\varepsilon = 0.05$, but in larger samples $\varepsilon = 0.02$ leads to a higher power if m is also relatively large. With regard to the correlation of the shifts, we find that the power of the test increases in small samples if shifts show a stronger correlation. In large samples the power slightly decreases if shifts are perfectly correlated.

Overall in the bivariate setup, the test shows good size and power properties and for an increasing bandwidth both size and power improve. Note however that a larger bandwidth also makes the test more prone to errors if short memory dynamics are present. Therefore, we will address the choice of ε and m in practice later.

4.2 The Effect of Increasing Dimensionality

Since the proposed MLWS test is multivariate and its limiting distribution is independent of the dimension q of the process, we now consider how its finite sample properties depend on the dimension q.

As before, our size DGP, $D(d_1,...,d_q)X_t = v_t$, is a fractionally integrated white noise. Motivated by our previous findings, we set $m = \lfloor T^{0.75} \rfloor$ and $\varepsilon = 0.05$ and consider only the effect of increasing the dimension q.

Since there is no other multivariate test against spurious long memory available in the literature, a practitioner has no other choice but to apply the Qu test to each of the q components of the process separately. We will use this approach as a benchmark procedure. To avoid Bonferroni errors, we apply the correction of Simes (1986) that consists in ordering the p-values in ascending order and then comparing them with α/q, $2\alpha/q$, ..., α. The null hypothesis is rejected if any of the ordered p-values exceeds its respective threshold.

Note that for $q = 1$ the MLWS test and the Qu test are identical. The left panel of Table 3 contains the results. We can observe that the MLWS test is quite conservative in small samples, but the size improves if the sample size increases. It also maintains approximately the same size independent of the dimension q and independent of the correlation among the components of v_t. For the repeated application of the Qu test we find that similar to the MLWS test it is conservative in small samples. In addition, the size tends to further decrease with increasing q and with increasing correlation ρ_v between the noise components, which is an effect of the Simes correction.

As in the bivariate setup, the power DGP, $Y_t = \mu_t + v_t$, is the sum of the q-dimensional white noise v_t and the q-dimensional multivariate random level shift model from equation (10). Similar to the size DGP, we restrict the correlations of shifts in the components as well as the correlation of the shift sizes to be the same among all components such that $\rho_{\pi,ab} = \rho_{e,ab} = \rho_\pi = \rho_e$ for all $a \neq b$.

If we consider the results on the right hand side in Table 3, we find that there are indeed large power gains compared to the repeated application of the Qu test. For $T = 100$ these can be more than 36 percent. While correlated shifts increase the power in smaller samples, the power reduction observed in the bivariate simulations increases with increasing q. We find that the power is increasing in q and T. For larger dimensions the power is already close to one with only 250 observations.

4.3 Short Memory Dynamics

So far we have considered the MLWS test applied directly to the observed series X_t. As discussed in Section 4, the performance of the local Whittle based methods can be negatively affected in finite samples if short memory dynamics are present. This is why we suggested a pre-whitening procedure based on the FIVARMA(1,d,1). Subsequently, the test is applied to the filtered series $\tilde{X}_t$. The performance of this procedure is analyzed in the following.

		Pre-whitened: $\tilde{X}_t$								Unfiltered: X_t			
DGP		1		2		3		4		1		2	
T	δ/ε	**.02**	**.05**	**.02**	**.05**	**.02**	**.05**	**.02**	**.05**	**.02**	**.05**	**.02**	**.05**
250	**.60**	.006	.008	.006	.007	.002	.003	.006	.006	.006	.013	.006	.013
	.65	.007	.010	.005	.007	.002	.004	.008	.014	.012	.020	.014	.030
	.70	.012	.008	.004	.005	.001	.005	.007	.011	.014	.016	.043	.083
	.75	.007	.010	.002	.002	.001	.005	.006	.010	.015	.025	.138	.217
500	**.60**	.013	.011	.013	.015	.009	.009	.010	.013	.016	.015	.010	.018
	.65	.011	.015	.011	.019	.005	.010	.010	.017	.017	.024	.017	.034
	.70	.016	.020	.010	.013	.003	.005	.015	.017	.023	.028	.059	.110
	.75	.015	.013	.003	.004	.001	.003	.008	.014	.025	.032	.255	.340
1000	**.60**	.017	.023	.017	.026	.020	.020	.016	.020	.018	.025	.014	.019
	.65	.020	.022	.026	.031	.017	.014	.016	.020	.027	.027	.016	.039
	.70	.018	.021	.021	.023	.013	.009	.022	.025	.028	.032	.068	.116
	.75	.019	.021	.017	.016	.004	.005	.021	.018	.034	.038	.356	.464
2000	**.60**	.019	.022	.014	.025	.016	.022	.015	.028	.018	.030	.012	.026
	.65	.019	.028	.026	.026	.024	.024	.023	.023	.024	.026	.022	.035
	.70	.027	.032	.027	.028	.014	.022	.025	.027	.033	.038	.081	.115
	.75	.021	.031	.024	.016	.006	.006	.023	.025	.039	.039	.450	.549

Table 4: Size of the MLWS test for FIVARMA(1,d,1): $(1-\phi_1 L)D(d_1,d_2)X_t = (I_q - \mathcal{M}_1 L)v_t$ with and without pre-whitening. Parameters values of the respective DGPs are given above. The bandwidth m is determined by $m = \lfloor T^{\delta} \rfloor$.

We consider four different DGPs for the size. These are given by:

$$
\begin{aligned}
&\textbf{DGP1:} \quad D(0.2,0.3)X_t = v_t \\
&\textbf{DGP2:} \quad (1-0.4L)D(0.2,0.3)X_t = v_t \\
&\textbf{DGP3:} \quad (1-0.6L)D(0.2,0.3)X_t = v_t \\
&\textbf{DGP4:} \quad (1-0.4L)D(0.2,0.3)X_t = \left[I_2 - \begin{pmatrix} 0.5 & 0 \\ 0 & 0.3 \end{pmatrix} L \right] v_t .
\end{aligned}
$$

For all processes we set $\sigma_v^2 = 1$, $\rho_v = 0.5$ and $p = 5/T$. All of these processes are special cases of the FIVARMA(1,d,1). DGP1 is a simple bivariate fractional white noise, while DGP2 and DGP3 contain autoregressive dynamics. Finally, DGP4 contains both autoregressive and moving average dynamics.

For the power studies we combine the respective size DGP X_t with the multivariate random level shift μ_t:

$$
\begin{aligned}
Y_t &= X_t + \mu_t \\
\mu_t &= (I_q - \Pi_t)\mu_{t-1} + \Pi_t e_t .
\end{aligned}
$$

		Pre-whitened: $\tilde{X}_t$								Unfiltered: X_t			
DGP		1		2		3		4		1		2	
T	δ/ε	**.02**	**.05**	**.02**	**.05**	**.02**	**.05**	**.02**	**.05**	**.02**	**.05**	**.02**	**.05**
250	**.60**	.028	.034	.025	.027	.011	.013	.027	.034	.041	.051	.013	.021
	.65	.060	.080	.045	.047	.015	.016	.066	.082	.093	.111	.019	.026
	.70	.113	.128	.057	.051	.014	.015	.116	.122	.192	.199	.023	.024
	.75	.177	.192	.056	.054	.013	.018	.190	.194	.312	.339	.030	.037
500	**.60**	.102	.103	.089	.082	.043	.030	.086	.095	.129	.127	.073	.060
	.65	.176	.203	.125	.153	.038	.047	.175	.212	.257	.297	.078	.096
	.70	.314	.326	.174	.175	.044	.039	.319	.344	.470	.499	.095	.094
	.75	.444	.446	.174	.149	.039	.035	.458	.455	.675	.683	.063	.073
1000	**.60**	.271	.292	.250	.247	.141	.135	.253	.290	.327	.352	.230	.234
	.65	.484	.484	.386	.374	.177	.155	.452	.483	.579	.594	.324	.303
	.70	.629	.637	.468	.397	.171	.131	.627	.627	.789	.781	.335	.267
	.75	.750	.729	.497	.414	.148	.098	.733	.736	.899	.891	.246	.190
2000	**.60**	.565	.598	.529	.547	.353	.345	.532	.555	.619	.647	.521	.536
	.65	.769	.757	.690	.662	.417	.346	.747	.752	.847	.857	.652	.611
	.70	.857	.818	.731	.691	.393	.287	.859	.831	.930	.934	.658	.599
	.75	.857	.796	.732	.617	.305	.146	.889	.824	.968	.957	.578	.409

