ELSEVIER

Journal of Econometrics 110 (2002) 113 – 133

JOURNAL OF Econometrics

www.elsevier.com/locate/econbase

Properties of nonlinear transformations of fractionally integrated processes

Ingolf Dittmann[a,*], Clive W. J. Granger[b]

[a]School of Business and Economics, Humboldt University, Berlin, 10178 Berlin, Germany
[b]Department of Economics, University of California, San Diego, La Jolla, CA 92093-0508, USA

Abstract

This paper shows that the properties of nonlinear transformations of a fractionally integrated process strongly depend on whether the initial series is stationary or not. Transforming a stationary Gaussian $I(d)$ process with $d>0$ leads to a long-memory process with the same or a smaller long-memory parameter depending on the Hermite rank of the transformation. Any nonlinear transformation of an antipersistent Gaussian $I(d)$ process is $I(0)$). For non-stationary $I(d)$ processes, every polynomial transformation is non-stationary and exhibits a stochastic trend in mean and in variance. In particular, the square of a non-stationary Gaussian $I(d)$ process still has long memory with parameter d, whereas the square of a stationary Gaussian $I(d)$ process shows less dependence than the initial process. Simulation results for other transformations are also discussed.

JEL classification: C22

Keywords: Long memory; Nonlinearity; Non-stationarity; Fractional integration

1. Introduction

Two of the most popular topics in univariate time-series analysis in recent years have concerned non-stationary series and nonlinear series. Both topics remained separate research areas, however, so that only little is known about non-stationary *and* nonlinear series. This paper ventures into this largely uncharted territory, by considering the properties of $g(X_t)$ for various nonlinear functions $g(.)$ where X_t is fractionally integrated, or $I(d)$. The advantage of studying fractionally integrated processes is that they can be stationary or non-stationary depending on the long-memory parameter d.

* Corresponding author.
E-mail address: dittmann@wiwi.hu-berlin.de (I. Dittmann).

0304-4076/02/$ - see front matter
PII: S0304-4076(02)00089-1

By simply increasing d, we can therefore move from stationarity to non-stationarity and, in particular, study the threshold between stationarity and non-stationarity.

To our knowledge, nonlinear transformations of fractionally integrated processes have been considered only in two papers before: Taqqu (1979) and Giraitis and Surgailis (1985) study nonlinear transformations of fractional Brownian motions, the continuous-time equivalent of fractionally integrated time-series. For $0 \leqslant d < 1/2$, they derive the long-memory properties of the transformed process. In this paper, we derive similar results for discrete time fractionally integrated processes and extend them to the whole stationary parameter space $d < 1/2$. In addition, we provide some novel results for the non-stationarity region. We also corroborate and extend our theoretical findings with Monte-Carlo simulations.

It turns out that the properties of the transformed series critically depend on the characteristics of the input series. As usual, we distinguish three main types of fractional processes:

(i) $d < 0$ are stationary and antipersistent.
(ii) $0 < d < 1/2$ are stationary and have long memory.
(iii) $1/2 \leqslant d < 1$ are non-stationary and have long memory.

For class (i), we find that the antipersistency property is lost for any nonlinear transformation $g(.)$. In finite samples, however, antipersistency is not completely lost but only reduced if $g(.)$ is an odd function. For class (ii) the long memory of $g(X_t)$ is sensitive to the Hermite rank of the transformation $g(.)$, with the higher the rank the lower the memory. For the third class we establish that an integer power transformation $g(x) = x^n$ generates trends in mean and in variance but does not change the process' long memory. In contrast, periodic transformations, e.g., $g(x) = \cos(x)$, markedly reduce the initial long memory of a non-stationary fractionally integrated process.

Our results can be well illustrated by considering the square of a Gaussian $I(d)$ process: If $d \leqslant 1/4$, the squared process is $I(0)$. If $1/4 < d < 1/2$, the squared time-series has long memory with parameter $2d - 1/2$. If $1/2 \leqslant d < 1$, the squared series has long-memory with parameter d. Furthermore, for $d \geqslant 1/2$, the mean of the squared series has a trend of the form t^{2d-1}, whereas the variance diverges with t^{4d-2}.

The paper is organized as follows: Section 2 considers the stationary cases (i) and (ii). The non-stationary case (iii) is studied in Sections 3 and 4 for integer power transformations. In particular, Section 3 is concerned with the trending behavior of mean and variance of $g(X_t)$, and Section 4 considers the autocorrelation structure of $g(X_t)$. Section 5 presents simulation results for periodic and exponential transformations. Finally, Section 6 briefly considers long-memory properties of transformations of processes with stochastic breaks in mean.

2. Transformations of stationary processes

We start this section with a couple of basic definitions which will be used throughout the paper.

Fractional integration: A time-series $\{X_t\}$ is called fractionally integrated with differencing parameter d, or short-hand $X_t \sim I(d)$, if

$$X_t = \sum_{j=0}^{\infty} c_j \varepsilon_{t-j} \quad \text{with } c_j = \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)} \quad \text{and} \quad \varepsilon_t \sim \text{iid}(0, \sigma^2). \tag{1}$$

If $\varepsilon_t \sim \text{iid N}(0, \sigma^2)$, $\{X_t\}$ is called Gaussian fractionally integrated or Gaussian $I(d)$.

Stationarity: We call a time-series $\{X_t\}$ stationary if its mean and variance are finite and do not depend on the time index t. If, in addition, the instantaneous (or theoretical) autocovariance function $\gamma(t,h) = \text{Cov}(X_t, X_{t-h})$ does not depend on t, $\{X_t\}$ is called covariance stationary.

It is important to note the difference between the instantaneous autocorrelation function $\gamma(t,h)$ and the empirical autocorrelation function

$$\tilde{\gamma}(h) = \frac{1}{T} \sum_{t=1}^{T-h} (X_t - \bar{X})(X_{t+h} - \bar{X}),$$

where T is the sample size and $\bar{X}$ is the sample mean. If $\{X_t\}$ is not covariance stationary, $\gamma(t,h)$ and $\tilde{\gamma}(h)$ are clearly different concepts. All our theoretical results will be about instantaneous autocorrelations, whereas the simulation results are based on empirical autocorrelations.

Spectral density at frequency zero: The spectral density of a covariance stationary time-series $\{X_t\}$ is given by

$$f(\lambda) = \gamma(0) + 2 \sum_{h=1}^{\infty} \gamma(h) \cos(\lambda h). \tag{2}$$

If $f(\lambda) \sim \lambda^{-2d}$ as λ approaches zero ($d < 1/2$), we write $X_t \sim \text{LM}(d)$. For $d \geqslant 1/2$, we define $X_t \sim \text{LM}(d) :\Leftrightarrow (1-L)^k X_t \sim \text{LM}(d-k)$ for $k = [d + 1/2]$, where $[x]$ denotes the largest integer smaller or equal to x. Hence, $\text{LM}(d)$ can be thought of as a generalization of $I(d)$, because $X_t \sim I(d) \Rightarrow X_t \sim \text{LM}(d)$.

Long memory: If $X_t \sim \text{LM}(d)$ with $0 < d < 1$, we call $\{X_t\}$ a long-memory process with parameter d. Note that any covariance stationary time-series with hyperbolically decreasing autocovariance function of the form $\gamma(h) \sim h^{2d-1}$, $0 < d < 1/2$, is $\text{LM}(d)$, i.e., the decay of the autocorrelation function uniquely determines the size of the process's long memory. If $d = 0$, $\{X_t\}$ is called a short-memory process.

Antipersistence: A covariance stationary time-series $\{X_t\}$ is called antipersistent if its spectral density at frequency zero is zero. Hence, an $I(d)$ process with $d < 0$ is antipersistent. Note, however, that the hyperbolical decay of the autocorrelations of such an $I(d)$ process is neither necessary nor sufficient for the presence of antipersistence.

Our argument for the stationary case is based on a Hermite polynomial decomposition (see Granger and Newbold, 1976, for a detailed introduction). Taqqu (1979) and Giraitis and Surgailis (1985) first use Hermite polynomial decompositions in conjunction with long-memory processes: They derive limit theorems for functionals of fractional Brownian motion.

