# Properties of Nonlinear Transformations of Fractionally Integrated Processes[#]


**Ingolf Dittmann**
Department of Economics
University of Dortmund
44221 Dortmund, Germany
i.dittmann@wiso.uni-dortmund.de

**Clive W. J. Granger**
Department of Economics
University of California, San Diego
La Jolla, CA 92093-0508
cgranger@ucsd.edu


April 2000


***Abstract:***

This paper shows that the properties of nonlinear transformations of a fractionally integrated process depend strongly 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 integer power transformation is non-stationary and exhibits a deterministic 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.


---


[#] The first author gratefully acknowledges the support of the German Academic Exchange Service (DAAD/HSP III).

## 1. Introduction

Two of the most popular topics in univariate time series analysis in recent years have concerned fractionally integrated series and nonlinear series. In this paper, we link the two topics by considering the properties of $f(X_t)$ for different functions $f$ (.) where $X_t$ is fractionally integrated, or I($d$). It is well known that there are three main types of fractional processes (the topics mentioned here are carefully defined in the next section):

i. $d < 0$ are stationary and antipersistent
ii. $0 < d < ½$ are stationary in integer moments greater than one, and are long-memory
iii. $d \geq ½$ which have non-stationary second and higher moments and are also long-memory.

It is found that for class (i), the anti-persistent property is lost if $f$(.) is not an odd function, that for class (ii) the long-memory is sensitive to the Hermite rank of the transformation, with the higher the rank the lower the memory. For the third class a variety of results are obtained. In particular, an integer power transformation $f(x) = x^n$ does not change the process' long memory, whereas periodic transformations, e.g., $f(x) = \cos(x)$ markedly reduce the initial long memory.

The next section considers cases (i) and (ii) while Section 4 is concerned with case (iii). In Section 3, the moments of fractionally integrated processes and thereby the trending behavior of integer power transformations are derived. Finally, Section 5 presents simulation results for periodic and exponential transformations.

## 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

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

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

**Stationarity.** A time series $\{X_t\}$ is called stationary if its mean and variance do not depend on the time index $t$. If, in addition, the instantaneous (or theoretical) autocovariance function $\gamma(t, h) \equiv \mathrm{Cov}(X_t, X_{t-h})$ 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

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

where $T$ is the sample size and $\overline{X}$ is the sample mean. If $\{X_t\}$ is not covariance stationary, $\gamma(t, h)$ and $\tilde{\gamma}(h)$ are clearly different concepts. All our theory 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 $\lambda$ approaches zero ($d < ½$), we write $X_t \sim \mathrm{LM}(d)$. For $d \geq ½$, we define $X_t \sim \mathrm{LM}(d) :\Leftrightarrow (1 - \mathrm{L})^k X_t \sim \mathrm{LM}(d - k)$ for $k = [d + ½]$, where $[x]$ denotes the largest integer smaller or equal to $x$. Hence, $\mathrm{LM}(d)$ can be thought of as a generalization of $\mathrm{I}(d)$, because $X_t \sim \mathrm{I}(d) \Rightarrow X_t \sim \mathrm{LM}(d)$.

**Long memory.** If $X_t \sim \mathrm{LM}(d)$ 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 $\gamma(h) \sim h^{2d-1}$, $0 < d < ½$, is $\mathrm{LM}(d)$, 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 $\mathrm{I}(d)$ process with $d < 0$ is antipersistent. Note, however, that the hyperbolical decay of the autocorrelations of such an $\mathrm{I}(d)$ process is neither necessary nor sufficient for the presence of antipersistence.

To derive the long-memory properties of a non-linear transformation of a stationary Gaussian fractionally integrated process, we decompose the transformation into a sum of Hermite polynomials. This method has first been applied to fractionally integrated processes by Gourieroux and Jasiak (1999). See Granger and Newbold (1976) for a more detailed introduction to this approach.

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

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

$H_j(x)$ is a polynomial of degree $n$; the first five Hermite polynomials are $H_0(x) = 1$, $H_1(x) = x$, $H_2(x) = (x^2 - 1)/\sqrt{2}$, $H_3(x) = (x^3 - 3x)/\sqrt{6}$, $H_4(x) = (x^4 - 6x^2 + 3)/\sqrt{24}$. The Hermite polynomials are closely connected to the standard normal distribution. Let $\phi(x)$ be the density function of the standard normal distribution and let $X$ be a standard normal random variable. Then (see, e.g., Cramér, 1946, pp. 131-133):

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

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

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

(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:

[1] For notational convenience, this definition slightly differs from the traditional definition of Hermite polynomials, which is $\tilde{H}_j(x) = \sqrt{j!}\, H_j(x)$.

