

# Inference on the Long-Memory Properties of Time Series with Non-Stationary Volatility\*

Matei Demetrescu<sup>†</sup> and Philipp Sibbertsen<sup>‡</sup>

## Abstract

Many time series exhibit unconditional heteroskedasticity, often in addition to conditional one. But such time-varying volatility of the data generating process can have rather adverse effects when inferring about its persistence; e.g. unit root and stationarity tests possess null distributions depending on the so-called variance profile. On the contrary, this is not the case in stationary autoregressions, and correctly sized inference is guaranteed if taking protective actions as simple as using White standard errors (which are employed anyway to deal with conditional heteroskedasticity). The paper explores the influence of time-varying volatility on fractionally integrated processes. Concretely, we discuss how to model long memory in the presence of time-varying volatility, and analyze the effects of such nonstationarity on several existing inferential procedures for the fractional integration parameter. Based on asymptotic arguments and Monte Carlo simulations, we show that periodogram-based estimators, such as the local Whittle or the log-periodogram regression estimator, remain consistent, but have asymptotic distributions whose variance depends on the variance profile. Time-domain, regression-based tests for fractional integration retain their validity if White standard errors are used. Finally, the modified range-scale statistic is only affected if the series require adjustment for deterministic components.

## Key words

Time-varying variance, Heteroskedasticity, Persistence, Fractional integration, Modulated process

## JEL Classification

C12 (Hypothesis Testing), C22 (Time-Series Models)

---

\*The authors would like to thank Jörg Breitung, Uwe Hassler, Liudas Giraitis and Maya Olivares for very helpful comments and suggestions. Philipp Sibbertsen gratefully acknowledges the financial support of the Deutsche Forschungsgemeinschaft.

<sup>†</sup>*Corresponding author:* Institute for Statistics and Econometrics, Christian-Albrechts-University of Kiel, Olshausenstr. 40-60, D-24118 Kiel, Germany, email: mdeme@stat-econ.uni-kiel.de.

<sup>‡</sup>Department of Economics, University of Hannover, Königswörther Platz 1, D-30167 Hannover, Germany. E-mail address: sibbertsen@statistik.uni-hannover.de.

# 1 Introduction

There is an increasing body of evidence indicating that a standard assumption of many test procedures, unconditional homoskedasticity, is violated for many types of data. Indeed, macroeconometricians could immediately name the Great Moderation as an example. Financial time series have been found to exhibit unconditional heteroskedasticity too, in addition to conditional heteroskedasticity (see e.g. Amado and Teräsvirta, 2008).

The effect of deterministic (unconditional) heteroskedasticity has already been analyzed for common econometric tools such as stationarity, unit root, and cointegration tests; see Cavaliere (2004), Cavaliere and Taylor (2005), and Cavaliere et al. (2010). The finding common to these papers is that the null distributions of virtually all examined test statistics are distorted whenever so-called global homoskedasticity is violated (see also Cavaliere and Taylor, 2009). The reason for such behavior is as follows. The asymptotic null distributions of the respective test statistics can be expressed as functionals of Wiener processes, relying on invariance principles for cumulated sums of stochastic processes. But under time-varying variance of the increments, modified invariance principles hold: intuitively, the changing variance transforms time in the limit, which in turn affects the limiting distributions. The exact mechanism can be seen e.g. in Hall (1977), who shows that weak convergence to Brownian motion of cumulative sums of square-integrable, possibly heteroskedastic, triangular martingale difference arrays holds only if normalizing appropriately (i.e. using an estimate of the quadratic variation).

Such examinations, however, are actually scarce in the context of fractionally integrated series ( $I(d)$ ). On the one hand, the modified range/scale statistic discussed by Lo (1991) has an asymptotic distribution expressed in terms of Wiener processes under unconditional homoskedasticity, so it could be conjectured that it is affected by time-varying variance in the same manner as stationarity and unit root tests. On the other hand, Kew and Harris (2009) provided an analysis of regression-based fractional integration tests applied to linear processes with heteroskedastic innovations (shocks). The efficient Wald test of Lobato and Velasco (2007) or the augmented LM [ALM] test of Demetrescu et al. (2008) fall in the examined category. Kew and Harris (2009) show that, if using White standard errors (recommended by Demetrescu et al. (2008) to deal with conditional heteroskedasticity), some types of deterministic heteroskedasticity are allowed for; their assumptions exclude, for instance, a simple break in the variance, but they conjecture that the results can be extended in that particular direction. Finally, Cavaliere et al. (2013) propose wild bootstrap implementations of Wald, LR and LM tests for the fractional integration parameter in a parametric ARFIMA( $p, d, q$ ) model (see Cavaliere and Taylor, 2007a, for the unit root case) and show them to be robust to both conditional and unconditional heteroskedasticity of unknown form in the shocks.

But the behavior of the equally popular frequency-domain procedures such as the local Whittle or the log-periodogram regression estimator is not known at all in the presence of time-varying variance. Provided that the variance varies smoothly enough, a possible approach to deal with time-varying variances is the concept of local stationarity as suggested in Dahlhaus (2000). Local versions of fractionally integrated processes are discussed among others in Beran (2009) or Palma and Olea (2010). These authors work with localized versions of linear time series and define the spectrum of the time series as an averaged local spectrum. This leads to a higher flexibility in the sense that model parameters can depend on time (which thus robustifies against time-varying variance) but introduces an additional bandwidth parameter in the analysis based on spectral densities. Concretely, the localized version of the local Whittle estimator relies on the usual bandwidth parameter determining the amount of frequencies used for the estimation but has a second smoothing parameter coming from the localization. As the choice of the bandwidth parameter is crucial in spectral-based long memory estimation, the effect of the introduction of a second bandwidth parameter is unclear. Practitioners are thus likely to choose to work with localized versions of the estimators only if there is compelling evidence against weak stationarity.

The paper therefore discusses the effect of time-varying variance on the (non-localized) estimation of the fractional integration parameter and onto the related tests. The focus is on finding procedures that are robust to time-varying volatility, i.e. that do not require a decision between the basic form of an estimator or test and its heteroskedasticity-corrected version. We emphasize the need for robustness, since, in order to apply some correction (if available), one has to be aware that the correction is necessary.

Concretely, our contributions are as follows. Section 2 describes our setting in detail and provides a discussion of the interactions of time-varying variance with the long memory properties of time series, as well as of the consequences of using different variants of the heteroskedastic long memory model we work with. In Section 3, we show that periodogram-based procedures are still asymptotically unbiased but have typically an inflated variance which affects inference based on the distributional properties of these estimators. At the same time, regression-based procedures are shown to be robust against time-varying variance as modeled here. The modified range-scale statistic turns out to be robust if adjustment for deterministic is not required, but is affected otherwise. The final section concludes, and the proofs have all been gathered in the appendix.

Let us set some notation before providing the framework of our analysis and the main findings and results. Denote by  $f_y$  the spectrum of a weakly stationary process  $y_t$ ,  $f_y(\lambda) = \frac{1}{2\pi} |E(y_t e^{i\lambda t})|^2$ , by  $I_y$  the periodogram of a series  $y_t$ ,  $I_y(\lambda) = \frac{1}{2\pi T} \left| \sum_{t=1}^T y_t e^{i\lambda t} \right|^2$ , and by  $\lambda_j$  the  $j^{\text{th}}$  Fourier frequency,  $\lambda_j = 2\pi j/T$ . Further, let  $C$  be a generic constant and  $\|\cdot\|_2$  the  $L_2$  norm of a random variable;  $\xrightarrow{d}$  denotes convergence in distribution, and  $\xrightarrow{p}$

denotes convergence in probability. Finally, the  $O_p$  symbol has its usual meaning of an upper bound for the magnitude order of a stochastic sequence.

## 2 Heteroskedasticity and long memory

A very popular model for long memory and antipersistence is given by the fractionally integrated process

$$y_t = \mu + (1 - L)^{-d} u_t \mathbb{I}(t > 0), \quad t \in \mathbb{Z} \quad (1)$$

where the fractional integration filter is given by the usual series expansion and  $u_t$  is a short memory component. The paper assumes without loss of generality that  $\mu = 0$ . Under regularity conditions for  $u_t$ —in particular, weak stationarity or asymptotically negligible departures from weak stationarity (see e.g. Phillips, 1987, in the unit root case)—, well-known invariance principles hold. If  $d \in (-0.5; 0.5)$ , suitably normalized cumulated sums of  $y_t$  converge weakly to fractional Brownian motion of type I; if  $d \in (0.5; 1.5)$ , it is  $y_t$  itself that, suitably normalized, converges weakly to fractional Brownian motion (of type II). For sets of sufficient conditions see e.g. Davydov (1970) or Marinucci and Robinson (2000), respectively.

Alternatively, the behavior of the spectrum of the process in a neighborhood of the origin can be used to characterize the memory properties

$$\lambda^{2d} f_y(\lambda) \rightarrow C \quad \text{as } \lambda \rightarrow 0.$$

If  $u_t$  has a continuous spectrum, bounded and bounded away from 0 at the origin, the spectrum of  $y_t$  from (1) behaves like in the above equation.

Motivated e.g. by the literature discussing changes in the persistence of processes with long-range dependence (see Leybourne et al., 2007, or the more recent contributions of Sibbertsen and Kruse, 2009, and Hassler and Scheithauer, 2011), one could allow for deterministic changes in  $d$ : they necessarily influence the variance of  $y_t$ . Such a model, however, is not too convenient because it mingles the variability of the variance with the persistence of the process.

Virtually all existing papers assume a component model for introducing heteroskedasticity, where the stochastic component is weakly stationary and the heteroskedasticity is induced by multiplication with a sequence of constants; typically, a triangular array structure is allowed for in order to model e.g. a structural break occurring at a fixed position relative to the sample size. This structure is nothing else than a uniformly modulated process in the terminology of Priestley (1988, p. 165).

Looking back at the unit root case, the natural choice for the nonstationary fractional case seems to be to focus on the short memory increments; for them, Cavaliere (2004)

assumes that  $u_t = \sigma_t v_t$  for each  $t = 1, \dots, T$ , where  $v_t$  is (short-memory) weakly stationary and the so-called variance pattern  $\sigma_t$  is given by the sampling at  $t/T$  of a piecewise continuous function with a finite number of discontinuities fulfilling a uniform Lipschitz condition at all continuity points.

But from an economic modeling perspective, say, the more convenient way to introduce heteroskedasticity relies on specifying the short-memory component  $u_t$  as a linear process. Under weak stationarity, such a linear Wold representation exists and the assumption is not restrictive. So one can require—as do Phillips and Xu (2006) and Cavaliere and Taylor (2007b)—that the innovations of the linear process be modulated, instead of the process itself,  $u_t = \sum_{j \geq 0} b_j \varepsilon_{t-j}$  with  $\varepsilon_t = \sigma_t \epsilon_t$ , where  $\epsilon_t$  is uncorrelated with constant variance and the coefficients  $b_j$  satisfy some summability condition.

The assumption of a linear process is quite general and is standard in the unit root literature; a very similar assumption is made e.g. by Marinucci and Robinson (2000) in the fractional integration context. It allows of course for finite-order ARMA processes, whose Wold coefficients decay exponentially. Such an assumption allows for the nicer interpretation that *innovations*, and not the entire process, have a changing variance in response to changing environment conditions instead of a change affecting all previous innovations as well. It also allows for a clear distinction between dynamics and volatility of the series.

By focusing on the short memory component, however, the discussion ignores the “intermediate position” of fractional integration between stationarity and integration: when dealing with *stationary* long memory, the process itself could be modulated, as opposed to having modulated innovations or modulated short-run components.

All these are different data generating processes [DGPs]. In practice, one may however be tempted to pick that DGP that can be more easily dealt with analytically. So the question relevant at this point is, what is the relation between them? Do they have different implications? More precise assumptions are required for answering these questions.

The variance pattern examined here relies on Cavaliere’s (2004) assumption but allows for unbounded, trending variances. See also Cavaliere and Taylor (2009) and Cavaliere et al. (2013). As pointed out by Cavaliere (2004), assuming the variance pattern to be deterministic is a simple way of ensuring it to be exogenous; if  $\sigma_t$  is random, but measurable w.r.t. the remote  $\sigma$ -field  $\bigcap_{s=-\infty}^{\infty} \{\epsilon_s, \epsilon_{s-1}, \dots\}$ , the results are not affected.