Table 5: Power of the MLWS test for $Y_t = \mu_t + X_t$, with $(1-\phi_1 L)D(d_1,d_2)X_t = (I_q - \mathcal{M}_1 L)v_t$ with and without pre-whitening. Parameters values of the respective DGPs are given above. The bandwidth m is determined by $m = \lfloor T^{\delta} \rfloor$.

In the multivariate random level shift process we use $\rho_\pi = \rho_e = 0.9$ and $\sigma_e = 2$.
To investigate the costs and benefits of the pre-whitening procedure, the simulations for DGP1 and DGP2 are conducted with and without pre-whitening. The results for the size simulations are given in Table 4.
DGP1 is the baseline case. By comparing the results for the tests applied to the pre-whitened series $\tilde{X}_t$ with the results of the unfiltered series X_t, we can observe that the empirical size becomes a bit more conservative. However, for an increasing T the size approaches its nominal level. By considering DGP2, we see that there are considerable over-rejections if moderate autoregressive dynamics are present and no pre-whitening is applied, whereas with pre-whitening these distortions are successfully removed.
The results for DGP3 and DGP4 show that the pre-whitening procedure works well in controlling the size for different forms of short memory dynamics. As before, the size is generally better if $\varepsilon = 0.05$ is used. With regard to the bandwidth selection, the best size is observed most often for $\delta = 0.65$ or $\delta = 0.7$. It is no longer strictly increasing in m.
Table 5 considers the power results. For the baseline case there is a considerable power reduction caused by the additional flexibility introduced through the pre-whitening procedure. But when comparing the results for DGP1 and DGP2 without pre-whitening, we

observe that the power also suffers severely if there are short memory dynamics but no pre-whitening is applied. Using the filtering procedure reduces this effect substantially. The power is increasing in T and generally also in m, but we observe some power drops for larger bandwidths, especially if the autoregressive dynamics become more persistent. Similarly to Section 4.1, in small samples $\varepsilon = 0.05$ tends to give better results, whereas $\varepsilon = 0.02$ gives better power results in large samples.

In view of the size and power results presented here, the rule of thumb to choose $\varepsilon = 0.05$ for $T \leq 500$, that is suggested by Qu (2011), still works well in the bivariate case if short memory dynamics are present. Similarly, Qu's approach to use $m = \lfloor T^{0.7} \rfloor$ for the bandwidth is a good rule in the bivariate case.

5 Empirical Example

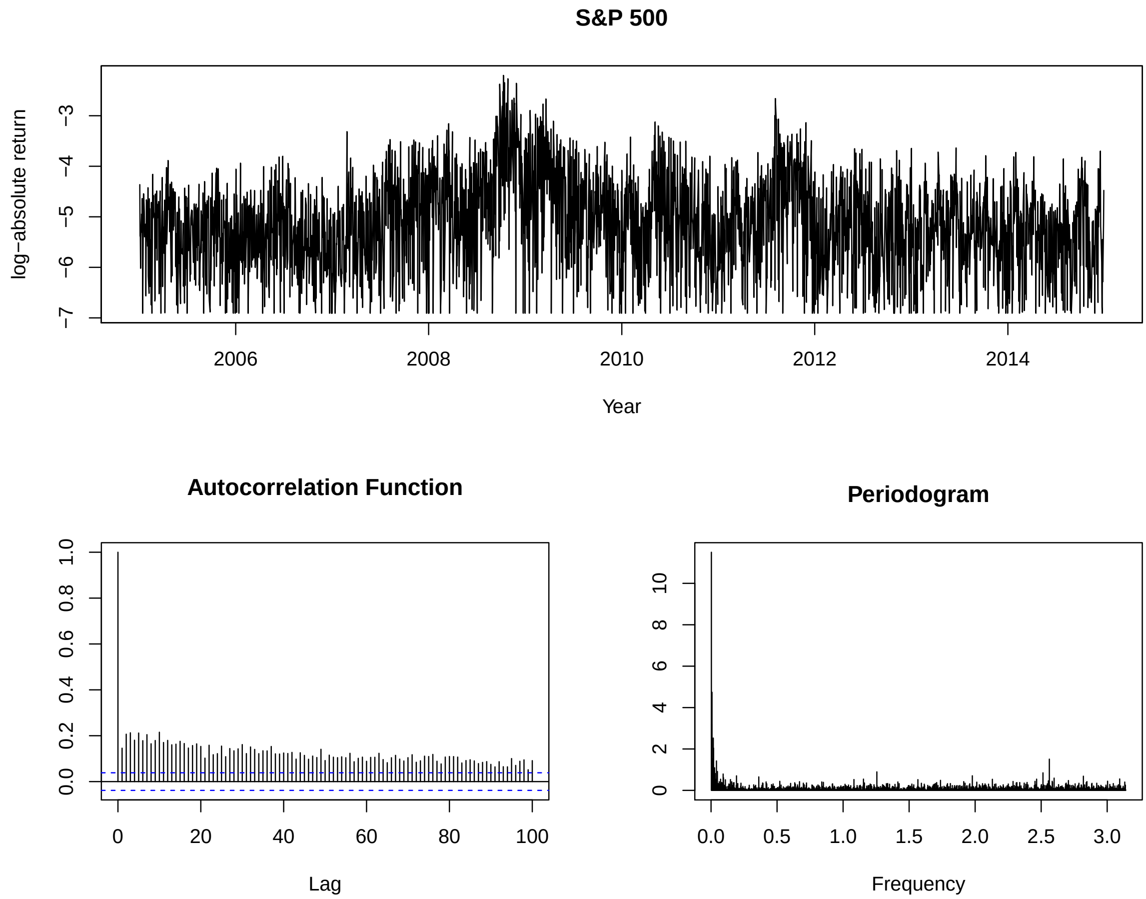

Figure 1: The log-absolute return of the S&P 500 series with the corresponding autocorrelation function and periodogram.

Volatility time series, such as log-absolute returns of stock market indices, are a typ-

	DAX	S&P 500	FTSE	NIKKEI
mean	-4.977	-5.156	-5.107	-4.924
median	-4.920	-5.133	-5.051	-4.803
std.dev.	0.924	0.937	0.900	1.004
skewness	-0.131	0.053	-0.087	-0.307
kurtosis	2.475	2.528	2.573	2.419

Table 6: Summary statistics of the log-absolute return series.

ical example in the spurious long memory literature. A frequently analyzed series in this context is the log-absolute return of the Standard & Poor 500 stock market index (hereafter S&P 500). The series is examined by Granger and Ding (1996) who find that it seems to follow a long-memory process. Nevertheless, they argue that long memory properties can be generated by other models than the standard $I(d)$ process. Granger and Hyung (2004) obtain a reduction of the estimated memory parameter by considering structural breaks in the series. Similarly, Varneskov and Perron (2011) consider a model allowing for both random level shifts and ARFIMA effects. Lu and Perron (2010) and Xu and Perron (2014) analyze the forecast performance of random level shift processes for the log-absolute returns of the S&P 500. In most cases, random level shift processes clearly outperform GARCH, FIGARCH and HAR models.