Hermite Polynomials: For $j = 0, 1, 2, \ldots,$ we define the Hermite polynomials $H_j(x)$ by

$$\left(\frac{\mathrm{d}}{\mathrm{d}x}\right)^j \mathrm{e}^{-x^2/2} = (-1)^j \sqrt{j!} H_j(x) \mathrm{e}^{-x^2/2}. \tag{3}$$

$H_j(x)$ is a polynomial of degree j; the first five Hermite polynomials are $H_0(x) = 1$, $H_1(x) = x$, $H_2(x) = (x^2 - 1)/\sqrt{2}$, $H_3(x) = (x^3 - 3x)/\sqrt{6}$, $H_4(x) = (x^4 - 6x^2 + 3)/\sqrt{24}$. Note that every continuous function on a compact interval can be approximated arbitrarily well by a sum of Hermite polynomials. Hermite polynomials are closely connected to the standard normal distribution. Let X be a random variable with standard normal density function $\phi(x)$. Then (see, e.g., Cramér, 1946, pp. 131–133)

$$\int_{-\infty}^{\infty} H_m(x) H_n(x) \phi(x)\, \mathrm{d}x = \begin{cases} 1, & \text{for } m = n, \\ 0, & \text{for } m \neq n, \end{cases} \tag{4}$$

which immediately implies $\mathrm{E}(H_j(X)) = 0$ and $\mathrm{Var}(H_j(X)) = 1$ for all $j > 0$. Moreover, if X_t and X_{t-h} are jointly normally distributed with $\mathrm{E}(X_t) = 0 = \mathrm{E}(X_{t-h})$, $\mathrm{Var}(X_t) = 1 = \mathrm{Var}(X_{t-h})$ and $\mathrm{Cov}(X_t, X_{t-h}) = \rho_h$, the joint density can be written as

$$f(x_t, x_{t-h}) = \phi(x_t)\phi(x_{t-h}) \left\{ 1 + \sum_{j=1}^{\infty} \rho_h^j H_j(x_t) H_j(x_{t-h}) \right\} \tag{5}$$

(see, e.g., Barrett and Lampard, 1955).

With this result, we can easily calculate the autocorrelation function of any transformation of a stationary Gaussian process which can be written as the sum of Hermite polynomials:

Lemma 1. *Let $\{X_t\}$ be a covariance stationary Gaussian process with standard normal marginal distribution and autocorrelation function $\rho_h = \mathrm{Corr}(X_t, X_{t-h})$. Let $g(.)$ be a univariate transformation which can be written as a sum of Hermite polynomials $H_j(.)$:*

$$g(x) = g_0 + \sum_{j=1}^{\infty} g_j H_j(x). \tag{6}$$

Then the autocorrelation function of $g(X_t)$ is given by

$$\mathrm{Corr}(g(X_t), g(X_{t+h})) = \frac{\sum_{j=1}^{\infty} g_j^2 \rho_h^j}{\sum_{j=1}^{\infty} g_j^2}. \tag{7}$$

Proof. With (5) and (6), we obtain: $\mathrm{E}(g(X_t), g(X_{t-h})) = \mathrm{E}(g(X_t))^2 + \sum_{j=1}^{\infty} g_j^2 \rho_h^j$, which implies $\mathrm{Cov}(g(X_t), g(X_{t-h})) = \sum_{j=1}^{\infty} g_j^2 \rho_h^j$. With $\mathrm{Var}(g(X_t)) = \sum_{j=1}^{\infty} g_j^2$ the result follows immediately. $\square$

In the Hermite polynomial decomposition (6), the number $J \in \mathbb{N}$ for which $g_j = 0 \ \forall 1 \leqslant j < J$ and $g_J \neq 0$ is called the *Hermite rank* of $g(.)$. Hence, the

Hermite rank is the number of the "smallest" non-constant Hermite polynomial needed in the decomposition of $g(x)$. The Hermite rank turns out to be of central importance for the long-memory properties of the transformed series:

Proposition 1. *Let $\{X_t\}$ be a stationary Gaussian $I(d)$ process with* $\mathrm{E}(X_t)=0$ *and* $\mathrm{Var}(X_t)=1$. *Let $g(.)$ be a univariate transformation which can be written as the finite sum of Hermite polynomials $H_j(.)$:*

$$g(x) = g_0 + \sum_{j=J}^{\bar{J}} g_j H_j(x) \quad \text{with } 1 \leqslant J < \bar{J} < \infty \quad \text{and} \quad g_J \neq 0. \tag{8}$$

(a) *If $0 < d < 0.5$, then $g(X_t)$ is a long-memory process* LM($\tilde{d}$) *with* $\tilde{d} = \max\{0, (d-0.5)J + 0.5\}$.
(b) *If $-1 < d < 0$ and if $g(.)$ is nonlinear, then $g(X_t)$ is a short-memory process* LM(0).

Proof.
(a) The autocorrelation function of $\{X_t\}$ decays hyperbolically, i.e., for large h, $\rho_h \approx [\Gamma(1-d)/\Gamma(d)]h^{2d-1}$. With Lemma 1 we therefore obtain

$$\mathrm{Corr}(g(X_t), g(X_{t-h})) = \frac{\sum_{j=J}^{\bar{J}} g_j^2 \rho_h^j}{\sum_{j=J}^{\bar{J}} g_j^2} \underset{\text{for large } h}{\approx} \sum_{j=J}^{\bar{J}} a_j h^{(2d-1)j} \tag{9}$$

$$\text{with } a_j = \frac{g_j^2}{\sum_{k=J}^{\bar{J}} g_k^2} \left(\frac{\Gamma(1-d)}{\Gamma(d)} \right)^J.$$

As h increases, the term $a_J h^{(2d-1)J}$ dominates all other terms $a_j h^{(2d-1)j}$, $j > J$, in (9), so the autocorrelations of $g(X_t)$ also decrease hyperbolically

$$\mathrm{Corr}(g(X_t), g(X_{t-h})) \underset{\text{for large } h}{\approx} a_J h^{(2d-1)J} = a_J h^{2\bar{d}-1} \quad \text{with } \bar{d} = (d-0.5)J + 0.5.$$

Note that $a_J h^{(2d-1)J}$ need not dominate the sum on the right-hand side of (9) if $\bar{J}$ is replaced by ∞. Therefore, Proposition 1(a) need not hold for transformations with infinite Hermite expansion. Section 5 presents a counter example.

Case 1: $\bar{d} \geqslant 0$. In this case, $g(X_t)$ is LM($\bar{d}$), since a long-memory process is exclusively determined by the decay pattern of its autocorrelations. It is this result, Taqqu (1979) established for functionals of fractional Brownian motion.
Case 2: $\bar{d} < 0$. In this case, $g(X_t)$ has the same autocorrelation decay pattern as an $I(\bar{d})$ process, but all autocorrelations of $g(X_t)$ are positive (because all autocorrelations of X_t are positive) so that

$$f_g(0) = \mathrm{Var}(g(X_t)) + 2\sum_{h=1}^{\infty} \mathrm{Cov}(g(X_t), g(X_{t-h})) > 0.$$

Hence, $g(X_t)$ is not antipersistent and thus an $I(0)$-process.

(b) If $-1 < d < 0$, then $\rho_h < 0$ for all $h > 0$ and $f(0) = 1 + 2\sum_{h=1}^{\infty} \rho_h = 0$. Therefore, $1 + 2\sum_{h=1}^{\infty} \rho_h^j > 0$ for all $j > 1$. With Lemma 1, we obtain $f_g(0) = \sum_{j=J}^{\bar{J}} g_j^2 + 2\sum_{h=1}^{\infty}\sum_{j=J}^{\bar{J}} g_j^2 \rho_h^j$, which is positive if there is a $j > 1$ with $g_j \neq 0$. Hence, $g(X_t)$ is not antipersistent but LM(0) if $g(.)$ is nonlinear. □

Proposition 1(a) states that every transformation $g(.)$ with Hermite rank larger than one reduces the long memory of the transformed stationary process. The larger the original long-memory parameter, the smaller is the reduction of the long memory caused by such a transformation. In the limit case $d = 1/2$, the size of long memory stays the same for any transformation. The non-stationary case ($d \geqslant 1/2$) will be considered in Section 4.

A special case of Proposition 1(a) is that the square of a Gaussian $I(d)$ process with $d \in (0, 0.25]$ is $I(0)$. For this case, a more intuitive proof can be given. First note that for any two jointly normally distributed random variables Z_1 and Z_2 with $\mathrm{E}(Z_1) = 0 = \mathrm{E}(Z_2)$ it can be shown that $\mathrm{Corr}(Z_1^2, Z_2^2) = [\mathrm{Corr}(Z_1, Z_2)]^2$. Hence, $\mathrm{Corr}(X_t^2, X_{t-h}^2) = \rho_h^2$ which decays like $h^{2(2d-0.5)-1}$ if X_t is $I(d)$. Only if $d > 0.25$ does X_t^2 still have the long-memory property (with parameter $2d - 0.5$).

The second part of Proposition 1 shows that antipersistence is a much more fragile property than long memory. In theory, it is immediately lost for any nonlinear transformation. In practice, we can expect this effect to be larger for even functions $g(.)$ than for odd functions. To see this, consider an $I(d)$ process with $-1 < d < 0$. This process has hyperbolically decaying negative autocorrelations which sum up to $-1/2$, so that the spectral density at frequency zero is zero. If we consider the square, which has Hermite rank 2, (or any other even transformation) of this process, all its autocorrelations are positive (due to (7)), so that the spectral density at frequency zero is considerably larger than zero. In contrast, the autocorrelations of an odd transformation, e.g., of the cube, which has Hermite rank 1, are all still negative. The spectral density at frequency zero is positive, but it can be expected to be much closer to zero than in the case of an even transformation. Therefore, an odd transformation of an antipersistent process might still seem antipersistent in practice.