**Assumption 1** *There exists a non-negative function  $\sigma(s)$  with real support satisfying a uniform Lipschitz condition at all but a finite number of (jump) discontinuity points, and  $\alpha \geq 0$ , such that  $\sigma_t = T^\alpha \sigma\left(\frac{t}{T}\right)$ .*

The assumption allows e.g. for deterministic jumps in the variance (with  $\alpha = 0$ ), but also for a linear trend in the variance (with  $\sigma(s) = s\mathbb{I}(s > 0)$  and  $\alpha = 1$ ). The

nonnegativity requirement keeps the interpretability of  $\sigma$  as an exogenous variance scaling factor. The assumption also implies  $\sigma(\cdot)$  to be bounded, with boundedness being helpful when considering  $t < 0$ . Assumption 1 basically requires that the properly scaled variance does not vary too wildly in the limit.

Let us now examine the differences between the two discussed DGPs. The proposition below examines the interchangeability of linear filtering and variance modulation as implied by Assumption 1.

**Proposition 1** *Let  $u_t = \sum_{j \geq 0} b_j \epsilon_{t-j} \sigma_{t-j}$ , and  $\tilde{u}_t = \sigma_t \left( \sum_{j \geq 0} b_j \epsilon_{t-j} \right)$  where  $\|\epsilon_t\|_2 < C < \infty \forall t \in \mathbb{Z}$  and  $\{b_j\}_{j \geq 0}$  is 1-summable. It then holds true under Assumption 1 that*

$$\|u_{[sT]} - \tilde{u}_{[sT]}\|_2 < CT^{\alpha-1}$$

at all continuity points  $s \in [0, 1]$  of  $\sigma(\cdot)$ .

**Proof:** See Appendix C.

The proposition is very general in that it does not require specific assumptions about the innovations  $\epsilon_t$  except uniform  $L_2$  boundedness, not even short memory or weak stationarity. As can be seen from the proof, the difference between  $u_{[sT]}$  and  $\tilde{u}_{[sT]}$  is  $O_p(T^\alpha)$  at the discontinuity points of  $\sigma(\cdot)$ . If  $\sigma_t$  obeys Assumption 1 with  $\alpha < 1$ ,  $u_{[sT]}$  converges in mean square to  $\tilde{u}_{[sT]}$  at a rate uniform in  $s$ . The rate depends on the trending behavior of the variance; if the variance trend is too strong, there is no equivalence. This can also be the case if  $\sigma_t$  induces “local heteroskedasticity,” under which the differences do not vanish in general; seasonal MA filters, for instance, never lead to asymptotic equivalence under local heteroskedasticity of the shocks. In contrast,  $b_j = 0 \forall j > 0$  is an obvious particular case where equivalence occurs.

The result assumes a short-memory type summability condition on the filter coefficients. For the case of more persistent filters, we have the following proposition.

**Proposition 2** *Let  $u_t$  be fractionally integrated of order  $d$ , and  $y_t = \sum_{j=0}^{t-1} \phi_j^{(d_1)} u_{t-j} \sigma_{t-j}$  and  $\tilde{y}_t = \sigma_t \sum_{j=0}^{t-1} \phi_j^{(d_1)} u_{t-j}$  where  $\phi_j^{(d_1)}$  are the coefficients of the fractional difference filter with parameter  $d_1$ , not necessarily equal to  $d$ . Then,*

$$\|y_{[sT]} - \tilde{y}_{[sT]}\|_2 \leq CT^{\alpha+(d-d_1)-0.5}.$$

at all continuity points  $s \in [0, 1]$  of  $\sigma(\cdot)$ .

**Proof:** See Appendix C.

Thus, the considered DGPs are equivalent in a certain sense. But do not know yet what the exact properties of the DGPs are with respect to serial dependence. The heteroskedastic structure of the series does influence the theoretical long memory properties:

since the series is not stationary, one has time-varying autocovariances,

$$\gamma_h^{t,T}(y_t) = \gamma_h \sigma_t \sigma_{t-h}.$$

Obviously, if the variance function is trending, the autocovariance  $\gamma_h^{t,T}(y_t)$  does not even have to vanish as  $h \rightarrow \infty$ .

In spite of this behavior of the autocovariances, the persistence properties of  $y_t$  are still characterized by  $d$ , as it is reviewed in Appendix B. The remaining question is, whether time-varying heteroskedasticity affects inference on  $d$ . Section 3 examines the effects of modulation on corresponding estimators and tests.

Moreover, Appendix A uses Propositions 1 and 2 to argue that the test statistics considered in Section 3.1 are themselves not affected asymptotically by the choice of the DGP when regarding short memory. For long memory, this is not warranted, however. The derivations are procedure-specific anyway, so one might want to check the effects of interchanging modulation and filtration in each specific case. Given the positive results in Appendix A, we pick the more convenient DGP and model the innovations  $\varepsilon_t$  as being modulated.

For the derivations in the following section related to the regression-based tests, the stationary component is required to be a stable, finite-order AR process with martingale difference innovations.

**Assumption 2** *Let  $u_t$  be a stable autoregression of order  $p$  driven by modulated shocks  $\varepsilon_t = \sigma_t \epsilon_t$  such that  $\sigma_t$  obeys Assumption 1 and  $\epsilon_t$  is a weakly stationary martingale difference sequence with absolutely summable  $8^{\text{th}}$ -order cumulants.*

The assumption allows for the incorporation of conditional heteroskedasticity in the model. In fact, even more general conditions for  $u_t$  are feasible (e.g. ARMA models, accommodated for by using autoregressive approximations), but the paper sticks with a finite autoregression for clarity of the exposition.

## 3 Inference on the fractional integration parameter

### 3.1 Procedures

When the interest lies in testing hypotheses about the fractional integration parameter  $d$ , one basically has the choice between time-domain based tests and frequency-domain based tests.

For the time domain, we shall examine here regression-based tests of the form

$$x_t = \hat{\phi}x_{t-1}^* + \sum_{j=1}^p \hat{a}_j x_{t-j} + \hat{\varepsilon}_t, \quad t = p + 1, \dots, T, \quad (2)$$

where  $x_{t-1}^* = \sum_{j=1}^{t-1} \psi_j x_{t-j}$  for suitably chosen weights  $\psi_j$  and  $x_t = \Delta^{d_0} y_t$ . This formulation encompasses the fractional Dickey-Fuller test of Dolado et al. (2002), the efficient Wald test of Lobato and Velasco (2007), or the ALM test due to Demetrescu et al. (2008), depending on the choice of the filter  $\{\psi_j\}_{j \geq 1}$ . For the ALM test for instance, one has  $\psi_j = 1/j$ , which is motivated by the LM principle. Dolado et al. suggest to choose  $\psi_j$  as the coefficients of the fractional integration filter having as integration parameter the (possibly estimated) distance between the null and the alternative. Lobato and Velasco modify the latter to achieve efficiency. Throughout the paper, fixed weights are assumed to keep the complexity of the proofs under control.

Under the null hypothesis  $d = d_0$ , one has  $x_t \equiv u_t$ , and, consequently,  $\phi = 0$ ; evidence in favor of  $\phi < 0$  points toward the alternative  $d < d_0$ , and evidence in favor of  $\phi > 0$  points toward the alternative  $d > d_0$ . Standard asymptotics apply for the  $t$  statistic of  $\phi$  as long as  $\psi_j$  is square summable; letting  $p \rightarrow \infty$  at an appropriate rate, one can even approximate linear processes in an autoregressive manner and maintain the standard asymptotics. The approximation order  $p$ , however, should be a deterministic function of  $T$ , in order to avoid problems associated with post-model selection inference. See Demetrescu et al. (2008) for a discussion of the ALM situation, discussion extending to square summable coefficients  $\psi_j$ .

Demetrescu et al. also suggest to use White standard errors, as they ensure correct inference under conditional heteroskedasticity of  $\varepsilon_t$  on the one hand, and there is little, if nothing, to lose under conditional homoskedasticity on the other hand. (This applies of course for the fractional Dickey-Fuller test or the efficient Wald test as well.)

In order to deal with deterministic components such as a non-zero mean, one simply removes from  $x_t = \Delta^{d_0} y_t$  the correspondingly *differenced* deterministic component before applying the ALM test; it would not affect the asymptotic standard normal distribution under homoskedasticity. See Demetrescu et al. (2008, Proposition 4), and it is straightforward to show that the result holds under unconditional heteroskedasticity as well.

Also in the time domain, one can resort to the modified range/scale statistic discussed by Lo (1991). It relies as well on building differences under the null hypothesis. Should the differences require no adjustment for deterministic components, the statistic is given by

$$Q = \frac{1}{\sqrt{T}} \frac{\max_{1 \leq t \leq T} S_t - \min_{1 \leq t \leq T} S_t}{\hat{\omega}}$$



where  $S_t$  are the partial sums of the series of interest differenced under the null,  $S_t = \sum_{j=1}^t x_j$ , and  $\widehat{\omega}^2$  is an estimator of the long-run variance of  $x_t$  (cf. Lo, 1991). Under homoskedasticity, the limiting distribution is

$$Q \xrightarrow{d} \max_{s \in [0,1]} W(s) - \min_{s \in [0,1]} W(s)$$

with  $W$  a standard Wiener process. In the case with adjustment for deterministic, say for the prototypical constant, the statistic  $Q$  is computed with demeaned series  $x_t - \bar{x}$  and the limiting distribution is given in terms of standard Brownian bridges rather than Wiener processes.

In the frequency domain, the paper looks at the local Whittle estimator  $\widehat{d}_{lw}$  and at the log-periodogram regression. The asymptotic distribution of the local Whittle estimator was first derived for  $d < 0.5$  by Robinson (1995a) assuming homoskedasticity:

$$\sqrt{4m} \left( \widehat{d}_{lw} - d_0 \right) \xrightarrow{d} \mathcal{N}(0, 1),$$

where  $m$  is the bandwidth,  $m/T \rightarrow 0$  as  $m, T \rightarrow \infty$  (regularity conditions assumed). The local Whittle estimator is given by

$$\widehat{d}_{lw} = \arg \min_d \left[ \log \left( \frac{1}{m} \sum_{j=1}^m \lambda_j^{2d} I_y(\lambda_j) \right) - \frac{2d}{m} \sum_{j=1}^m \log \lambda_j \right],$$

with  $I_y(\cdot)$  being the periodogram of  $y_t$  and  $\lambda_j = 2\pi j/T$  the  $j^{\text{th}}$  Fourier frequency. A non-zero mean is accommodated by construction. Shao and Wu (2007) show the asymptotic distribution above to be robust to a certain degree of conditional heteroskedasticity, and also study the behavior of  $\widehat{d}_{lw}$  for values of the fractional difference parameter up to  $d < 1.5$ .

For the log-periodogram regression of Geweke and Porter-Hudak (1983) [GPH],  $\widehat{d}_{gph}$  is obtained from the regression

$$\log I_y(\lambda_j) = C - 2d \log \lambda_j + \log \xi_j, \quad j = 1, \dots, m.$$

For  $\widehat{d}_{gph}$ , Hurvich et al. (1998) show that

$$\sqrt{m \frac{24}{\pi^2}} \left( \widehat{d}_{gph} - d \right) \xrightarrow{d} \mathcal{N}(0, 1),$$

again under weak stationarity of the short-memory component.

The robustness properties of these inferential procedures is analyzed in small samples in the following subsection, and analytical results are provided in Subsection 3.3.

### 3.2 Experimental evidence

The examined null hypotheses are  $d_0 \in \{-0.2, -0.1, 0, 0.1, 0.2\}$ , and the true fractional integration parameter  $d$  is taken to belong to  $\{-0.2, -0.1, 0, 0.1, 0.2\}$  as well, leading to 25 possible combinations. The ALM test examined is representative for the regression-based tests; for the entire class, one could restrict oneself to  $d_0 = 0$ , as the test is based on differences under the null. But all null hypotheses are looked at: slight differences in the behavior of the ALM test may appear, since the fractional difference operator is sample-size dependent in finite samples. To focus on the effects of nonstationary volatility, we let the short-memory component  $u_t$  be uncorrelated via independence of the Gaussian standardized innovations  $\epsilon_t$ .

