Discriminating between fractional integration and spurious long memory

Niels Haldrup$^{ac}$ and Robinson Kruse$^{bc}$  
Aarhus University$^a$, Leibniz Universität Hannover$^b$, CREATES$^c$

14 January 2013

Abstract. Fractionally integrated processes have become a standard class of models to describe the long memory features of economic and financial time series data. However, it has been demonstrated in numerous studies that structural break processes and processes with non-linear features can often be confused as being long memory. The question naturally arises whether it is possible empirically to determine the source of long memory as being either a truly fractionally integrated process or whether the long range dependance is of a different nature. In this paper we exploit a particular feature of stationary fractionally integrated Gaussian processes to suggest a testing procedure that helps discriminating such processes from spurious long memory processes. The idea is that nonlinear transformations of stationary fractionally integrated processes decrease the order of integration in a specific way determined by the Hermite rank of the transformation. In principle, a non-linear transformation of the series can make the series short memory I(0). The results can be easily extended to allow for non-stationary FI$(d)$ processes. We suggest using the Wald test of Shimotsu (2007) to test the null hypothesis of a vector time series of transformed variables to be I(0). The test is shown to have excellent size and power against a range of non-linear and break processes that are known to generate spurious long memory. An illustration considers realized volatility and correlation series for the IBM stock.

Keywords: Long memory, fractional integration, non-linear models, break processes

JEL Classification: C12, C2, C22

1. Introduction

Recent empirical research indicates that many time series in economics and finance have long memory and belong to the class of fractionally integrated processes. This is especially the case for high frequency financial time series such as log squared returns, implied and realized volatility, interest rate spreads etc., see e.g. Taylor (1986), Diebold and Rudebusch (1989), Ding, Granger and Engle (1993), Baillie, Bollerslev and Mikkelsen (1996), Comte and Renault (1998), Andersen, et al (2001, 2003), Christensen and Nielsen (2007), Bollerslev, Osterrieder, Sizova, and Tauchen(2012).

*Corresponding author: Niels Haldrup, Department of Economics and Business, Aarhus University, Fuglesangs Allé 4, DK-8210 Aarhus V. Denmark. Email addresses: nhaldrup@creates.au.dk and kruse@statistik.uni-hannover.de. This research is supported by CREATES funded by The Danish National Research Foundation.
A univariate time series process is said to have long memory if its autocorrelation function $\gamma_x(\tau)$ at long lags decays at a hyperbolic (rather than an exponential) rate, i.e. $\gamma_x(\tau) \simeq B\tau^{-2d-1}$ as $\tau \to \infty$. Alternatively, the spectral density of the series is proportional to $\lambda^{-2d}$ with $d \neq 0$ for low frequencies $\lambda \to 0$. This implies that distant observations tend to be highly correlated and hence the term long memory or long range dependence is used for such processes. Fractionally integrated processes have this property and are often used as a convenient model class for time series which exhibit seemingly long memory features. A vast literature exists in estimating the long memory parameter $d$. Many of these studies address the estimation problem in a semiparametric framework, e.g. Geweke and Porter-Hudak (1983), Künsch (1987), and Robinson (1995a, 1995b), but parametric models within the class of fractionally integrated processes have also attracted considerable attention, i.e. Fox and Taqqu (1986), Sowell (1992a,b), Nielsen (2005) and Johansen and Nielsen (2010, 2012). Recently Johansen and Nielsen (2010, 2012) and Lasak (2010, 2011) have developed estimation and testing procedures for co-fractional VAR models.

Despite its attractiveness for parsimonious model building, fractionally integrated processes are often found to be a too flexible model class in the sense that processes that are not truly fractional or long memory, can be fitted arbitrarily well to the data. For instance, the features of hyperbolic decay of autocorrelations and unboundedness of the spectral density at the origin can also be present when a short memory process is affected by a regime change, a smooth trend, or similar. Hence, in practice, it can be hard to distinguish these processes from long memory processes. Often this feature is referred to as "spurious" long memory, see e.g. Diebold and Inoue (2001), Gourieroux and Jasiak (2001), Granger and Ding (1996), Granger and Hyung (2001), Shimotsu (2006), Ohanissian, Russell and Tsay (2008), Perron and Qu (2011), and Qu (2011). A commonly used example is a short memory process, (a summable and invertible ARMA process for instance), which is subject to a random level shift where the shifts are governed by a Bernoulli process with a shift probability $p$, see for instance Diebold and Inoue (2001), Qu (2011) and Perron and Qu (2011). Diebold and Inoue (2001) show that if the shift probability is allowed to depend on the sample size in a particular way, then the process will mimic the long run behaviour of a long memory process. Even though fractionally integrated processes and the random level shift model for instance are similar in terms of long memory, fractionally integrated processes have the property of self-similarity whereas the random level shift model is not self-similar. The notion of a self-similar process is discussed in e.g. Mendelbrot and Van Ness (1968) and has the implication that the long memory properties of the process is unaffected by the sampling frequency.

A steadily growing literature has developed with emphasis on whether it is possible to empirically discriminate between true long memory processes (or fractionally integrated processes) and spurious long memory processes. This is an important question from both a statistical and an economic perspective. Estimation and inference under a stationary long memory model is quite different from structural break, non-linear, and non-stationary models and from an economic perspective the long lasting
impact of shocks in presence of long memory is rather different from models where rare structural breaks occur for example. A number of tests have been developed which attempt discriminating between true and spurious long memory. Ohanissian, Russell and Tsay (2008) exploit the self-similarity property of fractional processes to develop a test based on the invariance of the long memory parameter for temporal aggregates of the process under the null of true long memory. Shimotsu (2006) develops two tests. One test is based on the estimation of the long memory parameter using subsamples and comparing these estimates with the estimate for the full sample. Under the null of true long memory the memory parameters are identical. The second test suggested by Shimotsu (2006) is based on the idea of estimating \( d \) for the full sample and then testing under the null whether \( \Delta^d x_t \) is short memory, or I(0). This is a property which does not apply for a range of spurious long memory models under the alternative. Qu (2011) proposes a test for the null of true stationary long memory. The test is a frequency domain test based on the derivatives of the profiled local Whittle likelihood function in a shrinking neighborhood of the origin. For long memory processes the behaviour of the likelihood at the origin depends on the bandwidth parameter in a particular way which is a feature he uses as a basis for testing. Perron and Qu (2011) propose a test against the mean shift hypothesis based on the observation that under the alternative the estimate of \( d \) will depend on the number of frequencies included in the log-periodogram regression.

The present papers takes a different point of departure in testing for (true) long memory. More precisely, the null being tested is that a univariate time series is Gaussian, fractionally integrated of some order \( d \) with \( 0 < d < 1/2 \), i.e. the process is assumed to be a stationary long memory process. As for the second test of Shimotsu (2006) it is true that the \( d \)th difference is short memory I(0). However, there are other ways a Gaussian fractional long memory process can be made short memory. We exploit the feature that nonlinear transformations of a fractionally integrated process will have a lower order of memory than the original series, see e.g. Dittmann and Granger (2002) and Avarucci and Marinucci (2007). The actual reduction in memory is determined by the so-called Hermite rank of the transformation. An \( FI(d) \) series \( x_t \) can thus be made I(0) by an appropriately chosen nonlinear transformation, possibly in combination with a partial differencing of the series \( \Delta^\delta x_t \) with \( \delta < d \). In principle, a range of different transformations can be considered in constructing a vector series which is short memory I(0) under the null hypothesis. The combination of a non-linear transform and partial pre-differencing also allows a straightforward generalization to the case of non-stationary Gaussian \( FI(d) \) processes with \( d > 0.5 \). We suggest to use a multivariate Gaussian semiparametric estimator of long-range dependent processes due to Shimotsu (2007) to estimate the long memory parameters of a vector of transformed time series and to use a Wald-test to test the null of whether the vector time series is jointly I(0). The test is \( \chi^2 \) distributed under the null and by appropriate choice of tuning parameters, i.e. the choice of Hermite rank and partial differencing parameter \( \delta \), the test is shown to have excellent size and reasonable power.
Discriminating between fractional integration and spurious long memory
against non-stationary random level shift models, Markov-switching GARCH models, and monotonic trend models.