All these findings indicate spurious long memory in this volatility series. However, univariate tests are often not able to reject the null hypothesis of true long memory in volatility time series. Dolado et al. (2005), for example, apply their test to absolute and squared returns of the S&P 500, without being able to indicate spurious long memory. In view of the power gains of the multivariate procedure, demonstrated in Section 4.2, we revisit the log-absolute return series of the S&P 500 and additionally consider log-absolute returns of the DAX, FTSE and NIKKEI in a multivariate setup to test for spurious long memory using the MLWS test.

We analyze daily stock price indices from 2005/01/03 to 2014/12/31 (T=2608 observations) obtained from datastream. The log-returns are computed by first differencing the logarithm of the price index, $r_t = \ln(P_t) - \ln(P_{t-1})$. Subsequently, the log-absolute returns are calculated as $\ln(|r_t| + 0.001)$.[5] As an example, Figure 1 depicts the log-absolute returns of the S&P 500 series. Both, the autocorrelation function and the periodogram show the typical characteristics of a long-memory process. Since the series of the DAX,

[5]The constant 0.001 is added to avoid infinite values for zero returns, by following Lu and Perron (2010) and Xu and Perron (2014).

δ	DAX	S&P 500	FTSE	NIKKEI	partitions	coint.rank	$\hat{\beta}$	$\hat{d}_u$
0.60	0.379	0.472	0.393	0.295	(1,1,1,1)	1	(0.420, 0.421, -0.128)	0.224
0.65	0.338	0.411	0.362	0.290	(1,1,1,1)	1	(0.431, 0.419, -0.123)	0.210
0.70	0.328	0.359	0.303	0.285	(1,1,1,1)	1	(0.435, 0.415, -0.115)	0.183
0.75	0.264	0.300	0.290	0.252	(1,1,1,1)	1	(0.438, 0.422, -0.111)	0.139

Table 7: Fractional cointegration analysis based on local Whittle estimates of d with different bandwidths $m = \lfloor T^{\delta} \rfloor$.

FTSE and NIKKEI are highly correlated with that of the S&P 500, we omit plots of these series. Descriptive statistics for the dataset are given in Table 6. It can be seen that all four series have similar locations and standard deviations. With the exemption of the S&P 500, they are slightly negatively skewed and all series have lighter tails than the normal distribution.

Since the specification of the MLWS test depends on whether or not the series are fractionally cointegrated, we proceed by applying the semiparametric cointegrating rank estimation method of Robinson and Yajima (2002). The method consists of two steps. First, the vector series X_t is partitioned into subvectors with equal memory parameters using sequential tests for the equality of the d_a in each subvector. In the second step, the cointegrating rank of the relevant subvectors is estimated.

All results of this procedure are given in Table 7. The analysis is carried out for different bandwidths $m = \lfloor T^{\delta} \rfloor$ using the local Whittle estimator. It can be observed, that the estimates tend to decrease as the bandwidth increases, which indicates that the series indeed might be contaminated by level shifts.

Using the $\hat{T}_0$ statistic of Robinson and Yajima (2002) to test for the equality of the memory parameters, the null hypothesis cannot be rejected for any of the bandwidths, so that no further partitioning of X_t is necessary. Subsequently, the cointegrating rank of X_t is estimated. Again, the results are remarkably stable for different bandwidth choices. We find that there is one cointegrating relationship between the four volatility series.

As described in Section 3.2, the analysis than proceeds by estimating the cointegrating vector β in (8) using the FMNBLS estimator of Nielsen and Frederiksen (2011). The DAX series is specified to be the dependent variable. Subsequently, the transformed series $\hat{B}X_t$ are obtained. Additionally, we report the estimate $\hat{d}_u$ of the noise term in the last column of Table 7 to show that the condition of $0 \leq d_w < d_1 < 1/2$ is fulfilled.

To formally test for true long memory, we then apply the robustified multivariate local Whittle score-type test ($\widetilde{MLWS}$) to our system $\hat{B}X_t$. As a benchmark, we also apply the univariate test of Qu (2011) to each series separately. Because of the large number of observations the trimming parameter is set to $\varepsilon = 0.02$ for both tests. The corresponding

δ	Qu test DAX	S&P 500	FTSE	NIKKEI	MLWS ALL
0.60	0.521	0.515	0.446	0.860	1.334
	(0.862)	(0.871)	(0.949)	(0.314)	(0.030)
0.65	0.505	1.078	0.443	0.749	1.584
	(0.886)	(0.118)	(0.953)	(0.474)	(0.007)
0.70	0.395	1.100	0.739	0.519	1.630
	(0.983)	(0.107)	(0.492)	(0.865)	(0.004)
0.75	0.640	1.477	0.547	0.469	1.604
	(0.662)	(0.013)	(0.824)	(0.929)	(0.006)

Table 8: Test statistics of the Qu test applied to each series separately and the MLWS test applied to the multivariate series for different bandwidths $m = \lfloor T^{\delta} \rfloor$. p-values are given in brackets. Critical values are 1.252 and 1.374 for $\alpha = 5\%$ and $\alpha = 1\%$, respectively.

test statistics are given in Table 8, where the p-values are displayed in brackets. As one can see, Qu's univariate test fails to reject the null hypothesis of true long memory for nearly all bandwidths. The only exemption is the S&P 500 if the bandwidth is set to $m = \lfloor T^{0.75} \rfloor$. From these results one would conclude that the process is a pure long memory process that is not contaminated by level shifts. On the contrary, the MLWS test, that takes the information of the cross-spectrum into account, clearly rejects the null hypothesis in all cases, indicating the presence of spurious long memory due to level shifts. The application of the MLWS test therefore gives formal support to the arguments of Granger and Ding (1996) and Granger and Hyung (2004), among others, who argued that the memory in the log-absolute returns of the S&P 500 might be spurious.

6 Conclusion

This paper provides a multivariate score-type test for spurious long memory based on the objective function of the local Whittle estimator. The test statistic consists of a weighted sum of the partial derivatives of the concentrated local Whittle likelihood function. By introducing a suitable weighting scheme, the test statistic becomes pivotal and the limiting distribution becomes independent of the dimension of the data generating process. Our test encompasses the test of Qu (2011) as a special case for scalar processes. Consistency against multivariate random level shift processes and smooth trends is shown. Furthermore, we provide a modification of the test statistic in the case of fractionally cointegrated data which has the same asymptotic properties as the

original test statistic.
A Monte Carlo study shows that the test has good size and power properties in finite samples. These properties hold for different bandwidths, $m = \lfloor T^{\delta} \rfloor$, as well as for different trimming parameters ε. Furthermore, the size and power remain good if the dimensions of the data generating process increase and the test is robust against short memory dynamics if a pre-whitening procedure is applied.
Therefore, our test allows to consider multivariate problems in a multivariate setup instead of testing for spurious long memory in each series separately. By using information about the cross dependencies between time series we gain efficiency compared to a repeated application of a univariate test. This finding is exploited in our empirical example where we revisit the log-absolute returns of the S&P 500 together with the DAX, FTSE and NIKKEI stock indices in a multivariate framework. By applying our multivariate test, we find evidence of spurious long memory in those series. A simple application of the univariate Qu test to the univariate log-absolute returns, on the other hand, cannot reject the null hypothesis of true long memory.
As discussed in Section 5, several authors have pointed out that the log-absolute returns might follow a spurious long-memory process. Our empirical application adds to this literature by providing a formal rejection of pure long memory in the sense of a statistically significant test decision.