Heteroskedasticity is incorporated in the Monte Carlo study by suitably choosing  $\sigma_t$ . We study four variance patterns: a constant one,  $\sigma_t^2 = 1$ , as benchmark, a break in the variance,  $\sigma_t^2 = 1 + 4\mathbb{I}(t > T/2)$  and a linear trend in the variance, all three obeying Assumption 1, as well as a variance pattern switching periodically between 1 and 5 with period  $S = 2$ ,  $\sigma_t^2 = 3 + 2(-1)^t$ , as a counter-example.

The considered sample size is  $T = 500$ . The samples were actually generated with  $T+200$  observations, of which the first 200 were then discarded; a number of 10 000 Monte Carlo replications were run for each studied case. All tests are two-sided. The ALM test applied to the differences  $x_t = \Delta^{d_0} y_t$  with a lag order choice of  $p = 4[(T/100)^{0.25}]$ , for the modified R/S statistic relying on a long-run variance estimator with the quadratic spectral kernel and bandwidth  $b = 4[(T/100)^{0.25}]$ . The local Whittle test is computed with deterministic bandwidth  $m = 0.25 T^{0.8}$  and corrected variance

$$\left( 4 \sum_{j=1}^m \left( \ln \lambda_j - \frac{1}{m} \sum_{j=1}^m \ln \lambda_j \right)^2 \right)^{-1}, \quad (3)$$

as discussed in Robinson's (1995a) derivations; it is asymptotically equivalent to  $1/4m$ , but performs better in small samples. See Hurvich and Chen (2000) for the examination of a variant of (3) using  $2 \sin \lambda_j/2$  instead of  $\lambda_j$ . Finally, the GPH estimator is computed with bandwidth  $m = T^{0.79}$ . The results are reported in Table 1.

When there is no unconditional heteroskedasticity, we learn that the tests based on the local Whittle and the GPH estimators can be oversized, with the LW test having a rejection frequency between 6% and 7%, and the GPH test rejecting between 7% and 8% of the times. In contrast, the ALM and the modified R/S tests hold size. In terms of power, the GPH test dominates clearly, which is mostly due to the larger bandwidth. The local Whittle test is slightly more powerful than the ALM test, perhaps due to the overrejection under the null. Finally, the power of the ALM test is higher against alternatives of lower persistence (the asymmetry of the power function of many LM tests has been observed

Table 1: Size and power of fractional integration tests without adjustment for deterministic for different variance patterns in the innovations

$\sigma_t$	$d_0$	$d = -.2$				$d = -.1$				$d = 0$				$d = .1$				$d = .2$			
		ALM	RS	LW	GPH	ALM	RS	LW	GPH	ALM	RS	LW	GPH	ALM	RS	LW	GPH	ALM	RS	LW	GPH
co	-.2	5.4	5.2	6.7	7.2	13.3	17.2	16.3	45.3	32.1	39.3	51.1	92.8	51.9	62	83.9	99.9	63.1	78.5	97.1	100
	-.1	14.3	7.7	19.1	42.2	4.9	4.6	6.4	7.5	13.1	17.5	15.5	46.1	33.6	40.3	51.2	93.3	52.4	62.3	83.8	99.9
	0	46.1	23.7	54.4	93.7	15.6	7.5	20.8	43.6	4.9	4.8	6.2	7.2	13.1	17.9	16	45.8	35.4	41.6	51.5	93.2
	.1	83.4	54	87.2	100	46.7	24.4	55.3	94.2	15.6	7.9	21.3	43.4	4.8	4.9	6.3	6.9	14.1	18.1	16.8	46.4
	.2	97.8	81.6	98.5	100	83.5	53.7	87.7	100	47.5	24.2	55.4	94.3	14.8	7.9	20.3	42.7	4.8	4.5	6.2	7.7
br	-.2	5	4.5	10.3	11.1	10	14.4	20.1	46.2	24	34.7	50.5	91.1	37.8	55.2	80	99.7	46.4	70.8	95.4	100
	-.1	12	6.4	23.1	43.5	5.1	4.6	10.4	10.5	10.6	15.1	20.3	45.9	23.4	34	50.8	91.1	38.7	55.5	81.6	99.6
	0	36.9	18.2	55.9	91.8	12.5	6.4	24.4	44.1	5.1	4.5	10.1	10.5	10.1	14.7	20.2	47.1	24	35.3	50.6	91.6
	.1	70.2	37.4	85.6	99.9	35.6	17.5	55.4	91.6	13.2	6.2	25.4	44.8	5.1	4.5	9.9	9.8	10.1	15.3	19.8	46.9
	.2	92.5	60.9	97.6	100	69	37.9	85	99.9	36	18.2	56.9	92.2	11.8	6.1	22.4	43.5	4.9	4.2	10.1	10.6
tr	-.2	5.5	4.6	9.3	10.2	9.3	14.2	19.5	46.3	24.2	34.8	50.5	91.5	38.4	54.5	80.8	99.7	47.7	72.6	95.6	100
	-.1	13.1	7.2	22.2	43.7	5.3	4.3	9.6	10.3	10	13.7	18.6	46.4	24.2	34.3	49.7	91.5	37.9	55.6	81.1	99.7
	0	36.7	17.9	53.8	91.7	12.9	6.7	23.2	43.5	5.4	4.6	9.5	9.6	9.6	13.9	19.5	47.3	22.9	34.3	50.2	91.9
	.1	70.4	37.7	84.8	99.9	37.7	18.2	56	92.3	12.9	6.9	23.3	43.2	4.8	4.5	9.5	10	9.9	15	19.7	45.7
	.2	91.7	60.5	97.7	100	70.2	38.8	85.1	99.9	37.8	18.7	56.2	92	13.4	6.8	23.8	44.5	5	4.3	9.2	10.7
pe	-.2	5.5	5.6	6.5	9.2	13.8	17.9	17.5	47.5	34.7	40.4	50.9	92.7	54	62	83.1	99.8	66.4	78.9	96.7	100
	-.1	12.6	8.6	20	41.9	5.2	5.2	6.2	8.7	13.4	16.5	15.7	45.3	34.8	41.2	50.9	93	54.9	62.2	84.4	99.7
	0	39.6	24.7	55	92.4	14.1	9.1	20.7	43.2	4.6	4.5	5.7	8.9	13.4	17.2	16.3	47.1	34.8	42.3	51.4	92.8
	.1	72.9	49.3	87.3	99.9	39.8	25	54.7	92.9	13.4	8.6	20.9	43.2	5.1	5.2	6.3	9	13.6	18.1	17.1	47.5
	.2	92.9	71.1	98.6	100	73	50	88.1	99.9	40	25.1	56.7	92.6	13.8	8.8	21.1	42.9	4.8	4.9	6.2	8.9

11

**Note:** The nominal size is 5% and the variance patterns are denoted as co: constant, br: break, tr: linear trend, pe: periodic changes. The ALM test is computed with  $p = 4[(T/100)^{0.25}]$  and White standard errors, and the modified range-scale statistic uses as long-run variance estimator the periodogram at the origin smoothed with the quadratic spectral kernel and a bandwidth of  $p = 4[(T/100)^{0.25}]$ . The local Whittle test is computed with bandwidth  $m = 0.25T^{0.8}$  and variance as in (3), while the test based on the GPH estimator uses as bandwidth  $m = T^{0.79}$  and the asymptotic variance. The series  $y_t$  are fractional white noise integrated of order  $d$  with variance-modulated shocks, and the sample size is  $T = 500$ . For further details, in particular regarding the variance patterns, see the text.

before), whereas the exact opposite is observed for the modified R/S statistic, which can be explained by the different behavior of the long-run variance estimator under the respective alternatives (Teverovsky et al., 1999).

The picture changes dramatically for the local Whittle and the log-periodogram regression based tests under time-varying variance. Under the variance patterns with break or trend, their size can be as high as 12%, and is never below 9%. In contrast, the size of the ALM test is practically not affected by time-varying variance, in fact it is in most cases closer to the nominal level of 5% than in the homoskedastic case. The rejection frequencies under the considered alternatives do not change significantly when compared to the unconditionally homoskedastic case, with the exception of the ALM test for which the power drops a bit, compared to the unconditionally homoskedastic case. (Thus, the liberality does not even lead to higher rejection frequencies under the alternative.)

For the periodic pattern, the size of the LW test is slightly closer to the nominal 5% than in the benchmark case, while the GPH test reacts more sensitively. The empirical power of all tests does not change significantly either.

When the innovations are affected by heteroskedasticity, it appears that the ALM and the modified R/S tests are robust to heteroskedasticity, but the local Whittle test is robust to periodic heteroskedasticity only (the local Whittle estimator apparently confuses such patterns with conditional heteroskedasticity to some extent). The difference lies presumably in the “persistence” of the pattern, the locally homoskedastic one being still globally heteroskedastic, whereas for the periodic pattern fluctuations average out as  $t$  increases. This does not appear to hold for the GPH based test, at least not for this sample size.

We also examined the four tests under a DGP where the persistent process itself is modulated, rather than the shocks  $\varepsilon_t$ ; see Table 2.

As suggested by the arguments of Appendix A, the picture is essentially the same as in Table 1, up to Monte Carlo variability. There is one important difference, however: in the case of periodic variance changes not covered by 2, all four tests are seriously oversized for  $d_0 = d < 0$ . This puzzling fact has a simple explanation: the process can be written as the sum of two components, both integrated of negative order, but with one only being observed every second period. It is known (see e.g. Hassler, 2011) that overdifferenced processes become integrated of order zero upon skip sampling, shifting the DGP towards the alternative, so the tests reject as long as  $d_0 = d < 0$ .

For the local Whittle and the GPH procedures, additional experimental evidence suggests that the estimator  $\hat{d}$  itself is still consistent under both local homoskedasticity and periodic heteroskedasticity, although its distribution is distorted under local homoskedasticity. The results are reported in Table 3 for  $T = 500$ ,  $T = 1000$ ,  $T = 1500$  and  $T = 2000$ .

The bias of the two estimators for  $d$  disclosed in Table 3 are quite small, in fact the

Table 2: Size and power of fractional integration tests without adjustment for deterministic for different variance patterns in the levels