**Lemma 1:** Let $\{X_t\}$ be a covariance stationary Gaussian process with standard normal marginal distribution and autocorrelation function $\rho_h = corr(X_t, X_{t-h})$. 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

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

**Proof:** With (5) and (6), we obtain:

$$\begin{aligned}
E(g(X_t), g(X_{t-h})) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x_t) g(x_{t-h}) f(x_t, x_{t-h}) dx_t dx_{t-h} \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x_t) g(x_{t-h}) \phi(x_t)\phi(x_{t-h}) \{1 + \sum_{j=1}^{\infty} \rho_h^j H_j(x_t) H_j(x_{t-h})\} dx_t dx_{t-h} \\
&= \int_{-\infty}^{\infty} g(x_t)\phi(x_t) dx_t \int_{-\infty}^{\infty} g(x_{t-h})\phi(x_{t-h}) dx_{t-h} \\
&\quad + \sum_{j=1}^{\infty} \rho_h^j \int_{-\infty}^{\infty} g(x_t) H_j(x_t)\phi(x_t) dx_t \int_{-\infty}^{\infty} g(x_{t-h}) H_j(x_{t-h})\phi(x_{t-h}) dx_{t-h} \\
&= E(g(X_t))^2 + \sum_{j=1}^{\infty} g_j^2 \rho_h^j
\end{aligned}$$

$$\Rightarrow Cov(g(X_t), g(X_{t-h})) = \sum_{j=1}^{\infty} g_j^2 \rho_h^j$$

With $Var(g(X_t)) = \sum_{j=1}^{\infty} g_j^2$ the result follows immediately. ∎

Lemma 1 is a powerful tool that enables us to determine the long-memory properties of transformations of stationary Gaussian fractionally integrated processes:

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

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

$J$ is called the Hermite rank of $g(.)$.

(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 non-linear, 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

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

This implies that the autocorrelations of $g(X_t)$ decrease hyperbolically, too, i.e., $corr(g(X_t), g(X_{t-h})) \underset{\text{for large } h}{\approx} c \cdot h^{2\bar{d}-1}$ with a positive constant $c$ and $2\bar{d} - 1 = (2d-1)J$ $\Leftrightarrow \bar{d} = (d-0.5)J + 0.5$.

Case 1: $\bar{d} \geq 0$. 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.

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

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

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 non-linear. ■

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 $d = ½$, the size of the long memory stays the same for any transformation. The non-stationary case ($d \geq ½$) will be considered in Section 4. Moreover, Proposition 1 implies that it is not possible to generate long memory by transforming a Gaussian short-memory process.

Proposition 1(a) is not particularly surprising, as Taqqu (1979) derived the same result for stationary increments of fractional Brownian motions, which form another class of long-memory processes. Note, however, that increments of fractional Brownian motions are clearly distinct from fractionally integrated time series, because their autocorrelations and, equivalently, their spectra are not identical (see, e.g., Beran, 1994, pp. 52/53 and 63/64). Moreover, Taqqu (1979) only considers processes with $d > (1 - J^{-1})/2$, so that $\tilde{d} > 0$ by assumption.

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 non-linear 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 $-½$ , 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 3, 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 non-linear 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 non-linear transformation $g(X_t)$.

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

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

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, Deo and Brodsky, 1998). Each cell shows the average estimate of the long-memory parameter $d$, its empirical standard error in brackets and the corresponding theoretical value.

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 section, we derive the moments of stationary and, more interestingly, of non-stationary I(d) processes. 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 and to obtain more detailed results about the moments of long-memory processes, we consider finite-past processes in this and the next section:

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

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

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

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

(a) If $d < 0.5$, all moments of $X_t$ converge as $t \to \infty$.

(b) If $d > 0.5$, all even moments of $X_t$ diverge. More explicitly, for $m \in \{2,4,6,...\}$ $\mu_m^X(t) \sim O(t^{m(d-0.5)})$.

(c) If $m$ is odd and $d < (m+1)/2m$, the $m$th 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 $m \in \{3,5,7,...\}$
$\mu_m^X(t) \sim O(t^{m(d-0.5)-0.5})$.

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

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 & Ryzhik (1980).