The plan of the paper is as follows. In section 2, we define the class of models considered in the paper: $FI(d)$ processes and non-linear and break processes with apparent long memory features. In section 3 the vehicle underlying our suggested testing procedure is defined including a review of how fractionally integrated processes are affected when transforming these by Hermite polynomials or other transformations. Section 4 presents the suggested testing procedure and an extensive Monte Carlo study of the size and power properties are discussed in Section 5. Section 6 provides an empirical application to IBM stock market realized volatility and correlation series.

2. Fractional integration, structural breaks, and non-linear models

Consider a fractionally integrated time series process $x_t$ generated according to

$$(1 - L)^d x_t = u_t$$

where $u_t$ is a covariance stationary process with spectral density $f_u(\lambda)$ being bounded (and bounded away from zero) at frequency zero. We will assume that $0 \leq d < 1/2$ meaning that $x_t$ is stationary with long memory, except for $d = 0$ where $x_t$ is short memory. More generally, $-1/2 < d < 0$ means that the process is antipersistent whereas $1/2 < d < 1$ implies a non-stationary, mean-reverting, long memory process. Assume for instance that $x_t$ follows a summable and invertible ARFIMA($p, d, q$) process

$$\Phi(L)(1 - L)^d x_t = \Theta(L) \varepsilon_t$$

where $\Phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p$ and $\Theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_q L^q$ and $\varepsilon_t$ is white noise with variance $E(\varepsilon_t^2) = \sigma_\varepsilon^2$. Then the spectral density of $x_t$ in a neighborhood of zero approximates

$$f_x(\lambda) \simeq \frac{\sigma_\varepsilon^2 |\Theta(1)|^2}{2\pi |\Phi(1)|^2} \lambda^{-2d} \text{ as } \lambda \to 0_+$$

where "$\simeq$" signifies that the ratio on the right and left sides tends to unity. In general, a long memory process is defined to be a process with spectral density $f(\lambda) \simeq C\lambda^{-2d}$ as $\lambda \to 0_+$ for $-1/2 < d < 1/2$ and finite $C$. In other words, the spectral density of an ARFIMA process exhibits the shape of a long memory process around the origin and hence belongs to this class of models. We will denote a long memory process with memory parameter $d$ an $LM(d)$ process. Note that an $FI(d)$ process is $LM(d)$, but the reverse is not the case since $LM(d)$ processes is a more general class of processes.

In the present paper the class of long memory processes we want to discriminate from spurious long memory processes are of the Gaussian ARFIMA type with $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ and $0 \leq d < 1/2$.

To compare fractionally integrated processes with time series models that are frequently mistakenly found to generate long memory, we consider three processes:
Discriminating between fractional integration and spurious long memory

a) a non-stationary random level shift model (RLS-NS), b) a Markov-switching model with GARCH regimes (MS-GARCH) and c) a monotonic time trend model (MONO).

The RLS-NS model is defined as follows:

\[ x_t = \mu_t + \varepsilon_t \]
\[ \mu_t = \mu_{t-1} + \pi_t \eta_t \]

where \( \varepsilon_t \sim N(0, \sigma^2_\varepsilon) \). In every period a level shift occurs with probability \( P(\pi_t = 1) = p \) and the size of the level shift is random and drawn from a standard normal distribution, i.e. \( \eta_t \sim N(0, 1) \).

The MS-GARCH model reads:

\[ x_t = \log y_t^2 \]
\[ y_t = \sqrt{h_t} \varepsilon_t \]
\[ h_t = \omega_0 + \omega_1 s_t + \alpha x_{t-1}^2 + \beta h_{t-1} \]

with \( \varepsilon_t \sim N(0, 1) \) and \( s_t = \{0, 1\} \) following a Markov process with state transition probabilities \( p_{10}, p_{01} \). As seen, the GARCH equation exhibits regime switches in its intercept from \( \omega_0 \) to \( \omega_1 \). The remaining parameters are the same across the two regimes. For both the RLS-NS and MS-GARCH models to generate apparent long memory the shift and transition probabilities have to be rather small.

Finally, the MONO model is given as:

\[ x_t = ct^{(d-0.5)} + \varepsilon_t \]

with \( \varepsilon_t \sim N(0, 1) \). Hence, the series is a deterministic trend process plus noise.

These models have been analyzed previously as competing models to long memory models, see e.g. Ohannesian et al. (2008) and Qu (2011). The random level shift model has also been investigated in Chen and Tiao (1990) and Perron and Qu (2011). Smith (2005) analyzes the effects of occasional level shifts on the log-periodogram estimator by Geweke and Porter-Hudak (1983) and proposes suitable modifications. The problem of estimating the long memory parameter \( d \) in the presence of level shifts and trends is further treated in McCloskey and Perron (2011). This study also covers the case of monotonic deterministic trends. The Markov switching GARCH model dates back to Cai (1994), Hamilton and Susmel (1994) and has been further investigated in Dueker (2009) and Haas, Mittnik and Paolella (2009) amongst others.

The models above can be calibrated in such a way that when the long memory parameter is estimated using simulated data and by adopting a range of different estimators, strong evidence of long memory can be found. In Figure 1 four simulated series of length \( T = 3000 \) observations are displayed: An \( FI(0.40) \) process and representative processes (3), (4) and (6) calibrated to yield local Whittle estimates of \( d \) of a similar magnitude. Each series have been contaminated with an AR(1) component with \( \phi = 0.5 \). The semi-parametric local Whittle estimator of Künsch (1987) was used.
in these estimations. Figures 2 and 3 display the estimated autocorrelation function and the spectral density for these simulated processes. From visual inspection of these figures the apparent slow decay of the autocorrelation functions and the similar shape of the spectral density can be seen and hence indicating that discrimination between these processes may be difficult in practice.

Figures 1-3 about here

3. Nonlinear transformations of Gaussian fractionally integrated processes

It is well known that stationary Gaussian fractionally integrated processes can be transformed to reduce their order of long memory, see e.g. Dittmann and Granger (2001) and Avarucci and Marinucci (2007). We will exploit this property to develop a test for true fractional integration. This section reviews some properties of non-linear transformations and Hermite polynomial expansions and we present the main result upon which our test is based.

Consider the case of a zero mean Gaussian random variable \( x \) with variance \( \sigma^2 \) and some transformation given by the function \( G(\cdot) \) satisfying that \( EG^2(x_t) < \infty \). The function can be expanded in terms of a series of Hermite polynomials

\[
G(x) = \sum_{k=0}^{\infty} \frac{c_k}{k!} H_k(x)
\]  

(7)

where the Hermite polynomials read

\[
H_k(x) = (-1)^k \exp\left(\frac{-x^2}{2\sigma^2}\right) \frac{d^k}{dx^k} \left( \exp\left(\frac{-x^2}{2\sigma^2}\right) \right).
\]  

(8)

Apart from scaling, the first 5 Hermite polynomials are given by:

\[
\begin{align*}
H_0(x) &= 1 \\
H_1(x) &= x \\
H_2(x) &= x^2 - 1 , \\
H_3(x) &= x^3 - 3x \\
H_4(x) &= x^4 - 6x^2 + 3 \\
H_5(x) &= x^5 - 10x^3 + 15x \\
\end{align*}
\]  

(9)

The coefficients \( c_k \) are defined as

\[
c_k = E[G(x)H_k(x)] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} G(x)H_k(x) \exp\left(\frac{-x^2}{2\sigma^2}\right)dx
\]  

(10)
so for instance \( c_0 = \mathbb{E}[G(x)] \).

The Hermite rank of \( G(.) \) is the index \( J \) of the lowest non-zero coefficient \( c_k \), i.e.

\[
    c_k = 0 \quad \text{for} \quad k = 1, 2, \ldots, J - 1
\]

\[
    c_J = \mathbb{E}[G(x)H_J(x)] \neq 0
\]

Interestingly, we have that

\[
    \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} H_l(x)H_k(x) \exp\left(-\frac{x^2}{2\sigma^2}\right) dx = \left\{ \begin{array}{ll}
    1 & \text{for } l = k \\
    0 & \text{for } l \neq k
    \end{array}\right.
\]

such that the Hermite polynomials are orthogonal and hence, according to (10), a polynomial transformation \( G(x) \) with Hermite rank \( J \) can always be found by selecting Hermite polynomials \( G(x) = H_j(x) \), i.e. by construction transformations corresponding to Hermite polynomials of order \( J \) will have Hermite rank \( J \).