$\sigma_t$	$d_0$	$d = -.2$				$d = -.1$				$d = 0$				$d = .1$				$d = .2$			
		ALM	RS	LW	GPH	ALM	RS	LW	GPH	ALM	RS	LW	GPH	ALM	RS	LW	GPH	ALM	RS	LW	GPH
co	-.2	5.2	5.1	6.3	7.3	13.5	17.6	16.7	46.8	33.4	39.6	51.4	93.3	53	64.1	83.9	99.9	63.4	78.6	96.9	100
	-.1	14.7	8.1	19.6	42.7	4.9	4.8	5.9	7	12.7	16.4	15.9	45	34.4	40.9	51.2	93.1	53.7	63	83.9	99.9
	0	46.8	23.8	54.7	93.4	15	7.7	20.5	43.5	5	4.7	6.3	7.3	13.1	17	16	46	34.2	40.9	50.5	93.1
	.1	82.9	52.7	87.4	100	46.8	24.5	55	93.7	15.7	8.3	21.1	44.2	4.8	4.4	6.1	7.7	13.7	17.6	16.6	46.4
	.2	98.1	81	98.5	100	83.8	54.9	88.6	100	48	24.4	56.3	93.8	14.9	7.8	20.5	44.1	5.4	4.9	6.6	8.1
br	-.2	4.9	4.4	9.7	10.1	10.2	15.8	20	46.5	23.3	34	50	91.8	39.6	56.6	80.9	99.6	49.6	74	95.4	100
	-.1	12.5	6.4	22.9	43.2	4.7	4.1	9.7	10.4	10.4	15.1	19.8	47.1	25.1	35.9	50.1	91.3	41.1	58	82.1	99.7
	0	34.7	16.8	54	91.9	12.7	6.4	23.6	43.9	5.4	4.4	10	10.6	10.4	15.7	20.3	46.5	26.2	37.8	51.4	92.5
	.1	69	36.7	84.7	99.9	35.9	17.8	55.4	91.8	12.5	6.3	24.2	43.6	5	4.6	10.3	10.5	11	16.2	20.5	47.6
	.2	91.5	59.5	97.1	100	68.5	37.3	85.7	99.8	35.3	17.2	56.7	92.1	11.9	6.6	23.2	44.6	4.9	5.1	9.4	10.3
tr	-.2	5.3	4.1	9.4	10.3	9.2	13.1	19.4	46.1	23.5	34	51.5	91.4	39.7	56.5	81	99.6	51.2	74.1	95.7	100
	-.1	13.5	7	22.2	43.2	4.5	3.7	9.1	10.2	9.3	14.6	18.3	46.1	25.5	36.5	50.9	91.8	43.7	60.8	81.6	99.8
	0	37.8	19.4	54.3	91.8	13.9	6.9	23.8	44.7	4.8	4.4	8.9	10.4	10.1	15.7	19.1	46.8	26.9	39.3	51.4	91.4
	.1	71	38.9	84.9	99.9	37.6	18.7	55.7	92.3	12.7	6.8	23.2	43.2	5.4	5.2	9.5	10.1	11.7	18.3	19.4	46.6
	.2	92.3	60.2	97.6	100	70.8	39.1	85.8	99.9	38.2	18.8	56.9	92.6	13.4	6.6	23.5	44.8	5.3	5.5	9.3	10.1
pe	-.2	21.2	25.5	20.8	50.7	23.7	28.4	30.2	71.9	34.1	40.1	50.4	92.6	52.7	58.2	79.2	99.5	69	76.4	95.2	100
	-.1	6.4	8	6	8.9	7.2	9.1	7.2	16.7	12.8	16.9	16	46.7	31.2	36	42.7	85.5	55.3	59.8	77.8	99
	0	10.5	6.2	15.7	39.5	7.8	5.5	11	20.1	4.7	4.3	6.1	8.9	11.5	14.4	11.9	31.4	34.5	37.1	41.1	80.1
	.1	34.6	20.6	48.3	91.4	26.9	16.3	38.2	77.8	13.6	8.9	20.7	42.2	5.3	4.1	7.2	11.2	12.7	14.6	11	23.4
	.2	68.4	43.7	84.2	99.9	60.5	38.8	77.1	99.1	41	25.1	56	92.9	16.5	10.5	26.5	59.5	5.3	4.4	7.4	14.8

**Note:** The process itself rather than the shocks are modulated. See otherwise Table 1.

Table 3: Mean ( $m$ ) and standard deviation ( $s$ ) of the local Whittle and log-periodogram regression estimator for different variance patterns in the innovations

$\sigma_t$	$T$	$d = -.2$				$d = -.1$				$d = 0$				$d = .1$				$d = .2$			
		$m_{lw}$	$s_{lw}$	$m_{gph}$	$s_{gph}$	$m_{lw}$	$s_{lw}$	$m_{gph}$	$s_{gph}$	$m_{lw}$	$s_{lw}$	$m_{gph}$	$s_{gph}$	$m_{lw}$	$s_{lw}$	$m_{gph}$	$s_{gph}$	$m_{lw}$	$s_{lw}$	$m_{gph}$	$s_{gph}$
co	500	-0.21	0.101	-0.2	0.060	-0.11	0.102	-0.1	0.061	-0.01	0.101	0.0	0.061	0.09	0.102	0.10	0.061	0.19	0.105	0.20	0.061
	1000	-0.20	0.073	-0.2	0.045	-0.10	0.074	-0.1	0.045	-0.01	0.073	0.0	0.045	0.09	0.073	0.10	0.044	0.20	0.072	0.20	0.045
	1500	-0.20	0.060	-0.2	0.037	-0.10	0.060	-0.1	0.038	0.0	0.060	0.0	0.038	0.10	0.060	0.10	0.038	0.20	0.061	0.20	0.038
	2000	-0.20	0.053	-0.2	0.033	-0.10	0.052	-0.1	0.033	0.0	0.052	0.0	0.034	0.10	0.053	0.10	0.033	0.20	0.053	0.20	0.034
br	500	-0.21	0.117	-0.2	0.068	-0.11	0.117	-0.1	0.067	-0.01	0.118	0.0	0.067	0.09	0.116	0.10	0.066	0.19	0.114	0.20	0.067
	1000	-0.21	0.088	-0.2	0.051	-0.11	0.090	-0.1	0.051	-0.01	0.089	0.0	0.052	0.09	0.090	0.10	0.051	0.19	0.089	0.20	0.052
	1500	-0.20	0.076	-0.2	0.044	-0.11	0.078	-0.1	0.045	-0.01	0.076	0.0	0.045	0.10	0.076	0.10	0.044	0.20	0.076	0.20	0.045
	2000	-0.20	0.068	-0.2	0.040	-0.11	0.067	-0.1	0.040	-0.01	0.068	0.0	0.040	0.09	0.066	0.10	0.040	0.20	0.068	0.20	0.040
tr	500	-0.21	0.113	-0.2	0.066	-0.11	0.113	-0.1	0.066	-0.01	0.113	0.0	0.066	0.09	0.114	0.10	0.066	0.19	0.114	0.20	0.066
	1000	-0.21	0.086	-0.2	0.051	-0.11	0.086	-0.1	0.051	-0.01	0.086	0.0	0.051	0.09	0.087	0.10	0.050	0.19	0.087	0.20	0.051
	1500	-0.20	0.073	-0.2	0.044	-0.11	0.073	-0.1	0.044	-0.01	0.073	0.0	0.044	0.10	0.074	0.10	0.044	0.20	0.073	0.20	0.044
	2000	-0.20	0.065	-0.2	0.039	-0.10	0.066	-0.1	0.040	-0.01	0.065	0.0	0.039	0.10	0.065	0.10	0.039	0.20	0.066	0.20	0.039
pe	500	-0.20	0.104	-0.2	0.063	-0.11	0.101	-0.1	0.063	-0.01	0.102	0.0	0.063	0.09	0.104	0.10	0.063	0.19	0.103	0.20	0.064
	1000	-0.20	0.073	-0.2	0.044	-0.11	0.073	-0.1	0.045	-0.01	0.074	0.0	0.045	0.10	0.073	0.10	0.045	0.20	0.074	0.20	0.045
	1500	-0.20	0.061	-0.2	0.038	-0.10	0.061	-0.1	0.037	-0.01	0.060	0.0	0.038	0.10	0.061	0.10	0.038	0.20	0.061	0.20	0.037
	2000	-0.20	0.054	-0.2	0.034	-0.10	0.052	-0.1	0.033	0.0	0.052	0.0	0.033	0.10	0.053	0.10	0.033	0.20	0.053	0.20	0.033

**Note:** The sample sizes vary from  $T = 500$  to  $T = 2000$ . See otherwise Table 1

GPH estimator is practically unbiased. (This is not really surprising, given that there is no short-run dynamics.) Moreover, the bias, small as it is, decreases as  $T$  increases from 500 to 2000, as does the variance of the estimators. Most of the distortions of the LW test in the benchmark homoskedastic case (see Table 1) can be traced back to the bias of the local Whittle estimator, while the distortions of the GPH test are apparently due to the imperfect asymptotic approximation of its variance (its bias being practically 0). The additional distortions due to heteroskedasticity (for the break and the trend patterns) are most likely caused by the inflated variance of the estimator. For the periodic pattern, bias and variance are back at the values in the benchmark case (this suggests that it is the shape of the small-sample distribution of the GPH estimator that induces the oversizedness).

### 3.3 Limiting results

In the case of regression-based tests, the robustness illustrated in the previous subsection is easily explained for the case of square summable  $\psi_j$ . The asymptotic distribution of the  $t$  statistic is standard normal due to a martingale difference central limit theorem, and White standard errors used to capture serial dependence in the squares of  $\varepsilon_t$ , capture unconditional heteroskedasticity as well. See Phillips and Xu (2006) for an analysis of heteroskedastic, finite-order, autoregressions. This is shown in the following Proposition, which is intimately related to the result of Kew and Harris (2009).<sup>1</sup>

**Proposition 3** *Under the null hypothesis  $d = d_0$  and Assumptions 1 and 2, it holds as  $T \rightarrow \infty$  for the  $t$  statistic with White standard errors  $t_\phi$  from regression (2) that*

$$t_\phi \xrightarrow{d} \mathcal{N}(0, 1),$$

*provided that  $\sum_{j \geq 0} \psi_j^2 < \infty$ .*

**Proof:** *See Appendix C.*

**Remark 1** *The extension of the proposition to deterministic components is straightforward as outlined in Proposition 4 of Demetrescu et al. (2008). The extension for  $p \rightarrow \infty$ , or for filter coefficients  $\psi_j$  based on estimated fractional integration parameters arguably holds as well.*

Note, however, that White standard errors do not work always: the Dickey-Fuller test is one such case (cf. Demetrescu, 2010). More generally, when the asymptotic distribution

---

<sup>1</sup>We consider a triangular array structure of the modulated innovations, while Kew and Harris (2009) consider sequences of (fixed) heterogenous variances; the differences are mainly technical in nature but we include the result to keep this paper self-contained.

of the test statistic is expressed as a functional of (fractional) Brownian motions, as can be the case with particular choices of  $\psi_j$  for the fractional Dickey-Fuller test (Dolado et al., 2002), White standard errors fail in accounting for unconditional heteroskedasticity. For the same reason, distortions appear for the fractional cointegration test of Łasak (2010) as well, which is not the case for the regression-based tests for cointegration due to Avarucci and Velasco (2009) and Demetrescu et al. (2013).

The behavior of the modified R/S statistic depends under time-varying volatility on whether deterministic components are accounted for or not. Concretely, if no adjustment is necessary, the limiting distribution of the modified R/S statistic is invariant to the variance profile, whereas, under e.g. demeaning, it is not.

**Proposition 4** *Under the null hypothesis  $d = d_0$  and Assumptions 1 and 2, it holds as  $T \rightarrow \infty$  for the R/S statistic without demeaning that*

$$Q \xrightarrow{d} \max_{s \in [0,1]} W(s) - \min_{s \in [0,1]} W(s)$$

whereas under demeaning

$$Q \xrightarrow{d} \max_{s \in [0,1]} H(s) - \min_{s \in [0,1]} H(s)$$

with  $H(s) = W(\eta(s)) - sW(1)$ .

**Proof:** See Appendix C.

So the modified R/S statistic could only be used to robustly test the persistence of series that do not require adjustment for deterministic components. This is either the case when the mean function of the differenced series is known, or when the differenced deterministic component is zero or negligible. If e.g. the null hypothesis is for instance  $d = 1$ , a constant non-zero mean of the series is differenced away; and it can be shown that, as long as  $d > 0.5$  under the null, a constant non-zero mean is negligible after differencing so one can apply the R/S test without worrying about the differenced deterministic components.

In the case of the log-periodogram or the local Whittle estimator, the analysis is handicapped by the fact that the spectrum of  $y_t$  is not properly defined, since  $y_t$  is not weakly stationary, not even for  $d < 0.5$ . But examining the periodogram, which can still be computed, the *rate* at which it vanishes or explodes at the origin is not affected; see the following Proposition.

**Proposition 5** *Under Assumptions 1 and 2, it holds for  $y_t$  from (1) with uniformly bounded 4th order moments of  $\epsilon_t$  that*

$$\frac{T^{-2\alpha} I_y(\lambda)}{\lambda^{-2d}} = O_p(1)$$



as  $T \rightarrow \infty$  where the  $O_p(1)$  term is uniform for  $\lambda \in (0, \bar{\lambda})$  with fixed  $\bar{\lambda} > 0$ .

