Supplementary Appendix

This supplementary appendix contains additional proofs, simulation studies, and empirical results. The first part gives further details on the pre-whitening procedure. In Section B we provide the proofs for Lemma 1, Theorem 2 and Lemma 3. Section C presents several Monte Carlo simulations that cover the effect of pre-whitening, non-stationary long memory, breaks in the variance-covariance matrix, the power against other forms of low frequency contaminations and the effect of conditional heteroscedasticity and perturbations on the test itself and the pre-whitening procedure. Finally, Section D presents additional material on our empirical application - such as summary statistics and robustness checks.

A Pre-Whitening

Although the MLWS statistic in (7) is asymptotically independent of short memory dynamics, we need to apply a pre-whitening procedure to avoid negative effects on the size and power properties of the test in finite samples if short memory dynamics are present. Similar to Qu (2011), we do so by approximating the short memory dynamics of u_t in (2) with a low order VARMA process that is given by $u_t = \mathcal{A}(L)^{-1}\mathcal{M}(L)v_t$, where $\mathcal{A}(L) = I_q - \mathcal{A}_1 L - ... - \mathcal{A}_{p_A} L^{p_A}$ and $\mathcal{M}(L) = I_q - \mathcal{M}_1 L - ... - \mathcal{M}_{q_M} L^{q_M}$ are matrix lag polynomials and v_t is a $(q \times 1)$ white noise with zero mean and variance-covariance matrix Σ_v.

Since $\mathcal{A}(L)$ and $D(d_1,\ldots,d_q)$ are not commutative in the multivariate case, there are two ways to generalize the univariate ARFIMA process to a vector process. This has been pointed out by Lobato (1997). For the specification used here, Sela and Hurvich (2009) coined the acronym FIVARMA because the process $X_t = D(d_1,\ldots,d_q)^{-1}\mathcal{A}(L)^{-1}\mathcal{M}(L)v_t$ can be interpreted as a fractionally integrated VARMA process. The process $X_t = \mathcal{A}(L)^{-1}D(d_1,\ldots,d_q)^{-1}\mathcal{M}(L)v_t$ is a vector autoregression of a fractionally integrated vector moving average process and is referred to as VARFIMA. The FIVARMA model has recently been applied in Chiriac and Voev (2011) and has first been studied by Sowell (1989). As Lütkepohl (2007) points out, the parameters of the unrestricted VARMA model are not identified. Thus, we follow Chiriac and Voev (2011) and estimate the model in its final equation form, where $\mathcal{A}(L)$ is a scalar operator. The estimation is carried out using the conditional sum of squares estimator (cf. Beran (1995), Hualde et al. (2011) and Nielsen (2015)) which is based on an approximation of the AR(∞) representation of the FIVARMA process.[1]

[1]For the sample sizes considered here, this approach turns out to be faster than the method of Sela

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

An issue is the computational cost of estimating the FIVARMA model in larger samples and for larger q. In these situations one can resort to a univariate pre-whitening as proposed in Qu (2011). A Monte Carlo study on the performance of the test with pre-whitening is conducted in Section C below. It is found that the test is correctly sized and the power loss incurred is within a reasonable range. The same holds true after univariate pre-whitening, but compared to the multivariate pre-whitening there is an additional loss of power in larger samples.

B Proofs

Proof of Lemma 1:

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

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

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

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

and Hurvich (2009). For larger samples the computing time can be further reduced by conducting the ML estimation on subsamples and using the average of these subsample estimators, as suggested by Beran and Terrin (1994).

the matrix M, and M_a denoting the a-th column of the matrix M

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

From equations (20) and (27) in Shimotsu (2007). Furthermore, from his equation (22) and the fact that $m^{-1/2}\sum_{j=1}^{m}\nu_j Re\left[\Lambda_j(d)^{-1}I(\lambda_j)\Lambda_j^*(d)^{-1}\right]$ and $m^{-1/2}\sum_{j=1}^{m}Im\left[\Lambda_j(d)^{-1}I(\lambda_j)\Lambda_j^*(d)^{-1}\right]$ are $O_P(1)$

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

which proves our lemma. □

Proof of Theorem 2:

To prove the consistency of our test statistic we closely follow the arguments in the consistency proof of Qu (2011) with the obvious adaptions for the multivariate case. For ease of notation we suppress the dependence of $\hat{G}(\hat{d})=\frac{1}{m}\sum_{j=1}^{m}Re\left[\Lambda_j(\hat{d})^{-1}I(\lambda_j)\Lambda_j^*(\hat{d})^{-1}\right]$ on $\hat{d}$ and simply write G, instead.

For our test statistic we write

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

Let us consider the term I first. Note that ν_j is monotonically increasing in j with $\nu_1 < 0$ and $\nu_m > 0$ and define $j^* = \min\{j : \nu_j \geq 0\}$. Here, the proof uses the fact that divergence of the quantity in the MLWS statistic for at least one r implies convergence of the supremum over all r, so that it is sufficient to focus on the case $r = 1$.

From Qu's (2011) Lemma B.1 it follows that $j^* = Km$ for some constant K. Define now

$$A^I = \left\| \sum_{a=1}^{q} \eta_a \left(\sum_{j=1}^{m} \nu_j^2 \right)^{-1/2} \sum_{j=j^*}^{m} \nu_j \left({}_aG^{-1} \left[Re\left[\Lambda_j(\hat{d})^{-1} I(\lambda_j) \Lambda_j^*(\hat{d})^{-1} \right] \right]_a - 1 \right) \right\|$$

and

$$B^I = \left\| \sum_{a=1}^{q} \eta_a \left(\sum_{j=1}^{m} \nu_j^2 \right)^{-1/2} \sum_{j=1}^{j^*-1} \nu_j \left({}_aG^{-1} \left[Re\left[\Lambda_j(\hat{d})^{-1} I(\lambda_j) \Lambda_j^*(\hat{d})^{-1} \right] \right]_a - 1 \right) \right\|.$$

Applying the reverse triangle inequality to I gives $I \geq \max(A^I - B^I, B^I)$. Now, if $A^I \geq 2B^I$ than we have

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

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

$$A^I = \left\| \sum_{a=1}^{q} \eta_a \left(\sum_{j=1}^{m} \nu_j^2 \right)^{-1/2} \sum_{j=j^*}^{m} \nu_j \left({}_aG^{-1} \left[Re\left[\Lambda_j(d)^{-1} A(\lambda_j) I_{\varepsilon j} A^*(\lambda_j) \Lambda_j^*(d)^{-1} \right] \right]_a - 1 \right) \right\|.$$

Applying the reverse triangle inequality to A gives

$$A^I \geq \sum_{a=1}^{q} \eta_a \left(\sum_{j=1}^{m} \nu_j^2 \right)^{-1/2} \sum_{j=j^*}^{m} \nu_j - \sum_{a=1}^{q} \eta_a \sum_{j=j^*}^{m} \nu_j \left({}_aG^{-1} \left[Re\left[\Lambda_j(d)^{-1} A(\lambda_j) I_{\varepsilon j} A^*(\lambda_j) \Lambda_j^*(d)^{-1} \right] \right]_a \right).$$

For the first term of this inequality we have for each component

$$\left(\sum_{j=1}^{m} \nu_j^2 \right)^{-1/2} \sum_{j=j^*}^{m} \nu_j = m^{1/2} \int_K^1 (1 + \log s) ds + o\left(m^{1/2} \right)$$

which is strictly positive and of exact order $m^{1/2}$.

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

$$\nu_j\left({}_aG^{-1}\left[Re\left[\Lambda_j(d)^{-1}A(\lambda_j)I_{\varepsilon j}A^*(\lambda_j)\Lambda_j^*(d)^{-1}\right]\right]_a\right) = O(\log(m))O_p\left(\lambda_j^{\hat{d}_a-d_a^0+min_b(\hat{d}_b-d_b^0)}\right) = O_P(\log(m))$$

for all $a = 1,...,q$ and all $b = 1,...,q$. This holds because $\nu_j = O(\log(m))$ for $j^* \leq j \leq m$, $\hat{G}(d)$ is positive definite, $P(\hat{d}_a - d_a^0 \geq 0) \to 1$ for all a and $\lambda_j = o(1)$. Therefore, the second term is of lower order than $m^{1/2}$ and is dominated asymptotically by the first term. Thus, $A^I \xrightarrow{P} \infty$ if $T \to \infty$.

