A Multivariate Test Against Spurious Long Memory

Philipp Sibbertsen,[1] Christian Leschinski and Marie Busch

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

This version: April 20, 2017

Abstract

This paper provides a multivariate score-type test against spurious long memory. In particular, we prove the consistency of the test against the alternatives of random level shifts and smooth trends. The test statistic is based on the weighted sum of the partial derivatives of the multivariate local Whittle likelihood function. By choosing the weighting scheme accordingly, one can either test the complete spectral density matrix for a misspecification local to the origin, or one can focus on particular rows and columns. In the first case, we obtain a pivotal limiting distribution, whereas we can use the second weighting scheme in a subsequent step to evaluate which series of the multivariate system might cause a possible rejection.

To apply the test to fractionally cointegrated series, the test statistic is calculated for the linearly transformed system after estimating the cointegrating matrix. We derive the limiting distribution and show consistency under this procedure. A Monte Carlo analysis shows good finite sample properties of the test in terms of size and power.

To highlight the usefulness of the test in practice, we apply it to the log-absolute returns and the log-realized volatilities of the S&P 500, the DAX, the FTSE, and the NIKKEI. It is found that the log-absolute return of the S&P 500 is not correctly specified as a pure long memory process. In contrast to that, there is no indication of spurious long memory in the realized volatility series.

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, series with slowly decaying empirical autocorrelation functions are modeled as fractionally integrated processes. However, several authors point out that other data generating processes such as nonlinear time varying coefficient models, random level shift processes, STOPBREAK models, and markov switching models can generate similar autocovariance features. Examples of this literature include Granger and Ding (1996), Lobato and Savin (1998), Diebold and Inoue (2001), Granger and Hyung (2004) and Mikosch and Stărică (2004).

Motivated by these findings, several tests have been proposed to distinguish true and spurious long memory. Berkes et al. (2006) or Yau and Davis (2012), among others, suggest tests for the null hypothesis of spurious long memory. Tests for the null hypothesis of true long memory include Dolado et al. (2005), Shimotsu (2006), Ohanissian et al. (2008), Haldrup and Kruse (2014) and Davidson and Rambaccussing (2015).

Perron and Qu (2010) derive the properties of the periodogram of processes with short memory and 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 against a wide range of alternatives the Qu test has the best power among the tests suggested so far.

Closely related to our paper are also multivariate extensions of the local Whittle estimator. In particular, Lobato (1999) and Shimotsu (2007) extend the local Whittle estimator to a multivariate framework. Extensions of the local Whittle estimator to fractionally cointegrated systems have been considered by Nielsen (2007), Robinson (2008b) and Shimotsu (2012).

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. As Kechagias and

Pipiras (2015) show, the assumed form of the spectral density matrix local to the origin is specific to causal filters with hyperbolically decaying coefficients. Therefore, our test can be interpreted as a general test on the correct specification of a multivariate series as a causal long memory process. If one is willing to assume that the process is causal, a rejection of the test can be interpreted as evidence for low frequency contaminations. The limiting distribution of the test statistic is derived for general weights. However, by choosing the weighting scheme accordingly, one can obtain a pivotal distribution that coincides with that of the univariate Qu test. Furthermore, it is also possible to choose the weights so that one can gain further insights into which components of a multivariate process cause a rejection.

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 $\lfloor mr \rfloor \leq m$ Fourier frequencies with $r \in [\varepsilon, 1]$ is approximately equal to zero. 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. Nevertheless, we show that the test can easily be modified for the situation of fractionally cointegrated data.

In the empirical example we apply our test to the log-absolute returns and log-realized volatilities of four stock market indices: the Standard & Poor 500, the DAX, the FTSE and the NIKKEI. Even though especially 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 we can clearly reject the null hypothesis of a pure long memory process for the S&P 500. For realized volatility series of these four stock market indices, on the other hand, we do not find any evidence of spurious long memory.

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. A supplementary appendix is provided on the authors webpages. It contains some of the more standard proofs, details on the pre-whitening procedure, a series of additional Monte Carlo experiments and robustness checks for the empirical application.

2 Model Specification and Assumptions

The spectral density of a multivariate long-memory process X_t, with $d = (d_1, d_2, ..., d_q)'$ and $-1/2 < d_1, \dots, d_q < 1/2$ being the memory vector, is local to the origin given by

$$f(\lambda_j) \sim \Lambda_j(d) G \Lambda_j^*(d), \tag{1}$$

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, and $j = 1, \dots, \lfloor T/2 \rfloor$. 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 assumptions on G exclude fractional cointegration as they stand. We first derive our test statistic under this assumption and consider the case of fractionally cointegrated series afterwards in Section 3.3.
The spectral density representation in (1) 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 in row a and column b of the spectral matrix of X_t contain complex valued elements which are not vanishing at $\lambda = 0$ and which depend on the difference between the memory parameters d_a and d_b. These complex valued elements vanish if and only if the matrix G in (1) is diagonal or $d_a = d$ for all dimensions a.
A possible example is the multivariate q-dimensional FIVARMA 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 $t = 1, \dots, T$. This can alternatively be written as

$$D(d_1, ..., d_q)(X_t - EX_t) = u_t, \tag{2}$$

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 polynomial 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 process u_t in (2) is assumed to fulfill the local condition $f_u(\lambda) \sim G$, as $\lambda \to 0$. 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 weight matrices.

Furthermore, the periodogram of X_t evaluated at frequency λ is defined as $I(\lambda) = w(\lambda)w^*(\lambda)$, with $w(\lambda) = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^{T} X_t e^{it\lambda}$ and the superscript 0 denotes the true value of a parameter. We need to state the following assumptions which follow those in Shimotsu (2007):

Assumption 1. *For $\beta \in (0,2]$ and $a,b = 1,\ldots,q$ 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 $f(\lambda)$ 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 $t = 0, \pm1, \pm2, \ldots$ where $\mathfrak{F}_t$ denotes the σ-field generated by ε_s and I_q is an identity matrix, $s \leq t$. 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 $P(\|\varepsilon_t\|^2 > \tau) \leq KP(\varepsilon^2 > \tau)$. In addition, it holds for $a,b,c,d = 1,2$, $t = 0,\pm1,\pm2,\ldots$ that $E(\varepsilon_{at}\varepsilon_{bt}\varepsilon_{ct}|\mathfrak{F}_{t-1}) = \mu_{abc}$ a.s. and $E(\varepsilon_{at}\varepsilon_{bt}\varepsilon_{ct}\varepsilon_{dt}|\mathfrak{F}_{t-1}) = \mu_{abcd}$ 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 to 5 are satisfied by multivariate FIVARMA processes. Assumption 4 is slightly stronger than the assumption used in Qu (2011) for the univariate local Whittle estimator. It gives a sharp upper bound of $m = o(T^{0.8})$ for the number of frequencies m which can be used for the local Whittle estimator and thus for our test statistic. 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.

3 Testing for Spurious Long Memory

In this section we propose a multivariate test 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, trends or other forms of low frequency contaminations. Special use will be made of the fact that the slope of the spectral density of a process with these kind of contaminations is nearly zero for Fourier frequencies λ_j with $j > \sqrt{T}$.

3.1 The MLWS Statistic

To be specific, we are interested in testing the hypothesis that the spectral density local to the origin has the shape given in equation (1):

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

as $\lambda_j \to 0^+$ with $d_a \in (-1/2, 1/2)\ \forall\ a = 1, \ldots, q$. Thus, under the null hypothesis X_t is a multivariate causal long-memory process with phase $(d_a - d_b)(\pi - \lambda)/2$. The alternative is that the data cannot be described by this spectral density:

$$H_1 : f(\lambda_j) \nsim \Lambda_j(d) G \Lambda_j^*(d).$$

To motivate the test statistic, we discuss the properties of the periodogram under the alternative of low frequency contaminations for the example of random level shift processes and smooth trends. The multivariate random level shift model is defined by

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

where κ_t, $\Pi_t = diag(\pi_{1t},\ldots,\pi_{qt})$ 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 Σ_π for $i,j=1,...,q$. We consider a shift probability that is defined by $p = \tilde{p}/T$, where $\tilde{p}$ is the expected number of shifts in the sample. 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 is an *iid* sequence with $\kappa_t \sim N(0,\Sigma_\kappa)$. 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 autoregressive coefficient $0 \le \phi \le 1$ 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 example for a possible 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 $h_a(t/T)$ is a Lipschitz continuous function on $[0,1]$, $\forall\ a=1,...,q$. 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 four components by

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

By similar arguments as in Proposition 3 of Perron and Qu (2010) for $\lambda_j = o(1)$ the first summand is of order $O_P(T^{-1}\lambda_j^{-2})$, the second is of order $O_P(1)$, and the third and fourth term are 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 as in

(4). McCloskey and Perron (2013) show that these orders also hold for deterministic level shifts and fractional trends.
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 $\lfloor mr \rfloor$ 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).
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 $\hat{G}(d) = \frac{1}{m}\sum_{j=1}^{m} Re\left[\Lambda_j(d)^{-1}I(\lambda_j)\Lambda_j^*(d)^{-1}\right]$. Substituting this into $Q_m(G,d)$ and

$$\log\det\Lambda_j(d) + \log\det\Lambda_j^*(d) = \log\det\Lambda_j(d)\Lambda_j^*(d) = -2\sum_{a=1}^{q} d_a \log\lambda_j$$

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

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

To state our test statistic, we need to introduce an approximation of the first derivative of the objective function $R(d)$ that is used in Shimotsu (2007). For easier reference, we restate it in Lemma 1 below. Denote by $\eta = (\eta_1, \ldots, \eta_q)'$ a $(q \times 1)$ vector of real numbers with at least one $\eta_a \neq 0$ 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*

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

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:

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

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 that appear under long memory. Kechagias and Pipiras (2015) show that the phase will be given by $(d_a - d_b)\pi/2$ for every causal linear process with hyperbolically decaying coefficients in their Wold representation. The MLWS test will therefore also generate power against non-causal processes and can thus be interpreted as a general misspecification test. If one is willing to assume that the process is causal and has the required Wold representation (which is the case for the commonly used fractionally integrated model), than the test will be specifically against low frequency contaminations.