For higher moments, we proceed analogously:

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

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

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

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

$$\mu_5^X(t) = \sum_{j=0}^{t} c_j^5 \mu_5^\varepsilon + \binom{5}{2} \sum_{j1=0}^{t-1} \sum_{j2=j1+1}^{t} c_{j1}^2 \mu_2^\varepsilon c_{j2}^3 \mu_3^\varepsilon = (\mu_5^\varepsilon - 30 \mu_2^\varepsilon \mu_3^\varepsilon) \sum_{j=0}^{t} c_j^5 + 30 \sum_{j=0}^{t} c_j^2 \sum_{j=0}^{t} c_j^3 \mu_2^\varepsilon \mu_3^\varepsilon$$

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

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

$$+ \binom{6}{2,2,2} \sum_{j1} \sum_{j2} \sum_{j3} c_{j1}^2 c_{j2}^2 c_{j3}^2 (\mu_2^\varepsilon)^3$$

As the divergence rate of $\left(\sum c_j^2\right)^3$ is larger than the divergence rates of $\sum c_j^6$, $\sum c_j^4 \sum c_j^2$ and $\left(\sum c_j^3\right)^2$, the last term dominates the other three terms. Hence, $\mu_6^X(t) \sim O(t^{3(2d-1)}) = O(t^{6d-3})$. 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 $k$.

$$\mu_7^X(t) = \sum_{j=0}^{t} c_j^7 \mu_7^\varepsilon + \binom{7}{2} \sum_{j1} \sum_{j2} c_{j1}^2 c_{j2}^5 \mu_2^\varepsilon \mu_5^\varepsilon + \binom{7}{3} \sum_{j1} \sum_{j2} c_{j1}^3 c_{j2}^4 \mu_3^\varepsilon \mu_4^\varepsilon$$
$$+ \binom{7}{2,2,3} \sum_{j1} \sum_{j2} \sum_{j3} c_{j1}^2 c_{j2}^2 c_{j3}^3 (\mu_2^\varepsilon)^2 \mu_3^\varepsilon$$

Here, the last term has the largest divergence rate:

$$\mu_7^X(t) \sim O(t^{7d-6}) + O(t^{2d-1})O(t^{5d-4}) + O(t^{3d-2})O(t^{4d-3}) + O(t^{2d-1})O(t^{2d-1})O(t^{3d-2}) = O(t^{7d-4})$$

For higher odd moments $m > 3$, the same argument holds: The dominating terms will always be of the form $\left(\sum c_j^2 \mu_2^\varepsilon\right)^{(m-3)/2} \sum c_j^3 \mu_3^\varepsilon \sim O(t^{0.5(m-3)(2d-1)}) O(t^{3d-2}) = O(t^{m(d-0.5)-0.5})$. ■

From Proposition 2, we can immediately obtain the trend in mean of any power transformation (i.e., $X^2$, $X^3$, …) of a nonstationary $\tilde{I}(d)$ process. For instance, if $X_t \sim \tilde{I}(d)$, $X_t^2$ has a trend of the form $t^{2d-1}$ and $X_t^4$ has a trend of the form $t^{4d-2}$. Figure 1 illustrates 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.

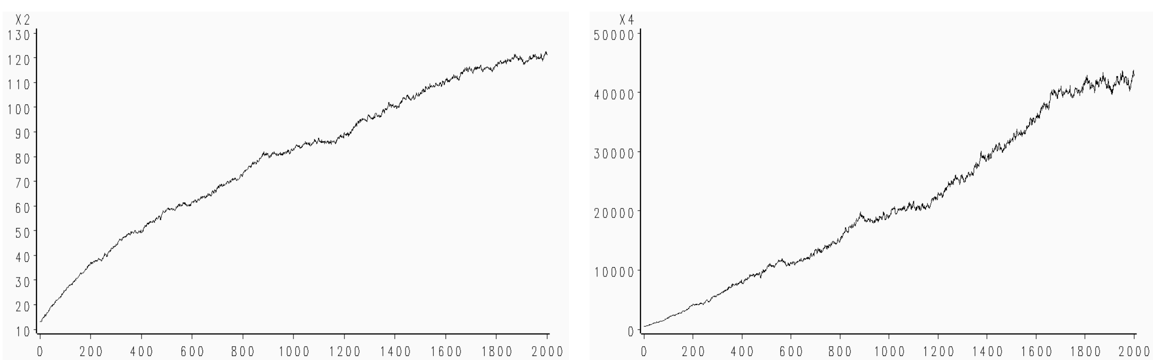


**Figure 1:** Average trend in mean of the transformations of 2,000 simulated $\tilde{I}(0.8)$ time series; left for $X_t^2$, right for $X_t^4$.