The treatment of term II is similar to that of term I. Write again

$$A^{II} = \left\|\sum_{a=1}^{q}\eta_a\left({}_aG^{-1}\right)\left(\sum_{j=1}^{m}\nu_j^2\right)^{-1/2}\sum_{j=j^*}^{[mr]}\frac{\lambda_j-\pi}{2}Im\left[\Lambda_j(\hat{d})^{-1}I(\lambda_j)\Lambda_j^*(\hat{d})^{-1}\right]_a\right\| \tag{16}$$

and

$$B^{II} = \left\|\sum_{a=1}^{q}\eta_a\left({}_aG^{-1}\right)\left(\sum_{j=1}^{m}\nu_j^2\right)^{-1/2}\sum_{j=1}^{j^*-1}\frac{\lambda_j-\pi}{2}Im\left[\Lambda_j(\hat{d})^{-1}I(\lambda_j)\Lambda_j^*(\hat{d})^{-1}\right]_a\right\|. \tag{17}$$

As before we obtain from the reverse triangle inequality that $II \geq A^{II}/2$. Now, we write A^{II} in the form

$$A^{II} = \left\|\left(\sum_{j=1}^{m}\nu_j^2\right)^{-1/2}\frac{\pi}{2}\sum_{j=j^*}^{m}Im\left[\Lambda_j(d)^{-1}A(\lambda_j)I_{\varepsilon j}A^*(\lambda_j)\Lambda_j^*(d)^{-1}\right]\right\| \tag{18}$$

By using exactly the same arguments as before for the second part of A^I the term II is of lower order than $m^{1/2}$, and therefore dominated asymptotically by term I. This proves the theorem. □

Proof of Lemma 3:

The proof of the lemma is similar to the proof of Corollary 2 in Qu (2011) after replacing his univariate coefficients of the short-memory model with our multivariate coefficients and using the respective norms. It is therefore omitted here. □

C Additional Monte Carlo Results

In the following we provide additional Monte Carlo simulations on the efficiency of the pre-whitening procedure, the performance of the test for non-stationary long memory series, its power against non-causal alternatives, its robustness under heteroscedasticity and breaks in the variance-covariance matrix, its power against other DGPs under

		Pre-whitened: $\tilde{X}_t$								Unfiltered: X_t			
DGP		1		2		3		4		1		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**
250	**0.60**	0.006	0.007	0.006	0.009	0.004	0.007	0.010	0.008	0.007	0.009	0.005	0.010
	0.65	0.012	0.013	0.006	0.009	0.003	0.007	0.007	0.009	0.012	0.017	0.011	0.024
	0.70	0.010	0.013	0.005	0.006	0.002	0.004	0.013	0.013	0.018	0.017	0.037	0.072
	0.75	0.009	0.010	0.001	0.002	0.002	0.003	0.011	0.012	0.018	0.020	0.127	0.209
500	**0.60**	0.012	0.015	0.013	0.013	0.009	0.008	0.013	0.012	0.012	0.015	0.007	0.013
	0.65	0.013	0.015	0.015	0.016	0.009	0.011	0.011	0.012	0.014	0.025	0.017	0.038
	0.70	0.013	0.017	0.013	0.015	0.004	0.004	0.016	0.015	0.021	0.021	0.061	0.105
	0.75	0.013	0.013	0.006	0.006	0.002	0.003	0.010	0.013	0.023	0.030	0.238	0.340
1000	**0.60**	0.014	0.021	0.016	0.024	0.016	0.018	0.017	0.018	0.017	0.018	0.011	0.025
	0.65	0.021	0.019	0.021	0.024	0.015	0.017	0.021	0.023	0.023	0.029	0.023	0.033
	0.70	0.020	0.019	0.025	0.019	0.011	0.007	0.016	0.018	0.031	0.032	0.075	0.112
	0.75	0.018	0.018	0.014	0.011	0.003	0.004	0.019	0.016	0.039	0.038	0.351	0.452
2000	**0.60**	0.016	0.027	0.017	0.024	0.016	0.028	0.017	0.026	0.020	0.027	0.017	0.027
	0.65	0.019	0.030	0.023	0.035	0.022	0.029	0.023	0.034	0.023	0.031	0.023	0.039
	0.70	0.024	0.022	0.029	0.024	0.018	0.017	0.026	0.027	0.030	0.036	0.075	0.106
	0.75	0.028	0.022	0.025	0.020	0.009	0.006	0.020	0.024	0.043	0.040	0.433	0.535

Table 9: Size of the MLWS test for FIVARMA(1,d,1): $(1-\phi_1 L)D(d_1,d_2)X_t = (I_q - \mathcal{M}_1 L)v_t$ with and without pre-whitening. Parameter values of the respective DGPs are given in the text. The bandwidth m is determined by $m = \lfloor T^{\delta} \rfloor$.

the alternative, univariate pre-whitening, robustness of the test and the pre-whitening procedure in perturbed series and pre-whitening in fractionally cointegrated series.

Short Memory Dynamics

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

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

$$\textbf{DGP1:} \quad D(0.2, 0.3)X_t = v_t$$

$$\textbf{DGP2:} \quad (1 - 0.4L)D(0.2, 0.3)X_t = v_t$$

$$\textbf{DGP3:} \quad (1 - 0.6L)D(0.2, 0.3)X_t = v_t$$

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

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

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

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

In the multivariate random level shift process we use $\rho_\pi = \rho_e = 0.9$ and $\sigma_e = 2$. To investigate the costs and benefits of the pre-whitening procedure, the simulations for DGP1 and DGP2 are conducted with and without pre-whitening.

The results for the size simulations are given in Table 9. DGP1 is the baseline case. By comparing the results for the tests applied to the pre-whitened series $\tilde{X}_t$ with the results of the unfiltered series X_t, we can observe that the empirical size becomes a bit more conservative. However, for an increasing T the size approaches its nominal level. By considering DGP2, we see that there are considerable over-rejections if moderate autoregressive dynamics are present and no pre-whitening is applied, whereas with pre-whitening these distortions are successfully removed.

The results for DGP3 and DGP4 show that the pre-whitening procedure works well in controlling the size for different forms of short memory dynamics. As before, the size is generally better if $\varepsilon = 0.05$ is used. With regard to the bandwidth selection, the best size is observed most often for $\delta = 0.65$ or $\delta = 0.70$. It is no longer strictly increasing in m.

Table 10 considers the power results. For the baseline case there is a considerable power reduction caused by the additional flexibility introduced through the pre-whitening procedure. But when comparing the results for DGP1 and DGP2 without pre-whitening, we

		Pre-whitened: $\tilde{X}_t$								Unfiltered: X_t			
DGP		1		2		3		4		1		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**
250	**0.60**	0.022	0.026	0.032	0.032	0.012	0.014	0.023	0.030	0.037	0.045	0.014	0.020
	0.65	0.052	0.066	0.050	0.062	0.017	0.020	0.053	0.077	0.099	0.102	0.024	0.028
	0.70	0.099	0.088	0.068	0.062	0.016	0.016	0.107	0.110	0.183	0.199	0.020	0.021
	0.75	0.154	0.159	0.066	0.059	0.020	0.016	0.161	0.168	0.315	0.322	0.025	0.048
500	**0.60**	0.083	0.076	0.092	0.081	0.044	0.041	0.080	0.081	0.130	0.121	0.075	0.063
	0.65	0.146	0.158	0.139	0.154	0.049	0.061	0.155	0.177	0.243	0.287	0.077	0.102
	0.70	0.254	0.266	0.194	0.194	0.054	0.052	0.284	0.300	0.462	0.508	0.096	0.087
	0.75	0.372	0.361	0.203	0.187	0.049	0.037	0.399	0.399	0.662	0.677	0.067	0.067
1000	**0.60**	0.226	0.231	0.241	0.234	0.134	0.128	0.232	0.245	0.308	0.339	0.232	0.223
	0.65	0.398	0.398	0.369	0.354	0.186	0.158	0.396	0.413	0.564	0.588	0.324	0.307
	0.70	0.543	0.518	0.463	0.393	0.177	0.099	0.566	0.569	0.766	0.772	0.327	0.273
	0.75	0.634	0.591	0.475	0.381	0.121	0.069	0.685	0.651	0.889	0.894	0.255	0.181
2000	**0.60**	0.502	0.514	0.524	0.526	0.345	0.345	0.511	0.534	0.611	0.639	0.509	0.518
	0.65	0.709	0.710	0.685	0.668	0.423	0.379	0.721	0.730	0.841	0.855	0.651	0.632
	0.70	0.777	0.749	0.749	0.698	0.404	0.285	0.829	0.807	0.931	0.930	0.679	0.604
	0.75	0.779	0.703	0.765	0.651	0.316	0.146	0.845	0.793	0.961	0.958	0.569	0.441

Table 10: Power of the MLWS test for $Y_t = \mu_t + X_t$, where $(1-\phi_1 L)D(d_1,d_2)X_t = (I_q - \mathcal{M}_1 L)v_t$ with and without pre-whitening. Parameter values of the respective DGPs are given in the text. The bandwidth m is determined by $m = \lfloor T^{\delta} \rfloor$.