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, $\odot$ is the Hadamard product and $\Rightarrow$ denotes weak convergence:

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

$$
\begin{aligned}
MLWS \quad \Rightarrow \quad & \frac{1}{2} \sup_{r \in [\epsilon,1]} \Bigg\| \int_0^r \Bigg[(1+\log s)\left(2\eta'\eta + 2\eta'\left(G^0 \odot \left(G^0\right)^{-1}\right)\eta\right)^{1/2} \\
& + i\left[\frac{\pi^2}{2}\left(\eta'\left(G^0 \odot \left(G^0\right)^{-1}\right)\eta - \eta'\eta\right)\right]^{1/2}\Bigg] dB(s) \\
- \quad & 2\eta' B(1) \int_0^r (1+\log s) ds \eta \\
- \quad & \int_0^1 \Bigg[(1+\log s)\left(2\eta' F(r)\Omega^{-1}(G^0 \odot (G^0)^{-1} + I_q)\Omega^{-1'} F(r)'\eta\right)^{1/2} \\
& + i\left(\frac{\pi^2}{2}\eta' F(r)\Omega^{-1}(G^0 \odot (G^0)^{-1} - I_q)\Omega^{-1'} F(r)'\eta\right)^{1/2}\Bigg] dB(s) \Bigg\|,
\end{aligned} \tag{8}
$$

where $\Omega = 2\left[G^0 \odot (G^0)^{-1} + I_q + \frac{\pi^2}{4}(G^0 \odot \left(G^0\right)^{-1} - I_q)\right]$ *and*

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

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. Furthermore, 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 $T \to \infty$*

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

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 with $\eta_a = 1/\sqrt{q}\ \forall\ a = 1,...,q$ 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).

Remark 4: It is assumed that $d_a^0 \in (-1/2, 1/2)\ \forall\ a = 1, \ldots, q$, i.e. that the process has stationary long memory. However, the simulation results in Table 11 in the supplementary appendix indicate that the test statistic remains valid for $d < 1$.

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). Assume that as $T \to \infty$, we have $m/T^{1/2} \to \infty$, $P(\hat{d}_a - d_a^0 \geq 0) \to 1$ for all $a \in \{1,...,q\}$, where $\hat{G}(\hat{d})$ is positive definite and Assumptions 1 to 5 hold. Then, $MLWS \xrightarrow{P} \infty$, as $T \to \infty$, for any $\|\eta\| > 0$.*

Note that (3) and (4) nest the cases, where only a subvector of X_t is subject to low frequency contaminations. Theorem 2 therefore does not assume, that all components of X_t are affected. Furthermore, the consistency result holds for every weight vector η - except for the trivial case when all elements are zero. The intuition behind this is discussed in detail in Section 3.2.

To robustify the test against the influence of short memory dynamics in finite samples, we proceed in analogy to Qu (2011), and apply the MLWS test to the filtered series $\tilde{X}_t = \hat{\mathcal{A}}(L)^{-1}\hat{\mathcal{M}}(L)(X_t - EX_t)$, where $\hat{\mathcal{A}}(L)$ and $\hat{\mathcal{M}}(L)$ are the estimated lag-polynomials from a low order FIVARMA model in final equation form selected using the AIC. Details on the implementation of the pre-whitening procedure and its performance in Monte Carlo studies can be found in the supplementary appendix.
To prove the validity of this procedure, we sharpen Assumption 2 and replace it by Assumption 6, which is a multivariate version of Assumption F in Qu (2011).

Assumption 6. *Assume that in addition to Assumption 2 we have $A_j = O(j^{-1/2-c})$ with $c > 0$ as $j \to \infty$.*

We then obtain the following result.

Lemma 3. *Assume that X_t satisfies Assumptions 1 to 6. Then, the MLWS test applied on the filtered series $\tilde{X}_t$ has the same limiting distribution as given in Theorem 1.*

Note that we do not assume that the short memory dynamics follow a VARMA-process. We only use it as a reasonable approximation to the true short memory dynamics in finite 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.

3.2 Testing for Low Frequency Contaminations in a Component of a Multivariate System

A rejection of the MLWS statistic indicates misspecifications in at least one of the components of the process. To gain further insights into which of the components of X_t cause the rejection, one can use the limiting distribution derived in Theorem 1 to test the hypothesis

$$H_0(a): \ S(a) \odot f(\lambda_j) \sim S(a) \odot \left(\Lambda_j(d) G \Lambda_j^*(d)\right), \tag{9}$$

as $\lambda_j \to 0$, where $S(a)$ is a selection matrix with ones in its a-th row and a-th column and zeros in all other elements. Such a test is obtained by setting the a-th element of η to one and all others to zero.

In this case the limiting distribution simplifies slightly to

$$\begin{aligned} MLWS \quad \Rightarrow \quad & \frac{1}{2} \sup_{r \in [\epsilon, 1]} \Bigg\| \int_0^r \Bigg[\sqrt{2}(1+\log s) \left(\frac{g_{aa} \det(G^0_{-aa})}{\det(G^0)} + 1 \right)^{1/2} \\ & + i \left[\frac{\pi^2}{2} \left(\frac{g_{aa} \det(G^0_{-aa})}{\det(G^0)} - 1 \right) \right]^{1/2} \Bigg] dB(s) - 2B(1) \int_0^r (1+\log s) ds \\ - \quad & \int_0^1 \Big[(1+\log s) \left(2F(r) \Omega^{-1} (G^0 \odot (G^0)^{-1} + I_q) \Omega^{-1'} F(r)' \right)_{aa}^{1/2} \\ & + i \left(\frac{\pi^2}{2} F(r) \Omega^{-1} (G^0 \odot (G^0)^{-1} - I_q) \Omega^{-1'} F(r)' \right)_{aa}^{1/2} \Big] dB(s) \Bigg\|. \end{aligned} \tag{10}$$

Since the distribution depends on G^0, it is not pivotal and the implementation of the test statistic requires the simulation of critical values for each $\hat{G}(\hat{d})$.

Under the null hypothesis the a-th row and column of the spectral density matrix correspond to those of a multivariate long memory process. In case of a low frequency contamination in component $b \neq a$, only one of the off-diagonal elements in the a-th row and the a-th column is affected, whereas all elements in the b-th row and column differ from the null hypothesis.

A rejection of $H_0(a)$ might therefore be due to a low frequency contamination in the b-th component. However, a non-rejection of $H_0(a)$ and a rejection of $H_0(b)$ can be interpreted as evidence for a contamination in component b only. Furthermore, the ordering of the p-values of the test statistics can be used as an indication for the relative probability of a contamination in the respective components.

If there is indeed only a contamination in the component of the series for which the test is applied, the test statistic for this component will have better power than the test using equal weights. Other tests for contaminations in a subset of the components of X_t can be constructed following the same logic, by using a suitable weighting vector. However, an optimal choice of the weighting vector requires a priori information on the components of X_t that may be contaminated. If contaminations are considered to be equally likely in each component, then an equal weighting will generate the best power. We therefore recommend (as the default procedure) to start with an equal weighting scheme. In case of a rejection for the whole system, one can proceed with testing for contaminations in each of the components separately.

After the components that are subject to contaminations are identified, one is left with the task to model the time variation in the mean. A discussion of the available methods in the univariate case is provided by Qu (2011). Recently, McCloskey and Perron (2013) and Hou and Perron (2014) proposed semiparametric estimators of the memory parameters that are robust to spurious long memory and can be used to determine the memory orders of the components. The estimated memory orders can in turn be used together with the methods of Lavielle and Moulines (2000) or Beran and Feng (2002) - depending on the assumptions about the nature of the time variation in the mean.

3.3 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. Let there be p_G cointegrating relationships between the components of X_t, where $1 \leq p_G < q$, and assume without loss of generality that these involve the first p_G components of X_t. Then $rank(G) = q - p_G$, the memory order of the first p_G components is $d_1 = d_2 = ... = d_{p_G}$, and there exists a cointegrating matrix B, such that $BX_t = w_t$, and w_t has a spectral density matrix as specified in (1), but with $\tilde{G} = BGB'$ and $\tilde{d} = (d_{p_G} - b_1, ..., d_{p_G} - b_{p_G}, d_{p_G+1}, ..., d_q)'$ instead of G and d, for $0 < b_a \leq d_a$. The matrix B is such that in w_t the first p_G elements of X_t are replaced by the cointegrating residual series so that $\tilde{G}$ has full rank. This is achieved, if the first p_G rows of B contain the cointegrating vectors, normalized so that the diagonal elements of B equal 1 and the remaining rows contain zeros on all the off-diagonal elements. In the bivariate case, for example, B takes the form

$$B = \begin{pmatrix} 1 & \beta \\ 0 & 1 \end{pmatrix}.$$

Consequently, the MLWS test for the hypothesis $H_0: f_{BX_t}(\lambda_j) \sim \Lambda_j(d)\tilde{G}\Lambda_j^*(d)$, as $\lambda_j \to 0^+$ can simply be carried out on the transformed series BX_t, if the cointegrating matrix B is known. To obtain a feasible procedure for unknown B, a consistent estimator for B has to be applied, that converges with a faster rate than $\sqrt{m}$, where m is the bandwidth used for the MLWS statistic. For this purpose multivariate local Whittle estimators of $\hat{B}_{MLW}$, such as those of Robinson (2008b) or Shimotsu (2012) can be used for which ${}_a\hat{B}_{MLW}$ converges to ${}_aB^0$ with rate $\sqrt{m}\lambda_m^{-b_a}$. This is asymptotically equivalent to constructing the test statistic as in (7), but using the concentrated local Whittle likelihood of the cointegrated system.[1]

The resulting test statistic is

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

where $\tilde{G}(d,B) = \frac{1}{m}\sum_{j=1}^m Re\left[\Lambda_j(d)^{-1}BI(\lambda_j)B'\Lambda_j^*(d)^{-1}\right]$ and $\tilde{I}(\lambda_j,B) = BI(\lambda_j)B'$. Here, we write $\widehat{MLWS}(B)$ as a function of B, to stress the dependence on the cointegrating matrix that is used. We will write B^0 and p_G^0 for the true cointegrating matrix and cointegrating rank and $\hat{B}(p_G^0)$ and $\hat{B}(\hat{p}_G)$ for estimates of B^0 that are either based on the (known) true cointegrating rank or an estimation of it. We then obtain the following result.