Whereas the trend can be clearly seen in the aggregates of Figure 1, this is much less so for the individual series as Figure 2 demonstrates. The reason is that $X_t^2$ 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 $(X_t)^m$, the trend in variance which is of order $t^{m(2d-1)}$ dominates the trends in all other moments in the sense that multiplying $(X_t)^m$ by $t^{-m(d-0.5)}$

removes the trend in all moments. Hence, the nonstationary process $(X_t)^m$ can be transformed into an asymptotically stationary process by simply multiplying it with $t^{-m(d-0.5)}$. Figure 3 demonstrates this "rescaling" for the two time series shown in Figure 2: 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 couldn't 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".

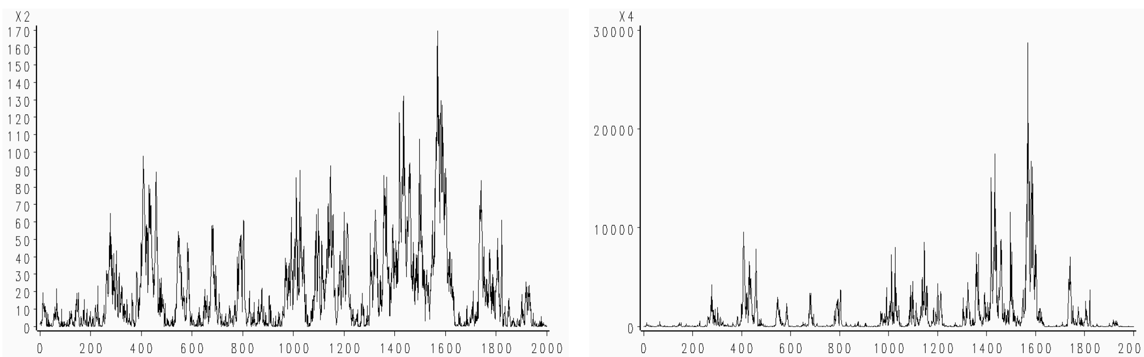


**Figure 2:** Square and fourth power of a simulated $\tilde{I}(0.8)$ time series.

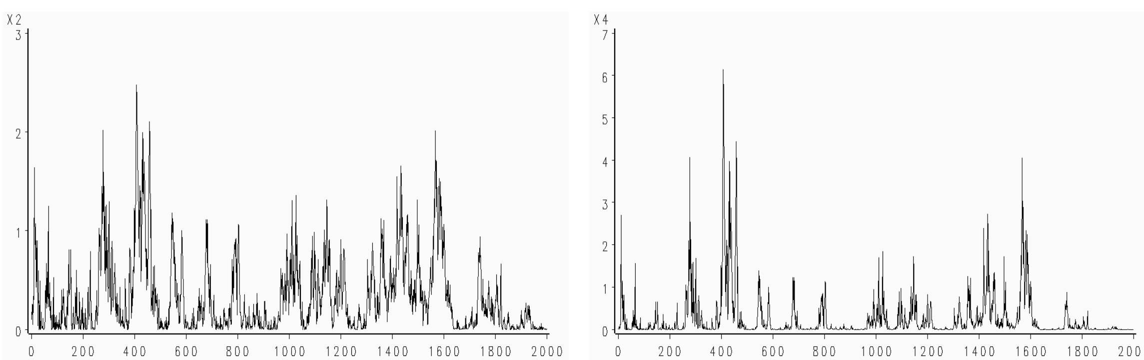


**Figure 3:** Square and fourth power of a simulated $\tilde{I}(0.8)$ time series rescaled by $t^{-0.6}$ and $t^{-1.2}$, respectively.

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 $\tilde{\mathrm{I}}(d)$ processes and the moments of power transformations of $\tilde{\mathrm{I}}(d)$ 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 $\tilde{\mathrm{I}}(d)$ process.

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

(a) Then $\mathrm{Corr}(X_t, X_{t-h}) \xrightarrow{t\to\infty} 1$.

(b) If $X_t$ is Gaussian, then $\mathrm{Corr}(X_t^m, X_{t-h}^m) \xrightarrow{t\to\infty} 1$ for all $m \in \{2,3,\ldots\}$.