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

Non-Stationary DGPs

Even though it is assumed that $d_a^0 \in (-1/2, 1/2)\ \forall\ a = 1, \ldots, q$, in practice one will encounter situations where the memory parameters are in the non-stationary but mean reverting range of $0.5 < d < 1$. To analyze the reliability of the MLWS statistic in these situations, we conduct additional Monte Carlo experiments. Hereafter, we only provide a short description of the simulation setups and the results in the text and the full DGPs are specified in the captions of the respective tables.

In Table 11 a bivariate long memory process is contaminated with a stationary random level shift process that is scaled so that its standard deviation is a multiple ξ of that of the sample standard deviation of the long memory component. The range of d covers stationary and non-stationary values. For $\xi = 0$, there is no contamination, so that we obtain the size. One can see that the size remains stable for non-stationary DGPs. Particularly remarkable is the fact that the size remains intact for $d = 0.8$, since the test is based on the local Whittle estimator and Velasco (1999) showed that the limit distribution of the univariate local Whittle estimator changes for $d > 0.75$.

For stationary long memory components it can be observed that the power is decreasing with increasing d. In the non-stationary range however, the power is increasing with increasing d. These results show that the MLWS test remains valid in situations with non-stationary but mean-reverting long memory.

Non-Causal Alternatives

To analyze the power of the MLWS test against non-causal alternatives, we build on the work of Kechagias and Pipiras (2015) who consider non-causal multivariate fractionally integrated processes. Table 12 shows the results. The scaling factor ξ determines to which degree the behaviour of the process is determined by the non-causal part. It can be seen that the test is correctly sized, but only develops power very slowly.

Breaks in the Variance-Covariance Matrix

Table 13 shows the effect of breaks in the variance-covariance matrix of a bivariate fractionally integrated process. Initially the innovations have unit variance and a correlation of $\rho_1 = 0.45$. After 100ξ percent of the sample, the variance and correlation switch. It can be seen that the size of the test is robust to unconditional breaks in the variance as well as breaks in the correlation between the innovations. Only for increases of the correlation after 20 percent of the sample, the size increases to about 7 percent. In all other combinations it remains close to or below the nominal level of 5 percent.

Robustness under Heteroscedasticity

Table 14 shows the effect of conditional heteroscedasticity on the size and power of the test. We find that the size increases slightly, but only for very persistent GARCH effects. The power remains largely unchanged. The MLWS test is therefore robust to conditional and unconditional heteroscedasticity.

		d	0		0.2		0.4		0.6		0.8	
ξ	T	δ/ε	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**
0	250	**0.60**	0.006	0.009	0.007	0.012	0.008	0.011	0.008	0.008	0.004	0.006
		0.65	0.005	0.013	0.008	0.017	0.008	0.015	0.010	0.018	0.007	0.013
		0.70	0.012	0.014	0.012	0.022	0.012	0.016	0.015	0.014	0.012	0.018
		0.75	0.012	0.017	0.016	0.025	0.014	0.021	0.017	0.023	0.020	0.023
	1000	**0.60**	0.012	0.018	0.013	0.020	0.014	0.024	0.014	0.023	0.012	0.019
		0.65	0.019	0.025	0.018	0.024	0.018	0.027	0.022	0.024	0.019	0.028
		0.70	0.021	0.032	0.023	0.030	0.023	0.032	0.025	0.026	0.021	0.032
		0.75	0.024	0.029	0.029	0.035	0.031	0.037	0.035	0.044	0.033	0.041
1	250	**0.60**	0.193	0.195	0.082	0.087	0.033	0.034	0.031	0.037	0.060	0.065
		0.65	0.340	0.376	0.146	0.168	0.065	0.078	0.078	0.084	0.154	0.190
		0.70	0.486	0.472	0.232	0.208	0.112	0.113	0.139	0.151	0.316	0.349
		0.75	0.597	0.555	0.314	0.284	0.177	0.179	0.271	0.274	0.550	0.596
	1000	**0.60**	0.817	0.823	0.519	0.485	0.148	0.155	0.107	0.112	0.249	0.278
		0.65	0.930	0.914	0.692	0.655	0.290	0.272	0.241	0.280	0.569	0.599
		0.70	0.958	0.950	0.811	0.761	0.426	0.401	0.485	0.494	0.861	0.871
		0.75	0.973	0.969	0.887	0.852	0.619	0.605	0.769	0.779	0.978	0.987
2	250	**0.60**	0.214	0.228	0.146	0.141	0.086	0.090	0.078	0.087	0.119	0.134
		0.65	0.403	0.452	0.302	0.308	0.187	0.226	0.187	0.217	0.266	0.323
		0.70	0.589	0.593	0.445	0.454	0.348	0.345	0.352	0.365	0.485	0.520
		0.75	0.727	0.695	0.587	0.565	0.508	0.496	0.568	0.571	0.727	0.734
	1000	**0.60**	0.839	0.845	0.707	0.726	0.488	0.478	0.391	0.405	0.546	0.582
		0.65	0.942	0.937	0.879	0.880	0.728	0.731	0.683	0.715	0.865	0.889
		0.70	0.974	0.966	0.951	0.932	0.899	0.879	0.908	0.909	0.983	0.989
		0.75	0.986	0.983	0.976	0.968	0.962	0.957	0.986	0.987	0.998	0.999

Table 11: Size and Power of the MLWS test against level shifts in long memory processes with increasing persistence. Here, $Y_t = X_t + \xi \hat{SD}(X_t)/\hat{SD}(\mu_t)\mu_t$, $D(d,d)X_t = v_t$, $v_t \sim N(0,((1,0.5),(0.5,1))')$ and μ_t is defined as in (3), with $\rho_e = \rho_\pi = 0.5$ and $\tilde{p} = 5$.

Other DGPS

Random level shifts are not the only data generating processes that cause spurious long memory. Similar effects are caused (among others) by linear or monotonous trends, non-monotonous trends or Markov switching models. Theoretically, it has been shown by McCloskey and Perron (2013) that the effect of these processes on the periodogram

	ζ	0		1		2	
T	δ/ϵ	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**
250	**0.60**	0.009	0.015	0.005	0.008	0.006	0.013
	0.65	0.014	0.018	0.008	0.012	0.011	0.024
	0.70	0.015	0.024	0.009	0.019	0.021	0.040
	0.75	0.018	0.027	0.011	0.020	0.030	0.058
500	**0.60**	0.012	0.016	0.010	0.017	0.011	0.024
	0.65	0.018	0.030	0.013	0.024	0.022	0.043
	0.70	0.022	0.033	0.014	0.032	0.049	0.076
	0.75	0.030	0.033	0.025	0.031	0.085	0.139
1000	**0.60**	0.016	0.029	0.014	0.021	0.020	0.038
	0.65	0.025	0.029	0.023	0.035	0.044	0.079
	0.70	0.029	0.038	0.027	0.048	0.103	0.136
	0.75	0.037	0.045	0.042	0.057	0.201	0.262
2000	**0.60**	0.025	0.032	0.019	0.040	0.029	0.060
	0.65	0.031	0.038	0.031	0.055	0.074	0.102
	0.70	0.031	0.044	0.051	0.074	0.174	0.222
	0.75	0.046	0.055	0.074	0.097	0.403	0.458

Table 12: Power of the MLWS test against non-causal alternatives of the form $X_t = D(0.2,0.3)^{-1}Q_{-}v_t + \tilde{D}(0.2,0.3)^{-1}\zeta Q_{-}v_t$, where $\tilde{D}(d_1,d_2) = diag((1-L^{-1})^{d_1},(1-L^{-1})^{d_2})$, $v_t \sim N(0,\Sigma_v)$, Σ_v is an identity matrix and $Q_{-} = ((1,0.7),(-0.5,1))$.

is approximately of the same order as that of random level shifts. Qu (2011) also demonstrates the power of his test against these alternatives. A similar analysis for the MLWS test is provided in Table 15. As can be seen, the test has good power against all of these alternatives.

Univariate Pre-Whitening

In our empirical application we opt to conduct the test after using a univariate pre-whitening. This is to avoid convergence issues of the numerical optimization involved in the maximum likelihood estimation of the FIVARMA model, that could arise due to the large number of free parameters. As one can see from Table 16, the size of the test is controlled using univariate pre-whitening as it is with multivariate pre-whitening. With regard to the power, in smaller samples with up to 1000 observations univariate pre-whitening even leads to improvements. For $T = 2000$, on the other hand, we observe a considerable power loss compared to the multivariate pre-whitening. This effect is particularly pronounced for DGPs 1 and 4, where we observe a non-monotonicity of the power in T. It is therefore advantageous to use the computationally more demanding multivariate pre-whitening in larger samples in terms of power. Nevertheless, if the computational costs become very large - as it is the case in our empirical example - the size is well controlled by the univariate pre-whitening procedure.