Theorem 3. *Let Assumption 1 to 5 hold, and let* ${}_a\hat{B}(p_G^0) = {}_aB^0 + O_P\left(\sqrt{m}\lambda_m^{-b_a}\right)$. *Then, for known* p_G^0, *the test statistic* $\widehat{MLWS}(\hat{B}(p_G^0))$ *has the null limiting distribution of the MLWS statistic in Theorem 1, but with* G^0 *replaced by* $\tilde{G}^0 = B^0G^0B^{0\prime}$.

Consequently, if fractional cointegration is present, the cointegrated variables can be removed from X_t and replaced by an estimate of the cointegration residuals. Since the consistency of $\hat{B}$ depends on the assumed cointegrating rank, Theorem 3 assumes that the true cointegrating rank p_G^0 is known. However, we can extend the $\widehat{MLWS}(B)$ statistic to the case when p_G^0 is unknown but can be estimated consistently.

[1]It should be noted that Robinson (2008b) and Shimotsu (2012) derive their results for bivariate processes. However, Robinson (2008b) states that this restriction is mainly for expositional purposes and the results can be extended straightforward (cf. Remark 12 in Robinson (2008b)).

Theorem 4. *Let Assumption 1 to 5 hold, let* ${}_a\hat{B}(p_G^0) = {}_aB^0 + O_P\left(\sqrt{m}\lambda_m^{-b_a}\right)$*, and let* $\lim_{T\to\infty} P(\hat{p}_G = p^0) = 1$*. Then* $\widetilde{MLWS}(\hat{B}(\hat{p}_G)) \xrightarrow{d} \widetilde{MLWS}(\hat{B}(p_G^0))$*.*

Note that multivariate local Whittle estimators and narrow band estimators of the cointegrating vectors are usually derived under the assumption of a known cointegrating rank (mostly $p_G = 1$). Estimates of the cointegrating rank $\hat{p}_G$ (as required in Theorem 4) are proposed in Robinson and Yajima (2002) and Robinson (2008a), but the effect of an estimated p_G on subsequent estimators or tests is usually not considered. By allowing for an estimated cointegrating rank, we therefore improve the theoretical justification for the empirical application of the $\widetilde{MLWS}(B)$ test relative to other procedures.

There is a close relationship between Theorems 3 and 4. Theorem 3 considers the distribution of the test statistic conditional on the use of the correct cointegrating rank. Here $\hat{B}$ is required to converge to B^0 with a rate faster than $\sqrt{m}$ in Theorem 3. This is due to the appearance of the partial sum in the test statistic.

Theorem 4, on the other hand, applies to the test statistic without conditioning. It shows that the distribution of the test statistic based on an estimate of the cointegrating rank converges uniformly to the distribution of the test statistic conditional on the true cointegrating rank p_G^0 for any consistent estimator $\hat{p}_G$. Therefore, the methods of Robinson and Yajima (2002) and Robinson (2008a) can be used in this context.

The consistency of the test statistic under fractional cointegration is established in the following theorem. Note that we do not require p_G^0 to be known or estimated consistently under the alternative.

Theorem 5. *Suppose that the process* BX_t *is generated by (3) or (4) and* $\hat{B}$ *has full rank. Assume that as* $T \to \infty$*, we have* $m/T^{1/2} \to \infty$*,* $P(\hat{\tilde{d}}_a - \tilde{d}_a^0 \geq 0) \to 1$ *for all* $a \in \{1, ..., q\}$*,* $\tilde{G}(\hat{\tilde{d}}, \hat{B})$ *is positive definite and Assumptions 1 to 5 hold. Then,* $\widetilde{MLWS}(\hat{B}(\hat{p}_G)) \xrightarrow{p} \infty$*, as* $T \to \infty$*.*

4 Monte Carlo Study

To analyze the finite sample properties of the MLWS test, we conduct a Monte Carlo analysis that consists of four 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. Afterwards, we consider the finite sample performance of the test under fractional cointegration. Finally, in the fourth part, we study the properties of the test for breaks in components that was proposed in Section 3.2.

			$\rho_v = -0.8$				$\rho_v = 0$				$\rho_v = 0.4$				$\rho_v = 0.8$			
		d_2	0		0.4		0		0.4		0		0.4		0		0.4	
T	d_1	δ/ε	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**
250	0	**0.60**	0.006	0.007	0.007	0.012	0.005	0.010	0.006	0.009	0.005	0.008	0.007	0.008	0.004	0.008	0.007	0.007
		0.65	0.009	0.014	0.011	0.013	0.008	0.015	0.010	0.015	0.009	0.014	0.007	0.013	0.011	0.015	0.012	0.012
		0.70	0.012	0.017	0.012	0.015	0.011	0.016	0.014	0.018	0.013	0.017	0.011	0.016	0.010	0.021	0.011	0.019
		0.75	0.013	0.015	0.016	0.023	0.014	0.019	0.014	0.019	0.016	0.023	0.012	0.018	0.015	0.021	0.013	0.021
	0.4	**0.60**	0.007	0.010	0.008	0.008	0.007	0.009	0.007	0.010	0.005	0.008	0.006	0.009	0.008	0.008	0.006	0.009
		0.65	0.011	0.016	0.007	0.013	0.009	0.016	0.011	0.015	0.010	0.015	0.011	0.014	0.009	0.018	0.011	0.013
		0.70	0.013	0.014	0.017	0.021	0.014	0.018	0.011	0.017	0.012	0.016	0.014	0.019	0.015	0.017	0.012	0.018
		0.75	0.015	0.025	0.017	0.023	0.016	0.024	0.019	0.019	0.014	0.022	0.021	0.022	0.021	0.019	0.016	0.022
2000	0	**0.60**	0.022	0.031	0.023	0.027	0.022	0.025	0.021	0.028	0.021	0.033	0.016	0.027	0.028	0.030	0.023	0.028
		0.65	0.028	0.035	0.025	0.035	0.026	0.029	0.022	0.035	0.023	0.037	0.024	0.033	0.020	0.029	0.026	0.032
		0.70	0.029	0.029	0.031	0.038	0.025	0.034	0.026	0.031	0.026	0.035	0.024	0.035	0.028	0.033	0.030	0.034
		0.75	0.031	0.041	0.043	0.042	0.036	0.040	0.034	0.040	0.032	0.040	0.031	0.038	0.033	0.039	0.042	0.039
	0.4	**0.60**	0.021	0.034	0.022	0.027	0.018	0.025	0.021	0.026	0.020	0.032	0.018	0.028	0.021	0.028	0.018	0.029
		0.65	0.024	0.028	0.026	0.038	0.024	0.035	0.023	0.034	0.028	0.031	0.028	0.032	0.023	0.035	0.026	0.032
		0.70	0.035	0.033	0.028	0.033	0.028	0.033	0.030	0.033	0.028	0.034	0.027	0.034	0.026	0.034	0.031	0.028
		0.75	0.036	0.044	0.035	0.042	0.032	0.036	0.033	0.043	0.034	0.043	0.038	0.039	0.037	0.046	0.038	0.043

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 $m = \lfloor T^\delta \rfloor$.

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 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 random level shift process for all power DGPs. Further simulation studies are discussed briefly in Section 4.5 and included in full length 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 (2), where the short memory component $u_t = v_t$ with $v_t \sim N(0, \Sigma_v)$ and $\Sigma_v = ((1, \rho_v), (\rho_v, 1))'$ 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 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

T	$\rho_\pi = \rho_e$	Stationary ($\phi = 1$)						Non-Stationary ($\phi = 0$)					
		0		0.5		1		0		0.5		1	
	δ/ε	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**
250	**0.60**	0.114	0.120	0.110	0.128	0.199	0.196	0.184	0.209	0.192	0.218	0.296	0.302
	0.65	0.264	0.308	0.259	0.299	0.350	0.377	0.395	0.441	0.407	0.465	0.485	0.502
	0.70	0.447	0.475	0.445	0.463	0.494	0.509	0.598	0.610	0.613	0.623	0.641	0.634
	0.75	0.632	0.640	0.638	0.645	0.634	0.633	0.766	0.764	0.764	0.768	0.743	0.730
2000	**0.60**	0.918	0.922	0.910	0.921	0.912	0.908	0.957	0.963	0.954	0.961	0.936	0.944
	0.65	0.984	0.984	0.982	0.982	0.959	0.958	0.991	0.993	0.988	0.986	0.963	0.964
	0.70	0.995	0.996	0.994	0.993	0.970	0.969	0.996	0.997	0.996	0.994	0.977	0.970
	0.75	0.998	0.996	0.995	0.994	0.976	0.974	0.999	0.998	0.995	0.994	0.979	0.977

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 $\mu_t = (I_q - \phi\Pi_t)\mu_{t-1} + \Pi_t e_t$. The bandwidth m is determined by $m = \lfloor T^\delta \rfloor$.