Assume that we have two zero mean Gaussian variables \( x \) and \( y \). Then

\[
    \mathbb{E}[H_k(x)H_l(y)] = \left\{ \begin{array}{ll}
    k! [\mathbb{E}(xy)]^k & \text{for } l = k \\
    0 & \text{for } l \neq k
    \end{array}\right.
\]

(11)

Now, consider the Gaussian process \( \Delta^d x_t \), \( 0 < d < 1/2 \), i.e. \( x_t \) is a fractionally integrated Gaussian process whereby the autocorrelation function declines according to \( \gamma_x(\tau) \simeq G_\tau^{-2d-1} \). It follows from (11) that

\[
    \mathbb{E}[H_k(x_t)H_k(x_{t-\tau})] = k! \gamma_x^k(\tau) \simeq C_\tau^{-k(2d-1)}, \quad \text{as } \tau \to \infty
\]

In other words, the sequence \( H_{k+1}(x_t) \) will exhibit less memory than the sequence \( H_k(x_t) \). More precisely, if \( x_t \sim \text{FI}(d) \), then the transformation \( H_k(x_t) \) will be a long memory series \( \text{LM}(d_k) \) with \( d_k < d \) where

\[
    d_k = \max \left\{ 0, (d - \frac{1}{2})k + \frac{1}{2} \right\}
\]

(12)

which follows from \( 2d_k - 1 = k(2d - 1) \).

Figure 4 displays the relationship (12) for transformations with Hermite rank \( J \). In practice, one may use the Hermite polynomials as transformations, i.e. \( G_{j=k}(x_t) = H_k(x_t) \). As seen, any transformation with Hermite rank bigger than one will reduce the long memory index of the original fractionally integrated process. For instance, for Hermite rank \( J = 2 \) an \( \text{I}(d) \) series can be made short memory \( \text{I}(0) \) as long as \( 0 < d < 0.25 \). For \( J = 3 \) the transformed series is \( \text{I}(0) \) when the original series has \( d \) in the range \( 0 < d < .33 \), and so forth. The memory range in which a series can be made \( \text{I}(0) \) increases with the Hermite rank of the transformation. However, as \( d \) increases towards the non-stationary region \( d = 0.5 \), it becomes difficult to make the transformed series \( \text{I}(0) \) without choosing a transformation with a very high Hermite rank. There is a way of circumventing this. Assume that the original series is \( \text{I}(d) \)
with \(d\) "close" to 0.5, say \(d = 0.45\). For a Hermite rank of \(J = 2\) it is not possible to make the series \(I(0)\). In fact, in this case \(d_2 = 0.4\) and only a minor reduction in memory results. However, it is possible to partially difference the series prior to non-linear transformation in order to magnify the reduction of memory order. In place of \(x_t\), consider for as an example the series \(y_t = \Delta^d x_t\), where \(d = 0.2\) such that \(y_t \sim FI(d - \delta = 0.25)\). Note that \(y_t\) is still a Gaussian process. A subsequent transformation \(H_J(y_t)\) with Hermite rank \(J = 2\) will thus make the original series \(I(0)\) after "partial" pre-differencing. We shall later see that such a combination of partial differencing and a non-linear transformation may be useful for testing purposes.

Even though we have assumed from the outset that \(x_t\) is a stationary \(FI(d)\) process, \(0 < d < 0.5\), the argument above shows that it is possible to extend the analysis to non-stationary processes with \(d \geq 0.5\). The idea is to consider an appropriately chosen value of \(\delta\), to partially pre-difference the series, and to apply the properties above to the process \(y_t = \Delta^\delta x_t\).

Figure 4 about here

As seen, an interesting feature of Gaussian \(FI(d)\) processes is that the series can be made short memory \(I(0)\) not only by taking a \(d'\)th difference of the series, but also by non-linear transformation. This has the consequence, that a series may "cointegrate" non-linearly with itself even though common long memory rather than cointegration seems to be the appropriate notion in this case, see Engle and Granger (1987) for the definition of cointegration. Assume for instance the case where \(x_t \sim FI(d)\) and consider the transformation \(H_3(x_t)\). It follows that \(H_3(x_t) = (x_t^3 - 3x_t)\) has memory \(LM(d_3)\), with \(d_3 = \max\{0, 3d - 1\} < d\). Note however, that the two terms \(x_t^3\) and \(3x_t\) both have Hermite rank \(J = 1\) and hence have memory \(LM(d)\). This shows that linear combinations of Gaussian \(FI(d)\) processes may have common memory features with non-linear transformations of itself and for certain values of \(d\) the linear combination of the series may even be short memory.

4. **Testing the null of fractional integration**

The properties of non-linear transformations of stationary \(FI(d)\) processes discussed in the previous section will now be used to develop a testing procedure to discriminate between stationary, fractionally integrated processes and spurious long memory processes.

The testing procedure consists of first estimating \(d\) for the original series, \(x_t\), which is \(FI(d)\) under the null hypothesis. We know that as long as the estimate of \(d\) is consistent, then in the limit \(\Delta^d x_t\) is short memory \(I(0)\), see Shimotsu (2006), and \(y_t = \Delta^{d-\delta} x_t\) is \(FI(\delta)\) for some target differencing \(\delta\). Consider next a Hermite polynomial transformation of \(y_t\), i.e. \(H_J(y_t)\) which is known theoretically to be short memory \(LM(0)\). In principle, a sequence of transformations with different Hermite ranks can be considered (and in different combinations) which leads to the vector time series \(Y_t = (\Delta^d x_t, H_{J_1}(y_t), H_{J_2}(y_t), ..., H_{J_m}(y_t))^T\) being short memory \(LM(0)\) under
the null hypothesis. The test statistic is based on a multivariate estimate of the long memory parameters of the vector series $Y_t$ and testing $Y_t \sim LM(0)$ using a Wald test due to Shimotsu (2007). Note that in the construction of the $y_t$ series using different Hermite polynomial transforms, the possibility of using different target values $\delta$ for each transform may be considered. In accordance with the discussion in section 3 a proper notation will include an index $J$ of the partially differenced series to indicate that the target differencing may depend on the Hermite ranks, i.e. $y_t^J = \Delta^{\delta - \delta_J} x_t$ is $FI(\delta_J)$. In section 5 we will use simulations to determine how the target differencing filter can be selected in practical applications.

Assume the dimension of $Y_t$ is $q \times 1$ with memory index $d = (d_1, d_2, ..., d_q)' = 0$ under the null, i.e.

$$H_0 : Y_t = (\Delta^{\delta} x_t, H_{J_1}(y_t^{J_1}), H_{J_2}(y_t^{J_2}), ..., H_{J_p}(y_t^{J_p}))' \sim LM(0).$$

In principle, a single element or a group of elements in $Y_t$ can be tested under the null. The test we suggest is based on Shimotsu’s (2007) multivariate Gaussian semiparametric estimator (GSE) of long memory processes and the associated Wald test to test the null hypothesis (13). The estimator is defined as

$$\hat{d} = \arg \min_{d \in \Theta} R(d)$$

where the objective function reads

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

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

The Fourier frequencies are given by $\lambda_j = 2\pi j / T$ with $j = 1, 2, ..., m$ and $m = o(T)$ is the bandwidth parameter. We also have $\Lambda_j(d) = \text{diag}(\Lambda_{j0}(d))$, $\Lambda_{j0}(d) = \lambda_j^{d_a} \exp(i(\pi - \lambda_j)d_a/2)$ and $\Lambda_j^*[d]$ is the conjugate transpose of $\Lambda_j(d)$. The admissible range of $d$ in minimizing the objective function (15) is $\Theta = [\Delta_1, \Delta_2]^q$, with $-\frac{1}{2} < \Delta_1 < \Delta_2 < \frac{1}{2}$. Nielsen (2011) extends the Shimotsu (2007) estimator to a expanded range of $d$ values given by $-\frac{1}{2} < d < \infty$.