**Proof:** See Appendix C.

Since a multiplicative factor  $T^{-2\alpha}$  does not affect the minimum of the local Whittle approximation or the log-periodogram regression, the proposition would explain the apparent consistency of the LW and GPH estimators. When modulating the process, however, the formulae for their variance are obviously not delivering the right numbers anymore; see Table 3. It is as if the wrong bandwidth were used with Equation (3); one could say that the distortions in the asymptotic distribution appear because the *effective* bandwidth changes due to the distortion of the time scale implied by the modulation.

**Proposition 6** *Under Assumptions 1 and 2, it holds for the log-periodogram regression estimator*

$$\sqrt{m \frac{24}{\pi^2}} (\hat{d}_{gph} - d) \xrightarrow{d} \mathcal{N}(0, \int_0^1 \sigma_t^2 dt / \sigma^2).$$

**Proof:** See Appendix C.

## 4 Concluding remarks

The paper discussed modeling and inference for long memory time series under unconditional heteroskedasticity.

Variance modulation has an effect on invariance principles for cumulative sums under Assumption 1, as is well-known in time series econometrics since the work of Cavaliere (2004). In the seasonal case, however, the limit is not affected (Burridge and Taylor, 2001). It is tempting to conjecture that, just as in the integer integration case, the limit in the fractional case is a time-transformed fractional Brownian motion. Dealing with this topic, however, goes beyond the scope of the paper and is left for further research.

It was found using Monte Carlo simulations that the ALM test is robust, as are related regression-based tests under weak additional conditions, but the tests based on the local Whittle or the log-periodogram estimators are not. The modified range/scale statistic is robust only if no adjustment for deterministic components is required after differencing under the null. Moreover, the paper provided asymptotic arguments as to why robustness is given for the time-domain but not for the frequency-domain tests.

For robustifying the modified range/scale statistic in the case where adjustment for deterministic components is required, one could resort to the solutions considered for unit root tests, since the limiting distribution is expressed just like for the latter, in terms of time-transformed Wiener processes. For robustifying frequency-domain tests, the wild

bootstrap suggests itself for getting a correct estimate of the variance of the estimators. Demodulation of the series (Cavaliere and Taylor, 2008) prior to estimating the long memory parameter would provide an alternative, but would require an estimation of the variance profile. Finally, inference based on averaged local periodogram as developed for locally stationary series would apply immediately, but at the cost of an additional bandwidth parameter. The question, which robustification strategy would be preferable, is currently under investigation.

## Appendix

### A Asymptotic equivalence of DGPs

#### A.1 Regression-based tests

We distinguish two situations. First, we show that modulation and filtration are exchangeable in the short memory component.

Denote by  $t_\phi$  the regression-based test statistic computed with  $x_t \equiv u_t$  under the null, and  $t_{\tilde{\phi}}$  the one computed with  $\tilde{x}_t \equiv \tilde{u}_t$ . Then  $t_{\tilde{\phi}} = t_\phi + o_p(1)$ .

Note that we may set  $\alpha = 0$  because  $T^\alpha$  cancels out in the expression of the t statistic. Note also that, as a consequence, outliers caused by discontinuities of  $\sigma$  do not have influence increasing with  $T$  and can be neglected.

With the notation of the proof of Proposition 3 we have to show that

$$\frac{1}{\sqrt{T}} \sum_{t=p+1}^T \tilde{\mathbf{x}}_{t-1} \tilde{x}_t - \frac{1}{\sqrt{T}} \sum_{t=p+1}^T \mathbf{x}_{t-1} x_t \xrightarrow{p} 0$$

and

$$\frac{1}{T} \sum_{t=p+1}^T \tilde{\mathbf{x}}_{t-1} \tilde{\mathbf{x}}'_{t-1} - \frac{1}{T} \sum_{t=p+1}^T \mathbf{x}_{t-1} \mathbf{x}'_{t-1} \xrightarrow{p} 0.$$

To this end note that the first difference is given by

$$\frac{1}{\sqrt{T}} \sum_{t=p+1}^T (\tilde{\mathbf{x}}_{t-1} - \mathbf{x}_{t-1}) x_t + \frac{1}{\sqrt{T}} \sum_{t=p+1}^T \mathbf{x}_{t-1} (\tilde{x}_t - x_t) + \frac{1}{\sqrt{T}} \sum_{t=p+1}^T (\tilde{\mathbf{x}}_{t-1} - \mathbf{x}_{t-1}) (\tilde{x}_t - x_t).$$

Let us examine the first term. Then, the Cauchy-Schwarz inequality leads for the first term and any  $j = 1, \dots, p$  to

$$\left| \frac{1}{\sqrt{T}} \sum_{t=p+1}^T (\tilde{x}_{t-j} - x_{t-j}) x_t \right| \leq \sqrt{\sum_{t=p+1}^T (\tilde{x}_{t-j} - x_{t-j})^2} \frac{1}{T} \sum_{t=p+1}^T x_t^2 \quad (4)$$

where  $\frac{1}{T} \sum_{t=p+1}^T x_t^2 = O_p(1)$ , as well as

$$\left| \frac{1}{\sqrt{T}} \sum_{t=p+1}^T (\tilde{x}_{t-1}^* - x_{t-1}^*) x_t \right| \leq \sqrt{\sum_{t=p+1}^T (\tilde{x}_{t-1}^* - x_{t-1}^*)^2 \frac{1}{T} \sum_{t=p+1}^T x_t^2} \quad (5)$$

For (4) and any  $j = 1, \dots, p$ , we have with Markov's inequality that  $\sum_{t=p+1}^T (\tilde{x}_{t-j} - x_{t-j})^2 = o_p(1)$  since  $\mathbb{E} \left( \sum_{t=p+1}^T (\tilde{x}_{t-j} - x_{t-j})^2 \right) \leq \sum_{t=p+1}^T \mathbb{E} ((\tilde{x}_{t-j} - x_{t-j})^2) \leq CT^{-1}$  due to Proposition 1, while for (5) it follows with the Cauchy-Schwarz inequality that

$$|\tilde{x}_{t-1}^* - x_{t-1}^*| \leq \sum_{j=1}^{t-1} |\psi_j| |\tilde{x}_{t-j} - x_{t-j}| \leq \sqrt{\sum_{j=1}^{t-1} \psi_j^2 \sum_{j=1}^{t-1} (\tilde{x}_{t-j} - x_{t-j})^2}$$

so

$$\mathbb{E} \left( (\tilde{x}_{t-1}^* - x_{t-1}^*)^2 \right) \leq \sum_{j=1}^{t-1} \psi_j^2 \sum_{j=1}^{t-1} \mathbb{E} ((\tilde{x}_{t-j} - x_{t-j})^2) = o(T^{-1})$$

thanks to the square summability of the series  $\psi_j$ . It follows with Markov's inequality that  $\sum_{t=p+1}^T (\tilde{x}_{t-1}^* - x_{t-1}^*)^2 = o_p(1)$  as required.

The remaining relations are established along the same lines, and so does the White expression for standard errors.

When it comes to exchangeability of long memory filtration and modulation, we note that Proposition 2 does not offer a tight enough bound on the difference between the two processes. In fact, the discussion in Hassler and Breitung (2006) suggests that the tests can't be asymptotically equivalent. If letting  $p \rightarrow \infty$  at suitable rates, however, Demetrescu et al. (2013) show that the difference between  $x_t$  and  $\tilde{x}_t$  is approximated away; we do not go into details here. The simulations in Section 3.2, where  $p$  is chosen as  $O(T^{1/4})$  for the ALM test (with  $\psi_j = j^{-1}$ ), support the intuition, however.

## A.2 The R/S statistic

Just like for regression-based tests, short memory processes pose no difficulty. To see why, assume that Proposition 1 applies and note that it suffices to show that

$$\frac{1}{\sqrt{T}} \sup_t |S_t - \tilde{S}_t| \xrightarrow{p} 0$$

with  $\tilde{S}_t = \sum_{j=1}^t \tilde{u}_j$  under the null. We have for all  $t < T$  that

$$|S_t - \tilde{S}_t| \leq \sum_{j=1}^t |u_j - \tilde{u}_j| \leq \sum_{t=1}^T |u_t - \tilde{u}_t|;$$

since  $\|u_t - \tilde{u}_t\|_1 \leq \|u_t - \tilde{u}_t\|_2$ , it follows with Markov's inequality that  $\sum_{t=1}^T |u_t - \tilde{u}_t| = O_p(T \|u_t - \tilde{u}_t\|_2)$  and consequently that

$$\frac{1}{\sqrt{T}} \sup_t |S_t - \tilde{S}_t| = O_p(T^{-0.5})$$

as required.

Note however that the bounds delivered by Proposition 2 do not suffice, and the result does not appear to be improvable ; see the proof of Proposition 2 below. Thus, it likely makes a difference whether a process is modulated then integrated of order  $d$  or the other way round, even if the simulations in Section 3.2 suggest that the difference is minor.

## B Long memory properties of modulated processes

Long memory is often defined in terms of the behavior of the spectrum of the process in the neighborhood of 0, or, equivalently, in terms of the asymptotic behavior of the autocovariance function. Neither is properly defined with nonstationary series as assumed here. On the other hand, looking at model (1), there is a strong tendency to say, ‘‘fractional filter, long memory.’’

Given the results in Propositions 1 and 2, we focus on modulated levels. Then we may state that  $E(\tilde{y}_t \tilde{y}_{t-h}) = E(\tilde{y}_t \tilde{y}_{t-h}) + o(1)$  as  $T \rightarrow \infty$ .

For now, we shall assume that  $\alpha=0$ , i.e. that the variance function  $\sigma_t^2$  is bounded. But if assuming the variance function to be bounded, we obtain immediately that

$$|\gamma_h| \sigma_{\min}^2 \leq \left| \gamma_h^{t,T}(y_t) \right| \leq |\gamma_h| \sigma_{\max}^2.$$

This implies that at least the decay rate of the autocovariance functions is not affected. So we could still speak of the same type of persistence in our situation. Similarly, one obtains for the spectrum that

$$f(\lambda) \sigma_{\min}^2 \leq f_{t,T}(\lambda) \leq f(\lambda) \sigma_{\max}^2.$$

Moreover, if allowing  $T \rightarrow \infty$ , some additional statements can be made about average spectra. (Nonstationary spectra have been discussed e.g. in Priestley, 1988, Chapter 6.) And what about neighboring autocovariances (spectra)? We have namely that

$$\gamma_h^{t+1,T}(y_t) - \gamma_h^{t,T}(y_t) = \gamma_h(\sigma_{t+1}\sigma_{t+1-h} - \sigma_t\sigma_{t-h}) = \gamma_h + O(T^{-1}),$$

at all continuity points of  $\sigma$  so the autocovariance sequence is changing smoothly enough (hopefully) to allow for average statements.

On the average, it holds that

$$\frac{1}{T} \sum_{t=h+1}^T \gamma_h^{t,T}(y_t) \rightarrow \gamma_h \int_{\kappa}^1 \sigma(s) \sigma(s - \kappa) ds$$

where  $\kappa = \lim h/T$ . This has the nice implication that the average autocorrelations of order  $h$ ,  $0 \leq h \leq H$ , converge to the stationary ones when  $H/T \rightarrow 0$ . And, more interestingly, the pseudospectrum considered as

$$\lim_{T, H, H/T \rightarrow 0} \sum_{-H}^H \gamma_h^{[sT], T}(y_t) \cos\{\lambda h\}$$

also has a limit proportional to  $f(\lambda)$  at all  $s$ . This allows us to state that the trending properties of the process are characterized by  $d$  alone and the modulation does not interfere with this property of the DGP.

## C Proofs

Before proving the main results of the paper, a preliminary lemma required for the proof of Proposition 3 is stated and proved.

**Lemma 1** *Under the assumptions of Proposition 3, it holds for some  $r > 2$  that*

$$a) \ E \left( |T^{-\alpha} \varepsilon_t|^{2r} \right) < C,$$

$$b) \ E \left( |T^{-\alpha} u_{t-j}|^4 \right) < C, \text{ and}$$

$$c) \ E \left( |T^{-\alpha} u_{t-1}^{**}|^4 \right) < C,$$

where  $u_{t-1}^{**} = \sum_{j \geq 1} \psi_j u_{t-j}$  and  $\sum_{j \geq 1} \psi_j^2 < \infty$ .