Another interesting implication of Proposition 1 is that a stationary Gaussian $I(d)$ process can be fractionally cointegrated with a nonlinear function of itself: Let $X_t \sim I(d)$, then $\{3X_t - X_t^3\}$ has a smaller long-memory parameter than $\{X_t\}$ itself. Note that such "cointegration with itself" is impossible for $I(1)$ series: Granger and Hallman (1991) show that $X_t \sim I(1)$ cannot be cointegrated with any nonlinear transformation $g(X_t)$.

Table 1 shows the results of a small simulation study. For four values of the long-memory parameter d (-0.4, -0.2, 0.2 and 0.4) we simulated 2,000 Gaussian $I(d)$ processes with 2,000 observations each, using the algorithm proposed by Hosking (1984). Then we transformed these series with several transformations $g(.)$ and estimated the long-memory parameter of the transformed series with a periodogram regression over the $m = [T^{0.8}]$ smallest Fourier frequencies (see Geweke and Porter–Hudak, 1983; Hurvich et al., 1998). Each cell shows the average estimate of the long-memory parameter d, its empirical standard error in parentheses and the corresponding theoretical value. The first column displays the transformations $g(.)$ and their respective

Table 1
Average estimated long-memory parameter of some transformations of 2,000 simulated stationary Gaussian $I(d)$ processes with 2,000 observations each

$g(X)$ and its Hermite rank		Long-memory parameter of the original series X			
		$d=-0.4$	$d=-0.2$	$d=0.2$	$d=0.4$
X	Theory	−0.4	−0.2	0.2	0.4
(rank 1)	Simulation	−0.40 (0.033)	−0.20 (0.032)	0.20 (0.032)	0.40 (0.032)
X^2	Theory	0	0	0	0.3
(rank 2)	Simulation	0.01 (0.032)	0.01 (0.032)	0.04 (0.037)	0.29 (0.060)
X^3	Theory	0	0	0.2	0.4
(rank 1)	Simulation	−0.13 (0.033)	−0.09 (0.032)	0.14 (0.033)	0.32 (0.045)
X^4	Theory	0	0	0	0.3
(rank 2)	Simulation	0.01 (0.033)	0.01 (0.032)	0.03 (0.039)	0.24 (0.071)
X^3-3X	Theory	0	0	0	0.2
(rank 3)	Simulation	−0.01 (0.033)	−0.00 (0.033)	0.01 (0.036)	0.19 (0.061)
X^4-6X^2	Theory	0	0	0	0.1
(rank 4)	Simulation	0.00 (0.32)	−0.00 (0.032)	0.00 (0.035)	0.12 (0.060)

Hermite ranks, which can be verified easily using the first five Hermite polynomials given above (e.g., $X^3=\sqrt{6}H_3(X)+3H_1(X)$ has Hermite rank 1).

The simulation results clearly confirm Proposition 1. The average estimates differ from the theoretical value by more than two standard errors only for negative values of d and the transformation $g(X)=X^3$. As we argued above, this is due to the fact that an odd transformation (especially one with Hermite rank 1) of an antipersistent process, disturbs (and thereby destroys) the antipersistence only slightly, so that it is likely to still look like an antipersistent process in finite samples. It is interesting to note that the standard errors are very small for negative values of d and for the identity transformation $g(X)=X$. (The asymptotic standard error of the estimation procedure is 0.031.) On the other hand, the standard error seems to increase with the size of the long memory for all other transformations and $d>0$.

3. Moments of fractionally integrated processes

In this and the following section, we turn to transformations of non-stationary fractionally integrated processes. In contrast to the stationary case considered in the previous section, the mean of the output series and the variance of both input and output series can exhibit stochastic trends. In this section, we derive the divergence properties of the moments of non-stationary fractionally integrated processes. With this result, the trends in mean and variance of polynomial transformations of fractionally integrated processes can be derived easily.

Note that the traditional definition of fractional integration, as given in (1), simply implies that the variance (and higher even moments) does not exist if $d>0.5$. The reason is that the process in (1) has an infinite past. To circumvent this problem, we consider finite-past processes in this and the next section:

Finite-past $I(d)$ process: A time-series $\{\tilde{X}_t\}$ is called fractionally integrated with finite past and differencing parameter d, if

$$\tilde{X}_t = \sum_{j=0}^{t} c_j \varepsilon_{t-j} \quad \text{with } c_j = \frac{\Gamma(j+d)}{\Gamma(j+1)\Gamma(d)} \quad \text{and} \quad \varepsilon_t \sim \text{iid}(0,\sigma^2). \tag{10}$$

We write $\tilde{X}_t \sim \tilde{I}(d)$. Note that for $\tilde{X}_t \sim \tilde{I}(d)$ and $Y_t \sim I(d)$, $\{\tilde{X}_t\} \xrightarrow{t\to\infty} \{Y_t\}$ (in distribution). If $\varepsilon_t \sim \text{iid}\,\mathrm{N}(0,\sigma^2)$, $\{\tilde{X}_t\}$ is called Gaussian $\tilde{I}(d)$.

Proposition 2. *Let $\tilde{X}_t \sim \tilde{I}(d)$ with innovations $\{\varepsilon_t\}$ whose moments $\mu_m^\varepsilon \equiv \mathrm{E}(\varepsilon_t^m)$ are finite.*

(a) *If $d < 0.5$, all moments of $\tilde{X}_t$ converge as $t \to \infty$.*
(b) *If $d > 0.5$, all even moments of $\tilde{X}_t$ diverge. More explicitly, for $m \in \{2,4,6,\dots\}$ $\mu_m^X(t) \sim O(t^{m(d-0.5)})$.*
(c) *If m is odd and $d < (m+1)/2m$, the mth moment $\mu_m^X(t)$ converges as $t \to \infty$.*
(d) *If $d > 2/3$ and $\mu_3^\varepsilon \neq 0$, all odd moments diverge, i.e., for $m \in \{3,5,7,\dots\}$ $\mu_m^X(t) \sim O(t^{m(d-0.5)-0.5})$.*

Proof.

$$\mu_2^X(t) = \mathrm{E}(\tilde{X}_t^2) = \mathrm{E}\left(\left(\sum_{j=0}^{t} c_j \varepsilon_{t-j}\right)^2\right)$$

$$= \sum_{j=0}^{t} c_j^2 \mu_2^\varepsilon \underset{\text{for large } t}{\approx} c + \frac{\mu_2^\varepsilon}{\Gamma(d)^2} \sum_{j=0}^{t} j^{2(d-1)} \sim c + O(t^{2d-1}),$$

where c is some constant. For large t, the approximation $c_t \approx t^{d-1}/\Gamma(d)$ has been used. The second approximation, $\sum_{k=1}^{n} k^q \sim O(n^{q+1})$, is an immediate implication of formula 0.121 given by Gradshteyn and Ryzhik (1980).

For higher moments, we proceed analogously

$$\mu_3^X(t) = \mathrm{E}(\tilde{X}_t^3) = \mathrm{E}\left(\left(\sum_{j=0}^{t} c_j \varepsilon_{t-j}\right)^3\right)$$

$$= \sum_{j=0}^{t} c_j^3 \mu_3^\varepsilon \underset{\text{for large } t}{\approx} c + \frac{\mu_3^\varepsilon}{\Gamma(d)^3} \sum_{j=0}^{t} j^{3(d-1)} \sim c + O(t^{3d-2}),$$

$$\mu_4^X(t) = \mathrm{E}(\tilde{X}_t^4) = \mathrm{E}\left(\left(\sum_{j=0}^{t} c_j \varepsilon_{t-j}\right)^4\right) = \sum_{j=0}^{t} c_j^4 \mu_4^\varepsilon + \binom{4}{2} \sum_{j1=0}^{t-1} \sum_{j2=j1+1}^{t} c_{j1}^2 \mu_2^\varepsilon c_{j2}^2 \mu_2^\varepsilon$$

$$= \sum_{j=0}^{t} c_j^4 \mu_4^\varepsilon + 6 \sum_{j1=1}^{t} \sum_{j2=1}^{t} c_{j1}^2 c_{j2}^2 (\mu_2^\varepsilon)^2 - 6 \sum_{j=0}^{t} c_j^4 (\mu_2^\varepsilon)^2$$

$$= (\mu_4^\varepsilon - 6(\mu_2^\varepsilon)^2)\sum_{j=0}^{t} c_j^4 + 6\left(\sum_{j=0}^{t} c_j^2\right)^2 (\mu_2^\varepsilon)^2$$

$$\sim c + O(t^{4d-3}) + O(t^{2(2d-1)}) = c + O(t^{4d-2}),$$

$$\mu_5^X(t) = \sum_{j=0}^{t} c_j^5\mu_5^\varepsilon + \binom{5}{2}\sum_{j1=0}^{t-1}\sum_{j2=j1+1}^{t} c_{j1}^2\mu_2^\varepsilon c_{j2}^3\mu_3^\varepsilon$$

$$= (\mu_5^\varepsilon - 30\mu_2^\varepsilon\mu_3^\varepsilon)\sum_{j=0}^{t} c_j^5 + 30\sum_{j=0}^{t} c_j^2\sum_{j=0}^{t} c_j^3\mu_2^\varepsilon\mu_3^\varepsilon$$

$$\sim c + O(t^{5d-4}) + O(t^{2d-1})O(t^{3d-2}) = c + O(t^{5d-3}),$$

$$\mu_6^X(t) = \sum_{j=0}^{t} c_j^6\mu_6^\varepsilon + \binom{6}{2}\sum_{j1}\sum_{j2} c_{j1}^2 c_{j2}^4\mu_2^\varepsilon\mu_4^\varepsilon + \binom{6}{3}\sum_{j1}\sum_{j2} c_{j1}^3 c_{j2}^3(\mu_3^\varepsilon)^2$$

$$+ \frac{6!}{2!2!2!}\sum_{j1}\sum_{j2}\sum_{j3} c_{j1}^2 c_{j2}^2 c_{j3}^2(\mu_2^\varepsilon)^3.$$