Shimotsu (2007) shows the consistency and asymptotic normality of this estimator. Also, he shows that the estimator is more efficient that the two step GSE estimator of Lobato (1999). It is important to note that Gaussianity is not assumed in the asymptotic theory and includes a general class of multivariate long-range dependent processes, including fractionally integrated processes. A test of the hypothesis $H_0 : d = 0$ is given by the Wald statistic

$$W = m \hat{d} \hat{\Omega} \hat{d}$$

17

19
Discriminating between fractional integration and spurious long memory

where
\[ \hat{\Omega} = 2 \left[ \hat{G}(\hat{d}) \odot \hat{G}(\hat{d})^{-1} + I_q + \frac{\pi^2}{4} (\hat{G}(\hat{d}) \odot \hat{G}(\hat{d})^{-1} - I_q) \right]. \]

The test has a limiting \( \chi^2(q) \) distribution under the null.

With reference to work by Hurvich and Chen (2000) concerning properties of the GSE estimator in univariate settings, Shimotsu (2007) proposes a correction factor of the Wald test that appears to yield better size properties. The modified test Wald test reads
\[ W_c(\delta_J, \mathbf{J}) = c_m \hat{d} \hat{\Omega} \hat{d} \]
where \( c_m = \sum_{j=1}^{m} v_j^2, v_j = \log \lambda_j - \frac{1}{m} \sum_{j=1}^{m} \log \lambda_j \). Because \( c_m/m \to 1 \) as \( m \to \infty \), the asymptotic distribution is unaffected by the correction factor. This is our preferred test to be used subsequently. The notation of the test indicates that an arbitrary vector \( Y_t \) can be chosen where \( \delta_J, \mathbf{J} \) are themselves vectors associated with \( \mathbf{J} = (J_1, J_2, ..., J_p) \).

5. A Monte Carlo study of power and size

In this section we will discuss the properties of our testing procedure through simulation. First, we will address the issue of optimally choosing the target difference \( \delta_J \), and second, we analyze the size of the test under a range of specifications in order to check robustness with respect to discrepancies from the Gaussian assumptions. Finally, we calculate power (rejection frequencies) for the range of competing models presented in section 2 which are known to exhibit seemingly long memory.

5.1. Test size and choice of \( \delta_J \). We start by analyzing how appropriate target values \( \delta_J \) can be found for transformations with Hermite rank \( J \). The selected transformations are simply the Hermite polynomials given in (9). In case of no estimation error, the relationship (12) would hold exactly and finding the appropriate differencing \( d - \delta_J \) that would make \( y_t^{d} \) exactly \( LM(0) \) would be trivial. However, when estimating \( d \) the estimation error should be accounted for. We have tried to empirically replicate the theoretical shapes displayed in figure 4 by estimating \( d \) for a range of \( FI(d) \) processes \( x_t \) with \( d \in [0, 0.5] \). Next, we have estimated the integration order \( d_k \) of the transformed series \( H_J(x_t) \). Figures 5-8 display mean values of the combinations of \( \hat{d} \) and \( \hat{d}_k \) using local Whittle estimates of the long memory parameter estimated with the commonly used bandwidth parameter \( m = [T^{0.65}] \), see Künsch (1987). The graphs are displayed for Hermite ranks 2-5 with \( T = 500 \) and \( T = 3000 \) observations. 1000 Monte Carlo replications were used to construct the graphs.

Figures 5-9 about here
As seen, the theoretical decline in long memory is only partially reflected in the empirical estimates. Generally, the reduction of memory is less steep than predicted by theory and the point where $\tilde{d}_k$ is approximately zero is lower than predicted by the theoretical results. This suggests that the target differencing $\delta_J$ should be chosen to be a smaller value than suggested by the theory. The simulations also confirm that $\delta_J$ should be chosen as an increasing function of $J$ whereas the sample size plays a minor role in the choice of $\delta_J$.

We will use empirical size calculations for various calibrations of the model to determine the desired values $\delta_J$. Figure 9 displays the size of the $W_c(\delta_J, J)$ test as a function of $\delta_J$ and for Hermite ranks $J = 2, 3, 4, 5$. The GSE was constructed with the bandwidth parameter $m = \lfloor T^{0.65} \rfloor$. The test is a $\chi^2(1)$ test of the null hypothesis

$$H_0 : Y_t = H_J(\Delta^{\delta_J} x_t) \sim LM(0)$$

that is, only a single element, the order of a particular Hermite polynomial transform, is considered in this case.

The sizes are reported for the case of $FI(d)$ processes with $d = .35$ and $d = .40$ and with $T = 500$ and 3000. Innovations are drawn from a normal distribution. All tests are conducted at a nominal 5% level. The idea is to see when the size as a function of $\delta_J$ is not being seriously distorted and to use this as a guideline for choice of $\delta_J$. As seen, the desired value of $\delta_J$ for $J = 2$ seems to lie in the range $[0.15-0.175]$, and for $J = 3, 4, 5$ the approximate range is $[0.175–0.20]$. Larger values of $\delta_J$ will result in size distortion whereas size is at the nominal level for smaller values of $\delta_J$. Note that the sample size plays only a minor role for the shape and location of the size curves.

To further examine the size and robustness of the test we conducted a number of experiments reported in Tables 1-4. Data was generated according to the process

$$\Delta^d x_t = u_t$$

with $u_t$ following a) standard normal distribution, b) a $t(6)$ distribution to allow for fat tails, c) an AR(1) process with $\phi = 0.4$ and 0.8, respectively and Gaussian errors, and finally d) a normal distribution where the process is contaminated with an additive measurement error term. In the latter case the variance of the measurement error is chosen such that the signal to noise ratio equals one. Several of these specifications are similar to those considered by Qu (2011) The Tables display the size of the $W_c(\delta_J, J)$ test of the hypothesis (18) for a range of target values $\delta_J$. In general, sizes are reasonable. In all cases with normally distributed errors, even with a fairly strong degree of autocorrelation, the tests appear to have size close to the nominal level. For the largest values of the Hermite rank the test happens to be slightly conservative when errors are drawn from a $t-$distribution. Especially, in the cases where the Hermite rank is odd, normal errors with measurement error give rise to a fairly severe size distortion. We believe that these distortion is a result of a fairly low signal to noise ratio which causes the fractionally integrated component signal to be excessively contaminated.
5.2. Power of test. We now examine the power properties of the $W_c(\delta_J, J)$ test with respect to the hypothesis (18) for the three non-linear model specifications outlined in section 2, i.e. the non-stationary random level shift model, RLS-NS, the Markov switching GARCH, MS-GARCH, and the monotonic trend, MONO, models. These models are also included in the studies of Ohanissian, Russell and Tsay (2008) and Qu (2011). For comparisons, the calibration of the models corresponds to the calibrations in Qu (2011) and we refer to his paper for the precise details. The configurations result in mean values of the estimated long memory parameters of around .25 for the RLS-NS model, .17 for the MS-GARCH, and .12 for the MONO model. These numbers indicate a relatively low value of the persistence parameter and hence we also conducted an experiment where the models were calibrated to yield $d$ estimates in the range [.40-.45].

In Tables 5-7 power calculations are reported for the three competing models. For each model the $W_c(\delta_J, J)$ test was calculated with $m = T^{0.65}$ and test rejection frequencies using 1000 replications were constructed for sample sizes $T = 500$ and $T = 3000$. Different combinations of Hermite ranks $J$ were considered in combination with target differencing parameter $\delta_J$ in the allowable range suggested by the size study. To compare with other tests we also report the power results for the Qu (2011) test (QU), the Ohanissian, Russell, and Tsay (2008) test, (ORT), the split-sample (S-SPLIT) and the difference (S-DIFF) tests of Shimotsu (2006). Finally, the mean-$t_d$ test (PQ) of Perron and Qu (2011) is compared with; these numbers are taken from Qu (2011).