**Proof:** Item a) is obvious given the properties of  $\sigma_t$  and  $\varepsilon_t$ . Item b) is a particular case of c) with absolutely summable filtering; but item c) requires some more care. Note that  $u_{t-1}^{**} = \sum_{j \geq 1} \tilde{\psi}_j \varepsilon_{t-j-1}$  where the coefficients sequence  $\tilde{\psi}_j$  is the (square summable) convolution of  $\{\psi_j\}$  and  $\{b_j\}$ . Then, the finiteness of  $E \left( \left| \sum_{j \geq 0} \tilde{\psi}_j \frac{\sigma_{t-j}}{T^\alpha} \varepsilon_{t-j-1} \right|^4 \right)$  reduces to finiteness of  $E \left( \left| \sum_{j \geq 0} \tilde{\psi}_j \varepsilon_{t-j-1} \right|^4 \right)$ , since  $\frac{\sigma_{t-j}}{T^\alpha}$  is uniformly bounded. Note that  $\tilde{\psi}_j = O(j^{-0.5})$ ; a tedious, yet straightforward modification of the proof of Lemma 8 in Demetrescu et al. (2008) leads to the desired preliminary result.

## Proof of Proposition 1

a) Examine first the case of no discontinuities of  $\sigma(s)$ . For this proof only, let  $t$  be shorthand for  $[sT]$ , and note that

$$\begin{aligned} \|u_t - \tilde{u}_t\|_2 &= \left\| \sum_{j \geq 0} b_j \epsilon_{t-j} (\sigma_t - \sigma_{t-j}) \right\|_2 \\ &\leq \sum_{j \geq 0} |b_j| |\sigma_t - \sigma_{t-j}| \|\epsilon_{t-j}\|_2. \end{aligned}$$

Due to the assumed Lipschitz condition on  $\sigma(\cdot)$ ,  $\exists C$  such that  $|\sigma_t - \sigma_{t-j}| \leq CT^\alpha |j/T|$ , so

$$\|u_t - \tilde{u}_t\|_2 \leq CT^{\alpha-1} \sum_{j \geq 0} |jb_j|$$

as required (recall, the coefficients  $b_j$  are 1-summable). Let now there be exactly one discontinuity of  $\sigma(\cdot)$ , at  $\lambda < 1$ . There are two possibilities,  $t < [\lambda T]$  and  $t > [\lambda T]$  (the third,  $t = [\lambda T]$ , is excluded by assumption). If  $t < [\lambda T]$ , the result from the case with no discontinuity applies. If  $t > [\lambda T]$ , we have that

$$\|u_t - \tilde{u}_t\|_2 \leq \sum_{j=0}^{t-[\lambda T]} |b_j| |\sigma_t - \sigma_{t-j}| \|\epsilon_{t-j}\|_2 + \sum_{j \geq t-[\lambda T]+1} |b_j| |\sigma_t - \sigma_{t-j}| \|\epsilon_{t-j}\|_2.$$

The arguments used in the case with no discontinuity apply directly for the first summand on the r.h.s., while, for the second summand, we make use of the uniform boundedness of  $\sigma(\cdot)$  and of  $\|\epsilon_{t-j}\|_2$  to arrive at

$$\sum_{j \geq t-\lambda T+1} |b_j| |\sigma_t - \sigma_{t-j}| \|\epsilon_{t-j}\|_2 \leq CT^\alpha \sum_{j \geq t-\lambda T+1} |b_j|.$$

Note now that  $\sum_{j \geq p} |jb_j| \leq p \sum_{j \geq p} |b_j|$  leading due to 1-summability of  $b_j$  to  $\sum_{j \geq p} |b_j| = O(p^{-1})$ , and the r.h.s. of the above equation is thus bounded by  $\frac{CT^\alpha}{t-\lambda T+1}$ . The desired result follows since  $t = [sT]$ . One proceeds similarly if there is more than one discontinuity of  $\sigma(s)$  (but a finite number thereof).

## Proof of Proposition 2

Assume first that there are no discontinuity points and let again  $t$  be shorthand for  $[sT]$ . Then, having assumed a zero mean,

$$\|y_{[sT]} - \tilde{y}_{[sT]}\|_2^2 = \text{Var}(y_t - \tilde{y}_t) = \sum_{j=0}^{t-1} \sum_{k=0}^{t-1} \phi_j^{(d)} \phi_k^{(d)} \gamma_{|j-k|} (\sigma_t - \sigma_{t-j}) (\sigma_t - \sigma_{t-k}).$$

With the Lipschitz property,  $\phi_j^{(d)} \sim \frac{1}{\Gamma(d)} j^{-1-d}$ , and  $\gamma_h \sim \frac{\Gamma(1-d)}{\Gamma(d)} h^{2d-1}$ , we have that, for a suitable  $C$ ,

$$\text{Var}(y_t - \tilde{y}_t) \leq \frac{C}{T^2} \sum_{j=0}^{t-1} \sum_{k=0}^{t-1} j^{-d_1} k^{-d_1} |j - k|^{2d-1}.$$

Rearrange the sum terms to obtain that

$$\text{Var}(y_t - \tilde{y}_t) \leq \frac{C}{T^2} \sum_{h=0}^{t-1} \sum_{i=0}^{t-h-1} h^{-d_1} (h+i)^{-d_1} h^{2d-1}.$$

If  $d_1 > 0$ ,  $h^{-d_1} (h+i)^{-d_1} < h^{-2d_1}$ , if  $d_1 < 0$  we have that  $h^{-d_1} (h+i)^{-d_1} < (h+i)^{-2d_1}$ . In the first case, we have that

$$\begin{aligned} \text{Var}(y_t - \tilde{y}_t) &\leq \frac{C}{T^2} \sum_{h=0}^{t-1} (t-h) h^{2(d-d_1)-1} \\ &\leq \frac{C}{T} \sum_{h=0}^{t-1} h^{2(d-d_1)-1} \\ &\leq CT^{2(d-d_1)-1}. \end{aligned}$$

In the second, we have that

$$\text{Var}(y_t - \tilde{y}_t) \leq \frac{C}{T^2} \sum_{h=0}^{t-1} h^{2d-1} \sum_{i=0}^{t-h-1} (h+i)^{-2d_1}.$$

We further have that  $\sum_{i=0}^{t-h-1} (h+i)^{-2d_1} \leq C(t-h)(t-1)^{-2d_1}$ . Hence,

$$\text{Var}(y_t - \tilde{y}_t) \leq CT^{-2d_1-1} \sum_{h=0}^{t-1} h^{2d-1}.$$

With the known approximation, we have the same

$$\text{Var}(y_t - \tilde{y}_t) \leq CT^{2(d-d_1)-1}$$

as required for the result. Discontinuities are handled as in the proof of Proposition 1.

### Proof of Proposition 3

The proof is completed in two steps.

First, it is shown that employing  $x_{t-1}^{**} = \sum_{j \geq 1} \psi_j x_{t-j}$  as a regressor instead of its finite-sample counterpart  $x_{t-1}^* = \sum_{j=1}^{t-1} \psi_j x_{t-j}$  has no asymptotic effect on the distribution of the estimators. This is accomplished by noting that the variance of the differences between the

two vanishes at an appropriate rate. It holds namely for the differences  $\tilde{\delta}_t = \sum_{j \geq t} \psi_j x_{t-j}$  that

$$\frac{1}{T^\alpha} \left( \tilde{\delta}_t \right) = \sum_{j \geq t} \psi_j \frac{1}{T^\alpha} u_{t-j};$$

use now Proposition 1 to conclude that the variance is given, up to an  $O_p(T^{-1})$  term, by  $\sum_{j \geq t} \sum_{k \geq t} \psi_j \psi_k \frac{\sigma_{t-j} \sigma_{t-k}}{T^{2\alpha}} \gamma_{j-k}$ , where  $\gamma_{j-k}$  is the autocovariance of  $\left( \sum_{j \geq 0} b_j \varepsilon_{t-j} \right)$ . The autocovariance is exponentially decaying, and tedious, yet straightforward algebra yields that  $\tilde{\delta}_t = O_p(t^{-0.5})$ , which does not affect the asymptotics as shown e.g. in Demetrescu et al. (2008) Lemma 2 for the ALM test. So let the regressor vector  $\mathbf{x}_{t-1}$  contain  $x_{t-1}^{**} = \sum_{j \geq 1} \psi_j x_{t-j}$ , i.e.  $\mathbf{x}_{t-1} = (x_{t-1}^{**}, x_{t-1}, \dots, x_{t-p})'$ .

Second, let us examine the expression

$$\frac{1}{T^{0.5+2\alpha}} \sum_{t=p+1}^T \mathbf{x}_{t-1} \varepsilon_t, \quad (6)$$

the numerator (so-to-say) of the estimators in regression (2) and of the corresponding  $t$  statistics. Since  $\mathbf{x}_{t-1} \varepsilon_t$  is a martingale difference, a central limit theorem applies under regularity conditions, see below. Then, what is needed for the regression-based test to be unaffected is to use the correct standard errors when computing the  $t$  statistic for  $\phi$ . Using a central limit theorem from Davidson (1994, p. 383), sufficient conditions for multivariate asymptotic normality of (6) are

- a)  $\max_{p+1 \leq t \leq T} \frac{1}{T^{0.5+2\alpha}} |\mathbf{x}_{t-1} \varepsilon_t| \xrightarrow{p} 0$
- b)  $\frac{1}{T^{1+4\alpha}} \sum_{t=p+1}^T \mathbf{x}_{t-1} \mathbf{x}'_{t-1} \varepsilon_t^2 \xrightarrow{p} \Sigma$  where  $\Sigma$  is some fixed positive semi-definite  $(p+1) \times (p+1)$  matrix.

The key to robustness is that one uses precisely this matrix  $\Sigma$  (or rather a sample counterpart) for building test statistics with White standard errors.

- a) We have due to Markov's inequality for some  $2 < r < 4$  and any positive  $\eta$

$$\begin{aligned} \Pr \left( \max_{p+1 \leq t \leq T} \frac{1}{T^{0.5+2\alpha}} |\mathbf{x}_{t-1} \varepsilon_t| > \eta \right) &\leq \sum_{t=p+1}^T \Pr \left( |T^{-2\alpha} \mathbf{x}_{t-1} \varepsilon_t| > \eta T^{0.5} \right) \\ &\leq \sum_{t=p+1}^T \frac{\mathbb{E} \left( |T^{-2\alpha} \mathbf{x}_{t-1} \varepsilon_t|^r \right)}{(\eta T^{0.5})^r}. \end{aligned}$$

Apply now Hölder's inequality to show that  $\mathbb{E} \left( |T^{-2\alpha} \mathbf{x}_{t-1} \varepsilon_t|^r \right)$  is uniformly bounded in  $t$  if  $\mathbb{E} \left( |T^{-\alpha} \varepsilon_t|^{4r/(4-r)} \right)$  and  $\mathbb{E} \left( |T^{-\alpha} \mathbf{x}_{t-1}|^4 \right)$  are finite; this is indeed the case, see Lemma 1.