As the divergence rate of $(\sum c_j^2)^3$ is larger than the divergence rates of $\sum c_j^6$, $\sum c_j^4 \sum c_j^2$ and $(\sum c_j^3)^2$, the last term dominates the other three terms. Hence, $\mu_6^X(t) \sim c + O(t^{3(2d-1)}) = c + O(t^{6d-3})$. For all higher even moments, the same argument holds: $\prod_k \sum c_j^{m(k)}$ given $\sum_k m(k) = m$ has the maximal divergence rate if $m(k) = 2$ for all k.

$$\mu_7^X(t) = \sum_{j=0}^{t} c_j^7\mu_7^\varepsilon + \binom{7}{2}\sum_{j1}\sum_{j2} c_{j1}^2 c_{j2}^5\mu_2^\varepsilon\mu_5^\varepsilon + \binom{7}{3}\sum_{j1}\sum_{j2} c_{j1}^3 c_{j2}^4\mu_3^\varepsilon\mu_4^\varepsilon$$

$$+ \frac{7!}{2!2!3!}\sum_{j1}\sum_{j2}\sum_{j3} c_{j1}^2 c_{j2}^2 c_{j3}^3(\mu_2^\varepsilon)^2\mu_3^\varepsilon.$$

Here, the last term has the largest divergence rate: $\mu_7^X(t) \sim c + O(t^{7d-6}) + O(t^{2d-1})$ $O(t^{5d-4}) + O(t^{3d-2})O(t^{4d-3}) + O(t^{2d-1})O(t^{2d-1})O(t^{3d-2}) = c + O(t^{7d-4})$. For higher odd moments $m > 3$, the same argument holds: The dominating terms will always be of the form $(\sum c_j^2\mu_2^\varepsilon)^{(m-3)/2}\sum c_j^3\mu_3^\varepsilon \sim c + O(t^{0.5(m-3)(2d-1)})O(t^{3d-2}) = c + O(t^{m(d-0.5)-0.5})$. □

From Proposition 2, we can immediately obtain the trend in mean of any power transformation (i.e., $X^2, X^3, \ldots$) of a non-stationary $\tilde{I}(d)$ process. For instance, if $\tilde{X}_t \sim \tilde{I}(d)$, $\tilde{X}_t^2$ has a trend of the form t^{2d-1} and $\tilde{X}_t^4$ has a trend of the form t^{4d-2}. Figs. 1 and 2 illustrate this for the $\tilde{I}(0.8)$ process. For these two plots, we simulated 2,000 $\tilde{I}(0.8)$ processes, transformed them by taking the second and fourth power, respectively, and then averaged over all 2,000 processes.

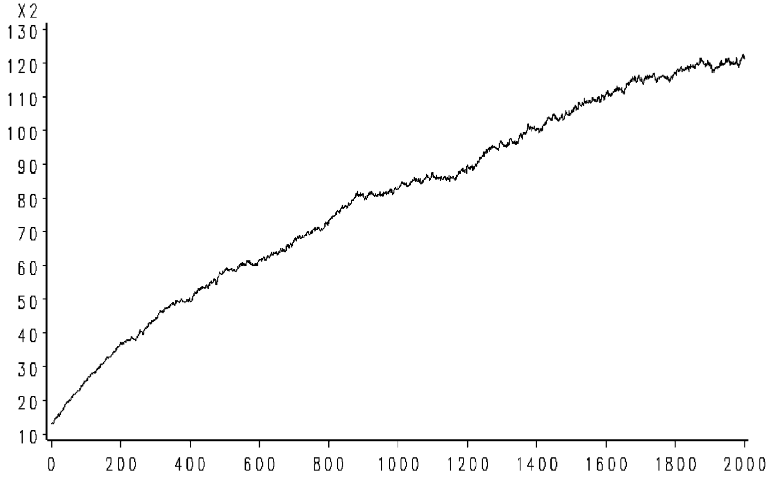

Fig. 1. Average trend in mean of 2,000 squared $\tilde{I}(0.8)$ time-series.

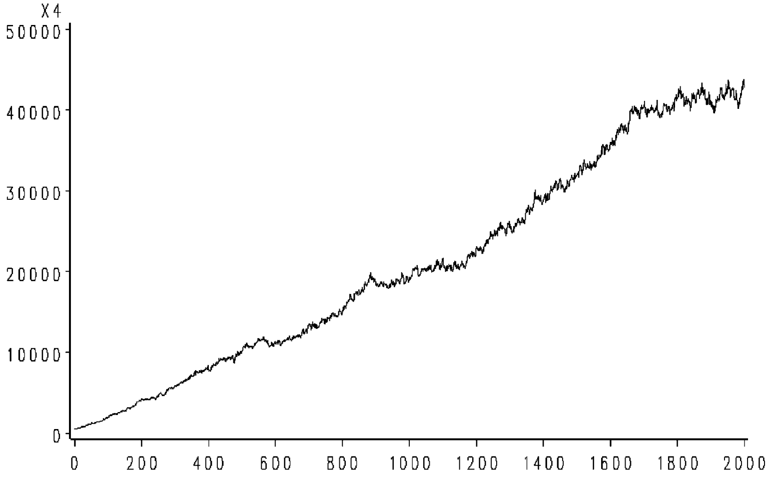

Fig. 2. Average trend in mean of 2,000 simulated $\tilde{I}(0.8)$ time-series raised to the fourth power.

Whereas the trend can be clearly seen in the aggregates of Figs. 1 and 2, this is much less so for the individual series as Figs. 3 and 4 demonstrate. The reason is that $\tilde{X}_t^2$ does not only have a trend in mean of the order t^{2d-1}, but it also has a trend in variance of the form t^{4d-2}. For all power transformations $\tilde{X}_t^m$, the trend in variance which is of order $t^{m(2d-1)}$ dominates the trends in all other moments in the sense that multiplying $\tilde{X}_t^m$ by $t^{-m(d-0.5)}$ removes the trend in all moments. Hence, the non-stationary process $\tilde{X}_t^m$ can be transformed into an asymptotically stationary process by simply multiplying it with $t^{-m(d-0.5)}$.

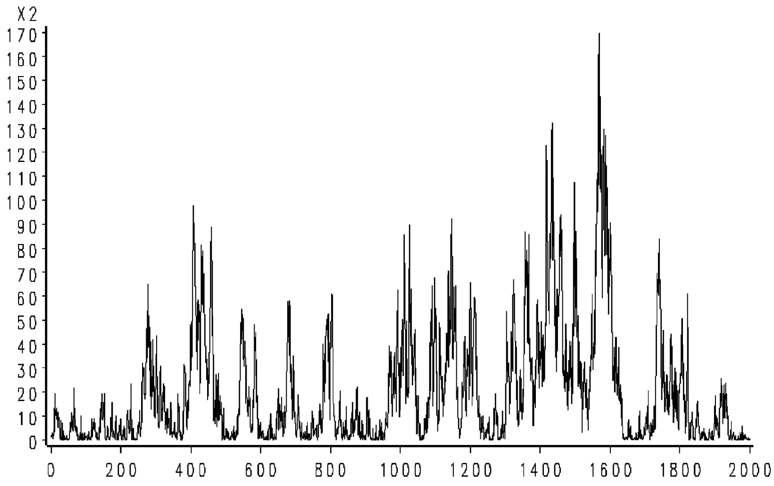

Fig. 3. Square of a single simulated $\tilde{I}(0.8)$ time-series.