Appendix

Proof of Lemma 1:

To prove the lemma, note the following arguments in Shimotsu (2007)

$$\begin{aligned}
\sum_{a=1}^{q}\eta_a\sqrt{m}\frac{\partial R(d)}{\partial d_a} &= \sum_{a=1}^{q}\eta_a\sqrt{m}\left[-\frac{2}{m}\sum_{j=1}^{m}\log\lambda_j + tr\left[\hat{G}(d)^{-1}\frac{\partial\hat{G}^r(d)}{\partial d_a}\right]\right]\\
&= \sum_{a=1}^{q}\eta_a\sqrt{m}\left[-\frac{2}{m}\sum_{j=1}^{m}\log\lambda_j + tr\left[\hat{G}(d)^{-1}\left[\frac{1}{m}\sum_{j=1}^{m}\log\lambda_j\right.\right.\right.\\
&\quad\times Re\left[(\Lambda_j(d))^{-1}(i_aI(\lambda_j)+I(\lambda_j)i_a)(\Lambda_j^*(d))^{-1}\right]\\
&\quad\left.\left.\left.+\frac{1}{m}\sum_{j=1}^{m}\frac{\lambda_j-\pi}{2}Im\left[(\Lambda_j(d))^{-1}(-i_aI(\lambda_j)+I(\lambda_j)i_a)(\Lambda_j^*(d))^{-1}\right]\right]\right]\right]\\
&= \sum_{a=1}^{q}\eta_a\sqrt{m}\left[-\frac{2}{m}\sum_{j=1}^{m}\log\lambda_j + tr\left[\hat{G}(d)^{-1}\times H_{1a}+H_{2a}\right]\right]
\end{aligned}$$

with $H_{1a}=\frac{1}{m}\sum_{j=1}^{m}\log\lambda_j Re[(\Lambda_j(d))^{-1}(i_aI(\lambda_j)+I(\lambda_j)i_a)(\Lambda_j^*(d))^{-1}]$ and $H_{2a}=\frac{1}{m}\sum_{j=1}^{m}\frac{\lambda_j-\pi}{2}$ $Im[(\Lambda_j(d))^{-1}(-i_aI(\lambda_j)+I(\lambda_j)i_a)(\Lambda_j^*(d))^{-1}]$.

Therefore, we can write with $\nu_j=\log\lambda_j-1/m\sum_{j=1}^{m}\log\lambda_j$, ${}_aM$ denoting the a-th row of the matrix M, and M_a denoting the a-th column of the matrix M

$$\begin{aligned}
\sum_{a=1}^{q}\eta_a\sqrt{m}\frac{\partial R(d)}{\partial d_a} &= \sum_{a=1}^{q}\eta_a\left[tr\left[\hat{G}(d)^{-1}\frac{2}{\sqrt{m}}\sum_{j=1}^{m}\nu_j Re\left[\Lambda_j(d)^{-1}I(\lambda_j)\Lambda_j^*(d)^{-1}\right]i_a\right.\right.\\
&\quad\left.\left.+\hat{G}(d)^{-1}\frac{1}{\sqrt{m}}\sum_{j=1}^{m}\frac{\lambda_j-\pi}{2}Im\left[(\Lambda_j(d))^{-1}(-i_aI(\lambda_j)+I(\lambda_j)i_a)(\Lambda_j^*(d))^{-1}\right]\right]\right]\\
&= \frac{2}{\sqrt{m}}\sum_{a=1}^{q}\eta_a\sum_{j=1}^{m}\nu_j\left({}_a\left(G^0\right)^{-1}\left[Re\left[\Lambda_j(d)^{-1}I(\lambda_j)\Lambda_j^*(d)^{-1}\right]\right]_a-1\right)+o_P(1)\\
&\quad+tr\left[\hat{G}(d)^{-1}\frac{1}{\sqrt{m}}\sum_{j=1}^{m}\frac{\lambda_j-\pi}{2}Im\left[(\Lambda_j(d))^{-1}(-i_aI(\lambda_j)+I(\lambda_j)i_a)(\Lambda_j^*(d))^{-1}\right]\right]
\end{aligned}$$

which proves our lemma. □

Proof of Theorem 1:

To prove the theorem we start with the Taylor expansion

$$\frac{\partial R^r(d)}{\partial d}\Big|_{\hat{d}} = \frac{\partial R^r(d)}{\partial d}\Big|_{d^0} + \frac{\partial^2 R^r(d)}{\partial d \partial d'}\Big|_{\bar{d}} (\hat{d} - d^0) \tag{11}$$

where $\bar{d}$ fulfills $\|\bar{d} - d^0\| \leq \|\hat{d} - d^0\|$ and the notation $R^r(d)$ indicates that the summation is done until $[mr]$ rather than m. For the first part of the right hand side of equation (11) we can write:

$$\begin{aligned}
\sum_{a=1}^{q} \eta_a \sqrt{m} \frac{\partial R^r(d)}{\partial d_a}\Big|_{d^0} &= \frac{2}{\sqrt{m}} \sum_{a=1}^{q} \eta_a \sum_{j=1}^{[mr]} \nu_j \left({}_a\left(G^0\right)^{-1} \left[Re\left[\Lambda_j^0(d)^{-1} I(\lambda_j) \Lambda_j^{0*}(d)^{-1}\right]\right]_a - 1 \right) \\
&+ o_P(1) + tr\left[\hat{G}\left(d^0\right)^{-1} \frac{1}{\sqrt{m}} \sum_{j=1}^{[mr]} \frac{\lambda_j - \pi}{2} Im\left[(\Lambda_j^0(d))^{-1} (-i_a I(\lambda_j)\right.\right. \\
&\left.\left. + I(\lambda_j) i_a)(\Lambda_j^{0*}(d))^{-1} \right]\right] \\
&= \frac{2}{\sqrt{m}} \sum_{a=1}^{q} {}_a G^0 \eta_a \sum_{j=1}^{[mr]} \nu_j \left[{}_a\left(G^0\right)^{-1'} {}_a\left(G^0\right)^{-1} \left[Re\left[\Lambda_j^0(d)^{-1} I(\lambda_j) \Lambda_j^{0*}(d)^{-1}\right]\right]_a \right. \\
&\left. -1 \right] - \frac{2 {}_a G^0}{m^{3/2}} \left(\sum_{j=1}^{[mr]} \nu_j \right) \sum_{j=1}^{m} \left[{}_a\left(G^0\right)^{-1'} {}_a\left(G^0\right)^{-1} \left[Re\left[\Lambda_j^0(d)^{-1} A(\lambda_j) I_{\varepsilon j} A^*(\lambda_j)\right.\right.\right. \\
&\left.\left.\left. \times \Lambda_j^{0*}(d)^{-1}\right]\right]_a - 1 \right] + o_P(1) + tr\left[\hat{G}(d_0)^{-1} \frac{1}{\sqrt{m}} \sum_{j=1}^{[mr]} \frac{\lambda_j - \pi}{2} Im\left[(\Lambda_j^0(d))^{-1}\right.\right. \\
&\left.\left. \times (-i_a I(\lambda_j) + I(\lambda_j) i_a)(\Lambda_j^{0*}(d))^{-1} \right]\right].
\end{aligned} \tag{12}$$