- 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 $T = 2000$, $m = \lfloor T^{0.75} \rfloor$ and $\varepsilon = 0.05$ for example, we find that the size is between 3.6 and 4.6 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 in larger samples and it is 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. Here, Y_t is the sum of the white noise sequence v_t and the multivariate random level shift process μ_t from equation (3):

$$\begin{aligned} Y_t &= \mu_t + v_t \\ \mu_t &= (I_q - \phi\Pi_t)\mu_{t-1} + \Pi_t e_t. \end{aligned} \tag{11}$$

For $\phi = 1$ the process is stationary and for $\phi = 0$ it is non-stationary. 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 could imply different 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

		Size							Power					
		MLWS			Qu				MLWS			Qu		
T	q/ρ_v	**0**	**0.4**	**0.8**	**0**	**0.4**	**0.8**	$q/\rho_\pi,\rho_e$	**0**	**0.5**	**1**	**0**	**0.5**	**1**
	1	0.011	0.010	0.011	0.014	0.013	0.012	**1**	0.098	0.093	0.095	0.093	0.098	0.094
	2	0.011	0.013	0.015	0.008	0.010	0.010	**2**	0.173	0.172	0.205	0.121	0.125	0.105
100	**3**	0.013	0.015	0.011	0.007	0.007	0.007	**3**	0.243	0.243	0.296	0.125	0.118	0.111
	4	0.013	0.014	0.009	0.007	0.010	0.008	**4**	0.295	0.309	0.328	0.128	0.126	0.120
	5	0.011	0.010	0.011	0.006	0.006	0.007	**5**	0.356	0.357	0.366	0.135	0.133	0.118
	1	0.027	0.026	0.025	0.027	0.027	0.026	**1**	0.751	0.742	0.752	0.742	0.743	0.747
	2	0.028	0.029	0.026	0.025	0.021	0.026	**2**	0.922	0.919	0.865	0.911	0.890	0.813
500	**3**	0.026	0.028	0.029	0.018	0.021	0.021	**3**	0.979	0.973	0.906	0.963	0.951	0.834
	4	0.029	0.029	0.029	0.025	0.026	0.021	**4**	0.996	0.987	0.917	0.983	0.974	0.849
	5	0.026	0.031	0.028	0.021	0.023	0.022	**5**	0.999	0.994	0.923	0.994	0.987	0.857

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 $v_t \sim N(0,\Sigma_v)$.

of the components of the series, whereas shifts always coincide in timing and size if $\rho_\pi = \rho_e = 1$.[2]

The results of this experiment are shown in Table 2. We find that the power is always increasing in the bandwidth and it is higher against non-stationary level shifts. For small sample sizes with weakly correlated shifts the 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, 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. In view of these results the rule of thumb to choose $\varepsilon = 0.05$ for $T \leq 500$, that is suggested by Qu (2011), still works well. The same holds true for using $m = \lfloor T^{0.7} \rfloor$ as the bandwidth.

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.

[2] 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.

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, some kind of size correction has to be employed. Since a standard Bonferroni correction is based on the assumption of independence between the test statistics, there is a considerable power loss. We therefore employ 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. As Sarkar (1998) shows, this approach is valid for processes X_t that are multivariate totally positive of order two, which is fulfilled in the Monte Carlo study, where the process is multivariate Gaussian. 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 (11). 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 24 percentage points. We find that the power is increasing in q and T. While correlated shifts increase the power in smaller samples, the power reduction observed in the bivariate simulations for correlated shifts in large samples increases with increasing q.

4.3 Testing Against Breaks in Fractionally Cointegrated Systems

In Section 3.3, we derived the limiting distribution of the $\widetilde{MLWS}(B)$ statistic calculated for a consistent estimate $\hat{B}(p_G^0)$ of the cointegrating matrix B^0, so that the test can be

		ρ_v	0		0.4		0.8		0		0.4		0.8	
			Size						**Power**					
d_2	T	δ/ε	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**
0.1	250	**0.60**	0.005	0.007	0.005	0.008	0.006	0.007	0.076	0.088	0.075	0.091	0.077	0.102
		0.65	0.008	0.016	0.010	0.010	0.009	0.013	0.179	0.223	0.178	0.198	0.187	0.227
		0.70	0.010	0.014	0.012	0.014	0.009	0.014	0.321	0.327	0.303	0.322	0.337	0.355
		0.75	0.012	0.021	0.011	0.019	0.011	0.016	0.471	0.465	0.459	0.455	0.509	0.517
	2000	**0.60**	0.018	0.025	0.015	0.026	0.021	0.026	0.854	0.862	0.864	0.861	0.862	0.872
		0.65	0.026	0.035	0.021	0.030	0.021	0.032	0.942	0.949	0.945	0.952	0.962	0.963
		0.70	0.028	0.037	0.028	0.036	0.028	0.031	0.972	0.972	0.974	0.966	0.979	0.979
		0.75	0.033	0.039	0.031	0.039	0.039	0.042	0.980	0.978	0.982	0.976	0.984	0.984
0.4	250	**0.60**	0.006	0.009	0.004	0.009	0.004	0.007	0.034	0.041	0.034	0.039	0.060	0.065
		0.65	0.008	0.013	0.007	0.013	0.008	0.012	0.075	0.091	0.077	0.102	0.145	0.156
		0.70	0.013	0.016	0.011	0.017	0.009	0.015	0.140	0.139	0.163	0.168	0.257	0.255
		0.75	0.018	0.022	0.014	0.021	0.015	0.022	0.235	0.227	0.262	0.262	0.401	0.400
	2000	**0.60**	0.020	0.031	0.016	0.032	0.020	0.029	0.582	0.605	0.665	0.667	0.745	0.766
		0.65	0.021	0.029	0.024	0.035	0.021	0.030	0.811	0.821	0.856	0.864	0.911	0.919
		0.70	0.026	0.034	0.022	0.038	0.026	0.033	0.912	0.902	0.932	0.933	0.966	0.964
		0.75	0.031	0.042	0.032	0.036	0.031	0.043	0.942	0.934	0.958	0.946	0.977	0.970

Table 4: Size and power of the MLWS test in a bivariate fractionally cointegrated system, where $D(0,d_2)BX_t = v_t$ with $B = ((1,0)', (-1,1)')$, $v_t \sim N(0, \Sigma_v)$ and $\Sigma_v = ((1,\rho_v),(\rho_v,1))'$. The bandwidth m is determined by $m = \lfloor T^\delta \rfloor$.

applied to the linearly transformed system $\hat{B}X_t$. To explore the finite sample performance of this approach, we conduct a simulation study where the DGP is

$$D(0,d_2)\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} X_t = v_t.$$

Here the components of X_t are fractionally cointegrated with cointegrating vector $(1,-1)'$. The parameter d_2 determines the memory of both components in X_t and since $d_1 = 0$, the memory in the linear combination is reduced to zero. By increasing d_2 the cointegration strength is increased. The correlation between the innovations to the linear combination and the common fractional trend is determined by ρ_v.

The results of this experiment are shown in Table 4. First, the size remains conservative for all parameter constellations. The power, on the other hand, is higher the higher the correlation ρ_v. Furthermore, one can observe that the power is decreasing with increasing strength of the cointegrating relationship. Since the convergence rate of the local Whittle estimator for the B matrix is faster if the cointegrating relationship is stronger, this effect cannot be attributed to the effect of the estimation error. Instead, the MLWS test has lower power to detect contaminations if the memory is stronger. For the Qu test this was pointed out by Kruse (2015), who advocates to apply the test to the fractionally differenced process. This is also visible in the results of Table 11 in the supplementary appendix.

4.4 Testing for Breaks in Components of a Multivariate System

In Section 3.2 we introduced a variation of the MLWS test where all components of the weight vector η are set to zero and only one takes the value 1. This allows to test for misspecifications in components of the spectral density matrix and can be used to gain further insights about the components of X_t that cause a rejection of the MLWS test with equal weights.

		$\eta = \frac{1}{\sqrt{2}}(1,1)'$				$\eta = (1,0)'$				$\eta = (0,1)'$			
	ξ	0		1		0		1		0		1	
T	δ/ε	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**
250	**0.60**	0.006	0.008	0.036	0.044	0.018	0.019	0.208	0.204	0.020	0.019	0.013	0.024
	0.65	0.010	0.017	0.089	0.100	0.023	0.025	0.351	0.367	0.021	0.024	0.019	0.032
	0.70	0.013	0.018	0.158	0.166	0.024	0.026	0.495	0.512	0.025	0.022	0.027	0.038
	0.75	0.016	0.021	0.273	0.268	0.024	0.023	0.633	0.629	0.024	0.032	0.030	0.042
2000	**0.60**	0.018	0.029	0.553	0.578	0.040	0.046	0.892	0.907	0.037	0.040	0.123	0.164
	0.65	0.025	0.034	0.826	0.827	0.045	0.052	0.959	0.962	0.044	0.052	0.184	0.239
	0.70	0.031	0.035	0.928	0.930	0.046	0.051	0.973	0.973	0.046	0.058	0.253	0.322
	0.75	0.033	0.039	0.959	0.952	0.051	0.055	0.976	0.976	0.054	0.052	0.279	0.340

Table 5: MLWS test for breaks in components using different weight vectors η for the DGP $Y_t = (\xi, 0)'\mu_t + v_t$, with $v_t \sim N(0, \Sigma_v)$ and $\Sigma_v = ((1, 0.4), (0.4, 1))'$. The bandwidth m is determined by $m = \lfloor T^{\delta} \rfloor$.

The performance of the MLWS test using the proposed weighting scheme is evaluated in Table 5. Here the DGP is a bivariate fractionally integrated process with stationary random level shifts only in its first component. It is given by

$$Y_t = \begin{pmatrix} \xi \\ 0 \end{pmatrix} \mu_t + v_t,$$

with $v_t \sim N(0, \Sigma_v)$, $\Sigma_v = ((1, 0.4), (0.4, 1))'$ and a shift variance of one. The parameter ξ controls the magnitude of the breaks. To determine the critical values based on an estimate $\hat{G}(\hat{d})$, we approximate the integrals in (10) by sums over 500 increments and we draw 1000 values. One can observe for $\xi = 0$ that the size is similar to the size of the global test. Also, as one would expect, the test generates better power if we specifically test for a contamination in the first component, compared to the baseline case with equal weights. Furthermore, if one specifically tests for a contamination in the second component, the test does not generate much power due to the misspecification of the off-diagonal components alone. A rejection therefore gives a good indication for a low frequency contamination in the respective component.

4.5 Further Simulations

A number of further simulations are provided in a supplementary appendix. First, we analyze the performance of the test if short memory dynamics exist. Without pre-whitening the size is no longer controlled for larger bandwidths $m = \lfloor T^{\delta} \rfloor$. With pre-whitening, on the other hand, the test remains conservative and the power loss is reasonable.
We also explore the impact of non-stationary long memory, perturbations, heteroscedasticity, breaks in the variance-covariance matrix of the innovations, the power against other alternative processes and the performance of the test, if the pre-whitening is conducted using univariate estimators. It is found, that the MLWS test is stable under all these complications. However, power against non-causal alternatives is only developed very slowly.