Let us first comment on powers for the RLS-NS specification in Table 5. For these situations the preferred target differencing is $\delta_J = 0.175$ with a Hermite rank of $J = 3$. The power for $T = 500$ is 0.198; only the S-DIFF test appears to have a better power. For $T = 0.674$ the power is 0.674 which is slightly exceeded by the QU and S-DIFF tests. Note that there is no general tendency that transformations with larger Hermite rank will be more powerful.

The MS-GARCH model, Table 6, reveals a similar pattern. For $T = 500$ all tests, including the competing tests, appear to have a fairly low power. However, for $T = 3000$, a test with $\delta_J = 0.175$ and $J = 3$ yields reasonable power compared with the competitors. It is remarkable, that the test has no power for $J = 4, 5$. We believe, however, that this is due to the design of the experiment. In Table 8, panel B, the model is calibrated to yield a much larger degree of persistence and as seen in this case the powers are excellent. In particular, power is also gained for $J = 4$ and 5 for both the small and large sample sizes.

The MONO specification, Table 7, also shows that power is excellent compared to the other tests. Again, only the S-DIFF test seems to dominate. It can be seen from Table 8 that remarkably large powers can be reached when the underlying process has a stronger degree of long memory.

The specifications chosen indicate that choosing $J = 3$ and $\delta_J = 0.175$ seem opti-
mal to maximize power. However, the actual specification is clearly dependent upon the alternative considered and further analysis where a range of other specifications, calibrations, and sample sizes are examined would enlighten this further.

As a suggestion for a future specification of the test that is likely to improve power, consider testing the null hypothesis

$$H_0 : Y_t = (\Delta^{\hat{d}} x_t, H_J(\Delta^{\hat{d}-s_j} x_t))^\prime \sim LM(0).$$

Testing this hypothesis extends the analysis conducted so far by considering the additional restriction $\Delta^{\hat{d}} x_t \sim LM(0)$. The Wald test will be $\chi^2(2)$ in this case. Interestingly, the extra hypothesis being tested corresponds to the null hypothesis of the S-DIFF test of Shimotsu (2006) which occurs to have excellent power in many cases. A combination of this hypothesis with the one considered in the present simulation study seems to suggest that potentially extended power gains can be achieved.

6. Empirical Illustration

As an empirical illustration we consider seven different volatility and correlation series for the IBM stock over the period from January 1, 2000 to June 30, 2008, yielding a total of $T = 2156$ daily observations per series. The first two series are 5-minute realized volatility and its logarithmic transformation. The last five series are realized correlation measures where correlations between the IBM stock and the following five other stocks are considered: American Express (AXP), Citi Group (Citi), General Electric (GE), Home Depot (HD) and JP Morgan (JPM). All these series display a typical hyperbolic decline of the autocorrelation function.

In order to study whether the data is generated from a fractionally integrated process or some other long memory process we apply the $W_c(0.175,3)$ test as well as Qu’s test with trimming parameter equal to 0.2. The results are reported in Table 9. Clearly, the $W_c(0.175,3)$ test rejects for all series at the nominal significance level of five percent whereas Qu’s test rejects in only two cases (i.e. $RV$ and $RC(IBM/AXP)$). We thus conclude that it is likely that these series are not generated by fractionally integrated processes.

7. Conclusion

We have suggested a new procedure to discriminate between Gaussian fractional integration and spurious long memory. The procedure is based on the feature that non-linear transformations of (stationary) Gaussian fractional processes will reduce the degree of long memory when the Hermite rank of the transformation exceeds one. By appropriate difference filtering combined with a non-linear transformation, the
Discriminating between fractional integration and spurious long memory

series can be made short memory I(0). This feature can be used as a framework for a range of different specifications of the null hypothesis. We suggest using the Wald test of Shimotsu (2007) based on multivariate Gaussian semi-parametric estimation of the vector of memory parameters. For a limited Monte Carlo experiment we have demonstrated that the test can be designed to have excellent size, and moreover, the test is robust to further short run dependence and fat-tailed error distributions which is likely to characterize many financial data. When compared to other tests in the literature, our test is shown to have excellent power against competing alternatives that are empirically relevant.

In future work a more extensive study of the testing procedure will be undertaken. In particular, our simulation results so far indicate that alternative ways of specifying the null hypothesis have the potential to further increase the power of our test. Also, we intend to generalize the analysis to fractionally integrated processes where $d > 1/2$. Nielsen (2011) extends the Gaussian semi-parametric estimator of Shimotsu to an extended parameter space and straightforward generalizations of our procedure seems straightforward.
REFERENCES


8. Tables and Figures
### Table 1

Size of Shimotsu (2007) Wald test of the process $Y(t)=H_j(y_t)$, where $y_t = \Delta^d x_t$. $X_t$ is generated according to an $FI(d)$ process governed by errors that are either Normal white noise, $t(6)$-distributed, Normal with AR(1) errors ($\Phi=0.4, 0.8$), or Normal with Measurement errors. Sample sizes are $T=500$, and $T=3000$. Hermite rank of transformation $J=2$.
Discriminating between fractional integration and spurious long memory

<table>
<thead>
<tr>
<th>J = 3</th>
<th>Normal</th>
<th>Normal</th>
<th>Normal</th>
<th>Normal</th>
<th>Normal</th>
<th>Normal</th>
<th>Normal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
</tr>
<tr>
<td></td>
<td>AR(.4)</td>
<td>AR(.4)</td>
<td>AR(.8)</td>
<td>AR(.8)</td>
<td>ME</td>
<td>ME</td>
<td>ME</td>
</tr>
<tr>
<td>T</td>
<td>500</td>
<td>3000</td>
<td>500</td>
<td>3000</td>
<td>500</td>
<td>3000</td>
<td>500</td>
</tr>
<tr>
<td>δ</td>
<td>0.050</td>
<td>0.350</td>
<td>0.054</td>
<td>0.354</td>
<td>0.055</td>
<td>0.356</td>
<td>0.048</td>
</tr>
<tr>
<td></td>
<td>0.054</td>
<td>0.356</td>
<td>0.058</td>
<td>0.358</td>
<td>0.060</td>
<td>0.360</td>
<td>0.053</td>
</tr>
<tr>
<td></td>
<td>0.060</td>
<td>0.360</td>
<td>0.055</td>
<td>0.355</td>
<td>0.057</td>
<td>0.357</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td>0.054</td>
<td>0.354</td>
<td>0.049</td>
<td>0.349</td>
<td>0.053</td>
<td>0.353</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td>0.047</td>
<td>0.347</td>
<td>0.055</td>
<td>0.355</td>
<td>0.057</td>
<td>0.357</td>
<td>0.055</td>
</tr>
<tr>
<td></td>
<td>0.047</td>
<td>0.347</td>
<td>0.052</td>
<td>0.352</td>
<td>0.054</td>
<td>0.354</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td>0.054</td>
<td>0.354</td>
<td>0.052</td>
<td>0.352</td>
<td>0.054</td>
<td>0.354</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td>0.054</td>
<td>0.354</td>
<td>0.060</td>
<td>0.360</td>
<td>0.058</td>
<td>0.358</td>
<td>0.058</td>
</tr>
<tr>
<td></td>
<td>0.058</td>
<td>0.358</td>
<td>0.064</td>
<td>0.364</td>
<td>0.062</td>
<td>0.362</td>
<td>0.062</td>
</tr>
<tr>
<td></td>
<td>0.062</td>
<td>0.362</td>
<td>0.060</td>
<td>0.360</td>
<td>0.060</td>
<td>0.360</td>
<td>0.060</td>
</tr>
<tr>
<td></td>
<td>0.060</td>
<td>0.360</td>
<td>0.063</td>
<td>0.363</td>
<td>0.063</td>
<td>0.363</td>
<td>0.063</td>
</tr>
<tr>
<td></td>
<td>0.063</td>
<td>0.363</td>
<td>0.124</td>
<td>0.485</td>
<td>0.112</td>
<td>0.485</td>
<td>0.112</td>
</tr>
<tr>
<td></td>
<td>0.116</td>
<td>0.501</td>
<td>0.116</td>
<td>0.501</td>
<td>0.116</td>
<td>0.501</td>
<td>0.116</td>
</tr>
<tr>
<td></td>
<td>0.519</td>
<td>0.519</td>
<td>0.519</td>
<td>0.519</td>
<td>0.519</td>
<td>0.519</td>
<td>0.519</td>
</tr>
<tr>
<td></td>
<td>0.540</td>
<td>0.600</td>
<td>0.600</td>
<td>0.720</td>
<td>0.600</td>
<td>0.720</td>
<td>0.600</td>
</tr>
<tr>
<td></td>
<td>0.156</td>
<td>0.160</td>
<td>0.156</td>
<td>0.160</td>
<td>0.156</td>
<td>0.160</td>
<td>0.156</td>
</tr>
<tr>
<td></td>
<td>0.350</td>
<td>0.495</td>
<td>0.495</td>
<td>0.501</td>
<td>0.495</td>
<td>0.501</td>
<td>0.495</td>
</tr>
<tr>
<td></td>
<td>0.495</td>
<td>0.501</td>
<td>0.501</td>
<td>0.501</td>
<td>0.501</td>
<td>0.501</td>
<td>0.501</td>
</tr>
<tr>
<td></td>
<td>0.720</td>
<td>0.730</td>
<td>0.730</td>
<td>0.730</td>
<td>0.730</td>
<td>0.730</td>
<td>0.730</td>
</tr>
<tr>
<td></td>
<td>0.732</td>
<td>0.732</td>
<td>0.732</td>
<td>0.732</td>
<td>0.732</td>
<td>0.732</td>
<td>0.732</td>
</tr>
</tbody>
</table>