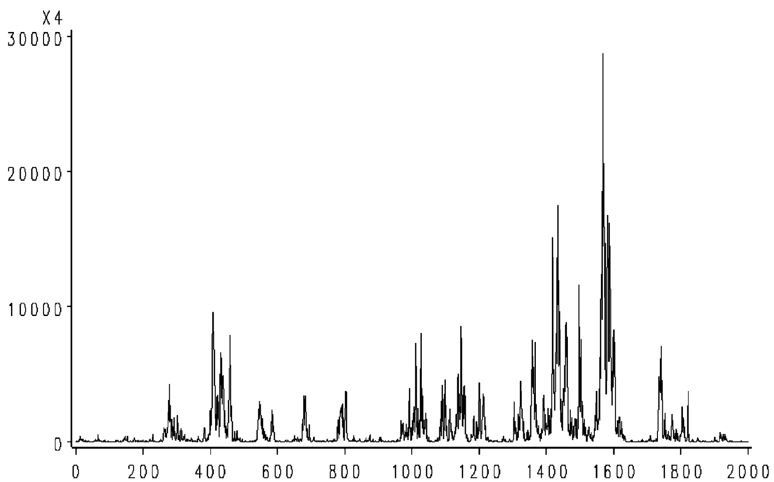

Fig. 4. A single simulated $\tilde{I}(0.8)$ time-series raised to the fourth power.

Figs. 5 and 6 demonstrate this "rescaling" for the two time-series shown in Figs. 3 and 4: The trends in mean and variance are removed successfully. It should be noted, however, that this rescaling method works well only if the true starting point of the time-series is known. If we rescale, e.g., with $(t-50)^{-m(d-0.5)}$ instead of $t^{-m(d-0.5)}$ (maybe because we could not observe the first 50 data points) the method still works for mid-sample and end-of-sample observations but the first couple of observations are not properly rescaled and remain "too large".

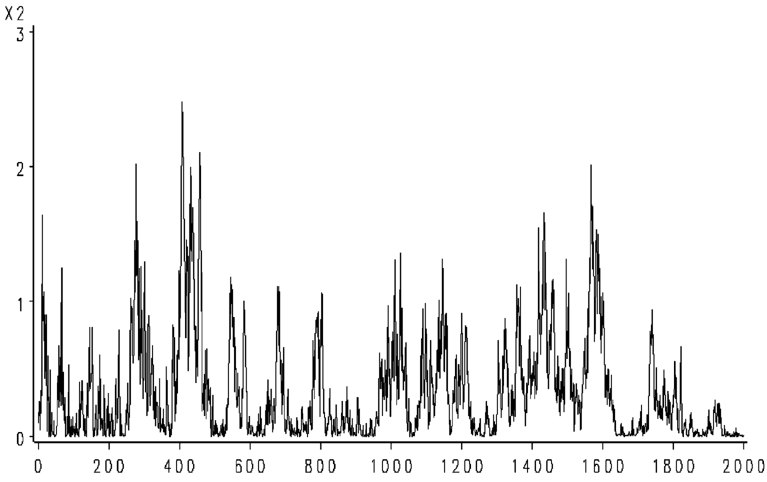

Fig. 5. Square of the simulated $\tilde{I}(0.8)$ time-series rescaled by $t^{-0.6}$.

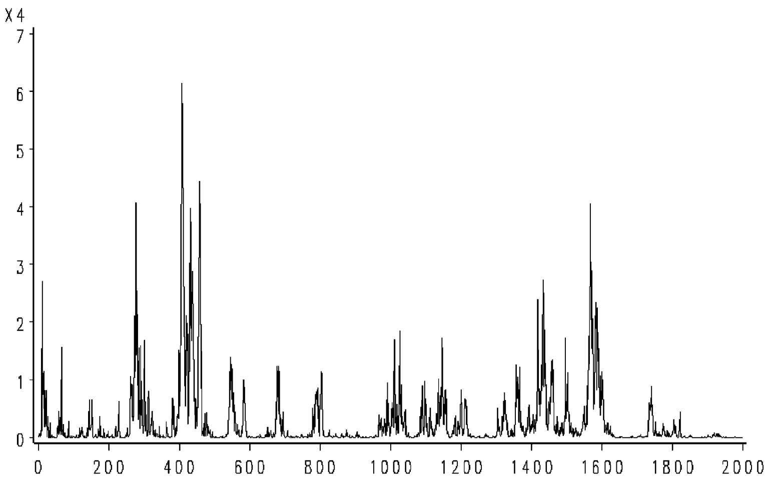

Fig. 6. Fourth power of the simulated $\tilde{I}(0.8)$ time-series rescaled by $t^{-1.2}$.

Finally, note that Proposition 2 also holds for the random walk ($d=1$) and for even more persistent series ($d>1$). For instance, the square of a random walk has a linear trend in mean and a quadratic trend in variance (cf. Granger, 1995).

4. Transformations of non-stationary processes

The last section was concerned with the moments of $\tilde{I}(d)$ processes and the moments of power transformations of $\tilde{I}(d)$ processes. In this section, we consider the autocorrelation structure of such processes, from which we can draw conclusions about their long-memory properties. In particular, we derive the long-memory parameter of the square of a non-stationary $\tilde{I}(d)$ process.

Proposition 3. *Let* $\tilde{X}_t \sim \tilde{I}(d)$ *with* $d > 0.5$:

(a) *Then* $\mathrm{Corr}(\tilde{X}_t, \tilde{X}_{t-h}) \xrightarrow{t\to\infty} 1$ *for all* $h \in \mathbb{N}$.
(b) *If* $\tilde{X}_t$ *is Gaussian, then* $\mathrm{Corr}(\tilde{X}_t^m, \tilde{X}_{t-h}^m) \xrightarrow{t\to\infty} 1$ *for all* $h \in \mathbb{N}$ *and* $m \in \{2, 3, \ldots\}$.

Proof.
(a)

$$\mathrm{Corr}(\tilde{X}_t, \tilde{X}_{t-h}) = \frac{\mathrm{E}(\tilde{X}_t, \tilde{X}_{t-h})}{\sqrt{\mathrm{E}(\tilde{X}_t^2)}\sqrt{\mathrm{E}(\tilde{X}_{t-h}^2)}} = \frac{\mathrm{E}(\sum_{j=0}^{t} c_j \varepsilon_{t-j} \sum_{i=0}^{t-h} c_i \varepsilon_{t-i})}{\sqrt{\sigma^2 \sum_{j=0}^{t} c_j^2}\sqrt{\sigma^2 \sum_{j=0}^{t-h} c_j^2}}$$

$$= \frac{\sigma^2 \sum_{j=0}^{t-h} c_j c_{j+h}}{\sqrt{\sigma^2 \sum_{j=0}^{t} c_j^2}\sqrt{\sigma^2 \sum_{j=0}^{t-h} c_j^2}} \geqslant \frac{\sum_{j=h}^{t} c_j^2}{\sqrt{\sum_{j=0}^{t} c_j^2}\sqrt{\sum_{j=0}^{t-h} c_j^2}}$$

$$= \left(1 - \frac{\sum_{j=0}^{h-1} c_j^2}{\sum_{j=0}^{t} c_j^2}\right)^{1/2} \left(1 - \frac{\sum_{j=0}^{h-1} c_j^2 + \sum_{j=t-h+1}^{t} c_j^2}{\sum_{j=0}^{t-h} c_j^2}\right)^{1/2} \overset{t\to\infty}{\to} 1.$$

Here, the inequality sign follows from $c_j > c_{j+h}$ and the limit holds as the sums in the denominators diverge whereas the sums in the numerators converge.

(b)

$$\mathrm{Corr}(\tilde{X}_t^m, \tilde{X}_{t-h}^m) = \frac{\mathrm{E}(\tilde{X}_t^m, \tilde{X}_{t-h}^m) - \mathrm{E}(\tilde{X}_t^m)\mathrm{E}(\tilde{X}_{t-h}^m)}{\sqrt{\mathrm{E}(\tilde{X}_t^{2m}) - \mathrm{E}(\tilde{X}_t^m)^2}\sqrt{\mathrm{E}(\tilde{X}_{t-h}^{2m}) - \mathrm{E}(\tilde{X}_{t-h}^m)^2}}$$

$$= \frac{(\mu_{mm}(t,h) - \mu_m^2) V_t^{m/2} V_{t-h}^{m/2}}{\sqrt{(\mu_{2m} - \mu_m^2) V_t^m}\sqrt{(\mu_{2m} - \mu_m^2) V_{t-h}^m}} = \frac{\mu_{mm}(t,h) - \mu_m^2}{\mu_{2m} - \mu_m^2},$$

where $V_t = \mathrm{Var}(X_t)$, μ_k is the kth moment of a standard normal variate and $\mu_{kk}(t,h) = E(\bar{X}_t^k, \bar{X}_{t-h}^k)$ with the normalization $\bar{X}_t = \tilde{X}_t / \sqrt{V_t}$. We now use the fact that, for given t and h, $(\bar{X}_t, \bar{X}_{t-h})$ are bivariate standard normal with correlation coefficient $\rho(t,h) = \mathrm{Corr}(\bar{X}_t, \bar{X}_{t-h}) = \mathrm{Corr}(\tilde{X}_t, \tilde{X}_{t-h}) \xrightarrow{t\to\infty} 1$. In particular, we show by induction that $\mu_{mm}(t,h) \xrightarrow{t\to\infty} \mu_{2m}$, which immediately implies result (b).