Pre-Whitening for Fractionally Cointegrated Systems

Another point that deserves further Monte Carlo evidence is the question whether the pre-whitening procedure still works for fractionally cointegrated systems, if the cointegrating matrix B is estimated first and the pre-whitening is then applied to the transformed system $\hat{B}X_t$. The results of this experiment are shown in Table 17. As one can see, the test maintains good size and power properties in this situation.

Perturbed Series

Since especially the log-absolute return series in the empirical application can be understood as the underlying volatility process perturbed by a measurement error, Table 18 evaluates the effect of perturbations on the size of the MLWS test. One can observe that the size is largely robust. The only exception is the situation when the bandwidth and the sample are large and the series are strongly correlated. In this case, the size can reach up to 10 percent, if also the variance of the perturbation is large compared to that of the long memory process. Apart from this special case, the test remains robust. The effect of perturbations on the performance of the pre-whitening procedure is explored in Table 19. It can be seen, that the size remains conservative for all DGPs. Interestingly, the size distortions that occur if no pre-whitening is applied are considerably reduced. One can therefore consider two possible ways to proceed. Either one relies on the positive effect of the pre-whitening procedure on the size of the test, or one omits the pre-whitening and chooses a smaller bandwidth since we observe a power loss after pre-whitening and an increase in power for the unfiltered series.

		σ_2^2			1						2			
		ρ_2	0		0.45		0.9		0		0.45		0.9	
ξ	**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**
0.2	250	**0.60**	0.006	0.011	0.007	0.007	0.012	0.015	0.004	0.008	0.006	0.010	0.008	0.010
		0.65	0.010	0.014	0.008	0.014	0.016	0.025	0.009	0.013	0.007	0.015	0.012	0.018
		0.70	0.014	0.018	0.012	0.015	0.020	0.035	0.011	0.021	0.015	0.019	0.018	0.023
		0.75	0.012	0.020	0.015	0.019	0.031	0.039	0.015	0.024	0.017	0.019	0.020	0.026
	1000	**0.60**	0.015	0.021	0.011	0.020	0.030	0.046	0.015	0.020	0.017	0.022	0.024	0.029
		0.65	0.021	0.030	0.017	0.027	0.039	0.057	0.021	0.024	0.019	0.031	0.032	0.041
		0.70	0.020	0.031	0.024	0.032	0.058	0.074	0.022	0.034	0.025	0.033	0.036	0.049
		0.75	0.031	0.040	0.031	0.037	0.064	0.077	0.032	0.039	0.033	0.043	0.045	0.051
0.5	250	**0.60**	0.005	0.011	0.006	0.008	0.010	0.016	0.008	0.011	0.009	0.011	0.011	0.015
		0.65	0.011	0.018	0.009	0.017	0.013	0.025	0.013	0.020	0.014	0.016	0.014	0.025
		0.70	0.016	0.018	0.012	0.018	0.019	0.026	0.016	0.024	0.014	0.022	0.017	0.028
		0.75	0.020	0.021	0.014	0.021	0.026	0.035	0.018	0.029	0.020	0.026	0.024	0.030
	1000	**0.60**	0.016	0.022	0.012	0.016	0.025	0.03	0.017	0.028	0.016	0.025	0.023	0.035
		0.65	0.023	0.024	0.022	0.022	0.031	0.044	0.028	0.035	0.024	0.030	0.031	0.039
		0.70	0.024	0.035	0.021	0.027	0.039	0.049	0.030	0.045	0.028	0.039	0.039	0.048
		0.75	0.033	0.043	0.025	0.034	0.046	0.053	0.040	0.050	0.033	0.049	0.044	0.055
0.8	250	**0.60**	0.006	0.01	0.007	0.007	0.008	0.011	0.008	0.01	0.005	0.011	0.005	0.012
		0.65	0.010	0.015	0.009	0.013	0.011	0.019	0.012	0.019	0.010	0.014	0.012	0.016
		0.70	0.014	0.017	0.016	0.018	0.016	0.023	0.017	0.024	0.012	0.020	0.015	0.023
		0.75	0.017	0.025	0.015	0.019	0.019	0.025	0.021	0.029	0.019	0.025	0.020	0.027
	1000	**0.60**	0.016	0.024	0.014	0.021	0.018	0.025	0.017	0.033	0.015	0.025	0.020	0.030
		0.65	0.019	0.029	0.017	0.026	0.025	0.034	0.029	0.040	0.020	0.036	0.023	0.034
		0.70	0.029	0.034	0.023	0.031	0.031	0.033	0.032	0.052	0.029	0.040	0.034	0.041
		0.75	0.034	0.038	0.027	0.037	0.032	0.043	0.046	0.060	0.038	0.043	0.035	0.050

Table 13: Size of the MLWS test in presence of breaks in the variance-covariance matrix. $D(0.2,0.3)X_t = v_t$, with $v_t \sim N(0,\Sigma_{v,t})$ and $\Sigma_{v,t} = ((1,0.45),(0.45,1))'\mathbb{I}_{(t/T<\xi)} + \sigma_2^2((1,\rho_2),(\rho_2,1))'(1-\mathbb{I}_{(t/T<\xi)})$.

		Size						Power					
	β_1	0.4		0.6		0.8		0.4		0.6		0.8	
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**	0.007	0.011	0.005	0.010	0.007	0.015	0.104	0.117	0.108	0.122	0.101	0.134
	0.65	0.011	0.015	0.010	0.015	0.015	0.026	0.247	0.313	0.251	0.296	0.257	0.295
	0.70	0.015	0.017	0.018	0.019	0.022	0.028	0.450	0.476	0.448	0.461	0.453	0.471
	0.75	0.017	0.022	0.020	0.027	0.035	0.039	0.639	0.638	0.634	0.643	0.629	0.654
1000	**0.60**	0.014	0.023	0.014	0.024	0.020	0.031	0.687	0.719	0.679	0.713	0.669	0.695
	0.65	0.020	0.026	0.023	0.027	0.033	0.046	0.901	0.909	0.898	0.911	0.890	0.902
	0.70	0.025	0.030	0.025	0.032	0.051	0.062	0.972	0.968	0.968	0.967	0.966	0.967
	0.75	0.027	0.036	0.031	0.038	0.069	0.087	0.990	0.987	0.988	0.988	0.991	0.985

Table 14: Size and power of the MLWS test for processes with GARCH effects. $D(0.2,0.3)X_t = v_t$, with $\Sigma_{v,t} = \left((\sigma^2_{1,t}, 0.8\sigma_{1,t}\sigma_{2,t}),(0.8\sigma_{1,t}\sigma_{2,t},\sigma^2_{2,t})\right)'$ and $\sigma^2_{a,t} = 1 + 0.15v^2_{a,t-1} + \beta_1\sigma^2_{a,t-1}$.

T	ζ	DGP δ/ε	mon 0.02	mon 0.05	lin 0.02	lin 0.05	sin 0.02	sin 0.05	ms 0.02	ms 0.05
250	0	**0.60**	0.004	0.010	0.006	0.010	0.006	0.008	0.008	0.008
		0.65	0.008	0.011	0.009	0.015	0.009	0.015	0.008	0.013
		0.70	0.009	0.017	0.012	0.017	0.013	0.017	0.014	0.020
		0.75	0.013	0.020	0.012	0.021	0.015	0.018	0.012	0.020
	1	**0.60**	0.068	0.070	0.125	0.123	0.973	0.987	0.103	0.106
		0.65	0.114	0.120	0.227	0.246	0.999	1.000	0.182	0.197
		0.70	0.167	0.131	0.342	0.279	1.000	1.000	0.296	0.286
		0.75	0.193	0.135	0.427	0.330	1.000	1.000	0.427	0.406
	2	**0.60**	0.216	0.232	0.225	0.282	1.000	1.000	0.116	0.129
		0.65	0.409	0.444	0.505	0.582	1.000	1.000	0.226	0.267
		0.70	0.609	0.545	0.772	0.756	1.000	1.000	0.395	0.403
		0.75	0.751	0.663	0.912	0.870	1.000	1.000	0.596	0.618
1000	0	**0.60**	0.015	0.020	0.014	0.024	0.015	0.017	0.011	0.022
		0.65	0.016	0.029	0.019	0.027	0.016	0.021	0.018	0.026
		0.70	0.023	0.029	0.024	0.030	0.023	0.029	0.021	0.033
		0.75	0.027	0.036	0.027	0.037	0.025	0.036	0.024	0.034
	1	**0.60**	0.819	0.751	0.919	0.882	1.000	1.000	0.612	0.640
		0.65	0.938	0.872	0.990	0.971	1.000	1.000	0.841	0.855
		0.70	0.975	0.912	0.998	0.988	1.000	1.000	0.938	0.939
		0.75	0.991	0.948	1.000	0.996	1.000	1.000	0.962	0.973
	2	**0.60**	0.988	0.987	0.992	0.992	1.000	1.000	0.578	0.615
		0.65	1.000	0.999	1.000	1.000	1.000	1.000	0.838	0.873
		0.70	1.000	1.000	1.000	1.000	1.000	1.000	0.959	0.966
		0.75	1.000	1.000	1.000	1.000	1.000	1.000	0.978	0.981

Table 15: Power against alternative DGPs, for $Y_t = \mu_t + X_t$, with $D(0.2, 0.3)X_t = v_t$, $\Sigma_v = ((1,0),(0,1))'$ and $\mu_t = \zeta\left(t^{-\frac{t}{5T}} - T^{-1}\sum_{t=1}^{T} t^{-\frac{t}{5T}}\right)$ (mon), $\mu_t = \zeta\,(t/T - 1/2)$ (lin), $\mu_t = \zeta \sin\left(\frac{4\pi t}{T}\right)$ (sin) or $\mu_t = \zeta_t - T^{-1}\sum_{t=1}^{T}\zeta_t$ (ms), where ζ_t is a markov sequence that takes the values ζ or 0 and has a switching probability of $5/T$.