By arguments as in Shimotsu (2007), we can write the first term plus the imaginary part as $\sum_{t=1}^{T} z_t + o_P(1)$ with $z_1 = 0$ and $z_t = \varepsilon_t' \sum_{s=1}^{t-1} [\Theta_{t-s} + \tilde{\Theta}_{t-s}] \varepsilon_s$. Here $\Theta_s = \frac{1}{\pi \sqrt{mT}} \sum_{j=1}^{m} \nu_j Re[\Psi_j + \Psi_j'] \cos(s\lambda_j)$, $\tilde{\Theta}_s = \frac{\pi}{2} \frac{1}{\pi \sqrt{mT}} \sum_{j=1}^{m} Re[\Psi_j - \Psi_j'] \sin(s\lambda_j)$, Ψ_j is defined by $\Psi_j = \sum_{a=1}^{q} \eta_a$ $[A^*(\lambda_j) \Lambda_j^{0*}(d)^{-1}]_{aa} (G^0)^{-1} \Lambda_j^0(d)^{-1} A(\lambda_j)$, $A(\lambda) = \sum_{j=0}^{\infty} A_j \varepsilon_{t-j}$ and A_j is given in Assumption 2. The asymptotic normality of z_t follows from Theorem 2 of Robinson (1995). To obtain the covariance of the z_t we have for $0 \leq r_1 \leq r_2 \leq 1$

$$\begin{aligned}
Cov\left(\sum_{t=1}^{T} z_{t,r_1}, \sum_{t=1}^{T} z_{t,r_2} \right) &= E\left(\sum_{t=1}^{T} \sum_{s=1}^{t-1} \left(\Theta_{t-s,r_1} + \tilde{\Theta}_{t-s,r_1}\right)' \varepsilon_s' \sum_{l=1}^{t-1} \left(\Theta_{t-s,r_2} + \tilde{\Theta}_{t-s,r_2}\right) \varepsilon_l \right) \\
&= \sum_{t=2}^{T} \sum_{s=1}^{t-1} tr\left[\left(\Theta_{t-s,r_1} + \tilde{\Theta}_{t-s,r_1}\right)' \left(\Theta_{t-s,r_2} + \tilde{\Theta}_{t-s,r_2}\right) \right] + o_P(1)
\end{aligned}$$

by using Lemma 2 and 3 from Lobato (1999). Now we have

$$\begin{aligned}\sum_{t=2}^{T}\sum_{s=1}^{t-1}\Theta'_{t-s,r_1}\tilde{\Theta}_{t-s,r_2} &= \frac{1}{2\pi mT^2}\sum_{t=1}^{T-1}\sum_{s=1}^{T-1}\sum_{j=1}^{[mr_1]}\sum_{l=1}^{[mr_2]}\nu_j\nu_l tr\left[Re\left[\Psi'_j+\Psi_j\right]Re\left[\Psi_l-\Psi'_l\right]\right] \\ &\quad \times\cos(s\lambda_j)\sin(s\lambda_j) \\ &= 0\end{aligned}$$

as $tr[(A'+A)(B-B')]=0$ for any real matrices A and B. Furthermore, we have

$$\begin{aligned}\sum_{t=2}^{T}\sum_{s=1}^{t-1}tr\left[\Theta'_{t-s,r_1}\Theta_{t-s,r_2}+\tilde{\Theta}'_{t-s,r_1}\tilde{\Theta}_{t-s,r_2}\right] &= \frac{1}{\pi^2mT^2}\sum_{t=1}^{T-1}\sum_{s=1}^{T-t}\sum_{j=1}^{[mr_1]}\nu_j^2 \\ &\quad \times tr\left[Re\left[\Psi'_j+\Psi_j\right]\, Re\left[\Psi_j+\Psi'_j\right]\right]\cos^2(s\lambda_j) \\ +\frac{1}{\pi^2mT^2}\sum_{t=1}^{T-1}\sum_{s=1}^{T-t}\sum_{j=1}^{[mr_1]}\sum_{l=1,l\neq j}^{[mr_2]}\nu_j\nu_l tr\left[Re\left[\Psi'_j+\Psi_j\right]\, Re\left[\Psi_l+\Psi'_l\right]\right]&\cos(s\lambda_j)\cos(s\lambda_l) \\ +\frac{\pi^2}{4}\frac{1}{\pi^2mT^2}\sum_{t=1}^{T-1}\sum_{s=1}^{T-t}\sum_{j=1}^{[mr_1]}tr\left[Re\left[\Psi'_j-\Psi_j\right]\, Re\left[\Psi_j-\Psi'_j\right]\right]&\sin^2(s\lambda_j) \\ +\frac{\pi^2}{4}\frac{1}{\pi^2mT^2}\sum_{t=1}^{T-1}\sum_{s=1}^{T-t}\sum_{j=1}^{[mr_1]}\sum_{l=1,l\neq j}^{[mr_2]}tr\left[Re\left[\Psi'_j\Psi_j\right]Re\,\left[\Psi_l-\Psi'_l\right]\right]&\sin(s\lambda_j)\sin(s\lambda_l).\end{aligned}$$

The second and fourth term of this sum are $o_P(1)$ by Lemma 3b) and 3d) in Shimotsu (2007). Applying Lemma 3a) in Shimotsu (2007) for the first term, we obtain for $\lambda_j\to 0$

$$\frac{1}{\pi^2mT^2}\sum_{t=1}^{T-1}\sum_{s=1}^{T-t}\sum_{j=1}^{[mr_1]}\nu_j^2 tr\left[Re\left[\Psi'_j+\Psi_j\right]Re\left[\Psi_j+\Psi'_j\right]\right]\cos^2(s\lambda_j)=$$

$$\frac{1}{m}\sum_{j=1}^{[mr_1]}\nu_j^2\frac{1}{4\pi^2}tr\left[Re\left[\Psi'_j+\Psi_j\right]Re\left[\Psi_j+\Psi'_j\right]\right]=\frac{1}{m}\sum_{j=1}^{[mr_1]}\nu_j^2\left(2\sum_{a=1}^{q}\eta_a^2+2\sum_{a=1}^{q}\sum_{b=1}^{q}\eta_a\eta_b G^0_{ab}\left(G^0\right)^{-1}_{ab}\right).$$

For the third term we have again for $\lambda_j\to 0$ by Shimotsu (2007) Lemma 3c)

$$\frac{\pi^2}{4}\frac{1}{\pi^2mT^2}\sum_{t=1}^{T-1}\sum_{s=1}^{T-t}\sum_{j=1}^{[mr_1]}tr\left[Re\left[\Psi'_j-\Psi_j\right]Re\left[\Psi_j-\Psi'_j\right]\right]\sin^2(s\lambda_j)=$$

$$\frac{\pi^2}{4m}\sum_{j=1}^{[mr_1]}\frac{1}{4\pi^2}tr\left[Re\left[\Psi'_j-\Psi_j\right]Re\left[\Psi_j-\Psi'_j\right]\right]=\frac{\pi^2}{4m}\sum_{j=1}^{[mr_1]}\left(2\sum_{a=1}^{q}\sum_{b=1}^{q}\eta_a\eta_b G^0_{ab}\left(G^0\right)^{-1}_{ab}-2\sum_{a=1}^{q}\eta_a^2\right).$$

From the Euler-Mc Laurin equality and Lemma B.1 in Qu (2011) it follows that $1/m\sum_{j=1}^{[mr_1]} v_j^2 \to \int_0^{r_1}(1+\log s)^2 ds$. We thus obtain altogether

$$\begin{aligned} Cov\left(\sum_{t=1}^{T} z_{t,r_1}, \sum_{t=1}^{T} z_{t,r_2}\right) \quad &\to \quad \int_0^{r_1}\left[(1+\log s)^2\left(2\sum_{a=1}^{q}\eta_a^2 + 2\sum_{a=1}^{q}\sum_{b=1}^{q}\eta_a\eta_b G_{ab}^0\left(G^0\right)_{ab}^{-1}\right)\right. \\ &+ \quad \left.\frac{\pi^2}{4}\left(2\sum_{a=1}^{q}\sum_{b=1}^{q}\eta_a\eta_b G_{ab}^0\left(G^0\right)_{ab}^{-1} - 2\sum_{a=1}^{q}\eta_a^2\right)\right] ds. \end{aligned}$$