Table 2. Size of Shimotsu (2007) Wald test of the process $Y(t) = H(Y_t)$, where $y_t = \Delta^d x_t$, $X_t$ is generated according to an F(d) process governed by errors that are either Normal white noise, t(6)-distributed, Normal with AR(1) errors (0.4, 0.8), or Normal with Measurement errors. Sample sizes are $T = 500$, and $T = 3000$. Hermite rank of transformation $J = 3$. 
Table 3. Size of Shimotsu (2007) Wald test of the process $Y(t) = H_y(y_t)$, where $y_t = \Delta^{d-\delta} x_t$, $x_t$ is generated according to an $H(d)$ process governed by errors that are either Normal white noise, t(6)-distributed, Normal with AR(1) errors ($\Phi=0.4, 0.8$), or Normal with Measurement errors. Sample sizes are $T=500$, and $T=3000$. Hermite rank of transformation $J=4$. 

<table>
<thead>
<tr>
<th>$J=4$</th>
<th>Normal</th>
<th>Normal</th>
<th>t(6)</th>
<th>t(6)</th>
<th>Normal</th>
<th>Normal</th>
<th>Normal</th>
<th>Normal</th>
<th>Normal</th>
<th>Normal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Normal</td>
<td>Normal</td>
<td>t(6)</td>
<td>t(6)</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
</tr>
<tr>
<td></td>
<td>AR(.4)</td>
<td>AR(.4)</td>
<td>AR(.8)</td>
<td>AR(.8)</td>
<td>ME</td>
<td>ME</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$T$</td>
<td></td>
<td>500</td>
<td>3000</td>
<td>500</td>
<td>3000</td>
<td>500</td>
<td>3000</td>
<td>500</td>
<td>3000</td>
<td>500</td>
</tr>
<tr>
<td>$\delta$</td>
<td></td>
<td>0.050</td>
<td>0.300</td>
<td>0.047</td>
<td>0.054</td>
<td>0.018</td>
<td>0.024</td>
<td>0.052</td>
<td>0.048</td>
<td>0.046</td>
</tr>
<tr>
<td></td>
<td>0.350</td>
<td>0.048</td>
<td>0.040</td>
<td>0.023</td>
<td>0.026</td>
<td>0.054</td>
<td>0.040</td>
<td>0.048</td>
<td>0.040</td>
<td>0.040</td>
</tr>
<tr>
<td></td>
<td>0.400</td>
<td>0.048</td>
<td>0.043</td>
<td>0.022</td>
<td>0.024</td>
<td>0.042</td>
<td>0.046</td>
<td>0.046</td>
<td>0.045</td>
<td>0.036</td>
</tr>
<tr>
<td></td>
<td>0.450</td>
<td>0.052</td>
<td>0.041</td>
<td>0.024</td>
<td>0.024</td>
<td>0.046</td>
<td>0.042</td>
<td>0.048</td>
<td>0.043</td>
<td>0.041</td>
</tr>
<tr>
<td>0.100</td>
<td></td>
<td>0.048</td>
<td>0.046</td>
<td>0.027</td>
<td>0.024</td>
<td>0.045</td>
<td>0.056</td>
<td>0.046</td>
<td>0.038</td>
<td>0.040</td>
</tr>
<tr>
<td></td>
<td>0.350</td>
<td>0.046</td>
<td>0.047</td>
<td>0.036</td>
<td>0.022</td>
<td>0.052</td>
<td>0.050</td>
<td>0.046</td>
<td>0.044</td>
<td>0.034</td>
</tr>
<tr>
<td></td>
<td>0.400</td>
<td>0.049</td>
<td>0.037</td>
<td>0.025</td>
<td>0.024</td>
<td>0.056</td>
<td>0.044</td>
<td>0.050</td>
<td>0.041</td>
<td>0.033</td>
</tr>
<tr>
<td></td>
<td>0.450</td>
<td>0.050</td>
<td>0.046</td>
<td>0.022</td>
<td>0.022</td>
<td>0.040</td>
<td>0.042</td>
<td>0.056</td>
<td>0.040</td>
<td>0.046</td>
</tr>
<tr>
<td>0.150</td>
<td></td>
<td>0.045</td>
<td>0.050</td>
<td>0.026</td>
<td>0.019</td>
<td>0.046</td>
<td>0.053</td>
<td>0.044</td>
<td>0.056</td>
<td>0.043</td>
</tr>
<tr>
<td></td>
<td>0.350</td>
<td>0.048</td>
<td>0.053</td>
<td>0.027</td>
<td>0.025</td>
<td>0.046</td>
<td>0.050</td>
<td>0.047</td>
<td>0.044</td>
<td>0.042</td>
</tr>
<tr>
<td></td>
<td>0.400</td>
<td>0.052</td>
<td>0.046</td>
<td>0.022</td>
<td>0.026</td>
<td>0.052</td>
<td>0.048</td>
<td>0.047</td>
<td>0.058</td>
<td>0.048</td>
</tr>
<tr>
<td></td>
<td>0.450</td>
<td>0.044</td>
<td>0.056</td>
<td>0.029</td>
<td>0.027</td>
<td>0.056</td>
<td>0.052</td>
<td>0.050</td>
<td>0.048</td>
<td>0.047</td>
</tr>
<tr>
<td>0.175</td>
<td></td>
<td>0.050</td>
<td>0.048</td>
<td>0.030</td>
<td>0.022</td>
<td>0.050</td>
<td>0.048</td>
<td>0.051</td>
<td>0.050</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td>0.350</td>
<td>0.055</td>
<td>0.058</td>
<td>0.022</td>
<td>0.024</td>
<td>0.057</td>
<td>0.052</td>
<td>0.046</td>
<td>0.054</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td>0.400</td>
<td>0.059</td>
<td>0.054</td>
<td>0.022</td>
<td>0.026</td>
<td>0.052</td>
<td>0.054</td>
<td>0.056</td>
<td>0.047</td>
<td>0.052</td>
</tr>
<tr>
<td></td>
<td>0.450</td>
<td>0.058</td>
<td>0.046</td>
<td>0.024</td>
<td>0.020</td>
<td>0.046</td>
<td>0.051</td>
<td>0.057</td>
<td>0.052</td>
<td>0.060</td>
</tr>
<tr>
<td>0.200</td>
<td></td>
<td>0.053</td>
<td>0.043</td>
<td>0.023</td>
<td>0.028</td>
<td>0.052</td>
<td>0.058</td>
<td>0.052</td>
<td>0.052</td>
<td>0.074</td>
</tr>
<tr>
<td></td>
<td>0.350</td>
<td>0.066</td>
<td>0.053</td>
<td>0.028</td>
<td>0.024</td>
<td>0.052</td>
<td>0.058</td>
<td>0.056</td>
<td>0.058</td>
<td>0.064</td>
</tr>
<tr>
<td></td>
<td>0.400</td>
<td>0.057</td>
<td>0.058</td>
<td>0.027</td>
<td>0.020</td>
<td>0.059</td>
<td>0.066</td>
<td>0.052</td>
<td>0.062</td>
<td>0.057</td>
</tr>
<tr>
<td></td>
<td>0.450</td>
<td>0.058</td>
<td>0.048</td>
<td>0.024</td>
<td>0.027</td>
<td>0.054</td>
<td>0.058</td>
<td>0.050</td>
<td>0.053</td>
<td>0.066</td>
</tr>
</tbody>
</table>
### Table 4