5 Empirical Example

Log-absolute returns of stock market indices are a typical example in the spurious long memory literature - in particular that of the Standard & Poor's 500 (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 log-absolute return series. However, univariate tests are often not able to reject the null hypothesis of true long memory. 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.
Due to the increased availability of high frequency data, the focus in the more recent literature has shifted to the modelling of realized volatility. Especially the heterogenous

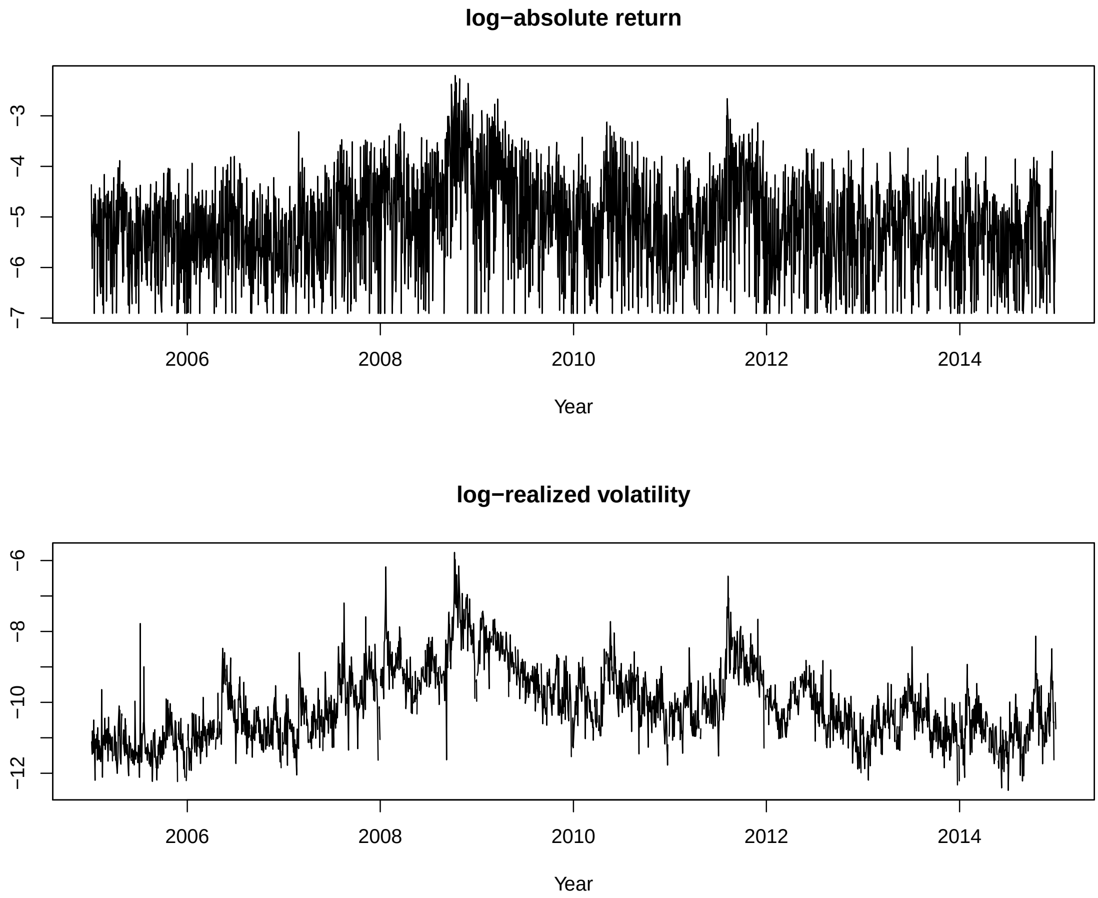

Figure 1: The log-absolute return and the log-realized volatility of the S&P 500.

autoregressive model of Corsi (2009) and its extensions have become very popular. As for the log-absolute return series existing tests against spurious long memory tend not to reject their null hypothesis if applied to these realized volatility series. An example is the application in Qu (2011), who finds no evidence for low frequency contaminations in the realized volatility of the exchange rate between Japanese Yen and US Dollar.

In view of the power gains of the multivariate procedure demonstrated in Section 4.2, we revisit these variation series of the S&P 500 and additionally consider those of the DAX, FTSE and NIKKEI in a multivariate setup to test for spurious long memory using the MLWS test. The analysis is conducted for both - the log-absolute return and the log-realized volatility.

We analyze the period from 2005/01/03 to 2014/12/31 (T=2608 observations). Data on daily stock price indices is obtained from Thomson Reuters 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 $\ln(|r_t| + 0.001)$.[3] Realized volatil-

[3] The constant 0.001 is added to avoid infinite values for zero returns, which is customary in the literature

δ	DAX	NIKKEI	S&P 500	FTSE	partitions	coint.rank	$\hat{\beta}$	$\hat{\mathbf{d}}_{\mathbf{w}}$
0.60	0.379	0.295	0.472	0.393	(1,1,1,1)	1	(0.155, 0.066, -1.446)	0.260
0.65	0.338	0.290	0.411	0.362	(1,1,1,1)	1	(-0.119, 0.079, -1.014)	0.236
0.70	0.328	0.285	0.359	0.303	(1,1,1,1)	1	(-0.043, -0.144, -1.153)	0.194
0.75	0.264	0.252	0.300	0.290	(1,1,1,1)	1	(-0.074, -0.037, -0.954)	0.139

Table 6: Fractional cointegration analysis for the log-absolute return series based on local Whittle estimates of d with different bandwidths $m = \lfloor T^{\delta} \rfloor$.

ities calculated from 5 minute returns are obtained from the Oxford-Man Realized Library.

As an example, Figure 1 depicts the log-absolute return and log-realized volatility of the S&P 500 series. Both series show the typical features of long memory time series, with local trends and cycles. This is also confirmed by the autocorrelation functions and the periodograms given in Figures 2 and 3 in the supplementary appendix. Since the series of the DAX, 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 20 (in the supplementary appendix). It can be seen that all four series have similar locations and standard deviations if the same variation measure is used. With the exception of the S&P 500, the distributions of the log-absolute return series are slightly negatively skewed and all log-absolute return series have lighter tails than the normal distribution. The realized volatility series on the other hand are positively skewed and have excess kurtosis.

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 Tables 6 and 7. The analysis is carried out for different bandwidths $m = \lfloor T^{\delta} \rfloor$ using the local Whittle estimator. For both variation measures 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. Since the log-absolute return series is considered to be a noisy estimate of the underlying variation process and perturbations cause a downward bias in the local Whittle estimator, we include further results using different specifications of the LPWN estimator of

(cf. for example Lu and Perron (2010) and Xu and Perron (2014)).

δ	DAX	NIKKEI	S&P 500	FTSE	partitions	coint.rank	$\hat{\beta}_{13}$	$\hat{\beta}_{23}$	$\hat{\mathbf{d}}_{w_{13}}$	$\hat{\mathbf{d}}_{w_{23}}$
0.60	0.642	0.631	0.635	0.637	(1,1,1,1)	2	-0.931	-1.084	0.464	0.596
0.65	0.605	0.612	0.642	0.570	(1,1,1,1)	2	-0.728	-0.978	0.463	0.514
0.70	0.594	0.611	0.633	0.568	(1,1,1,1)	2	-0.868	-1.068	0.400	0.483
0.75	0.563	0.573	0.588	0.540	(1,1,1,1)	2	-0.910	-1.172	0.368	0.447

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

Frederiksen et al. (2012) and the robust estimator of Hou and Perron (2014) in Table 21 in the supplementary appendix. It can be observed that there is a downward bias for the log-absolute return series. Nevertheless, as discussed in Section 4, the MLWS test is fairly robust to perturbations. Also the Hou-Perron estimator is lower for the S&P 500, which is a further indication of spurious long memory. Apart from that, all estimates turn out to be very stable.

It should be noted that the estimated memory parameters of the log-realized volatility series are in the lower non-stationary region, which is not covered by the assumptions under which the test statistic is derived. However, our simulation results indicate that the test statistic remains valid for non-stationary long memory processes. We therefore proceed with the analysis and provide an additional robustness check with the test carried out on the fractionally differenced series in the supplementary appendix.

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 stable for different bandwidth choices. We find that there is one cointegrating relationship between the four log-absolute return series and there are two relationships between the realized volatility series.

As described in Section 3.3, the analysis than proceeds by estimating the cointegrating matrix B using the multivariate local Whittle estimator of Robinson (2008b) with the phase set to $(d_a - d_b)(\pi - \lambda)/2$.

In the case of the log-absolute return series the DAX series is specified to be the variable that is replaced by the linear combination. For the log-realized volatility series we assume pairwise relationships of the DAX and the NIKKEI with the S&P 500. Subsequently, the transformed series $\hat{B}X_t$ are obtained. Additionally, we report the estimate $\hat{d}_w$ of the noise term in the last column of Table 6 and the last two columns of Table 7 to show that the memory in the linear combination is reduced. When the cointegrating rank analysis is repeated on the transformed series there is no evidence for a cointegrating

δ	Qu test DAX	NIKKEI	S&P 500	FTSE	Components DAX	NIKKEI	S&P 500	FTSE	MLWS ALL
					$\log(\lvert r_t \rvert + 0.001)$				
0.60	0.521	0.860	0.515	0.446	1.245	0.877	0.765	1.085	1.233
	(0.862)	(0.314)	(0.871)	(0.949)	(0.494)	(0.665)	(0.854)	(0.912)	(0.054)
0.65	0.505	0.749	1.078	0.443	1.052	0.793	1.681	0.619	1.470
	(0.886)	(0.474)	(0.118)	(0.953)	(0.220)	(0.495)	(0.013)	(0.979)	(0.014)
0.70	0.395	0.519	1.100	0.739	1.370	0.814	1.726	0.557	1.448
	(0.983)	(0.865)	(0.107)	(0.492)	(0.179)	(0.395)	(0.004)	(0.998)	(0.016)
0.75	0.640	0.469	1.477	0.547	1.322	0.918	1.843	0.598	1.413
	(0.662)	(0.929)	(0.013)	(0.824)	(0.057)	(0.257)	(0.002)	(0.948)	(0.019)
					$\log RV_t$				
0.60	0.317	0.425	1.179	0.544	0.700	0.452	1.283	0.548	0.662
	(0.999)	(0.966)	(0.071)	(0.829)	(0.600)	(0.978)	(0.173)	(0.992)	(0.621)
0.65	0.445	0.641	0.807	0.929	1.241	0.965	1.140	0.985	1.465
	(0.950)	(0.661)	(0.387)	(0.236)	(0.136)	(0.187)	(0.279)	(0.425)	(0.014)
0.70	0.406	0.657	0.700	0.670	0.617	0.539	1.039	0.807	0.643
	(0.977)	(0.632)	(0.554)	(0.607)	(0.729)	(0.881)	(0.331)	(0.635)	(0.656)
0.75	0.597	1.062	0.724	1.022	0.488	0.397	0.614	1.199	0.683
	(0.740)	(0.129)	(0.513)	(0.152)	(0.946)	(0.999)	(0.919)	(0.082)	(0.585)

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.

relationship anymore, supporting the selection of the estimated cointegrating relations. It should be noted, that the rank-estimation procedure of Robinson and Yajima (2002) operates under the assumption of a multivariate long memory series. In the presence of low frequency contaminations, on the other hand, it may no longer be consistent. The estimates of the cointegrating relations should therefore not be interpreted unless the MLWS test fails to reject.