Then, the sum  $\sum_{t=p+1}^T \frac{\mathbb{E} \left( |T^{-2\alpha} \mathbf{x}_{t-1} \varepsilon_t|^r \right)}{(\eta T^{0.5})^r}$  is of order  $O(T^{1-r/2}) = o(1)$  for any  $\eta$ , as required for the central limit theorem.



b) This is only slightly more complicated. Let  $\tilde{\psi}_j = (\tilde{\psi}_j, b_j, \dots, b_{j-p})'$  with  $\{\tilde{\psi}_j\}$  the convolution of  $\{\psi_j\}$  and  $\{b_j\}$  and  $b_j = 0$  for  $j > 0$ . Then,

$$\begin{aligned} & \frac{1}{T^{1+4\alpha}} \sum_{t=p+1}^T \left( \sum_{j \geq 0} \tilde{\psi}_j \varepsilon_{t-j-1} \right) \left( \sum_{k \geq 0} \tilde{\psi}_k \varepsilon_{t-k-1} \right)' \varepsilon_t^2 \\ &= \frac{1}{T} \sum_{t=p+1}^T \sum_{j \geq 0} \sum_{k \geq 0} \tilde{\psi}_j \tilde{\psi}_k' \frac{\sigma_{t-k-1} \sigma_{t-j-1} \sigma_t^2}{T^{4\alpha}} \varepsilon_{t-k-1} \varepsilon_{t-j-1} \varepsilon_t^2. \end{aligned}$$

Adding and subtracting the unconditional expectation  $\kappa_{jkt} = \mathbb{E}(\varepsilon_{t-k-1} \varepsilon_{t-j-1} \varepsilon_t^2)$ , we obtain two components of the sum,

$$\frac{1}{T} \sum_{t=p+1}^T \sum_{j \geq 0} \sum_{k \geq 0} \tilde{\psi}_j \tilde{\psi}_k' \frac{\sigma_{t-k-1} \sigma_{t-j-1} \sigma_t^2}{T^{4\alpha}} (\varepsilon_{t-k-1} \varepsilon_{t-j-1} \varepsilon_t^2 - \kappa_{jkt}),$$

and

$$\frac{1}{T} \sum_{t=p+1}^T \sum_{j \geq 0} \sum_{k \geq 0} \tilde{\psi}_j \tilde{\psi}_k' \frac{\sigma_{t-k-1} \sigma_{t-j-1} \sigma_t^2}{T^{4\alpha}} \kappa_{jkt}.$$

For the first, note that the quotient  $\frac{\sigma_{t-k-1} \sigma_{t-j-1} \sigma_t^2}{T^{4\alpha}}$  is uniformly bounded, and we can use the absolute summability of 8<sup>th</sup> order cumulants and the arguments of Demetrescu et al. (2008, proof of Proposition 1) to show the first summand to disappear, while, for the second, we need to show that a limit exists. To this end, recall e.g. from Demetrescu et al. (2008) that

$$\frac{1}{T} \sum_{t=p+1}^T \sum_{j \geq 0} \sum_{k \geq 0} \tilde{\psi}_j \tilde{\psi}_k' \kappa_{jkt} \tag{7}$$

does have a proper limit. Knowing that  $\frac{\sigma_{t-k-1} \sigma_{t-j-1} \sigma_t^2}{T^{4\alpha}}$  is bounded, it would be no surprise if the limit existed indeed. To establish the existence of the limit, we first need to show that  $\sum_{j \geq 0} \sum_{k \geq 0} \tilde{\psi}_j \tilde{\psi}_k' \frac{\sigma_{t-k-1} \sigma_{t-j-1} \sigma_t^2}{T^{4\alpha}} \kappa_{jkt}$  converges to a function of  $t$ , say  $\phi(t)$ . This follows from (7) and the piecewise Lipschitz property of  $\sigma(\cdot)$ .

Then we show that

$$\frac{1}{T} \sum_{t=p+1}^T \phi(t)$$

has a proper limit. Note that this is a Riemann sum, which converges if the total variation of  $\phi$  on  $[0, 1]$  is bounded.

The average of  $\phi(t)$  will converge if  $\phi(t)$  varies slowly enough. Write  $\frac{\sigma_{t-k-1} \sigma_{t-j-1} \sigma_t^2}{T^{4\alpha}} = \sigma\left(\frac{t-k-1}{T}\right) \sigma\left(\frac{t-j-1}{T}\right) \sigma^2\left(\frac{t}{T}\right)$ . Due to the smoothness of  $\sigma$ ,  $\sigma\left(\frac{t-k-1}{T}\right) \sigma\left(\frac{t-j-1}{T}\right) \sigma^2\left(\frac{t}{T}\right) = \sigma\left(\frac{t-k-2}{T}\right) \sigma\left(\frac{t-j-2}{T}\right) \sigma^2\left(\frac{t-1}{T}\right) + o(1)$ , of course uniformly with the exception where  $\sigma$  has discontinuities. But  $\sigma$  is integrable, so the discontinuities won't matter in the end. Thus,

$\phi(t)$  has the same smoothness properties as  $\sigma(t)$ , and as such the average will converge.

It is easily shown along the lines of b) that

$$\frac{1}{T^{1+2\alpha}} \sum_{t=p+1}^T \mathbf{x}_{t-1} \mathbf{x}'_{t-1} \xrightarrow{p} \Sigma_{\mathbf{x}}$$

where the probability limit  $\Sigma_{\mathbf{x}}$  is a constant, positive definite, matrix; hence,

$$T^{0.5+\alpha} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, (\Sigma_{\mathbf{x}})^{-1} \Sigma (\Sigma_{\mathbf{x}})^{-1}).$$

But looking at the expression of the White standard errors, and at the probability limits of its components derived above, the heteroskedasticity-robust estimator of the covariance matrix of  $\widehat{\boldsymbol{\beta}}$  is seen to converge to precisely  $(\Sigma_{\mathbf{x}})^{-1} \Sigma (\Sigma_{\mathbf{x}})^{-1}$ , as required for the result.

## Proof of Proposition 4

The R/S statistic is given in the first case by

$$Q = \frac{1}{\sqrt{T}} \frac{\max_{1 \leq t \leq T} S_t - \min_{1 \leq t \leq T} S_t}{\widehat{\omega}}.$$

Under time-varying variance, the  $x_t = u_t$  is generated according to Assumptions 1 and 2. Letting  $u_t$  have for simplicity unity long-run variance, we have that

$$\frac{1}{\bar{\sigma} T^{0.5+\alpha}} S_{[sT]} \Rightarrow W(\eta(s)),$$

where  $\bar{\sigma}^2 = \int_0^1 \sigma^2(s) ds$ ,  $\eta(s) = \bar{\sigma}^{-2} \int_0^s \sigma^2(r) dr$  and  $W$  is a standard Wiener process;  $W(\eta(s))$  is a so-called time-transformed Wiener process; see Cavaliere (2004). Considering the linearity of  $\widehat{\omega}$  in  $T^\alpha$ , we can use Cavaliere (2004) to conclude that

$$\frac{1}{T^{2\alpha}} \widehat{\omega}^2 \xrightarrow{p} \bar{\sigma}^2.$$

Then,

$$Q \xrightarrow{d} \max_{s \in [0,1]} W(\eta(s)) - \min_{s \in [0,1]} W(\eta(s)).$$

For any given path of  $W$ , the value the extremum points does not change; it is just the location that depends on the variance profile. So the distribution of the modified RS statistic is not affected by the time transformation. Note too that, since convergence is weak, it is the limiting distribution of the modified RS statistic that is invariant to  $\eta$ , and not the statistic itself.

Considering the case with demeaning, prototypical for adjusting series for deterministic components, the situation changes dramatically. The partial sums are built over demeaned  $x$ s,

$$S_t = \sum_{j=1}^t (x_j - \bar{x})$$

and consequently

$$\frac{1}{\sigma T^{0.5+\alpha}} S_{[rT]} \Rightarrow W(\eta(s)) - sW(1).$$

Under global homoskedasticity, the limit is nothing else than the first-order Brownian bridge, say  $B_1(s)$ . The invariance from the case with no demeaning does not carry over, however, since, under heteroskedasticity, the above limit is not the time-transformed Brownian bridge:

$$B_1(\eta(s)) = W(\eta(s)) - \eta(s)W(1) \neq W(\eta(s)) - sW(1),$$

where  $W(\eta(s)) - sW(1)$  cannot be expressed as a time-transformed process.

## Proof of Proposition 5

Assume first for Assumption 1 that  $\alpha = 0$ , case in which  $\sigma_t$  is uniformly bounded. The proof then modifies the proof of Theorem 6.2.2 from Priestley (1981) suitably. Denote by  $\zeta_x(\lambda)$  the finite Fourier transform of some process  $x_t$ ,  $\zeta_x(\lambda) = 1/\sqrt{2\pi T} \sum_{t=1}^T x_t e^{-i\lambda t}$  for  $\lambda \in [-\pi, \pi]$ . The periodogram  $I_x(\lambda)$  of  $x_t$  is then given by  $I_x(\lambda) = \zeta_x(\lambda) \zeta_x^*(\lambda)$ , where  $*$  denotes the complex conjugate. Denote by  $\varphi_j^{(d)}$  the coefficients of the fractional integration filter with parameter  $d$  and  $|\Gamma_d(\lambda)|$  the corresponding transfer function. Recall that  $\varphi_j^{(d)} \leq C j^{d-1}$  and  $|\Gamma_d(\lambda)| \sim \lambda^{-d}$  as  $\lambda \rightarrow 0$ .

It shall be first shown that

$$\zeta_y(\lambda) = \Gamma_d(\lambda) \zeta_u(\lambda) + O_p(T^d),$$

where the  $O_p$  term is uniform in  $\lambda$ . To this end, note that

$$\zeta_y(\lambda) = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^T y_t e^{-i\lambda t} = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^T \left( \sum_{j=0}^{t-1} \varphi_j^{(d)} u_{t-j} \right) e^{-i\lambda t},$$

by rearranging the terms of the sum and reindexing correspondingly, one obtains

$$\begin{aligned}
\zeta_y(\lambda) &= \frac{1}{\sqrt{2\pi T}} \sum_{k=0}^{T-1} \varphi_k^{(d)} \left( \sum_{t=1}^{T-k} u_{t-j} e^{-i\lambda(t+k)} \right) \\
&= \left( \sum_{k=0}^{T-1} \varphi_k^{(d)} e^{-i\lambda k} \right) \left( \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^T u_{t-j} e^{-i\lambda t} \right) + R_T \\
&= \Gamma_d(\lambda) \zeta_u(\lambda) + R_T,
\end{aligned}$$

where  $R_T = \frac{1}{\sqrt{2\pi T}} \sum_{k=0}^{T-1} \varphi_k^{(d)} e^{-i\lambda k} \left( \sum_{t=T-k+1}^T u_{t-j} e^{-i\lambda t} \right)$ . In order to examine the behavior of  $R_T$ , note that, due to the short memory of  $u_t$  it holds that

$$\mathbb{E} \left| \sum_{t=T-k+1}^T u_{t-j} e^{-i\lambda t} \right|^2 \leq Ck$$

uniformly in  $\lambda$ . Then,

$$\begin{aligned}
\mathbb{E} |R_T| &\leq \frac{1}{\sqrt{2\pi T}} \sum_{k=0}^{T-1} \mathbb{E} \left| \varphi_k^{(d)} e^{-i\lambda k} \left( \sum_{t=T-k+1}^T u_{t-j} e^{-i\lambda t} \right) \right| \\
&\leq \frac{C}{\sqrt{T}} \sum_{k=0}^{T-1} |\varphi_k^{(d)}| \sqrt{\mathbb{E} \left( \sum_{t=T-k+1}^T u_{t-j} e^{-i\lambda t} \right)^2} \leq \frac{C}{\sqrt{T}} \sum_{k=0}^{T-1} |\varphi_k^{(d)}| \sqrt{k} \\
&= O_p(T^d).
\end{aligned}$$

Consider now

$$I_y(\lambda) = \zeta_y(\lambda) \zeta_y^*(\lambda) = (\Gamma_d(\lambda) \zeta_u(\lambda) + O_p(T^d)) (\Gamma_d(\lambda) \zeta_u(\lambda) + O_p(T^d))^*$$

at the harmonic frequencies  $\lambda_j = j/T$ . Note at this point that  $\zeta_u(\lambda) \leq C\zeta_{\sigma\epsilon}(\lambda)$  uniformly in  $\lambda$ , since the filter  $b_j$  is 1-summable; the finite Fourier transform  $\zeta_{\sigma\epsilon}(\lambda)$  of the modulated innovations,  $\sigma_t \epsilon_t$ , behaves itself as follows

$$\begin{aligned}
\mathbb{E} |\zeta_{\sigma\epsilon}(\lambda)|^2 &= \mathbb{E} \left[ \left( \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^T \sigma_t \epsilon_t e^{-i\lambda t} \right) \left( \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^T \sigma_t \epsilon_t e^{-i\lambda t} \right)^* \right] \\
&= \frac{1}{2\pi T} \sum_{t=1}^T \sigma_t^2 \mathbb{E}(\epsilon_t^2) \leq C
\end{aligned}$$

due to the uncorrelatedness of the innovations  $\epsilon_t$  and the uniform boundedness of their variance. Since  $T^d > (T/j)^d$ , and considering the behavior of the transfer function  $|\Gamma_d(\lambda)|$  in the neighborhood of the origin, it follows that  $I_y(\lambda) = |\Gamma_d(\lambda)|^2 |\zeta_u(\lambda)|^2 + O_p(\lambda^{-2d})$  at the harmonic frequencies  $j/T$ .