		Size											
		Pre-whitened: $\tilde{X}_t$								Unfiltered: X_t			
	DGP	1		2		3		4		1		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**
250	**0.60**	0.007	0.010	0.011	0.011	0.009	0.008	0.008	0.011	0.005	0.011	0.005	0.01
	0.65	0.007	0.013	0.009	0.016	0.009	0.012	0.011	0.016	0.010	0.016	0.014	0.028
	0.70	0.010	0.012	0.010	0.012	0.008	0.007	0.016	0.015	0.015	0.018	0.040	0.070
	0.75	0.008	0.010	0.008	0.007	0.006	0.007	0.011	0.013	0.016	0.019	0.127	0.197
500	**0.60**	0.012	0.014	0.017	0.016	0.015	0.015	0.013	0.015	0.012	0.014	0.011	0.013
	0.65	0.013	0.022	0.015	0.020	0.012	0.020	0.019	0.024	0.014	0.022	0.019	0.037
	0.70	0.018	0.025	0.021	0.021	0.014	0.015	0.025	0.026	0.018	0.028	0.064	0.104
	0.75	0.023	0.019	0.015	0.018	0.011	0.010	0.028	0.028	0.025	0.028	0.240	0.317
1000	**0.60**	0.018	0.021	0.018	0.026	0.018	0.022	0.021	0.022	0.015	0.022	0.012	0.022
	0.65	0.023	0.027	0.026	0.026	0.027	0.023	0.028	0.030	0.026	0.030	0.022	0.037
	0.70	0.026	0.027	0.026	0.023	0.022	0.019	0.039	0.033	0.033	0.029	0.069	0.111
	0.75	0.027	0.025	0.025	0.027	0.018	0.015	0.038	0.035	0.040	0.043	0.356	0.471
2000	**0.60**	0.019	0.026	0.019	0.024	0.016	0.031	0.018	0.024	0.018	0.025	0.013	0.028
	0.65	0.025	0.034	0.023	0.033	0.026	0.031	0.024	0.034	0.026	0.034	0.022	0.038
	0.70	0.028	0.031	0.026	0.033	0.030	0.023	0.036	0.036	0.030	0.034	0.072	0.109
	0.75	0.033	0.036	0.033	0.029	0.022	0.016	0.049	0.043	0.041	0.046	0.442	0.525

		Power											
		Pre-whitened: $\tilde{X}_t$								Unfiltered: X_t			
	DGP	1		2		3		4		1		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**
250	**0.60**	0.038	0.044	0.053	0.056	0.038	0.039	0.040	0.046	0.041	0.042	0.013	0.015
	0.65	0.077	0.091	0.089	0.105	0.051	0.059	0.090	0.105	0.090	0.111	0.020	0.027
	0.70	0.127	0.126	0.119	0.111	0.056	0.044	0.150	0.143	0.175	0.180	0.024	0.023
	0.75	0.170	0.170	0.118	0.106	0.047	0.033	0.198	0.179	0.325	0.329	0.024	0.037
500	**0.60**	0.109	0.100	0.118	0.102	0.071	0.066	0.101	0.107	0.120	0.120	0.066	0.059
	0.65	0.197	0.217	0.176	0.195	0.093	0.105	0.189	0.221	0.249	0.281	0.083	0.100
	0.70	0.319	0.330	0.263	0.255	0.115	0.116	0.325	0.343	0.470	0.506	0.092	0.083
	0.75	0.411	0.410	0.272	0.233	0.118	0.089	0.423	0.416	0.672	0.673	0.068	0.063
1000	**0.60**	0.201	0.202	0.261	0.261	0.143	0.150	0.184	0.189	0.311	0.337	0.227	0.219
	0.65	0.378	0.362	0.400	0.391	0.201	0.167	0.334	0.325	0.577	0.589	0.325	0.298
	0.70	0.508	0.468	0.484	0.425	0.198	0.129	0.463	0.438	0.779	0.773	0.323	0.275
	0.75	0.577	0.552	0.486	0.407	0.156	0.097	0.537	0.497	0.895	0.887	0.246	0.190
2000	**0.60**	0.239	0.234	0.534	0.553	0.349	0.345	0.267	0.268	0.611	0.643	0.523	0.516
	0.65	0.329	0.320	0.699	0.697	0.415	0.380	0.389	0.352	0.838	0.842	0.656	0.643
	0.70	0.389	0.366	0.751	0.707	0.377	0.269	0.430	0.380	0.939	0.928	0.678	0.603
	0.75	0.448	0.404	0.765	0.631	0.287	0.121	0.469	0.380	0.965	0.959	0.586	0.432

Table 16: Size and power of the MLWS test after univariate pre-whitening for DGPs 1 to 4 described in Section C with and without pre-whitening. The bandwidth m is determined by $m = \lfloor T^{\delta} \rfloor$.

		Size											
		Pre-whitened: $\tilde{X}_t$								**Unfiltered:** X_t			
	DGP	1		2		3		4		1		2	
T	$\boldsymbol{\delta/\epsilon}$	**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	0.009	0.007	0.007	0.003	0.008	0.007	0.006	0.005	0.008	0.005	0.011
	0.65	0.005	0.010	0.007	0.011	0.006	0.009	0.006	0.008	0.008	0.012	0.013	0.028
	0.70	0.010	0.010	0.007	0.007	0.005	0.007	0.006	0.009	0.011	0.013	0.037	0.078
	0.75	0.008	0.010	0.004	0.008	0.005	0.010	0.008	0.010	0.013	0.013	0.135	0.220
1000	**0.60**	0.016	0.019	0.014	0.023	0.015	0.017	0.017	0.019	0.014	0.021	0.011	0.022
	0.65	0.018	0.022	0.015	0.023	0.016	0.021	0.015	0.024	0.017	0.024	0.019	0.038
	0.70	0.018	0.024	0.022	0.022	0.011	0.015	0.016	0.024	0.025	0.031	0.075	0.123
	0.75	0.017	0.027	0.014	0.018	0.009	0.013	0.014	0.023	0.030	0.040	0.362	0.463

		Power											
		Pre-whitened: $\tilde{X}_t$								**Unfiltered:** X_t			
	DGP	1		2		3		4		1		2	
T	$\boldsymbol{\delta/\epsilon}$	**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**	0.030	0.027	0.046	0.047	0.029	0.027	0.028	0.028	0.034	0.038	0.019	0.020
	0.65	0.054	0.066	0.081	0.091	0.045	0.049	0.057	0.066	0.081	0.116	0.027	0.035
	0.70	0.092	0.098	0.107	0.115	0.051	0.047	0.108	0.109	0.193	0.202	0.027	0.026
	0.75	0.158	0.160	0.126	0.113	0.060	0.054	0.181	0.193	0.351	0.361	0.030	0.029
1000	**0.60**	0.239	0.242	0.339	0.330	0.268	0.284	0.182	0.197	0.329	0.359	0.318	0.331
	0.65	0.419	0.436	0.552	0.554	0.413	0.387	0.373	0.363	0.610	0.644	0.498	0.481
	0.70	0.596	0.582	0.710	0.683	0.511	0.449	0.550	0.543	0.844	0.849	0.592	0.534
	0.75	0.717	0.705	0.801	0.776	0.595	0.488	0.692	0.666	0.947	0.944	0.577	0.486

Table 17: Size and power of the MLWS test after pre-whitening in a bivariate fractionally cointegrated system, where $X_t = B^{-1}Z_t$ and $Y_t = B^{-1}Z_t + \mu_t$, with $B = ((1,0)', (-1,1)')$ and where Z_t is generated by DGPs 1-4 as described in Section C, but with $d_1 = 0$ and $d_2 = 0.4$.