For the second term of the second equality of (12) we have by similar arguments as in Qu (2011) that

$$\frac{{}_aG^0}{m^{1/2}}\sum_{j=1}^{m}\left[{}_a\left(G^0\right)^{-1'}{}_a\left(G^0\right)^{-1}\left[Re\left[\Lambda_j^0(d)^{-1}I(\lambda_j)\Lambda_j^{0*}(d)^{-1}\right]\right]_a - 1\right] \to B(1)$$

where $B(s)$ denotes a standard Brownian motion. As before we have from Lemma B.1 in Qu (2011) that

$$\frac{1}{m}\sum_{j=1}^{[mr_1]} v_j \to \int_0^{r_1}(1+\log s)ds.$$

It remains the last part of the Taylor decomposition. For the second derivative of the objective function $R(d)$ we obtain

$$\frac{\partial^2 R^r(d)}{\partial d_a \partial d_b} = tr\left[-\hat{G}^{-1}(d)\frac{\partial \hat{G}^r(d)}{\partial d_a}\hat{G}^{-1}(d)\frac{\partial \hat{G}^r(d)}{\partial d_b} + \hat{G}^{-1}(d)\frac{\partial^2 \hat{G}^r(d)}{\partial d_a \partial d_b}\right].$$

Thus,

$$\frac{\partial \hat{G}^r(d)}{\partial d_a} = \frac{1}{m}\sum_{j=1}^{[mr]}(\log\lambda_j)\hat{G}_{1a}(d) + o_P\left((logT)^{-1}\right)$$

with

$$\hat{G}_{1a}(d) = i_a\hat{G}(d) + \hat{G}(d)i_a.$$

Furthermore, it is

$$\frac{\partial^2 \hat{G}^r(d)}{\partial d_a \partial d_b} = \frac{1}{m}\sum_{j=1}^{[mr]}(\log\lambda_j)^2\hat{G}_{2ab}(d) + \frac{\pi^2}{4}\hat{G}_{3ab}(d) + o_P(1)$$

with

$$G_{2ab}(d) = i_a i_b\hat{G}(d) + i_a\hat{G}(d)i_b + i_b\hat{G}(d)i_a + \hat{G}(d)i_a i_b$$

and

$$G_{3ab}(d) = -i_a i_b \hat{G}(d) + i_a \hat{G}(d) i_b + i_b \hat{G}(d) i_a - \hat{G}(d) i_a i_b.$$

Furthermore, it holds that

$$tr\left[\hat{G}(d)^{-1}\hat{G}_{1a}(d)\hat{G}(d)^{-1}\hat{G}_{1b}(d)\right] = tr\left[\hat{G}(d)^{-1}\hat{G}_{2ab}(d)\right].$$

Altogether this gives

$$\begin{aligned}
\frac{\partial^2 R^r(d)}{\partial d_a \partial d_b} &= tr\left[-\hat{G}(d)^{-1}\left(\frac{1}{m}\sum_{j=1}^{[mr]}(\log\lambda_j)\hat{G}_{1a}(d) + o_P\left((logT)^{-1}\right)\right)\right. \\
&\quad \times\hat{G}(d)^{-1}\left(\frac{1}{m}\sum_{j=1}^{[mr]}(\log\lambda_j)\hat{G}_{1b}(d) + o_P\left((logT)^{-1}\right)\right) \\
&\quad \left. + \hat{G}(d)^{-1}\left(\frac{1}{m}\sum_{j=1}^{[mr]}(\log\lambda_j)^2\hat{G}_{2ab}(d) + \frac{\pi^2}{4}\hat{G}(d)^{-1}\hat{G}_{3ab}(d) + o_P(1)\right)\right] \\
&= tr\left[\left(\frac{1}{m}\sum_{j=1}^{[mr]}\nu_j^2\right)\hat{G}(d)^{-1}\hat{G}_{2ab}(d) + \frac{\pi^2}{4}\hat{G}(d)^{-1}\hat{G}_{3ab}(d)\right] + o_P\left(log^2 T\right) \\
&\to tr\left[\int_0^r (1+\log s)^2 ds \left(G^0\right)^{-1} G^0_{2ab} + \frac{\pi^2}{4}\left(G^0\right)^{-1} G^0_{3ab}\right] \\
&= 2\int_0^r (1+\log s)^2 ds (G^0 \odot \left(G^0\right)^{-1} + I_q) + \frac{\pi^2}{4}(G^0 \odot \left(G^0\right)^{-1} - I_q).
\end{aligned}$$

By denoting $F(r) = \int_0^r (1+\log s)^2 ds (G^0 \odot \left(G^0\right)^{-1} + I_q) + \frac{\pi^2}{4}(G^0 \odot \left(G^0\right)^{-1} - I_q)$ we have

$$\eta' \frac{\partial^2 R^r(d)}{\partial d \partial d'}\sqrt{m}(\hat{d} - d_0) = \eta' F(r)\eta 2\int_0^1 (1+\log s) dW(s).$$

Like Qu (2011), we use Theorem 13.5 of Billingsley (2009) to prove tightness. Thus, we show that for every m and $r_1 \leq r \leq r_2$

$$E\left(\left|\sum_{t=1}^T z_{t,r} - \sum_{t=1}^T z_{t,r_1}\right|^2 \left|\sum_{t=1}^T z_{t,r_2} - \sum_{t=1}^T z_{t,r}\right|^2\right) \leq K(\psi_m(r_2) - \psi_m(r_1))^2$$

where K is some constant and $\psi_m(\cdot)$ is a function on $[0,1]$ which is finite, nondecreasing and fulfills

$$\lim_{\delta\to 0}\limsup_{m\to\infty}|\psi_m(s+\delta) - \psi_m(s)| \to 0$$

uniformly in $s \in [0,1]$. Here we denote $z_t(s,r) = z_{t,r} - z_{t,s}$. Denote also $c_t(r,s) = c_{t,r} - c_{t,s}$ and $c_t = tr[\Theta_t + \tilde{\Theta}_t]$. Using this notation we can use Qu's (2011) Lemma B.8 to show that $E(|\sum_{t=1}^T z_{t,r} - \sum_{t=1}^T z_{t,r_1}|^2 |\sum_{t=1}^T z_{t,r_2} - \sum_{t=1}^T z_{t,r}|^2)$ is bounded from above by

$$K\left(\sum_{t=1}^{T}\sum_{s=1}^{t-1} c_{t-s}(r_1,r)^2\right)\left(\sum_{t=1}^{T}\sum_{h=1}^{t-1} c_{t-h}(r,r_2)^2\right)$$

where K is some positive constant. By similar arguments as in Qu (2011) we obtain furthermore

$$\begin{aligned}\sum_{t=1}^{T}\sum_{s=1}^{t-1} c_{t-s}(r_1,r)^2 &\leq \left(\frac{1}{Tm}\sum_{j=[mr_1]+1}^{[mr]}\sum_{k\neq j}^{[mr]}\left(v_j^2+v_k^2\right)+\frac{1}{m}\sum_{j=[mr_1]+1}^{[mr]} v_j^2\right)\\ &\quad\times\left(2\sum_{a=1}^{q}\eta_a^2+2\sum_{a=1}^{q}\sum_{b=1}^{q}\eta_a\eta_b G^0_{ab}\left(G^0\right)^{-1}_{ab}\right)\\ &\leq \frac{3}{m}\sum_{j=[mr_1]+1}^{[mr]} v_j^2\left(2\sum_{a=1}^{q}\eta_a^2+2\sum_{a=1}^{q}\sum_{b=1}^{q}\eta_a\eta_b G^0_{ab}\left(G^0\right)^{-1}_{ab}\right).\end{aligned}$$