To formally test for true long memory, we then apply the $\widetilde{MLWS}(\hat{B}(\hat{p}_G))$ to the system $\hat{B}X_t$. The asymptotic validity of this approach is established in Theorems 4 and 5. In addition to that, some simulations for finite samples with parameter constellations similar to those found in the log-absolute return series and log-realized volatility series are provided in the supplementary materials. These show that the test maintains its size in the situation at hand.

The test for contaminations in components of the system discussed in Section 3.2 is applied to analyze which components of the series might cause a rejection of the $\widetilde{MLWS}$

test. 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 test statistics are given in Table 8, where the p-values are displayed in brackets.[4]

As one can see, Qu's univariate test fails to reject the null hypothesis of true long memory for each country, all bandwidth specifications, and both variation measures. The only exception is the log-absolute return series of the S&P 500, if the bandwidth is set to $m = \lfloor T^{0.75} \rfloor$. This would lead to the conclusion that there are no low frequency contaminations in the variation of stock returns. Similarly, the $\widetilde{MLWS}$ test calculated for the realized volatility series also fails to reject - except for $m = \lfloor T^{0.75} \rfloor$. For the log-absolute return series, on the other hand, the $\widetilde{MLWS}$ statistic rejects for all but one bandwidth.

If one considers the tests for contaminations in components of the spectral density matrix, we find that the test rejects for the S&P 500 series if the bandwidth parameter is $\delta \in \{0.65, 0.70, 0.75\}$ for the log-absolute return series, but not for the realized volatility series. The application of the $\widetilde{MLWS}$ 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. We find that one would falsely conclude that the process is not contaminated, if only the univariate test is used. In contrast to that, there is little evidence for low frequency contaminations in the log-realized volatility series. The Qu test as well as the $\widetilde{MLWS}$ test for contaminations in a specific component generate no rejection at the 95-percent level. Only the test with an equal weighting scheme generates a single rejection for $\delta = 0.65$. We therefore conclude that the realized volatility series are well modelled as long memory processes.

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. Consistency against multivariate random level shift processes and smooth trends is shown.

[4] Due to the large number of free parameters in the 4-dimensional example, the pre-whitening is carried out for each series separately. Monte Carlo results supporting the validity of this approach are provided in the supplementary appendix.

Our test encompasses the test of Qu (2011) as a special case for scalar processes. Apart from the generalization to vector valued series, we consider several issues that are unique to the multivariate case. First, we provide a modification of the test statistic in the case of fractionally cointegrated series. Second, by changing the weighting scheme, the multivariate test statistic can be used to gain insights about which components of the multivariate series cause a rejection.
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. Likewise, the $\widetilde{MLWS}(B)$ statistic and the test for contaminations in components of the process perform well in finite samples.
In our empirical example we consider the log-absolute returns and the log-realized volatilities of the S&P 500 together with those of the DAX, the FTSE and the NIKKEI in a multivariate framework. By applying our multivariate test, we find evidence of spurious long memory in the log-absolute returns of the S&P 500. A simple application of the univariate Qu test to the 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. For realized volatilities, on the other hand, no such evidence is found.

Appendix

Proof of Theorem 1:

To prove the theorem we start with the Taylor expansion

$$\sqrt{m}\eta' \frac{\partial R^r(d)}{\partial d}\bigg|_{\hat{d}} = \sqrt{m}\eta' \frac{\partial R^r(d)}{\partial d}\bigg|_{d^0} + \sqrt{m}\eta' \frac{\partial^2 R^r(d)}{\partial d \partial d'}\Big|_{\bar{d}} (\hat{d} - d^0) \qquad (12)$$

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

(12) we can write:

$$\sum_{a=1}^{q}\eta_a\sqrt{m}\frac{\partial R^r(d)}{\partial d_a}|_{d^0} = \frac{2}{\sqrt{m}}\sum_{a=1}^{q}\eta_a\sum_{j=1}^{[mr]}\nu_j\left({}_a\left(G^0\right)^{-1}\left[Re\left[\Lambda_j^0(d)^{-1}I(\lambda_j)\Lambda_j^{0*}(d)^{-1}\right]\right]_a-1\right) \tag{13}$$

$$+o_P(1)+tr\left[\hat{G}\left(d^0\right)^{-1}\frac{1}{\sqrt{m}}\sum_{j=1}^{[mr]}\frac{\lambda_j-\pi}{2}Im\left[(\Lambda_j^0(d))^{-1}(-i_aI(\lambda_j)\right.\right.$$

$$\left.\left.+I(\lambda_j)i_a)(\Lambda_j^{0*}(d))^{-1}\right]\right]$$

$$= \frac{2}{\sqrt{m}}\sum_{a=1}^{q}{}_aG^0\eta_a\sum_{j=1}^{[mr]}\nu_j\left[{}_a\left(G^0\right)^{-1'}{}_a\left(G^0\right)^{-1}\left[Re\left[\Lambda_j^0(d)^{-1}I(\lambda_j)\Lambda_j^{0*}(d)^{-1}\right]\right]_a-1\right]$$

$$-\frac{2{}_aG^0}{m^{3/2}}\left(\sum_{j=1}^{[mr]}\nu_j\right)\sum_{j=1}^{m}\left[{}_a\left(G^0\right)^{-1'}{}_a\left(G^0\right)^{-1}\left[Re\left[\Lambda_j^0(d)^{-1}A(\lambda_j)I_{\varepsilon j}A^*(\lambda_j)\right.\right.\right.$$

$$\left.\left.\times\Lambda_j^{0*}(d)^{-1}\right]\right]_a-1\Big]+o_P(1)+tr\left[\hat{G}(d_0)^{-1}\frac{1}{\sqrt{m}}\sum_{j=1}^{[mr]}\frac{\lambda_j-\pi}{2}Im\left[(\Lambda_j^0(d))^{-1}\right.\right.$$

$$\left.\left.\times(-i_aI(\lambda_j)+I(\lambda_j)i_a)(\Lambda_j^{0*}(d))^{-1}\right]\right].$$

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 $\Theta_s=\frac{1}{\pi\sqrt{mT}}\sum_{j=1}^{m}\nu_jRe[\Psi_j+\Psi_j^{'}]\cos(s\lambda_j)$, $\tilde{\Theta}_s=\frac{\pi}{2}\frac{1}{\pi\sqrt{mT}}\sum_{j=1}^{m}Re[\Psi_j-\Psi_j^{'}]\sin(s\lambda_j)$, 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 $0\le r_1\le r_2\le 1$

$$Cov\left(\sum_{t=1}^{T}z_{t,r_1},\sum_{t=1}^{T}z_{t,r_2}\right) = E\left(\sum_{t=1}^{T}\sum_{s=1}^{t-1}\left(\Theta_{t-s,r_1}+\tilde{\Theta}_{t-s,r_1}\right)^{'}\varepsilon_s^{'}\sum_{l=1}^{t-1}\left(\Theta_{t-s,r_2}+\tilde{\Theta}_{t-s,r_2}\right)\varepsilon_l\right)$$

$$= \sum_{t=2}^{T}\sum_{s=1}^{t-1}tr\left[\left(\Theta_{t-s,r_1}+\tilde{\Theta}_{t-s,r_1}\right)^{'}\left(\Theta_{t-s,r_2}+\tilde{\Theta}_{t-s,r_2}\right)\right]+o_P(1)$$

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

$$\sum_{t=2}^{T}\sum_{s=1}^{t-1}\Theta_{t-s,r_1}^{'}\tilde{\Theta}_{t-s,r_2} = \frac{1}{2\pi mT^2}\sum_{t=1}^{T-1}\sum_{s=1}^{T-1}\sum_{j=1}^{[mr_1]}\sum_{l=1}^{[mr_2]}\nu_j\nu_l tr\left[Re\left[\Psi_j^{'}+\Psi_j\right]Re\left[\Psi_l-\Psi_l^{'}\right]\right]$$

$$\times\cos(s\lambda_j)\sin(s\lambda_j)=0$$

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

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

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 $\lambda_j\to 0$

$$\begin{aligned}
& \frac{1}{\pi^2 mT^2}\sum_{t=1}^{T-1}\sum_{s=1}^{T-t}\sum_{j=1}^{[mr_1]} \nu_j^2 tr\left[Re\left[\Psi^{'}_j+\Psi_j\right]Re\left[\Psi_j+\Psi^{'}_j\right]\right]\cos^2(s\lambda_j) \\
= & \frac{1}{m}\sum_{j=1}^{[mr_1]}\nu_j^2\frac{1}{4\pi^2} tr\left[Re\left[\Psi^{'}_j+\Psi_j\right]Re\left[\Psi_j+\Psi^{'}_j\right]\right] \\
= & \frac{1}{m}\sum_{j=1}^{[mr_1]}\nu_j^2\left(2\sum_{a=1}^{q}\eta_a^2+2\sum_{a=1}^{q}\sum_{b=1}^{q}\eta_a\eta_b G^0_{ab}\left(G^0\right)^{-1}_{ab}\right).
\end{aligned}$$