Size of Shimotsu (2007) Wald test of the process $Y(t) = H_j(y_t)$, where $y_t = \Delta^{\delta d} x_t$, $X_t$ is generated according to an $Fi(d)$ process governed by errors that are either Normal white noise, $t(6)$-distributed, Normal with AR(1) errors ($\Phi=0.4, 0.8$), or Normal with Measurement errors.

Sample sizes are $T=500$, and $T=3000$. Hermite rank of transformation $J=5$.

<table>
<thead>
<tr>
<th>$J=5$</th>
<th>Normal</th>
<th>Normal</th>
<th>t(6)</th>
<th>t(6)</th>
<th>Normal</th>
<th>Normal</th>
<th>Normal</th>
<th>Normal</th>
<th>Normal</th>
<th>Normal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$T$</td>
<td>500</td>
<td>3000</td>
<td>500</td>
<td>3000</td>
<td>500</td>
<td>3000</td>
<td>500</td>
<td>3000</td>
<td>500</td>
<td>3000</td>
</tr>
<tr>
<td>$\delta$</td>
<td>0.050</td>
<td>0.300</td>
<td>0.040</td>
<td>0.053</td>
<td>0.020</td>
<td>0.030</td>
<td>0.052</td>
<td>0.051</td>
<td>0.060</td>
<td>0.047</td>
</tr>
<tr>
<td></td>
<td>0.350</td>
<td>0.052</td>
<td>0.045</td>
<td>0.022</td>
<td>0.024</td>
<td>0.024</td>
<td>0.045</td>
<td>0.046</td>
<td>0.050</td>
<td>0.052</td>
</tr>
<tr>
<td></td>
<td>0.400</td>
<td>0.053</td>
<td>0.042</td>
<td>0.024</td>
<td>0.022</td>
<td>0.022</td>
<td>0.050</td>
<td>0.043</td>
<td>0.050</td>
<td>0.038</td>
</tr>
<tr>
<td></td>
<td>0.450</td>
<td>0.052</td>
<td>0.036</td>
<td>0.026</td>
<td>0.022</td>
<td>0.022</td>
<td>0.050</td>
<td>0.038</td>
<td>0.053</td>
<td>0.047</td>
</tr>
<tr>
<td>$\delta$</td>
<td>0.100</td>
<td>0.300</td>
<td>0.042</td>
<td>0.041</td>
<td>0.024</td>
<td>0.022</td>
<td>0.056</td>
<td>0.042</td>
<td>0.050</td>
<td>0.042</td>
</tr>
<tr>
<td></td>
<td>0.350</td>
<td>0.049</td>
<td>0.042</td>
<td>0.030</td>
<td>0.023</td>
<td>0.023</td>
<td>0.060</td>
<td>0.048</td>
<td>0.043</td>
<td>0.049</td>
</tr>
<tr>
<td></td>
<td>0.400</td>
<td>0.050</td>
<td>0.048</td>
<td>0.022</td>
<td>0.024</td>
<td>0.024</td>
<td>0.046</td>
<td>0.044</td>
<td>0.052</td>
<td>0.040</td>
</tr>
<tr>
<td></td>
<td>0.450</td>
<td>0.049</td>
<td>0.038</td>
<td>0.022</td>
<td>0.024</td>
<td>0.024</td>
<td>0.044</td>
<td>0.034</td>
<td>0.048</td>
<td>0.041</td>
</tr>
<tr>
<td>$\delta$</td>
<td>0.150</td>
<td>0.300</td>
<td>0.052</td>
<td>0.043</td>
<td>0.030</td>
<td>0.016</td>
<td>0.044</td>
<td>0.044</td>
<td>0.050</td>
<td>0.046</td>
</tr>
<tr>
<td></td>
<td>0.350</td>
<td>0.047</td>
<td>0.043</td>
<td>0.028</td>
<td>0.026</td>
<td>0.026</td>
<td>0.050</td>
<td>0.043</td>
<td>0.054</td>
<td>0.040</td>
</tr>
<tr>
<td></td>
<td>0.400</td>
<td>0.044</td>
<td>0.048</td>
<td>0.020</td>
<td>0.019</td>
<td>0.021</td>
<td>0.053</td>
<td>0.046</td>
<td>0.042</td>
<td>0.045</td>
</tr>
<tr>
<td></td>
<td>0.450</td>
<td>0.050</td>
<td>0.050</td>
<td>0.027</td>
<td>0.031</td>
<td>0.031</td>
<td>0.056</td>
<td>0.048</td>
<td>0.056</td>
<td>0.046</td>
</tr>
<tr>
<td>$\delta$</td>
<td>0.175</td>
<td>0.300</td>
<td>0.048</td>
<td>0.048</td>
<td>0.028</td>
<td>0.022</td>
<td>0.053</td>
<td>0.046</td>
<td>0.051</td>
<td>0.050</td>
</tr>
<tr>
<td></td>
<td>0.350</td>
<td>0.048</td>
<td>0.043</td>
<td>0.024</td>
<td>0.027</td>
<td>0.027</td>
<td>0.047</td>
<td>0.041</td>
<td>0.051</td>
<td>0.042</td>
</tr>
<tr>
<td></td>
<td>0.400</td>
<td>0.040</td>
<td>0.044</td>
<td>0.023</td>
<td>0.023</td>
<td>0.023</td>
<td>0.052</td>
<td>0.039</td>
<td>0.054</td>
<td>0.044</td>
</tr>
<tr>
<td></td>
<td>0.450</td>
<td>0.056</td>
<td>0.040</td>
<td>0.023</td>
<td>0.023</td>
<td>0.023</td>
<td>0.052</td>
<td>0.050</td>
<td>0.059</td>
<td>0.040</td>
</tr>
<tr>
<td>$\delta$</td>
<td>0.200</td>
<td>0.300</td>
<td>0.042</td>
<td>0.042</td>
<td>0.026</td>
<td>0.026</td>
<td>0.044</td>
<td>0.048</td>
<td>0.046</td>
<td>0.040</td>
</tr>
<tr>
<td></td>
<td>0.350</td>
<td>0.054</td>
<td>0.050</td>
<td>0.023</td>
<td>0.024</td>
<td>0.024</td>
<td>0.048</td>
<td>0.046</td>
<td>0.049</td>
<td>0.048</td>
</tr>
<tr>
<td></td>
<td>0.400</td>
<td>0.050</td>
<td>0.039</td>
<td>0.026</td>
<td>0.022</td>
<td>0.022</td>
<td>0.064</td>
<td>0.048</td>
<td>0.050</td>
<td>0.054</td>
</tr>
<tr>
<td></td>
<td>0.450</td>
<td>0.050</td>
<td>0.039</td>
<td>0.020</td>
<td>0.025</td>
<td>0.025</td>
<td>0.056</td>
<td>0.042</td>
<td>0.049</td>
<td>0.038</td>
</tr>
<tr>
<td>RLS-NS</td>
<td>(T=500)</td>
<td></td>
<td></td>
<td></td>
<td>(T=3000)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\delta)</td>
<td>0.1</td>
<td>0.15</td>
<td>0.175</td>
<td>0.2</td>
<td>0.1</td>
<td>0.15</td>
<td>0.175</td>
<td>0.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(W_c(\delta, J=2))</td>
<td>0.065</td>
<td>0.087</td>
<td></td>
<td></td>
<td>0.069</td>
<td>0.092</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(W_c(\delta, J=3))</td>
<td>0.054</td>
<td>0.108</td>
<td>0.198</td>
<td></td>
<td></td>
<td>0.146</td>
<td>0.493</td>
<td>0.674</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(W_c(\delta, J=4))</td>
<td>0.051</td>
<td>0.065</td>
<td>0.086</td>
<td>0.114</td>
<td></td>
<td>0.060</td>
<td>0.076</td>
<td>0.092</td>
<td>0.135</td>
<td></td>
</tr>
<tr>
<td>(W_c(\delta, J=5))</td>
<td>0.037</td>
<td>0.064</td>
<td>0.099</td>
<td>0.120</td>
<td></td>
<td>0.051</td>
<td>0.122</td>
<td>0.185</td>
<td>0.324</td>
<td></td>
</tr>
<tr>
<td>QU</td>
<td></td>
<td></td>
<td></td>
<td>0.190</td>
<td></td>
<td>0.760</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORT</td>
<td></td>
<td></td>
<td></td>
<td>0.090</td>
<td></td>
<td>0.260</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S-SPLIT</td>
<td></td>
<td></td>
<td></td>
<td>0.170</td>
<td></td>
<td>0.510</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S-DIFF</td>
<td></td>
<td></td>
<td></td>
<td>0.370</td>
<td></td>
<td>0.740</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QP</td>
<td></td>
<td></td>
<td></td>
<td>0.190</td>
<td></td>
<td>0.510</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5. Finite sample power of tests at a 5% nominal level for the RLS-NS model. \(W_c(\delta, J)\) is the Wald test suggested in the paper for differencing parameter \(\delta\) and Hermite rank \(J\), QU is the test of Qu (2011) with a trimming parameter \(\varepsilon=0.02\), ORT is the test of Ohanissian, Russell and Tsay (2008), S-SPLIT and S-DIFF are the sample split and differencing tests of Shimotsu (2006), and QP is the Qu and Perron (2011) test. The model has been calibrated in accordance with Qu (2006).
### Table 6