T	ρ_ϑ	σ^2_ϑ	0								1								2							
		ρ_v	-0.8		0		0.4		0.8		-0.8		0		0.4		0.8		-0.8		0		0.4		0.8	
		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**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**	**0.02**	**0.05**
250	-0.8	**0.60**	0.006	0.010	0.007	0.009	0.005	0.008	0.005	0.010	0.009	0.009	0.006	0.013	0.009	0.011	0.013	0.014	0.009	0.011	0.007	0.011	0.010	0.011	0.012	0.014
		0.65	0.009	0.015	0.008	0.016	0.007	0.014	0.008	0.018	0.015	0.018	0.016	0.021	0.013	0.019	0.017	0.030	0.015	0.020	0.014	0.019	0.019	0.025	0.019	0.025
		0.70	0.013	0.016	0.013	0.018	0.011	0.019	0.011	0.013	0.021	0.023	0.023	0.025	0.022	0.025	0.025	0.028	0.018	0.024	0.022	0.024	0.023	0.025	0.029	0.031
		0.75	0.016	0.024	0.016	0.024	0.017	0.020	0.015	0.021	0.024	0.029	0.025	0.027	0.030	0.028	0.037	0.036	0.025	0.031	0.030	0.029	0.028	0.030	0.034	0.041
	0	**0.60**	0.005	0.009	0.005	0.008	0.010	0.008	0.007	0.008	0.012	0.011	0.008	0.008	0.008	0.013	0.012	0.010	0.009	0.010	0.009	0.011	0.008	0.011	0.010	0.012
		0.65	0.008	0.015	0.007	0.015	0.010	0.015	0.009	0.012	0.016	0.021	0.012	0.017	0.014	0.017	0.019	0.022	0.016	0.021	0.013	0.018	0.013	0.018	0.016	0.019
		0.70	0.012	0.018	0.013	0.018	0.013	0.016	0.012	0.017	0.022	0.026	0.021	0.020	0.018	0.020	0.024	0.026	0.024	0.028	0.019	0.022	0.020	0.024	0.022	0.025
		0.75	0.017	0.023	0.016	0.019	0.018	0.021	0.014	0.019	0.026	0.033	0.025	0.029	0.024	0.027	0.032	0.033	0.025	0.027	0.025	0.027	0.028	0.031	0.027	0.030
	0.4	**0.60**	0.005	0.011	0.008	0.007	0.004	0.011	0.005	0.009	0.009	0.011	0.006	0.010	0.005	0.013	0.008	0.011	0.010	0.012	0.009	0.012	0.009	0.010	0.009	0.010
		0.65	0.010	0.013	0.009	0.015	0.009	0.014	0.011	0.015	0.014	0.023	0.011	0.014	0.010	0.015	0.013	0.022	0.022	0.023	0.015	0.022	0.017	0.018	0.015	0.020
		0.70	0.014	0.016	0.014	0.018	0.013	0.018	0.015	0.017	0.026	0.024	0.018	0.020	0.019	0.021	0.020	0.023	0.027	0.026	0.018	0.024	0.023	0.020	0.022	0.024
		0.75	0.013	0.023	0.012	0.020	0.016	0.021	0.016	0.023	0.034	0.035	0.024	0.023	0.023	0.027	0.026	0.028	0.033	0.033	0.030	0.031	0.022	0.026	0.030	0.027
	0.8	**0.60**	0.007	0.012	0.005	0.011	0.006	0.010	0.006	0.008	0.010	0.012	0.010	0.013	0.010	0.009	0.010	0.013	0.014	0.013	0.011	0.015	0.008	0.010	0.009	0.009
		0.65	0.010	0.015	0.009	0.015	0.008	0.014	0.011	0.015	0.021	0.024	0.013	0.022	0.015	0.016	0.016	0.017	0.018	0.025	0.015	0.020	0.013	0.019	0.016	0.019
		0.70	0.014	0.019	0.012	0.016	0.013	0.017	0.012	0.016	0.028	0.029	0.021	0.021	0.016	0.024	0.018	0.019	0.030	0.026	0.021	0.023	0.021	0.025	0.020	0.024
		0.75	0.014	0.024	0.014	0.026	0.016	0.022	0.017	0.020	0.037	0.038	0.029	0.027	0.024	0.026	0.028	0.027	0.035	0.038	0.027	0.033	0.026	0.029	0.027	0.025
1000	-0.8	**0.60**	0.014	0.019	0.012	0.021	0.015	0.021	0.013	0.017	0.020	0.023	0.016	0.021	0.022	0.027	0.028	0.031	0.026	0.027	0.023	0.027	0.023	0.031	0.031	0.035
		0.65	0.018	0.026	0.018	0.026	0.019	0.025	0.019	0.025	0.036	0.035	0.024	0.035	0.032	0.037	0.046	0.046	0.045	0.045	0.039	0.042	0.044	0.042	0.050	0.049
		0.70	0.020	0.030	0.025	0.027	0.023	0.031	0.023	0.028	0.044	0.050	0.045	0.043	0.046	0.049	0.064	0.069	0.065	0.057	0.056	0.053	0.067	0.057	0.077	0.067
		0.75	0.029	0.035	0.026	0.036	0.030	0.034	0.029	0.037	0.067	0.074	0.057	0.066	0.068	0.073	0.100	0.093	0.093	0.097	0.077	0.086	0.095	0.093	0.100	0.102
	0	**0.65**	0.014	0.021	0.013	0.019	0.014	0.018	0.013	0.020	0.027	0.031	0.020	0.026	0.020	0.024	0.029	0.032	0.027	0.029	0.021	0.024	0.021	0.024	0.032	0.029
		0.65	0.015	0.026	0.019	0.025	0.017	0.023	0.021	0.026	0.039	0.041	0.025	0.033	0.031	0.034	0.041	0.046	0.043	0.045	0.035	0.036	0.039	0.041	0.041	0.043
		0.70	0.023	0.030	0.022	0.031	0.021	0.027	0.026	0.029	0.062	0.059	0.040	0.048	0.043	0.045	0.060	0.056	0.068	0.063	0.052	0.05	0.053	0.058	0.064	0.059
		0.75	0.028	0.036	0.029	0.038	0.029	0.034	0.029	0.037	0.085	0.086	0.063	0.071	0.070	0.075	0.092	0.094	0.100	0.095	0.078	0.088	0.083	0.082	0.093	0.093
	0.4	**0.60**	0.014	0.018	0.014	0.021	0.013	0.019	0.013	0.017	0.029	0.033	0.018	0.023	0.018	0.024	0.026	0.027	0.031	0.033	0.024	0.026	0.021	0.028	0.026	0.030
		0.65	0.023	0.025	0.020	0.023	0.017	0.026	0.019	0.026	0.043	0.042	0.032	0.032	0.029	0.035	0.037	0.043	0.047	0.052	0.036	0.035	0.039	0.035	0.045	0.046
		0.70	0.023	0.027	0.023	0.031	0.025	0.031	0.023	0.032	0.057	0.058	0.041	0.043	0.044	0.043	0.050	0.057	0.063	0.068	0.052	0.050	0.052	0.057	0.059	0.065
		0.75	0.027	0.034	0.029	0.033	0.025	0.038	0.029	0.035	0.090	0.087	0.059	0.067	0.065	0.069	0.082	0.080	0.097	0.093	0.078	0.085	0.085	0.090	0.092	0.094
	0.8	**0.60**	0.017	0.021	0.017	0.022	0.012	0.020	0.013	0.021	0.030	0.031	0.018	0.023	0.018	0.020	0.019	0.025	0.030	0.036	0.022	0.026	0.022	0.028	0.026	0.031
		0.65	0.023	0.025	0.016	0.025	0.020	0.026	0.019	0.025	0.051	0.048	0.026	0.036	0.029	0.035	0.029	0.039	0.044	0.050	0.035	0.041	0.036	0.038	0.044	0.048
		0.70	0.022	0.029	0.022	0.033	0.024	0.025	0.023	0.029	0.062	0.058	0.040	0.039	0.044	0.044	0.051	0.046	0.071	0.062	0.050	0.050	0.051	0.053	0.058	0.058
		0.75	0.030	0.037	0.025	0.035	0.028	0.037	0.026	0.037	0.100	0.095	0.065	0.068	0.062	0.062	0.069	0.072	0.103	0.103	0.087	0.085	0.078	0.083	0.095	0.094

Table 18: Size of the MLWS test for perturbed processes, where $Y_t = X_t + \vartheta_t$, $\vartheta_t \sim N(0,\Sigma_\vartheta)$, with $\Sigma_\vartheta = \sigma^2_\vartheta((1,\rho_\vartheta),(\rho_\vartheta,1))'$ and $D(0.2,0.3)X_t = v_t$ with $v_t \sim N(0,\Sigma_v)$ and $\Sigma_v = ((1,\rho_v),(\rho_v,1))'$.