As $(2\sum_{a=1}^q \eta_a^2 + 2\sum_{a=1}^q \sum_{b=1}^q \eta_a\eta_b G^0_{ab}(G^0)^{-1}_{ab}) \leq K$ for some constant K we set $\psi_m(s) = 1/m$ $\sum_{j=1}^{[ms]} v_j^2$. This satisfies the condition as

$$\lim_{\delta\to 0}\limsup_{m\to\infty}|\psi_m(s+\delta)-\psi_m(s)| = \lim_{\delta\to 0}\int_s^{s+\delta}(1+\log x)^2 dx \quad\to\quad 0.$$

This proves the theorem. □

Proof of Lemma 2:

To prove the lemma, it is enough to show that $2\eta'\eta + 2\eta'(G^0 \odot (G^0)^{-1})\eta = 4$. For this denote

$$G^0 = \begin{pmatrix} g_{11} & g_{12} & \cdots & g_{1q}\\ g_{21} & g_{22} & \cdots & g_{2q}\\ \vdots & \ddots & & \vdots\\ g_{q1} & g_{q2} & \cdots & g_{qq}\end{pmatrix}.$$

Thus, we obtain for the inverse matrix

$$\left(G^0\right)^{-1} = \frac{1}{\det(G^0)} \begin{pmatrix} \det(G^0_{11}) & -\det(G^0_{21}) & \dots & (-1)^{1+q}\det(G^0_{q1}) \\ -\det(G^0_{12}) & \det(G^0_{22}) & \dots & (-1)^{2+q}\det(G^0_{q2}) \\ \vdots & \ddots & & \vdots \\ (-1)^{1+q}\det(G^0_{1q}) & (-1)^{2+q}\det(G^0_{2q}) & \dots & \det(G^0_{qq}) \end{pmatrix}$$

by using Cramer's rule. Therefore, by applying Laplace's formula and using that $g_{ij} = g_{ji}$ we have

$$G^0 \odot \left(G^0\right)^{-1} = \frac{1}{\det(G^0)} \begin{pmatrix} g_{11}\det(G^0_{11}) & -g_{12}\det(G^0_{21}) & \dots & (-1)^{1+q}g_{1q}\det(G^0_{q1}) \\ -g_{21}\det(G^0_{12}) & g_{22}\det(G^0_{22}) & \dots & (-1)^{2+q}g_{2q}\det(G^0_{q2}) \\ \vdots & \vdots & \ddots & \vdots \\ (-1)^{1+q}g_{q1}\det(G^0_{1q}) & (-1)^{2+q}g_{q2}\det(G^0_{2q}) & \dots & g_{qq}\det(G^0_{qq}) \end{pmatrix}.$$

Therefore

$$\begin{aligned} G^0 \odot \left(G^0\right)^{-1}\eta &= \frac{1}{\det(G^0)} \begin{pmatrix} \frac{1}{\sqrt{q}}\sum_{i=1}^{q}(-1)^{1+i}g_{i1}\det(G^0_{i1}) \\ \frac{1}{\sqrt{q}}\sum_{i=1}^{q}(-1)^{2+i}g_{i2}\det(G^0_{i2}) \\ \vdots \\ \frac{1}{\sqrt{q}}\sum_{i=1}^{q}(-1)^{q+i}g_{iq}\det(G^0_{iq}) \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{\sqrt{q}} \\ \frac{1}{\sqrt{q}} \\ \vdots \\ \frac{1}{\sqrt{q}} \end{pmatrix} \end{aligned}$$

and thus finally

$$\eta' \left(G^0 \odot \left(G^0\right)^{-1}\right)\eta = 1$$

which proves the Lemma. □

Proof of Theorem 2:

To prove the consistency of our test statistic, we use the property that $I(\lambda_j) = O_P(1)$ when $jT^{-1/2} \to \infty$. Note that ν_j is monotonically increasing in j with $\nu_1 < 0$ and $\nu_m > 0$. For our test statistic we write

$$\begin{aligned} MLWS &= \frac{1}{2}\sup_{r\in[\varepsilon,1]}\left\| \frac{2}{\sqrt{\sum_{j=1}^m \nu_j^2}}\sum_{a=1}^{q}\eta_a\sum_{j=1}^{[mr]}\nu_j\Big({}_a g\Big[Re\Big[\Lambda_j(\hat{d})^{-1}I(\lambda_j)\Lambda_j^*(\hat{d})^{-1}\Big]\Big]_a - 1\Big)\right. \\ &\left. +tr\left[G(\hat{d})^{-1}\frac{1}{\sqrt{\sum_{j=1}^m \nu_j^2}}\sum_{j=1}^{[mr]}\frac{\lambda_j-\pi}{2}Im\Big[(\Lambda_j(\hat{d}))^{-1}(-i_a I(\lambda_j)+I(\lambda_j)i_a)(\Lambda_j^*(\hat{d}))^{-1}\Big]\right]\right\| \\ &= \sup_{r\in[\varepsilon,1]}\|I+II\|. \end{aligned}$$

Let us consider the term I first. Define $j^* = \min\{j : \nu_j \geq 0\}$. From Qu's (2011) Lemma B.1 it follows that $j^* = Km$ for some constant K. Define now

$$A = \left\|\sum_{a=1}^{q}\eta_a\sum_{j=j^*}^{m}\nu_j\Big({}_a g\Big[Re\Big[\Lambda_j(\hat{d})^{-1}I(\lambda_j)\Lambda_j^*(\hat{d})^{-1}\Big]\Big]_a - 1\Big)\right\|$$

and

$$B = \left\|\sum_{a=1}^{q}\eta_a\sum_{j=1}^{j^*-1}\nu_j\Big({}_a g\Big[Re\Big[\Lambda_j(\hat{d})^{-1}I(\lambda_j)\Lambda_j^*(\hat{d})^{-1}\Big]\Big]_a - 1\Big)\right\|.$$

Applying the reverse triangle inequality to I gives

$$I \geq \max(A-B, B).$$

Now, if $A \geq 2B$ than we have

$$\begin{aligned} I &\geq A - B \\ &= \frac{A}{2} + \Big(\frac{A}{2} - B\Big) \\ &\geq \tfrac{A}{2}. \end{aligned}$$

On the other hand, if $A < 2B$ than $I \geq B > A/2$. Altogether, we have $I \geq A/2$. Thus, we have to show that $A \xrightarrow{P} \infty$ if $T \to \infty$. To do so, we write A in the form

$$A = \left\|\sum_{a=1}^{q}\eta_a\sum_{j=j^*}^{m}\nu_j\Big({}_a g\Big[Re\Big[\Lambda_j(d)^{-1}A(\lambda_j)I_{\varepsilon j}A^*(\lambda_j)\Lambda_j^*(d)^{-1}\Big]\Big]_a - 1\Big)\right\|.$$

Applying the reverse triangle inequality to A gives

$$A \geq \sum_{a=1}^{q} \eta_a \sum_{j=j^*}^{m} \nu_j - \sum_{a=1}^{q} \eta_a \sum_{j=j^*}^{m} \nu_j \left({}_a g \left[Re \left[\Lambda_j(d)^{-1} A(\lambda_j) I_{\varepsilon j} A^*(\lambda_j) \Lambda_j^*(d)^{-1} \right] \right]_a \right).$$