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

$$\begin{aligned}
& \frac{\pi^2}{4}\frac{1}{\pi^2 mT^2}\sum_{t=1}^{T-1}\sum_{s=1}^{T-t}\sum_{j=1}^{[mr_1]} tr\left[Re\left[\Psi^{'}_j-\Psi_j\right]Re\left[\Psi_j-\Psi^{'}_j\right]\right]\sin^2(s\lambda_j) \\
= & \frac{\pi^2}{4m}\sum_{j=1}^{[mr_1]}\frac{1}{4\pi^2} tr\left[Re\left[\Psi^{'}_j-\Psi_j\right]Re\left[\Psi_j-\Psi^{'}_j\right]\right] \\
= & \frac{\pi^2}{4m}\sum_{j=1}^{[mr_1]}\left(2\sum_{a=1}^{q}\sum_{b=1}^{q}\eta_a\eta_b G^0_{ab}\left(G^0\right)^{-1}_{ab}-2\sum_{a=1}^{q}\eta_a^2\right).
\end{aligned}$$

From the Euler-Mc Laurin equality and Lemma B.1 in Qu (2011) it follows that $1/m\sum_{j=1}^{[mr_1]} \nu_j^2 \to \int_0^{r_1}(1+\log s)^2 ds$. 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 (13) we have by similar arguments as in Qu (2011) that

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

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

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

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

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

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 $\hat{G}_{1a}(d) = i_a\hat{G}(d) + \hat{G}(d)i_a$ and the superscript r denoting again that the sum goes up to $[mr]$ rather than m. 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 $G_{3ab}(d) = -i_a i_b\hat{G}(d) + i_a\hat{G}(d)i_b + i_b\hat{G}(d)i_a - \hat{G}(d)i_a i_b$. It also 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

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

so that

$$\begin{aligned}
\frac{\partial^2 R^r(d)}{\partial d\partial d'} &\to 2\int_0^r\left((1+\log s)^2(G^0\odot\left(G^0\right)^{-1}+I_q)+\frac{\pi^2}{4}(G^0\odot\left(G^0\right)^{-1}-I_q)\right)ds \\
&:= F(r). \qquad (14)
\end{aligned}$$

From the mean value theorem, we have $\sqrt{m}(\hat{d}-d^0)=\sqrt{m}\left(\frac{\partial^2 R(d)}{\partial d\partial d'}|_{\bar{d}}\right)^{-1}\frac{R(d)}{\partial d}|_{d^0}$. Since from Shimotsu (2007) $\frac{\partial^2 R(d)}{\partial d\partial d'}|_{\bar{d}}\to\Omega$, with $\Omega=2\left[G^0\odot(G^0)^{-1}+I_q+\frac{\pi^2}{4}(G^0\odot\left(G^0\right)^{-1}-I_q)\right]$, we have $\sqrt{m}(\hat{d}-d^0)\to\sqrt{m}\Omega^{-1}\frac{\partial R(d)}{\partial d}|_{d^0}$ and finally using the result from (14)

$$\eta'\frac{\partial^2 R^r(d)}{\partial d\partial d'}\sqrt{m}(\hat{d}-d^0)\to\eta' F(r)\Omega^{-1}\sqrt{m}\frac{\partial R(d)}{\partial d}|_{d^0}. \qquad (15)$$

Now, $\sqrt{m}\frac{\partial R(d)}{\partial d}|_{d^0}$ can be treated as before. Thus, the right hand side of (15) has the covariance $\int_0^1(1+\log s)^2 2\eta' F(r)\Omega^{-1}(G^0\odot(G^0)^{-1}+I_q)\Omega^{-1'}F(r)'\eta+\frac{\pi^2}{2}\eta' F(r)\Omega^{-1}(G^0\odot(G^0)^{-1}-I_q)\Omega^{-1'}F(r)'\eta ds$. Altogether, we obtain

$$\begin{aligned}
\eta'\frac{\partial^2 R^r(d)}{\partial d\partial d'}\sqrt{m}(\hat{d}-d^0) &\Rightarrow \int_0^1\left[(1+\log s)\left(2\eta' F(r)\Omega^{-1}(G^0\odot(G^0)^{-1}+I_q)\Omega^{-1'}F(r)'\eta\right)^{1/2}\right. \\
&\quad \left.+i\left(\frac{\pi^2}{2}\eta' F(r)\Omega^{-1}(G^0\odot(G^0)^{-1}-I_q)\Omega^{-1'}F(r)'\eta\right)^{1/2}\right]dB(s).
\end{aligned}$$

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

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

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

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

This proves the theorem. □

Proof of Lemma 2:

To prove the lemma, we first need to show that $\eta'(G^0\odot(G^0)^{-1})\eta=1$, if $\eta=(1/\sqrt{q},...,1/\sqrt{q})'$. 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, by using Cramer's rule we obtain for the inverse matrix

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

where G_{-ab} denotes the matrix G with the a-th row and b-th column omitted. 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,

$$G^0 \odot \left(G^0\right)^{-1}\eta = \frac{1}{\det(G^0)} \begin{pmatrix} \frac{1}{\sqrt{q}}\sum_{a=1}^{q}(-1)^{1+a} g_{1a}\det(G^0_{-a1}) \\ \frac{1}{\sqrt{q}}\sum_{a=1}^{q}(-1)^{2+a} g_{2a}\det(G^0_{-a2}) \\ \vdots \\ \frac{1}{\sqrt{q}}\sum_{a=1}^{q}(-1)^{q+a} g_{qa}\det(G^0_{-aq}) \end{pmatrix} = \frac{1}{\det(G^0)} \begin{pmatrix} \frac{\det(G^0)}{\sqrt{q}} \\ \frac{\det(G^0)}{\sqrt{q}} \\ \vdots \\ \frac{\det(G^0)}{\sqrt{q}} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{q}} \\ \frac{1}{\sqrt{q}} \\ \vdots \\ \frac{1}{\sqrt{q}} \end{pmatrix}$$

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

From this we can conclude that $(2\eta'\eta + 2\eta'(G^0 \odot (G^0)^{-1})\eta)^{1/2} = 2$ and $(2\eta'\eta - 2\eta'(G^0 \odot (G^0)^{-1})\eta)^{1/2} = 0$, which shows that the first term in (8) has the desired form. The second term of (8) equals the second term of the limiting distribution of Qu (2011) anyway, so it remains to consider the third term.

We first show that $\left(2\eta' F(r)\Omega^{-1}(G^0 \odot (G^0)^{-1} + I_q)\Omega^{-1\prime} F(r)'\eta\right)^{1/2} = 2\int_0^r (1+\log s)^2 ds$. To see this note that

$$\begin{aligned} F(r)'\eta &= 2\int_0^r (1+\log s)^2 \left(G^0 \odot (G^0)^{-1} + I_q\right)' \eta ds \\ &= 4\eta \int_0^r (1+\log s)^2 ds \end{aligned}$$

as $\left(G^0 \odot (G^0)^{-1} + I_q\right)'\eta = 2\eta$ by the same arguments as before. By denoting with η^{-1} the pseudo inverse defined by the equality $A\eta^{-1}\eta = A$ for every matrix A, we have $\eta^{-1} = \eta'$.

Consequently, $\Omega^{-1'}\eta = \left(\eta^{-1}\Omega'\right)^{-1} = (\Omega\eta)'^{-1}$. Now, $\Omega\eta = 4\eta$, since again $(G^0 \odot (G^0)^{-1})\eta = \eta$, so that $\Omega^{-1'}\eta = 1/4\eta$.
Applying the same arguments to the term $\eta' F(r)\Omega^{-1}$ on the left side gives us altogether

$$
\begin{aligned}
2\eta' F(r)\Omega^{-1}(G^0 \odot (G^0)^{-1} + I_q)\Omega^{-1'} F(r)'\eta &= 2\left(\int_0^r (1+\log s)^2 ds\right)^2 \eta'(G^0 \odot (G^0)^{-1} + I_q)\eta \\
&= 4\left(\int_0^r (1+\log s)^2 ds\right)^2.
\end{aligned}
$$

Now applying the same arguments we can furthermore conclude that $\eta' F(r)\Omega^{-1}(G^0 \odot (G^0)^{-1} - I_q)\Omega^{-1'} F(r)'\eta = 0$, which proves the lemma. □

Proof of Theorem 3:

The test statistic is given by

$$
\begin{aligned}
\widetilde{MLWS}\left(\hat{B}(p_G^0)\right) = \frac{1}{2}\sup_{r\in[\varepsilon,1]} \Bigg\| & \frac{2}{\sqrt{\sum_{j=1}^m \nu_j^2}} \sum_{a=1}^q \eta_a \sum_{j=1}^{[mr]} \nu_j \left({}_a\tilde{G}^{-1}(\hat{d},\hat{B}) Re\left[\Lambda_j(\hat{d})^{-1}\tilde{I}(\lambda_j,\hat{B})\Lambda_j^*(\hat{d})^{-1}\right]_a - 1\right) \\
&+ \frac{1}{\sqrt{\sum_{j=1}^m \nu_j^2}} \sum_{a=1}^q \eta_a \left({}_a\tilde{G}^{-1}(\hat{d},\hat{B})\right) \sum_{j=1}^{[mr]} \frac{(\lambda_j - \pi)}{2} Im\left[\Lambda_j(\hat{d})^{-1}\tilde{I}(\lambda_j,\hat{B})\Lambda_j^*(\hat{d})^{-1}\right]_a \Bigg\|.
\end{aligned}
$$