Size

		Pre-whitened: $\tilde{X}_t$								Unfiltered: X_t			
	DGP	1		2		3		4		1		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**
250	**0.60**	0.007	0.007	0.006	0.007	0.005	0.009	0.008	0.011	0.008	0.011	0.005	0.011
	0.65	0.007	0.014	0.006	0.010	0.006	0.010	0.008	0.010	0.015	0.024	0.007	0.016
	0.70	0.007	0.012	0.008	0.009	0.006	0.006	0.009	0.009	0.019	0.027	0.015	0.020
	0.75	0.011	0.008	0.004	0.004	0.003	0.005	0.013	0.013	0.024	0.033	0.020	0.043
500	**0.60**	0.012	0.013	0.016	0.015	0.016	0.011	0.010	0.013	0.019	0.020	0.011	0.013
	0.65	0.014	0.016	0.014	0.022	0.009	0.016	0.012	0.019	0.027	0.032	0.012	0.023
	0.70	0.014	0.015	0.014	0.011	0.006	0.007	0.014	0.017	0.038	0.045	0.020	0.039
	0.75	0.012	0.014	0.005	0.005	0.004	0.003	0.015	0.012	0.054	0.056	0.039	0.069
1000	**0.60**	0.015	0.021	0.024	0.029	0.025	0.023	0.019	0.021	0.024	0.032	0.019	0.024
	0.65	0.022	0.024	0.035	0.027	0.027	0.022	0.019	0.024	0.049	0.042	0.021	0.028
	0.70	0.024	0.015	0.025	0.019	0.014	0.015	0.017	0.019	0.064	0.060	0.025	0.049
	0.75	0.020	0.020	0.017	0.015	0.005	0.004	0.018	0.017	0.088	0.086	0.068	0.095
2000	**0.60**	0.016	0.031	0.021	0.033	0.022	0.037	0.021	0.030	0.027	0.039	0.019	0.028
	0.65	0.037	0.033	0.034	0.041	0.033	0.046	0.030	0.033	0.049	0.066	0.020	0.032
	0.70	0.027	0.026	0.041	0.035	0.037	0.031	0.029	0.027	0.079	0.081	0.031	0.044
	0.75	0.025	0.019	0.029	0.024	0.019	0.009	0.024	0.016	0.139	0.132	0.083	0.118

Power

		Pre-whitened: $\tilde{X}_t$								Unfiltered: X_t			
	DGP	1		2		3		4		1		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**
	0.60	0.034	0.033	0.028	0.030	0.018	0.017	0.038	0.037	0.074	0.081	0.030	0.033
	0.65	0.061	0.079	0.041	0.050	0.019	0.027	0.062	0.074	0.171	0.196	0.050	0.066
	0.70	0.114	0.117	0.061	0.055	0.020	0.012	0.114	0.117	0.323	0.332	0.078	0.068
	0.75	0.181	0.186	0.066	0.057	0.017	0.011	0.176	0.172	0.494	0.496	0.113	0.097
500	**0.60**	0.110	0.107	0.081	0.078	0.071	0.052	0.106	0.104	0.227	0.209	0.115	0.091
	0.65	0.172	0.207	0.119	0.140	0.061	0.076	0.170	0.209	0.408	0.460	0.164	0.189
	0.70	0.282	0.296	0.159	0.160	0.062	0.049	0.290	0.305	0.650	0.677	0.237	0.243
	0.75	0.358	0.337	0.169	0.143	0.047	0.022	0.367	0.346	0.818	0.811	0.286	0.254
1000	**0.60**	0.297	0.315	0.236	0.213	0.165	0.162	0.304	0.306	0.477	0.491	0.304	0.298
	0.65	0.443	0.423	0.331	0.313	0.202	0.184	0.457	0.460	0.736	0.746	0.450	0.440
	0.70	0.538	0.484	0.382	0.327	0.181	0.115	0.556	0.513	0.888	0.888	0.544	0.483
	0.75	0.581	0.488	0.405	0.294	0.115	0.054	0.586	0.537	0.947	0.941	0.573	0.514
2000	**0.60**	0.620	0.638	0.481	0.497	0.382	0.390	0.615	0.640	0.763	0.784	0.597	0.617
	0.65	0.776	0.760	0.626	0.602	0.469	0.439	0.781	0.768	0.917	0.922	0.745	0.740
	0.70	0.780	0.728	0.653	0.568	0.448	0.298	0.802	0.755	0.967	0.963	0.802	0.752
	0.75	0.733	0.623	0.641	0.448	0.323	0.109	0.781	0.670	0.980	0.979	0.810	0.728

Table 19: Size and power of the MLWS test after pre-whitening (as in Tables 9 and 10), but for perturbed series, where $X_t = Z_t + \vartheta_t$ and $Y_t = Z_t + \mu_t + \vartheta_t$, $\vartheta_t \sim N(0, \Sigma_\vartheta)$, with $\Sigma_\vartheta = ((1,0),(0,1))'$ and Z_t is given by DGPs 1 to 4 as described in Section C.

D Empirical Example

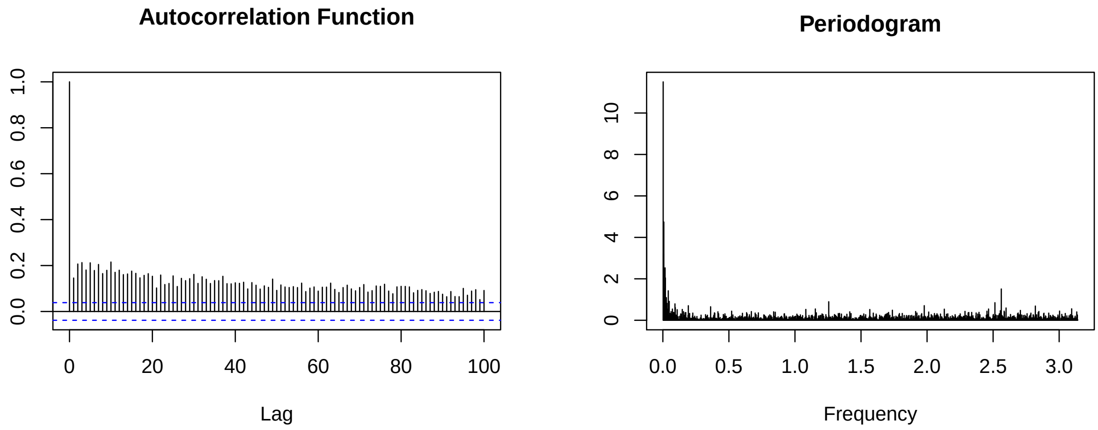

Figure 2: Autocorrelation function and periodogram of the log-absolute return series of the S&P 500.

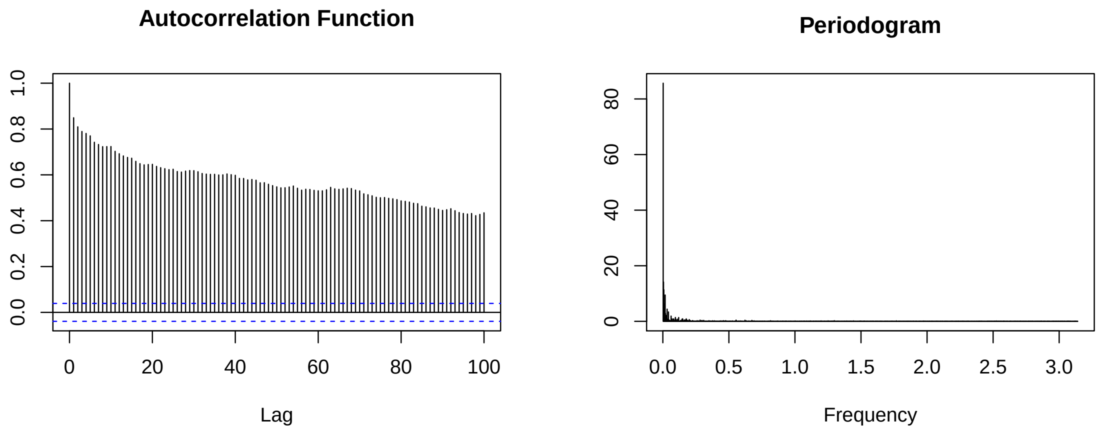

Figure 3: Autocorrelation function and periodogram of the log-realized volatility series of the S&P 500.

This section provides additional material on the empirical application in Section 5. As can be seen from Figures 2 and 3, both series show the typical behavior of long memory time series, with hyperbolically decaying autocorrelation functions and poles at the origin of the periodogram.