(1) $m = 1$: $\mu_{11}(t,h) = \rho(t,h) \xrightarrow{t\to\infty} 1 = \mu_2$.
(2) $m=2$: $\mu_{22}(t,h) = 1 + 2\rho^2(t,h) \xrightarrow{t\to\infty} 3$ (see Kendall et al., 1987, p. 103). Moreover, a formula in Johnson and Kotz (1970, p. 47) yields $\mu_4 = (4-1)(4-3) = 3$.
(3) $m > 2$: Johnson and Kotz (1972, p. 91) provide the following formula: $\mu_{mm}(t,h) = (2m-1)\rho(t,h)\mu_{m-1,m-1}(t,h) + (m-1)^2(1 - \rho^2(t,h))\mu_{m-2,m-2}(t,h)$. With

$\mu_{m-1,m-1}(t,h) \xrightarrow{t\to\infty} \mu_{2(m-1)}$, we obtain $\mu_{mm}(t,h) \xrightarrow{t\to\infty} (2m-1)\mu_{2(m-1)}$. On the other hand, $\mu_{2m} = (2m-1)\mu_{2(m-1)}$, according to Johnson and Kotz (1970). □

Proposition 3(a) states that $\tilde{I}(d)$ time-series with $d > 1/2$ have time-dependent autocorrelations which individually converge to 1. According to Proposition 3(b), this property is maintained for a Gaussian process under any power transformation. In this sense, a power transformation of a Gaussian $\tilde{I}(d)$ process with $d > 1/2$ still has long memory with some $d' > 1/2$.

Note in particular that rescaling the power transformation $\tilde{X}_t^m$—as discussed in the previous section—does not change the persistency property established in Proposition 3(b): The autocorrelations of $\tilde{Y}_t = \alpha(t)\tilde{X}_t^m + \beta(t)$ still converge to 1 for all functions $\alpha(\cdot)$ and $\beta(\cdot)$.[1] As a consequence, we can construct *stationary* processes (i.e., processes with converging mean and variance) which have the autocorrelations of a non-stationary long-memory process. Consider, e.g., $t^{0.5-d}\tilde{X}_t$ if $\tilde{X}_t \sim \tilde{I}(d)$ with $d > 1/2$. Such a process converges to a stationary process whose autocorrelations are 1 at all leads and lags. Clearly, the limit process is neither fractionally integrated, nor Gaussian.

Proposition 4.

(a) *If* $\tilde{X}_t \sim \tilde{I}(d)$ *with* $d > 1/2$, *then* $\tilde{X}_t$ *is asymptotically Gaussian*.
(b) *If* $\tilde{X}_t \sim \tilde{I}(d)$ *with* $1/2 < d < 1$, *then* $\tilde{X}_t^2$ *is asymptotically* LM(d).

Proof.

(a) According to Granger (1988) it suffices to show that $(\sum_{j=0}^t c_j^3)^{1/3}(\sum_{j=0}^t c_j^2)^{-1/2} \xrightarrow{t\to\infty} 0$. Note that $(\sum_{j=0}^t c_j^3)^{1/3} \underset{\text{for large } t}{\approx} (c + \sum_{j=0}^t j^{3d-3})^{1/3} \sim [O(t^{3d-2})]^{1/3} = O(t^{d-2/3})$, and $(\sum_{j=0}^t c_j^2)^{1/2} \underset{\text{for large } t}{\approx} (c+\sum_{j=0}^t j^{2d-2})^{1/2} \sim [O(t^{2d-1})]^{1/2} = O(t^{d-1/2})$. Hence, for $d > 1/2$, $(\sum_{j=0}^t c_j^2)^{1/2}$ diverges faster than $(\sum_{j=0}^t c_j^3)^{1/3}$, so that the ratio converges to zero.
(b) Here, we show that for $\tilde{X}_t \sim$ Gaussian $\tilde{I}(d)$ with $1/2 < d < 1$, $\tilde{X}_t^2$ is asymptotically LM(d). Proposition 4(b) then follows immediately from 4(a).
Let $Y_t \equiv \tilde{X}_t - \tilde{X}_{t-1}$ and $Z_t \equiv \tilde{X}_t + \tilde{X}_{t-1}$. Then (Y_t, Z_t) are jointly normally distributed with $\mathrm{E}(Y_t) = 0 = \mathrm{E}(Z_t)$, and $\Delta(\tilde{X}_t^2) = \tilde{X}_t^2 - \tilde{X}_{t-1}^2 = Y_t Z_t$. Using a formula for covariances of products of multivariate normal variables (see Bohrnstedt and Goldberger, 1969), we obtain

$$\mathrm{Cov}(Y_t Z_t, Y_{t-h} Z_{t-h}) = \mathrm{Cov}(Y_t, Y_{t-h})\mathrm{Cov}(Z_t, Z_{t-h}) + \mathrm{Cov}(Y_t, Z_{t-h})\mathrm{Cov}(Z_t, Y_{t-h}).$$

[1] Note that this is true only for the theoretical, or instantaneous, autocorrelations. In contrast, the empirical autocorrelations are sensitive to rescaling, so that rescaling might influence the estimated long-memory parameter.

We first show that the second term of this expression, $\mathrm{Cov}(Y_t, Z_{t-h})\mathrm{Cov}(Z_t, Y_{t-h})$, converges.

$$
\begin{aligned}
&\mathrm{Cov}(Y_t, Z_{t-h})\\
&= \sum_{j=0}^{t-h} c_j c_{j+h}\sigma^2 + \sum_{j=0}^{t-h-1} c_j c_{j+h+1}\sigma^2 - \sum_{j=0}^{t-h} c_j c_{j+h-1}\sigma^2 - \sum_{j=0}^{t-h-1} c_j c_{j+h}\sigma^2\\
&= \sigma^2 \left[\sum_{j=0}^{t-h-1} c_j (c_{j+h+1} - c_{j+h-1}) + c_{t-h}(c_t - c_{t-1}) \right]\\
&= \sigma^2 \left[\sum_{j=0}^{t-h-1} c_j c_{j+h-1} \left(\frac{j+h+d}{j+h+1}\,\frac{j+h+d-1}{j+h} - 1 \right) + c_{t-h}(c_t - c_{t-1}) \right]. \quad (*)
\end{aligned}
$$

In the last step, we used the recursive relationship between c_j and c_{j-1} (see, e.g., Hosking, 1984). Note that $c_j \sim O(j^{d-1})$ and

$$
\begin{aligned}
&1 - \frac{j+h+d}{j+h+1}\,\frac{j+h+d-1}{j+h}\\
&= \frac{(j+h+1)(j+h) - (j+h+d)(j+h+d-1)}{(j+h+1)(j+h)}\\
&= \frac{(2-2d)j + h(h+1) - (h+d)(h+d-1)}{(j+h+1)(j+h)} \sim O(j^{-1}).
\end{aligned}
$$

Consequently, $c_j c_{j+h-1}([(j+h+d)/(j+h+1)][(j+h+d-1)/(j+h)] - 1) \sim O(j^{2d-3})$, so that $(*)$ converges as $t \to \infty$ if $d < 1$. With exactly the same argument we can show that $\mathrm{Cov}(Z_t, Y_{t-h})$ and hence the product $\mathrm{Cov}(Y_t, Z_{t-h})$ $\mathrm{Cov}(Z_t, Y_{t-h})$ converges as $t \to \infty$.

Now consider the variance, which can analogously be written as

$$
\mathrm{Var}(Y_t Z_t) = \mathrm{Var}(Y_t)\mathrm{Var}(Z_t) + \mathrm{Cov}(Y_t, Z_t)^2,
$$

$$
\begin{aligned}
\mathrm{Cov}(Y_t, Z_t) &= E[(\tilde{X}_t - \tilde{X}_{t-1})(\tilde{X}_t + \tilde{X}_{t-1})] = \mathrm{E}(\tilde{X}_t^2) - \mathrm{E}(\tilde{X}_{t-1}^2)\\
&= \sum_{j=0}^{t} c_j^2\sigma^2 - \sum_{j=0}^{t-1} c_j^2\sigma^2 = c_t^2\sigma^2 \xrightarrow{t\to\infty} 0.
\end{aligned}
$$

Hence, $\mathrm{Var}(Y_t Z_t) \xrightarrow{t\to\infty} \mathrm{Var}(Y_t)\mathrm{Var}(Z_t)$, and we obtain for the autocorrelations

$$
\begin{aligned}
\mathrm{Corr}(Y_t Z_t, Y_{t-h} Z_{t-h}) &= \frac{\mathrm{Cov}(Y_t, Y_{t-h})\mathrm{Cov}(Z_t, Z_{t-h})}{\sqrt{\mathrm{Var}(Y_t Z_t)}\sqrt{\mathrm{Var}(Y_{t-h} Z_{t-h})}}\\
&\quad + \frac{\mathrm{Cov}(Y_t, Z_{t-h})\mathrm{Cov}(Z_t, Y_{t-h})}{\sqrt{\mathrm{Var}(Y_t Z_t)}\sqrt{\mathrm{Var}(Y_{t-h} Z_{t-h})}}.
\end{aligned}
$$