Now, define $C(\lambda_j,\hat{d},\hat{B}) = \Lambda_j(\hat{d})^{-1}\tilde{I}(\lambda_j,\hat{B})\Lambda_j^*(\hat{d})^{-1} = \Lambda_j(\hat{d})^{-1}\hat{B}I(\lambda_j)\hat{B}'\Lambda_j^*(\hat{d})^{-1}$ and from $\hat{B} - B^0 = O_P(m^{-1/2}\Delta_m^{-1})$, where $\Delta_m = diag(\lambda_m^{-b_1}, \ldots, \lambda_m^{-b_{p_G}}, 0, \ldots, 0)\, S_B$ and S_B is a selection matrix that specifies the position of the free parameters in $\hat{B}$

$$
\begin{aligned}
C(\lambda_j,\hat{d},\hat{B}) &= \Lambda_j(\hat{d})^{-1}B^0 I(\lambda_j)B^{0'}\Lambda_j^*(\hat{d})^{-1} \\
&+ \Lambda_j(\hat{d})^{-1}B^0 I(\lambda_j) O_P\left(m^{-1/2}\Delta_m^{-1}\right)'\Lambda_j^*(\hat{d})^{-1} \\
&+ \Lambda_j(\hat{d})^{-1} O_P\left(m^{-1/2}\Delta_m^{-1}\right) I(\lambda_j)B^{0'}\Lambda_j^*(\hat{d})^{-1} \\
&+ \Lambda_j(\hat{d})^{-1} O_P\left(m^{-1/2}\Delta_m^{-1}\right) I(\lambda_j) O_P\left(m^{-1/2}\Delta_m^{-1}\right)'\Lambda_j^*(\hat{d})^{-1}.
\end{aligned}
$$

Therefore, $C(\lambda_j,\hat{d},\hat{B})_{ab} = C(\lambda_j,\hat{d},B^0)_{ab} + O_P(\lambda_j^{\hat{d}_a - d_a^0 + \hat{d}_b - d_b^0}) O_P\left(m^{-1/2}\Delta_{m,ab}^{-1}\right)$ so that $plim\, C(\lambda_j,\hat{d},\hat{B})_{ab} = C(\lambda_j,\hat{d},B^0)_{ab} + O_P\left(m^{-1/2}\Delta_{m,ab}^{-1}\right)$, due to the consistency of $\hat{d}$. Now,

$$
\begin{aligned}
\tilde{G}(\hat{d},\hat{B}) = m^{-1}\sum_{j=1}^m Re\left[C(\lambda_j,\hat{d},\hat{B})\right] &\xrightarrow{P} m^{-1}\sum_{j=1}^m Re\left[C(\lambda_j,\hat{d},B^0) + O_P(m^{-1/2}\Delta_m^{-1})\right] \\
&= \tilde{G}(\hat{d},B^0) + O_P(m^{-1/2}\Delta_m^{-1}).
\end{aligned}
$$

Plugging these results into the $\widetilde{MLWS}(B)$ statistic and straightforward algebra give

$$\widetilde{MLWS}(\hat{B}(p_G^0)) \Rightarrow \widetilde{MLWS}(B^0) + \frac{\sum_{j=1}^{[mr]} O_P(m^{-1/2}\Delta_m^{-1})}{\sqrt{\sum_{j=1}^{m} v_j^2}},$$

and since $\sum_{j=1}^{m} v_j^2 \to m$, we have $\widetilde{MLWS}(\hat{B}(p_G^0)) \Rightarrow \widetilde{MLWS}(B^0) + O_P(\Delta_m^{-1})$, for $m = \lfloor T^\delta \rfloor$. The proof for the limit distribution of $\widetilde{MLWS}(B^0)$ proceeds as that of Theorem 1, but with $\hat{G}(d)$ and $I(\lambda_j)$ replaced by $\tilde{G}(d, B^0)$ and $\tilde{I}(\lambda_j, B^0)$. □

Proof of Theorem 4:

Let $F_{\widetilde{MLWS}(\hat{B}(\hat{p}_G))}(x)$ denote the finite sample distribution function of $\widetilde{MLWS}(\hat{B}(\hat{p}_G))$ for $x \in \Re \cup \{-\infty, \infty\}$. To prove the theorem, we show uniform convergence of the distribution of $\widetilde{MLWS}(\hat{B}(\hat{P}_G))$ to its counterpart for a known cointegration rank that is covered in Theorem 3. This is:

$$\sup_T \| F_{\widetilde{MLWS}(\hat{B}(\hat{p}_G))}(x) - F_{\widetilde{MLWS}(\hat{B}(p_G^0))}(x) \| = o_P(1).$$

Now, the estimator $\hat{p}_G$ of the cointegration rank can either estimate the correct, or a false cointegrating rank. By the law of total probability we can therefore decompose the distribution into the sum of the conditional distribution functions multiplied by the probabilities for these two possible events.

$$\begin{aligned}
&\| F_{\widetilde{MLWS}(\hat{B}(\hat{p}_G))}(x) - F_{\widetilde{MLWS}(\hat{B}(p_G^0))}(x) \| \\
=& \| F_{\widetilde{MLWS}(\hat{B}(\hat{p}_G))}(x|\hat{p}_G = p^0) P(\hat{p}_G = p^0) + F_{\widetilde{MLWS}(\hat{B}(\hat{p}_G))}(x|\hat{p}_G \neq p^0) P(\hat{p}_G \neq p^0) - F_{\widetilde{MLWS}(\hat{B}(p_G^0))}(x) \| \\
\leq& \| F_{\widetilde{MLWS}(\hat{B}(\hat{p}_G))}(x|\hat{p}_G = p^0) P(\hat{p}_G = p^0) - F_{\widetilde{MLWS}(\hat{B}(p_G^0))}(x) \| \\
&+ \| F_{\widetilde{MLWS}(\hat{B}(\hat{p}_G))}(x|\hat{p}_G \neq p^0) P(\hat{p}_G \neq p^0) \| \\
\leq& \| F_{\widetilde{MLWS}(\hat{B}(\hat{p}_G))}(x|\hat{p}_G = p^0) P(\hat{p}_G = p^0) - F_{\widetilde{MLWS}(\hat{B}(p_G^0))} \| + \| P(\hat{p}_G \neq p^0) \| = \| \Delta F \|,
\end{aligned}$$

where the first inequality is due to the triangle inequality and the second holds since $0 \leq F_{\widetilde{MLWS}(\hat{B}(\hat{p}_G))}(x|\hat{p}_G < p^0)$, $F_{\widetilde{MLWS}(\hat{B}(\hat{p}_G))}(x|\hat{p}_G > p^0) \leq 1$. Using $\lim_{T\to\infty} P(\hat{p}_G = p^0) = 1$, which implies $\| P(\hat{p}_G \neq p^0) \| = o_P(1)$, we have

$$\| \Delta F \| = \| (F_{\widetilde{MLWS}(\hat{B}(\hat{p}_G))}(x|\hat{p}_G = p^0) - F_{\widetilde{MLWS}(\hat{B}(p_G^0))}(x) \| + o_P(1) = o_P(1),$$

since $F_{\widetilde{MLWS}(\hat{B}(\hat{p}_G))}(x|\hat{p}_G = p^0) = F_{\widetilde{MLWS}(\hat{B}(p_G^0))}(x)$, by definition. Due to the continuity of the *sup*, the result follows from the continuous mapping theorem. The distribution $F_{\widetilde{MLWS}(\hat{B}(p_G^0))}(x)$ of $\widetilde{MLWS}(\hat{B}(p_G^0))$ is given in Theorem 3. □

Proof of Theorem 5:

We now prove the consistency of the $\widetilde{MLWS}(\hat{B}(\hat{p}_G))$ statistic calculated based on an estimate $\hat{p}_G$ of the cointegrating rank. For ease of notation we will write $\hat{B}$ instead of $\hat{B}(\hat{p}_G)$. There are two possible scenarios

1. $\hat{p}_G \geq p_G^0$, thus the cointegration rank has been estimated correctly or overestimated.
2. $\hat{p}_G < p_G^0$, thus the cointegration rank has been underestimated

In the second case $\lim_{T\to\infty} \widetilde{G}(\hat{d}, \hat{B})$ does not have full rank and is singular. Therefore, $\lim_{T\to\infty} \widetilde{G}(\hat{d}, \hat{B})^{-1}$ does not exist and divergence of the test statistic is obvious.
In the first case $\lim_{T\to\infty} \widetilde{G}(\hat{d}, \hat{B})$ has full rank and the inverse exists. Our procedure corresponds to an application of the MLWS test to the modified random level shift process $\hat{B}X_t = \hat{B}\mu_t + \hat{B}\kappa_t$. For the mean part we have $\hat{B}\mu_t$, which is a linear combination of random level shift processes. The periodogram of this low frequency contamination is therefore $BI_\mu(\lambda_j)B'$, which is of the same order $(O_P(\lambda^{-2}T^{-1}))$ as that of μ_t itself.
Consequently, the proof proceeds in the same way as that of Theorem 2 (in the supplementary appendix), but with $I(\lambda_j)$, $G(d)$, and d replaced by $\widetilde{I}(\lambda_j, \hat{B})$, $\widetilde{G}(d, \hat{B})$ and $\tilde{d}$.
□

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. and Feng, Y. (2002). SEMIFAR models? a semiparametric approach to modelling trends, long-range dependence and nonstationarity. *Computational Statistics & Data Analysis*, 40(2):393–419.

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

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.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2):174–196.

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.

Frederiksen, P., Nielsen, F. S., and Nielsen, M. Ø. (2012). Local polynomial whittle estimation of perturbed fractional processes. *Journal of Econometrics*, 167(2):426–447.

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.

Hou, J. and Perron, P. (2014). Modified local Whittle estimator for long memory processes in the presence of low frequency (and other) contaminations. *Journal of Econometrics*, 182(2):309–328.

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.

Kruse, R. (2015). A modified test against spurious long memory. *Economics Letters*, 135:34–38.

Lavielle, M. and Moulines, E. (2000). Least-squares estimation of an unknown number of shifts in a time series. *Journal of Time Series Analysis*, 21(1):33–59.

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. (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.

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

McCloskey, A. and Perron, P. (2013). Memory parameter estimation in the presence of level shifts and deterministic trends. *Econometric Theory*, 29(06):1196–1237.

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).

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. (2008a). Diagnostic testing for cointegration. *Journal of Econometrics*, 143(1):206–225.

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

Robinson, P. M. (2008b). Multiple local Whittle estimation in stationary systems. *The Annals of Statistics*, 36(5):2508–2530.

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

Sarkar, S. K. (1998). Some probability inequalities for ordered mtp2 random variables: a proof of the simes conjecture. *Annals of Statistics*, pages 494–504.

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.

Shimotsu, K. (2012). Exact local whittle estimation of fractionally cointegrated systems. *Journal of Econometrics*, 169(2):266–278.

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

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.