Finite sample power of tests at a 5% nominal level for the MS-GARCH model. $W_c (d,J)$ is the Wald test suggested in the paper for differencing parameter $\delta$ and Hermite rank $J$, QU is the test of Qu(2011) with a trimming parameter $\varepsilon=.02$, ORT is the test of Ohanissian, Russell and Tsay (2008), S-SPLIT and S-DIFF are the sample split and differencing tests of Shimotsu (2006), and QP is the Qu and Perron (2011) test. The model has been calibrated in accordance with Qu (2006).
### Table 7.

Finite sample power of tests at a 5% nominal level for the MONO model. $W_c(\delta,J)$ is the Wald test suggested in the paper for differencing parameter $\delta$ and Hermite rank $J$, QU is the test of Qu (2011) with a trimming parameter $\varepsilon=.02$, ORT is the test of Ohanissian, Russell and Tsay (2008), S-SPLIT and S-DIFF are the sample split and differencing tests of Shimotsu (2006), and QP is the Qu and Perron (2011) test. The model has been calibrated in accordance with Qu (2006).

<table>
<thead>
<tr>
<th>MONO</th>
<th>T=500</th>
<th>T=3000</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta$</td>
<td>0.1</td>
<td>0.15</td>
</tr>
<tr>
<td>$W_c(\delta,J=2)$</td>
<td>0.278</td>
<td>0.290</td>
</tr>
<tr>
<td>$W_c(\delta,J=3)$</td>
<td>0.106</td>
<td>0.160</td>
</tr>
<tr>
<td>$W_c(\delta,J=4)$</td>
<td>0.079</td>
<td>0.124</td>
</tr>
<tr>
<td>QU</td>
<td>0.020</td>
<td>0.130</td>
</tr>
<tr>
<td></td>
<td>T=500</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
<td>-------</td>
</tr>
<tr>
<td>δ</td>
<td>0.1</td>
<td>0.15</td>
</tr>
<tr>
<td>RLS-NS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$W_c(\delta,J=2)$</td>
<td>0.065</td>
<td>0.075</td>
</tr>
<tr>
<td>$W_c(\delta,J=3)$</td>
<td>0.042</td>
<td>0.095</td>
</tr>
<tr>
<td>$W_c(\delta,J=4)$</td>
<td>0.050</td>
<td>0.064</td>
</tr>
<tr>
<td>$W_c(\delta,J=5)$</td>
<td>0.033</td>
<td>0.052</td>
</tr>
<tr>
<td>MS-GARCH</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$W_c(\delta,J=2)$</td>
<td>0.297</td>
<td>0.422</td>
</tr>
<tr>
<td>$W_c(\delta,J=3)$</td>
<td>0.126</td>
<td>0.326</td>
</tr>
<tr>
<td>$W_c(\delta,J=4)$</td>
<td>0.129</td>
<td>0.219</td>
</tr>
<tr>
<td>$W_c(\delta,J=5)$</td>
<td>0.073</td>
<td>0.173</td>
</tr>
<tr>
<td>MONO</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$W_c(\delta,J=2)$</td>
<td>0.466</td>
<td>0.872</td>
</tr>
<tr>
<td>$W_c(\delta,J=3)$</td>
<td>0.069</td>
<td>0.102</td>
</tr>
<tr>
<td>$W_c(\delta,J=4)$</td>
<td>0.660</td>
<td>0.277</td>
</tr>
<tr>
<td>$W_c(\delta,J=5)$</td>
<td>0.933</td>
<td>0.683</td>
</tr>
</tbody>
</table>

Table 8. Finite sample power of tests at a 5% nominal level for the RLS-NS model. $W_c(\delta,J)$ is the Wald test suggested in the paper for differencing parameter $\delta$ and Hermite rank $J$, QU is the test of Qu(2011) with a trimming parameter $\varepsilon=.02$, ORT is the test of Ohanissian, Russell and Tsay (2008), S-_SPLIT and S-DIFF are the sample split and differencing tests of Shimotsu (2006), and QP is the Qu and Perron (2011) test. The model has been calibrated to have high persistence, see main text.
Table 9. W(0.175, 3) and Qu(0.02) test for fractional long memory. The volatility and correlation series considered are for the IBM stock, January 1 2000 - June 30 2008: the realized correlation with realized volatility and log realized volatility, as well as American Express (AXP), Citi Group (Citi), General Electric (GE), Home Depot (HD) and J.P. Morgan (JPM)
Figure 1: Example of 4 simulated series: FI(d), RLS-NS, MS-GARCH, and MONO, calibrated to have d=0.4 for the FI(d) process, and a similar degree of long memory of the remaining series.
Figure 2: Autocorrelation function of the four series displayed in Figure 1.
Figure 3: Estimated spectral density of the four series displayed in Figure 1

Figure 4: Figure 4. Plot of the function $d_J = \max\{0, (d - 0.5)J + 0.5\}$ for transformations with Hermite ranks $J = 1, 2, 3, 4, 5$. 
Figure 5: Mean values of local Whittle estimates of FI(d) processes, x-axis, and the estimate of d(J), y-axis, for a Hermite polynomial transform with J=2. The curves are based on 1000 MC replications.

Figure 6: Mean values of local Whittle estimates of FI(d) processes, x-axis, and the estimate of d(J), y-axis, for a Hermite polynomial transform with J=3. The curves are based on 1000 MC replications.
Figure 7: Mean values of local Whittle estimates of FI(d) processes, x-axis, and the estimate of d(J), y-axis, for a Hermite polynomial transform with J=4. The curves are based on 1000 MC replications.
Figure 8: Mean values of local Whittle estimates of FI(d) processes, x-axis, and the estimate of d(J), y-axis, for a Hermite polynomial transform with J=5. The curves are based on 1000 MC replications.
Figure 9: Size of W test as a function of delta for sample sizes T=500, 3000, and d=.35, .45. J=2,3,4,5.