Summary statistics for both series are provided in Table 20. We can observe that the kurtosis is higher for the log-realized volatility and the skewness is positive instead of slightly negative.

	$\log(\lvert r_t\rvert + 0.001)$				$\log RV_t$			
	DAX	NIKKEI	S&P 500	FTSE	DAX	NIKKEI	S&P 500	FTSE
mean	-4.977	-4.924	-5.156	-5.107	-9.789	-9.432	-10.055	-9.665
median	-4.920	-4.803	-5.133	-5.051	-9.946	-9.543	-10.183	-9.741
std.dev.	0.924	1.004	0.937	0.900	1.126	0.964	1.017	0.885
skewness	-0.131	-0.307	0.053	-0.087	0.744	0.662	0.639	0.747
kurtosis	2.475	2.419	2.528	2.573	3.666	3.751	3.295	4.301

Table 20: Summary statistics of the log-absolute returns and the log-realized volatility series.

δ	DAX	NIKKEI	S&P 500	FTSE	DAX	NIKKEI	S&P 500	FTSE	DAX	NIKKEI	S&P 500	FTSE
	$\log(\lvert r_t\rvert + 0.001)$											
	d_{ELW}				d_{HP}				$d_{HP+Noise}$			
0.60	0.365	0.290	0.476	0.396	0.350	0.204	0.436	0.334	0.354	0.545	0.436	0.334
0.65	0.341	0.289	0.415	0.368	0.292	0.233	0.327	0.306	0.488	0.446	0.562	0.306
0.70	0.337	0.289	0.363	0.310	0.291	0.247	0.242	0.230	0.427	0.247	0.608	0.484
0.75	0.273	0.258	0.307	0.299	0.187	0.205	0.136	0.231	0.534	0.380	0.648	0.415
	$d_{LPWN}(0,0)$				$d_{LPWN}(1,0)$				$d_{LPWN}(0,1)$			
0.60	0.436	0.545	0.472	0.446	0.452	0.540	0.477	0.461	0.515	0.553	0.472	0.631
0.65	0.488	0.446	0.582	0.462	0.431	0.459	0.521	0.477	0.440	0.575	0.481	0.538
0.70	0.452	0.373	0.608	0.531	0.476	0.397	0.551	0.416	0.448	0.535	0.558	0.387
0.75	0.534	0.408	0.648	0.489	0.412	0.428	0.600	0.493	0.413	0.399	0.602	0.490
	$\log RV_t$											
	d_{ELW}				d_{HP}				$d_{HP+Noise}$			
0.60	0.625	0.640	0.642	0.622	0.639	0.546	0.637	0.629	0.639	0.546	0.637	0.661
0.65	0.593	0.646	0.577	0.608	0.567	0.597	0.558	0.588	0.648	0.597	0.660	0.668
0.70	0.598	0.640	0.577	0.612	0.558	0.595	0.546	0.583	0.620	0.595	0.621	0.639
0.75	0.580	0.601	0.554	0.583	0.515	0.531	0.498	0.538	0.642	0.673	0.630	0.662
	$d_{LPWN}(0,0)$				$d_{LPWN}(1,0)$				$d_{LPWN}(0,1)$			
0.60	0.642	0.679	0.637	0.661	0.619	0.683	0.550	0.664	0.642	0.802	0.637	0.681
0.65	0.653	0.652	0.660	0.670	0.641	0.654	0.655	0.655	0.644	0.652	0.644	0.640
0.70	0.643	0.652	0.623	0.643	0.637	0.661	0.616	0.652	0.643	0.665	0.623	0.661
0.75	0.646	0.677	0.632	0.662	0.650	0.672	0.615	0.659	0.652	0.675	0.621	0.663

Table 21: Estimated memory parameters for different bandwidths $m = \lfloor T^{\delta} \rfloor$ using the estimators of Shimotsu and Phillips (2005) (d_{ELW}), Hou and Perron (2014) (d_{HP} and $d_{HP+Noise}$) and Frederiksen et al. (2012) (d_{LPWN}).

Table 21 shows estimates of the memory parameters in our empirical application in Section 5. The exact local Whittle estimates (d_{ELW}) are very close to the standard local Whittle estimates presented in Tables 6 and 7 of the paper. For the log-absolute returns the estimates using the LPWN estimator are considerably higher, as reported by Frederiksen et al. (2012). Estimates using the estimator of Hou and Perron (2014), on the other hand, are slightly reduced. When considering the log-realized volatility series,

δ	Qu test DAX	NIKKEI	S&P 500	FTSE	Components DAX	NIKKEI	S&P 500	FTSE	MLWS ALL
					$\log RV_t$				
0.60	0.327	0.390	0.438	0.546	0.774	0.378	0.790	0.516	0.598
	(0.999)	(0.985)	(0.957)	(0.825)	(0.511)	(0.997)	(0.811)	(0.997)	(0.737)
0.65	0.421	0.539	0.382	0.925	0.604	1.103	0.927	0.835	0.736
	(0.969)	(0.837)	(0.988)	(0.240)	(0.914)	(0.096)	(0.550)	(0.709)	(0.495)
0.70	0.339	0.426	0.538	0.654	0.612	0.883	0.839	0.689	0.565
	(0.997)	(0.965)	(0.837)	(0.637)	(0.778)	(0.313)	(0.642)	(0.862)	(0.794)
0.75	0.434	0.583	0.533	0.980	0.847	0.770	0.538	1.064	0.752
	(0.959)	(0.763)	(0.845)	(0.190)	(0.380)	(0.633)	(0.977)	(0.207)	(0.470)

Table 22: 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.

the estimates are extremely robust - irrespective of the bandwidth and the choice of the estimator.

Since the estimated memory parameters for the log-realized volatility series are in the lower non-stationary region, but the limit distribution of the test statistic is derived under the assumption of stationarity, a robustness check of these empirical results is conducted in Table 22. Here, the tests are applied to the series after fractional differentiation with $d = 0.4$, so that the memory of the cointegrating errors is close to zero and that of the series themselves is reduced to the region around 0.2.

	log-absolute			log-realized		
δ/δ_1	**0.65**	**0.70**	**0.75**	**0.65**	**0.70**	**0.75**
0.50	0.033	0.040	0.049	0.036	0.060	0.085
0.55	0.035	0.044	0.051	0.041	0.059	0.081
0.60	0.036	0.041	0.055	0.038	0.053	0.085

Table 23: Size of the MLWS test applied to cointegrated systems after estimation of the cointegrating rank. DGPs match parameter values of empirical application: $T = 2600$, $v_T = T^{-0.3}$, $\varepsilon = 0.02$, $m = \lfloor T^{\delta} \rfloor$, $m_1 = \lfloor T^{\delta_1} \rfloor$ for both DGPs and $B = ((1,0,0,-1),(0,1,0,0),(0,0,1,0),(0,0,0,1))'$, $d_1 = 0.2$, and $d_2 = d_3 = d_4 = 0.35$ (for **log-absolute** returns) and $B = ((-1,0,1,0),(0,-1,1,0),(0,0,1,0),(0,0,0,1))'$, $d_1 = 0.4$, $d_2 = 0.5$, and $d_3 = d_4 = 0.6$ (for **log-realized** volatility).

We can observe that the test statistics are generally stable or slightly reduced when compared to those in Table 8 of the paper. There is no rejection of the null hypothesis of true long memory - irrespective of the series, test and bandwidth. Therefore, the results from Section 5 are clearly confirmed.

The asymptotic validity of the $\widetilde{MLWS}(\hat{B}(\hat{p}_G))$ test is established in Theorem 4. As an additional robustness check, we run a small simulation study to explore the finite sample effect of estimating the cointegrating rank on the size of the test. The parameter values are selected so that they closely resemble the situation in the empirical application. Table 23 shows the results of this exercise. As one can see, the size is only elevated slightly for larger bandwidths in the setup resembling the situation of the realized volatility series, where the MLWS test did not reject. In the first setup (log-absolute), the size is unaffected by the previous estimation of the cointegrating rank. Therefore, there is no evidence that the test results might be affected by the estimation of the cointegrating rank.

References

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

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

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

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.

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.

Hualde, J., Robinson, P. M., et al. (2011). Gaussian pseudo-maximum likelihood estimation of fractional time series models. *The Annals of Statistics*, 39(6):3152–3181.

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.

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

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

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

Nielsen, M. Ø. (2015). Asymptotics for the conditional-sum-of-squares estimator in multivariate fractional time-series models. *Journal of Time Series Analysis*, 36(2):154–188.

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

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

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

Shimotsu, K. and Phillips, P. C. (2005). Exact local whittle estimation of fractional integration. *The Annals of Statistics*, 33(4):1890–1933.

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

Velasco, C. (1999). Gaussian semiparametric estimation of non-stationary time series. *Journal of Time Series Analysis*, 20(1):87–127.