Table 2
Average estimated long-memory parameter of some transformations of 2,000 simulated non-stationary Gaussian $I(d)$ processes with 2,000 observations each

$g(X)$ and its Hermite rank		Long-memory parameter of the original series X		
		$d=0.6$	$d=0.8$	$d=1$
X (rank 1)	Original series	0.60 (0.032)	0.80 (0.033)	1.00 (0.032)
X^2	Original series	0.56 (0.053)	0.78 (0.045)	0.99 (0.042)
(rank 2)	Rescaled series	0.56 (0.053)	0.78 (0.052)	0.98 (0.061)
X^3	Original series	0.55 (0.060)	0.78 (0.057)	0.99 (0.054)
(rank 1)	Rescaled series	0.55 (0.060)	0.76 (0.069)	0.97 (0.078)
X^4	Original series	0.52 (0.081)	0.76 (0.073)	0.98 (0.067)
(rank 2)	Rescaled series	0.52 (0.082)	0.74 (0.090)	0.95 (0.096)
X^3-3X	Original series	0.54 (0.069)	0.77 (0.058)	0.99 (0.054)
(rank 3)	Rescaled series	0.54 (0.069)	0.76 (0.071)	0.97 (0.078)
X^4-6X^2	Original series	0.50 (0.093)	0.75 (0.075)	0.98 (0.067)
(rank 4)	Rescaled series	0.50 (0.094)	0.73 (0.092)	0.95 (0.096)

As $t \to \infty$, the second term converges to zero, because the numerator converges and the denominator diverges. Hence,

$$\mathrm{Corr}(Y_tZ_t, Y_{t-h}Z_{t-h}) \overset{\text{for large } t}{\to} \frac{\mathrm{Cov}(Y_t, Y_{t-h})}{\sqrt{\mathrm{Var}(Y_t)}\sqrt{\mathrm{Var}(Y_{t-h})}} + \frac{\mathrm{Cov}(Z_t, Z_{t-h})}{\sqrt{\mathrm{Var}(Z_t)}\sqrt{\mathrm{Var}(Z_{t-h})}}$$

$$= \mathrm{Corr}(Y_t, Y_{t-h}).$$

Here, the equality sign follows with Proposition 3(a), using that $Z_t \sim I(d)$. As $Y_t \sim I(d-1)$, $\Delta(\tilde{X}_t^2) = Y_tZ_t \sim \mathrm{LM}(d-1)$ and thus $\tilde{X}_t^2 \sim \mathrm{LM}(d)$. $\square$

Proposition 4 states that taking the square of a non-stationary long-memory process does not change the size of the long-memory parameter. This is in obvious contrast to our findings for stationary Gaussian $I(d)$ processes (see Proposition 1), for which taking the square *reduces* the amount of long memory in the series. Note that our results only hold for $d<1$. For $d=1$, Granger (1995) shows that the square of a random walk is a random walk with drift, which has a variance that is quadratic in t.

For higher powers than the square of a non-stationary $I(d)$ process, we could not establish any theoretical results. This is due to the fact that the autocorrelations of a non-stationary process are not informative as they converge to one for any long-memory parameter $d > 1/2$ (see Proposition 3). As a consequence, we have to consider first differences in the proof of Proposition 4. First differences of a nonlinear transformation of a linear process can become very complicated, however. The substantial simplifications in the case of the square of a Gaussian time-series do not apply for higher power transformations.

Table 2 illustrates the findings of the previous two propositions. It contains the average estimated long-memory parameters of some polynomial transformations of three non-stationary $\tilde{I}(d)$ processes with $d = 0.6$, 0.8 and 1. Here, we estimated the long-memory parameters from the first differences of the transformed series, in order to guarantee consistency. The first line of each cell shows the average estimate for the "original" transformed series, while the second line displays the average estimate for the rescaled transformed series as discussed in the previous section. The numbers in parentheses are the empirical standard errors.

The simulations confirm the findings in Proposition 4: The square has the same long-memory parameter as the original process for all three values of d. Moreover, Table 2 suggests that the same holds for higher power transformations. Although the average long-memory parameter is consistently slightly smaller than the initial d, the difference is never smaller than the estimated standard error. Another important observation is that the Hermite rank of the transformation has virtually no influence on the long-memory properties of the processes any more. Instead, the estimates seem to depend mainly on the order of the polynomial transformation. Note that the standard errors increase with the order of the polynomial and that they are generally larger than in the stationary case.

In addition, Table 2 shows that rescaling the transformed series does not lead to any improvements of the estimates. On the contrary, the standard errors are larger for the rescaled series than for the original series. Moreover, standard errors become much worse if the transformed series are rescaled using a wrong time index (not shown in Table 2).

5. Transcendental transformations

The previous section presented results for polynomial transformations of non-stationary fractionally integrated time-series. In this section we discuss some simulation results for sine, cosine, exponential and logistic transformations. This discussion will confirm the fundamental differences between antipersistence, stationary long-memory and non-stationary long-memory found in earlier sections.

Table 3 contains average estimates for the long-memory parameter of four transcendental transformations: $\sin(X)$, $\cos(X)$, $\exp(X)$ and $(1 + \exp(X^{-1}))^{-1}$ if applied to seven different $\tilde{I}(d)$ processes with $d = -0.2$, -0.4, 0.2, 0.4, 0.6, 0.8 and 1. We report two Geweke–Porter–Hudak estimates: the first using the $m = [T^{0.8}] = 437$ smallest Fourier frequencies in the periodogram regression, the second using only $m = [T^{0.6}] = 95$ regression points. In the previous tables, we did not include the estimates for $m = [T^{0.6}]$, because they were very similar to (and by construction less precise than) those with $m = [T^{0.8}]$.

Table 3 shows that antipersistence (i.e., $d < 0$) is partly preserved under odd transformations (sine, logistic) but that it disappears under non-odd transformations, such as the cosine or the exponential function. In contrast, the size of the long memory of transformations of stationary long-memory processes ($0 < d < 1/2$) mainly depends on the Hermite rank of the transforming function, as Proposition 1 suggests. Note, however, that none of the four functions considered in Table 3 can be written as a *finite* sum of Hermite polynomials, which is a condition maintained in Proposition 1.

Table 3
Average estimated long-memory parameter of some transformations of 2,000 simulated Gaussian $I(d)$ processes with 2,000 observations each

Long-memory parameter of the series X	Number of regression points used	Transformation $g(X)$ and its Hermite rank			
		$\sin(X)$ (rank 1)	$\cos(X)$ (rank 2)	$\exp(X)$ (rank 1)	$(1+\exp(X^{-1}))^{-1}$ (rank 1)
$d=-0.4$	437	−0.23 (0.033)	0.02 (0.032)	−0.04 (0.069)	−0.36 (0.032)
	95	−0.14 (0.070)	−0.00 (0.070)	−0.02 (0.071)	−0.32 (0.071)
$d=-0.2$	437	−0.15 (0.031)	0.01 (0.033)	−0.04 (0.055)	−0.19 (0.032)
	95	−0.12 (0.070)	−0.00 (0.073)	−0.03 (0.078)	−0.19 (0.070)
$d=0.2$	437	0.18 (0.032)	0.04 (0.035)	0.09 (0.064)	0.20 (0.032)
	95	0.19 (0.073)	0.03 (0.074)	0.09 (0.100)	0.20 (0.072)
$d=0.4$	437	0.37 (0.036)	0.28 (0.058)	0.32 (0.060)	0.40 (0.032)
	95	0.38 (0.075)	0.27 (0.106)	0.31 (0.103)	0.40 (0.072)
$d=0.6$	437	0.41 (0.030)	0.41 (0.031)	0.42 (0.070)	0.58 (0.046)
	95	0.36 (0.069)	0.37 (0.067)	0.38 (0.109)	0.58 (0.090)
$d=0.8$	437	0.37 (0.030)	0.37 (0.031)	0.42 (0.112)	0.75 (0.086)
	95	0.16 (0.072)	0.16 (0.070)	0.25 (0.156)	0.71 (0.146)
$d=1.0$	437	0.29 (0.031)	0.29 (0.031)	0.38 (0.151)	0.85 (0.176)
	95	0.04 (0.071)	0.04 (0.070)	0.15 (0.166)	0.75 (0.260)

Indeed, the results of Proposition 1 do not hold for the exponential transformation, which can be written as $\exp(x)=\exp(0.5)\sum_{j=0}^{\infty} H_j(x)/\sqrt{j!}$ (see Cramér, 1946, p. 133). With Lemma 1 and $g_j = 1/\sqrt{j!}$ we then obtain

$$\operatorname{Corr}(g(X_t), g(X_{t-h})) = \frac{\sum_{j=1}^{\infty} g_j^2 \rho_h^j}{\sum_{j=1}^{\infty} g_j^2} = \frac{\sum_{j=1}^{\infty} \frac{1}{j!}\rho_h^j}{\sum_{j=1}^{\infty} \frac{1}{j!}} = \frac{\exp(\rho_h)-1}{\exp(1)-1}, \tag{11}$$

which converges to zero as h approaches infinity. However, the rate of convergence is clearly not h^{2d-1}, as in Proposition 1, but rather $\exp(h^{2d-1})-1$, which is faster than h^{2d-1}. This explains why the estimates of the long-memory parameter of $\exp(X)$ are smaller (although not significantly smaller) than Proposition 1 suggests.