Note that all derivations above are linear, so premultiplying the finite Fourier transforms with  $T^\alpha$  in the case  $\alpha > 0$  leads to the desired result.

## Proof of Proposition 6

The GPH-estimator is obtained as the least squares estimator for the parameter  $d$  in the linear regression model

$$\log(I_j) = \log f_0^* - C - 2dX_j + \log\left(\frac{f_j^*}{f_0^*}\right) + \varepsilon_j, \quad j = 1, 2, \dots, m$$

with  $\varepsilon_j = \log(I_j/f_j) + C$  and  $C = -0.577216\dots$  is Euler's constant. Furthermore,  $X_j$  denotes the  $j$ -th Fourier frequency and  $I_j = \frac{1}{2\pi n} \left| \sum_{t=0}^{T-1} y_t \exp\left(\frac{i2\pi jt}{T}\right) \right|^2$  is the periodogram of the underlying series  $y_t$ . The term  $f^*$  originates from the spectral density of a stationary long-memory model given by

$$f(\lambda) = \sigma^2 |1 - \exp(-i\lambda)|^{-2d} f^*(\lambda), \quad -\pi \leq \lambda \leq \pi.$$

In this set-up the remaining spectral density  $f^*(\lambda)$  is scale independent. This proves to be handy later as the variance is assumed to be time dependent in our model. Due to the time varying variance, however, the process  $y_t$  is not a stationary long-memory process. Proving the proposition comes now from fitting a stationary model to a non-stationary process and evaluating the consequences. This is done by employing the approach of Dahlhaus (1997) by application of the local spectral density or local periodogram respectively. Denoting  $u = t/T$  and  $T$  being the sample size the local spectral density of our modulated long-memory process is

$$f(u, \lambda) = \sigma(u)^2 f(\lambda).$$

Following the arguments in Robinson (1995b) and Hurvich et al. (1998) the asymptotic normality of the log-periodogram regression is derived from the term

$$U_j = \varepsilon_j + \log\left(\frac{f^*(\lambda_j)}{f^*(0)}\right) - 2d \log\left(\frac{|1 - \exp(-i\lambda_j)|}{\lambda_j}\right).$$

Fitting a stationary model to a non-stationary means that  $f(u, \lambda) = f(\lambda)$  independent

of  $u$ . Thus we obtain for  $U_j$ :

$$\begin{aligned}
U_j &= \log \left( \frac{\int_0^1 I(u, \lambda_j) du}{\sigma^2 f(\lambda_j)} \right) + C + \log \left( \frac{f^*(\lambda_j)}{f^*(0)} \right) - 2d \log \left( \frac{|1 - \exp(-i\lambda_j)|}{\lambda_j} \right) \\
&= \log \left( \frac{\int_0^1 \sigma(u)^2 du \int_0^1 I(\lambda_j) du}{\sigma^2 f(\lambda_j)} \right) + C + \log \left( \frac{f^*(\lambda_j)}{f^*(0)} \right) - 2d \log \left( \frac{|1 - \exp(-i\lambda_j)|}{\lambda_j} \right) \\
&= \log \left( \frac{\int_0^1 \sigma(u)^2 du I(\lambda_j)}{\sigma^2 f(\lambda_j)} \right) + C + \log \left( \frac{f^*(\lambda_j)}{f^*(0)} \right) - 2d \log \left( \frac{|1 - \exp(-i\lambda_j)|}{\lambda_j} \right)
\end{aligned}$$

To obtain the limit distribution of the log-periodogram regression we have to consider the term

$$\frac{1}{m^{1/2}} \sum_{j=1}^m a_j \varepsilon_j = T_1 + T_2 + T_3$$

with  $a_j = X_j - \bar{X}$ . The terms  $T_1$  to  $T_3$  are given in Hurvich et al. (1998) equation (A 9). From Hurvich et al. (1998) equation (A 10) and (A 11) we directly have  $T_1 = o_P(1)$  and  $T_2 = o_P(1)$ . The remaining term can as in Hurvich et al. (1998) equation (A 12) be written as  $T_3 = T_{31} + T_{32} + T_{33}$  where the terms  $T_{32}$  and  $T_{33}$  are not effected our modulation of the time series. Thus, we can directly conclude from Hurvich et al. (1998) equation (A 14) and (A 15) that  $T_{32} = o(1)$  and  $T_{33} = o(1)$ .

As the modulation does not effect the coefficients  $a_j$  they still fulfill condition (5.15) in Robinson (1995b) and thus we have

$$T_{31} \xrightarrow{D} N \left( 0, \int_0^1 \sigma_u^2 du \frac{\pi^2}{6\sigma^2} \right)$$

which proves the proposition.

## References

- Amado, C. and T. Teräsvirta (2008). Modelling Conditional and Unconditional Heteroskedasticity with Smoothly Time-Varying Structure. CREATES Research Papers 2008-08.
- Avarucci, M. and C. Velasco (2009). A Wald Test for the Cointegration Rank in Nonstationary Fractional Systems. *Journal of Econometrics* 151 (2), 178–189.
- Beran, J. (2009). On Parameter Estimation for Locally Stationary Long-Memory Processes. *Journal of Statistical Planning and Inference* 139 (3), 900–915.
- Burrige, P. and A. M. R. Taylor (2001). On Regression-Based Tests for Seasonal Unit

- Roots in the Presence of Periodic Heteroscedasticity. *Journal of Econometrics* 104(1), 91–117.
- Cavaliere, G. (2004). Unit Root Tests under Time-Varying Variances. *Econometric Reviews* 23(3), 259–292.
- Cavaliere, G., M. Ø. Nielsen, and A. M. R. Taylor (2013). Bootstrap Fractional Integration Tests in Heteroskedastic ARFIMA Models. Technical report.
- Cavaliere, G., A. Rahbek, and A. M. R. Taylor (2010). Testing for Co-Integration in Vector Autoregressions with Non-Stationary Volatility. *Journal of Econometrics* 158(1), 7–24.
- Cavaliere, G. and A. M. R. Taylor (2005). Stationarity Tests under Time-Varying Second Moments. *Econometric Theory* 21(06), 1112–1129.
- Cavaliere, G. and A. M. R. Taylor (2007a). Bootstrap Unit Root Tests For Time Series With Nonstationary Volatility. *Econometric Theory* 24(1), 43–71.
- Cavaliere, G. and A. M. R. Taylor (2007b). Testing for Unit Roots in Time Series Models with Non-Stationary Volatility. *Journal of Econometrics* 140(2), 919–947.
- Cavaliere, G. and A. M. R. Taylor (2008). Time-Transformed Unit Root Tests for Models with Non-Stationary Volatility. *Journal of Time Series Analysis* 29(2), 300–330.
- Cavaliere, G. and A. M. R. Taylor (2009). Heteroskedastic Time Series with a Unit Root. *Econometric Theory* 25(5), 1228–1276.
- Dahlhaus, R. (1997). Fitting Time Series Models to Nonstationary Processes. *The Annals of Statistics* 25(1), 1–37.
- Dahlhaus, R. (2000). A Likelihood Approximation for Locally Stationary Processes. *The Annals of Statistics* 28(6), 1762–1794.
- Davidson, J. (1994). *Stochastic Limit Theory*. Oxford University Press.
- Davydov, Y. (1970). The Invariance Principle for Stationary Processes. *Theory of Probability and Its Applications* 15(3), 487–498.
- Demetrescu, M. (2010). On the Dickey–Fuller Test with White Standard Errors. *Statistical Papers* 51(1), 11–25.
- Demetrescu, M., V. Kuzin, and U. Hassler (2008). Long Memory Testing in the Time Domain. *Econometric Theory* 24(1), 176–215.
- Demetrescu, M., V. Kuzin, and N. Salish (2013). Regression-Based Fractional Cointegration Testing under Estimated Degree of Integration. Mimeo.

- Dolado, J. J., J. Gonzalo, and L. Mayoral (2002). A Fractional Dickey-Fuller Test for Unit Roots. *Econometrica* 70(5), 1963–2006.
- Geweke, J. and S. Porter-Hudak (1983). The Estimation and Application of Long Memory Time Series Models. *Journal of Time Series Analysis* 4(4), 221–237.
- Hall, P. (1977). Martingale Invariance Principles. *The Annals of Probability* 5(6), 875–887.
- Hassler, U. (2011). Estimation of Fractional Integration under Temporal Aggregation. *Journal of Econometrics* 162(2), 240–247.
- Hassler, U. and J. Breitung (2006). A Residual-Based LM-Type Test against Fractional Cointegration. *Econometric Theory* 22(6), 1091–1111.
- Hassler, U. and J. Scheithauer (2011). Testing Against a Change from Short to Long Memory. *Statistical Papers* 52(4), 847–870.
- Hurvich, C. M. and W. W. Chen (2000). An Efficient Taper for Potentially Overdifferenced Long-Memory Time Series. *Journal of Time Series Analysis* 21(2), 155–180.
- Hurvich, C. M., R. Deo, and J. Brodsky (1998). The Mean Squared Error of Geweke and Porter-Hudak’s Estimator of the Memory Parameter of a Long-Memory Time Series. *Journal of Time Series Analysis* 19(1), 19–46.
- Kew, H. and D. Harris (2009). Heteroskedasticity-Robust Testing for a Fractional Unit Root. *Econometric Theory* 25(6), 1734–1753.
- Łasak, K. (2010). Likelihood Based Testing for no Fractional Cointegration. *Journal of Econometrics* 158(1), 67–77.
- Leybourne, S., A. M. R. Taylor, and T. H. Kim (2007). CUSUM of Squares-Based Tests for a Change in Persistence. *Journal of Time Series Analysis* 28(3), 408–433.
- Lo, A. W. (1991). Long-Term Memory in Stock Market Prices. *Econometrica* 59(5), 1279–1313.
- Lobato, I. N. and C. Velasco (2007). Efficient Wald Tests for Fractional Unit Roots. *Econometrica* 75(2), 575–589.
- Marinucci, D. and P. M. Robinson (2000). Weak Convergence of Multivariate Fractional Processes. *Stochastic Processes and their Applications* 86(1), 103–120.
- Palma, W. and R. Olea (2010). An Efficient Estimator for Locally Stationary Gaussian Long-Memory Processes. *The Annals of Statistics* 38(5), 2958–2997.



- Phillips, P. C. B. (1987). Time Series Regression with a Unit Root. *Econometrica* 55(2), 277–301.
- Phillips, P. C. B. and K. L. Xu (2006). Inference in Autoregression Under Heteroskedasticity. *Journal of Time Series Analysis* 27(2), 289–308.
- Priestley, M. B. (1981). *Spectral analysis and time series*. Elsevier.
- Priestley, M. B. (1988). *Non-linear and non-stationary time series analysis*. Academic Press London.
- Robinson, P. M. (1995a). Gaussian Semiparametric Estimation of Long Range Dependence. *The Annals of statistics* 23(5), 1630–1661.
- Robinson, P. M. (1995b). Log-Periodogram Regression of Time Series with Long Range Dependence. *The Annals of statistics* 23(3), 1048–1072.
- Shao, X. and W. B. Wu (2007). Local Whittle Estimation of Fractional Integration for Nonlinear Processes. *Econometric Theory* 23(5), 899–929.
- Sibbertsen, P. and R. Kruse (2009). Testing for a Break in Persistence under Long-Range Dependencies. *Journal of Time Series Analysis* 30(3), 263–285.
- Teverovsky, V., M. S. Taqqu, and W. Willinger (1999). A Critical Look at Lo’s Modified R/S Statistic. *Journal of Statistical Planning and Inference* 80(1-2), 211–227.