For the first term of this inequality we have

$$\sum_{a=1}^{q} \eta_a \sum_{j=j^*}^{m} \nu_j = m^{1/2} \int_{\kappa}^{1} (1 + \log s) ds + o\left(m^{1/2}\right)$$

which is strictly positive and of exact order $m^{1/2}$.
For the second term we can conclude the following. As $m/T^{1/2} \to \infty$ it holds that $j^*/T^{1/2} \to \infty$. Thus, for every $j^* \leq j \leq m$ it follows $I(\lambda_j) = O_P(1)$. Furthermore, we have

$$\nu_j \left({}_a g \left[Re \left[\Lambda_j(d)^{-1} A(\lambda_j) I_{\varepsilon j} A^*(\lambda_j) \Lambda_j^*(d)^{-1} \right] \right]_a \right) = O_P\left(\lambda_j^{2\hat{d}}\right) = O_P\left(\lambda_j^{2\epsilon}\right) = o_P(1)$$

for every a. This holds because $\nu_j = O_P(1)$ for $j^* \leq j \leq m$, $\hat{G}(d) = 1/m \sum_{j=1}^{m} Re[\Lambda_j(d)^{-1} I(\lambda_j) \Lambda_j^*(d)^{-1}]$, $P(\hat{d}_j > \epsilon) \to 1$ and $\lambda_j = o(1)$. Therefore, the second term is of lower order than $m^{1/2}$ and is dominated asymptotically by the first term. Thus, $A \xrightarrow{P} \infty$ if $T \to \infty$. By using exactly the same arguments as before the term II is of lower order than $m^{1/2}$, and therefore dominated asymptotically by term I. This proves the theorem. □

Acknowledgements

Financial support of the Deutsche Forschungsgemeinschaft (DFG) is gratefully acknowledged. We are also grateful to Jan Beran, Niels Haldrup, Uwe Hassler, Liudas Giraitis, Robinson Kruse, Michael Massmann and Morten Nielsen as well as the participants of the Frankfurt workshop on Time Series Econometrics 2014 and the NBER-NSF Time Series Conference 2014 in St. Louis for inspiring discussions on the paper. Additionally, we thank Roxana Halbleib for sharing her codes for estimating a FIVARMA model with us.

References

Beran, J. (1995). Maximum likelihood estimation of the differencing parameter for invertible short and long memory autoregressive integrated moving average models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57:659–672.

Beran, J., Feng, Y., Ghosh, S., and Kulik, R. (2013). *Long memory processes: Probabilistic properties and statistical methods*. Springer London, Limited.

Beran, J. and Terrin, N. (1994). Estimation of the long-memory parameter, based on a multivariate central limit theorem. *Journal of Time Series Analysis*, 15(3):269–278.

Berkes, I., Rorvath, L., Kokoszka, P., and Shao, Q.-M. (2006). On discriminating between long-range dependence and changes in mean. *The Annals of Statistics*, 34(3):1140–1165.

Billingsley, P. (2009). *Convergence of probability measures*. Wiley.

Chiriac, R. and Voev, V. (2011). Modelling and forecasting multivariate realized volatility. *Journal of Applied Econometrics*, 26(6):922–947.

Davidson, J. and Rambaccussing, D. (2015). A test of the long memory hypothesis based on self-similarity. Unpublished Manuscript, Department of Economics, University of Exeter.

Diebold, F. X. and Inoue, A. (2001). Long memory and regime switching. *Journal of Econometrics*, 105(1):131–159.

Dolado, J. J., Gonzalo, J., and Mayoral, L. (2005). What is what?: A simple time-domain test of long-memory vs. structural breaks. Unpublished Manuscript, Department of Economics, Universidad Carlos III de Madrid.

Engle, R. F. and Smith, A. D. (1999). Stochastic permanent breaks. *Review of Economics and Statistics*, 81(4):553–574.

Giraitis, L., Koul, H., and Surgailis, D. (2012). *Large sample inference for long memory processes*. World Scientific Publishing Company Incorporated.

Granger, C. W. and Ding, Z. (1996). Varieties of long memory models. *Journal of Econometrics*, 73(1):61–77.

Granger, C. W. and Hyung, N. (2004). Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns. *Journal of Empirical Finance*, 11(3):399–421.

Haldrup, N. and Kruse, R. (2014). Discriminating between fractional integration and spurious long memory. Unpublished Manuscript, Department of Economics, University of Aarhus.

Kechagias, S. and Pipiras, V. (2015). Definitions and representations of multivariate long-range dependent time series. *Journal of Time Series Analysis*, 36(1):1–25.

Leccadito, A., Rachedi, O., and Urga, G. (2015). True versus spurious long memory: Some theoretical results and a Monte Carlo comparison. *Econometric Reviews*, 34(4):452–479.

Lobato, I. N. (1997). Consistency of the averaged cross-periodogram in long memory series. *Journal of Time Series Analysis*, 18(2):137–155.

Lobato, I. N. (1999). A semiparametric two-step estimator in a multivariate long memory model. *Journal of Econometrics*, 90(1):129–153.

Lobato, I. N. and Savin, N. E. (1998). Real and spurious long-memory properties of stock-market data. *Journal of Business & Economic Statistics*, 16(3):261–268.

Lu, Y. K. and Perron, P. (2010). Modeling and forecasting stock return volatility using a random level shift model. *Journal of Empirical Finance*, 17(1):138–156.

Lütkepohl, H. (2007). *New introduction to multiple time series analysis*. Springer.

Marinucci, D. and Robinson, P. M. (1999). Alternative forms of fractional Brownian motion. *Journal of Statistical Planning and Inference*, 80(1):111–122.

Mikosch, T. and Stărică, C. (2004). Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects. *Review of Economics and Statistics*, 86(1):378–390.

Nielsen, M. Ø. (2007). Local Whittle analysis of stationary fractional cointegration and the implied–realized volatility relation. *Journal of Business & Economic Statistics*, 25(4).

Nielsen, M. Ø. and Frederiksen, P. (2011). Fully modified narrow-band least squares estimation of weak fractional cointegration. *The Econometrics Journal*, 14(1):77–120.

Ohanissian, A., Russell, J. R., and Tsay, R. S. (2008). True or spurious long memory? A new test. *Journal of Business & Economic Statistics*, 26(2):161–175.

Perron, P. and Qu, Z. (2010). Long-memory and level shifts in the volatility of stock market return indices. *Journal of Business & Economic Statistics*, 28(2):275–290.

Qu, Z. (2011). A test against spurious long memory. *Journal of Business & Economic Statistics*, 29(3):423–438.

Robinson, P. M. (1995). Gaussian semiparametric estimation of long range dependence. *The Annals of Statistics*, 23(5):1630–1661.

Robinson, P. M. and Yajima, Y. (2002). Determination of cointegrating rank in fractional systems. *Journal of Econometrics*, 106(2):217–241.

Sela, R. J. and Hurvich, C. M. (2009). Computationally efficient methods for two multivariate fractionally integrated models. *Journal of Time Series Analysis*, 30(6):631–651.

Shimotsu, K. (2006). Simple (but effective) tests of long memory versus structural breaks. Unpublished Manuscript, Department of Economics, Queens's University.

Shimotsu, K. (2007). Gaussian semiparametric estimation of multivariate fractionally integrated processes. *Journal of Econometrics*, 137(2):277–310.

Simes, R. J. (1986). An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, 73(3):751–754.

Sowell, F. (1989). Maximum likelihood estimation of fractionally integrated time series models. Unpublished Manuscript, Department of Economics, Carnegie Mellon University.

Varneskov, R. T. and Perron, P. (2011). Combining long memory and level shifts in modeling and forecasting the volatility of asset returns. Unpublished Manuscript, Department of Economics, Boston University.

Xu, J. and Perron, P. (2014). Forecasting return volatility: Level shifts with varying jump probability and mean reversion. *International Journal of Forecasting*, 30(3):449–463.

Yau, C. Y. and Davis, R. A. (2012). Likelihood inference for discriminating between long-memory and change-point models. *Journal of Time Series Analysis*, 33(4):649–664.