For non-stationary long-memory processes ($1/2 < d < 1$), we find again a completely different behavior: For these processes, neither the symmetry nor the Hermite rank of the transformation influence the long memory of the transformed series. Instead, these time-series are most sensitive to periodic transformations. For $d > 1/2$, the long-memory parameter of the sine or cosine transformation is smaller, the larger the initial d is. For $d = 1$, the average estimated long-memory parameter of $\sin(X_t)$ and $\cos(X_t)$ is 0.29 if we use 437 regression points and 0.04 if we use 95 regression points. The heavy dependence of the estimated long-memory parameter on the number of periodogram regression points implies that the series have strong short-term correlations but no long memory. Indeed, Granger and Hallman (1988) show that the sine and cosine of a random walk can be written as an AR(1) process with heteroskedastic errors. The results for $d = 0.6$ and 0.8 suggest that periodic transformations of non-stationary long-memory processes still have some long memory, but that more and more of the initial long memory is transformed into short memory as d increases.

Surprisingly, the pattern of estimates for the exponential transformation resembles the pattern of the periodic functions. The long-memory parameter of the transformed series seems to decrease as d approaches 1 while short-run correlations seem to become stronger. Granger and Hallman (1988) show that the exponential transformation of a random walk has the correlogram of a stationary AR(1) process. Furthermore, they show that this process has exponentially increasing variance, which can explain the exceptionally large standard errors for $\exp(X)$ in Table 3.

The logistic transformation, $(1 + \exp(X^{-1}))^{-1}$, exactly retains the long memory of stationary long-memory processes and distorts values of $d < 0$ only slightly. For $d > 1/2$, the long memory of the output is less than that of the input, but still larger than 1/2. Moreover, the long memory of the output series increases strictly with the long memory of the input—in contrast to the other three functions in Table 3. However, the most remarkable fact concerning the logistic transformation is that it is bounded. It thereby demonstrates that long-memory processes—even with long-memory parameter $d > 1/2$—can be bounded. Note that the logistic transformation of a random walk has all the dominant properties of a random walk except for the variance which is linear in t for the input series but constant for the transformed series (see Granger, 1995).

6. Nonlinear transformations of break processes

It is well known that processes with breaks in their mean look like long-memory processes, even if they are $I(0)$ between breaks; see, Granger and Hyung (2001) or Diebold and Inoue (1999). In this section, we briefly investigate the long-memory properties of nonlinear transformations of break processes and compare them with our results for long-memory processes from previous sections.

We consider the following break process:

$$X_t = \mu_t + \varepsilon_t, \tag{12}$$

where $\varepsilon_t \sim \text{iid}\,\mathrm{N}(0,1)$ and μ_t is the mean which is subject to stochastic breaks. Let p denote the probability of a break in each period and assume that the different levels of the mean are iid $\mathrm{N}(0,2)$. Hence, $\mu_t = \mu_{t-1}$ if no break occurs (which happens with probability $1 - p$). Otherwise, μ_t is drawn from a $\mathrm{N}(0,2)$ distribution.

Table 4 shows estimates of the long-memory parameter of various transformations of four break processes with break probabilities $p = 0.15$, 0.08, 0.045 and 0.01. The break probabilities have been determined by trial and error, so that the average estimated long-memory parameter is, respectively, 0.2, 0.4, 0.6 and 0.8. Therefore, this table can be directly compared with Tables 1 and 2 which contain the corresponding results for long-memory processes with these parameters. In accordance with the theory (Diebold and Inoue, 1999; Granger and Hyung, 2001), the estimated long-memory parameter increases as the break probability decreases. The reason is that autocorrelations over

Table 4
Average estimated long-memory parameter of some transformations of 2,000 simulated break processes with break probability p

$g(X)$ and its Hermite rank	Probability of a break each time period			
	$p = 0.15$	$p = 0.08$	$p = 0.045$	$p = 0.01$
X (rank 1)	0.21 (0.079)	0.39 (0.080)	0.60 (0.089)	0.79 (0.081)
X^2 (rank 2)	0.20 (0.094)	0.36 (0.108)	0.56 (0.124)	0.69 (0.119)
X^3 (rank 1)	0.19 (0.106)	0.36 (0.124)	0.55 (0.147)	0.71 (0.121)
X^4 (rank 2)	0.18 (0.123)	0.33 (0.151)	0.50 (0.172)	0.61 (0.154)
$X^3 - 3X$ (rank 3)	0.19 (0.112)	0.35 (0.133)	0.53 (0.159)	0.67 (0.139)
$X^4 - 6X^2$ (rank 4)	0.17 (0.128)	0.31 (0.160)	0.48 (0.184)	0.57 (0.171)

long horizons are large and positive, if breaks occur infrequently. The more often the process breaks, the smaller are these high-order autocorrelations and the smaller is the estimated long-memory parameter.

Table 4 demonstrates that the estimated long-memory parameter decreases only slightly with the order of the polynomial transformation. In particular, this estimate seems to be independent of the Hermite rank of the transformation. A comparison with Table 2 reveals that break processes with $p = 0.01$ and especially $p = 0.045$ have the same long-memory pattern under nonlinear transformations as non-stationary long-memory processes with $d = 0.8$ and 0.6, respectively. In contrast, break processes with $p = 0.15$ and 0.08 exhibit a completely different pattern compared to the corresponding stationary long-memory processes with $d = 0.2$ and 0.4 in Table 1. For instance, the square of a Gaussian $I(0.2)$ process is $I(0)$, whereas the estimated long-memory parameter of the square of a break process with $p = 0.15$ is still 0.2.

Intuitively, the apparent long memory in break processes stems from the size and the persistence of the shifts in their mean. Taking the square of such a process does not reduce the number of shifts and changes the size of the shifts only slightly. However, there are obviously many other ways to construct a process with breaks in mean. If we assume for instance that μ_t in (12) can only take on the values 1 and -1, the square of this process is trivially $I(0)$ for any break probability p.

Acknowledgements

We thank Christian Kleiber, Michael Rockinger and two referees for helpful comments. Ingolf Dittmann gratefully acknowledges the support of the German Academic Exchange Service (DAAD/HSP III).

References

Barrett, J.F., Lampard, D.G., 1955. An expansion for some second-order probability distributions and its application to noise problems. IRE Transactions—Information Theory 1, 10–15.

Bohrnstedt, G.W., Goldberger, A.S., 1969. On the exact covariance of products of random variables. Journal of the American Statistical Association 64, 1439–1442.

Cramér, H., 1946. Mathematical Methods of Statistics. Princeton University Press, Princeton, NJ.

Diebold, F.X., Inoue, A., 1999. Long memory and regime switching. Journal of Econometrics 105, 131–159.

Geweke, J., Porter–Hudak, S., 1983. The estimation and application of long memory time-series models. Journal of Time Series Analysis 4, 221–238.

Giraitis, L., Surgailis, D., 1985. Central limit theorems and other limit theorems for functionals of Gaussian processes. Probability Theory and Related Fields 70, 191–212.

Gradshteyn, I.S., Ryzhik, I.M., 1980. Table of Integrals, Series, and Products. Academic Press, New York.

Granger, C.W.J., 1988. Models that generate trends. Journal of Time Series Analysis 9, 329–343.

Granger, C.W.J., 1995. Modelling nonlinear relationships between extended-memory variables. Econometrica 63, 265–279.

Granger, C.W.J., Hallman, J., 1988. The algebra of $I(1)$. Finance and Economics Discussion Series 45, Board of Governors of the Federal Reserve System.

Granger, C.W.J., Hallman, J., 1991. Nonlinear transformations of integrated time-series. Journal of Time Series Analysis 12, 207–224.

Granger, C.W.J., Hyung, N., 2001. Occasional Breaks and Long Memory. UCSD discussion paper.

Granger, C.W.J., Newbold, P., 1976. Forecasting transformed series. Journal of the Royal Statistical Society B 38, 189–203.

Hosking, J.R.M., 1984. Modeling persistence in hydrological time-series using fractional differencing. Water Resources Research 20, 1898–1908.

Hurvich, C.M., Deo, R., Brodsky, J., 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, 19–46.

Johnson, N.I., Kotz, S., 1970. Distributions in Statistics: Continuous Univariate Distributions 1. Houghton Mifflin, Boston.

Johnson, N.I., Kotz, S., 1972. Distributions in Statistics: Continuous Multivariate Distributions. Wiley, New York.

Kendall, M., Stuart, A., Ord, J.K., 1987. Kendall's Advanced Theory of Statistics, Vol. 1 (Distribution Theory). Oxford University Press, New York.

Taqqu, M.S., 1979. Convergence of integrated processes of arbitrary Hermite rank. Probability Theory and Related Fields 50, 53–83.