**Proof:**

(a) $$\mathrm{Corr}(X_t, X_{t-h}) = \frac{E(X_t, X_{t-h})}{\sqrt{E(X_t^2)}\sqrt{E(X_{t-h}^2)}} = \frac{E\left(\sum_{j=0}^{t} c_j \varepsilon_{t-j} \sum_{i=0}^{t-h} c_i \varepsilon_{t-i}\right)}{\sqrt{\sigma^2 \sum_{j=0}^{t} c_j^2}\sqrt{\sigma^2 \sum_{j=0}^{t-h} c_j^2}} = \frac{\sigma^2 \sum_{j=0}^{t-h} c_j c_{j+h}}{\sqrt{\sigma^2 \sum_{j=0}^{t} c_j^2}\sqrt{\sigma^2 \sum_{j=0}^{t-h} c_j^2}}$$

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

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) $$\mathrm{Corr}(X_t^m, X_{t-h}^m) = \frac{E(X_t^m, X_{t-h}^m) - E(X_t^m)E(X_{t-h}^m)}{\sqrt{E(X_t^{2m}) - E(X_t^m)^2}\sqrt{E(X_{t-h}^2) - E(X_{t-h}^m)^2}}$$

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

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

(1) <u>$m = 1$</u>: $\mu_{11}(t,h) = \rho(t,h) \xrightarrow{t\to\infty} 1 = \mu_2$.

(2) <u>$m = 2$</u>: $\mu_{22}(t,h) = 1 + 2\rho^2(t,h) \xrightarrow{t\to\infty} 3$ (see Kendall, Stuart & Ord, 1987, p.103). Moreover, a formula in Johnson and Kotz (1970, p. 47) yields $\mu_4 = (4-1)(4-3) = 3$.

(3) <u>$m > 2$</u>: 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). ■

An immediate implication of Proposition 3 (b) is that the autocorrelations of $Y_t = \alpha(t)X_t^m + \beta(t)$ with Gaussian $X_t$ and any functions $\alpha(\cdot)$ and $\beta(\cdot)$ converge to 1 as well. Therefore, rescaling the process does not change its autocorrelations. Note that this is true only for the theoretical, or instantaneous, autocorrelations Corr($X_t$, $X_{t-h}$). In contrast, the empirical autocorrelations are sensitive to rescaling, so that rescaling might influence the estimated long-memory parameter.

Proposition 3 states that $\tilde{\mathrm{I}}(d)$ time series with $d > ½$ have time dependent autocorrelations which converge to 1. This property is maintained for a Gaussian process under any power transformation, even if it is detrended or rescaled afterwards. In this sense, any power transformation of an $\tilde{\mathrm{I}}(d)$ process with $d > ½$ has still long memory with some $d' > ½$. In particular, the rescaling considered in the previous section does not change this persistence property. As a consequence, we can construct *stationary* processes which have long-memory properties associated with a long-memory parameter $d > ½$, e.g., $t^{0.5-d}X_t$ if $X_t \sim \tilde{\mathrm{I}}(d)$ with $d > ½$. 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 $X_t$ ~ Gaussian $\tilde{\mathrm{I}}(d)$ with ½ < $d$ < 1, then $X_t^2$ is asymptotically LM($d$).

(b) If $X_t$ ~ $\tilde{\mathrm{I}}(d)$ with $d$ > ½, then $X_t$ is asymptotically Gaussian.

(c) If $X_t$ ~ $\tilde{\mathrm{I}}(d)$ with ½ < $d$ < 1, then $X_t^2$ is asymptotically LM($d$).

**Proof:**

(a) Let $Y_t \equiv X_t - X_{t-1}$ and $Z_t \equiv X_t + X_{t-1}$. Then ($Y_t$, $Z_t$) are jointly normally distributed with E($Y_t$) = 0 = E($Z_t$), and $\Delta(X_t^2) = Y_t Z_t$. Using a formula for covariances of products of multivariate normal variables (see Bohrnstedt and Goldberger, 1969), we obtain:

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

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:

$$\begin{aligned}
\mathrm{Cov}(Y_t, Z_{t-h}) &= \mathrm{E}[(X_t - X_{t-1})(X_{t-h} + X_{t-h-1})] \\
&= \mathrm{E}(X_t X_{t-h}) + \mathrm{E}(X_t X_{t-h-1}) - \mathrm{E}(X_{t-1} X_{t-h}) - \mathrm{E}(X_{t-1} X_{t-h-1}) \\
&= \mathrm{E}\left(\sum_{i=0}^{t} c_i \varepsilon_{t-i} \sum_{j=0}^{t-h} c_j \varepsilon_{t-h-j}\right) + \mathrm{E}\left(\sum_{i=0}^{t} c_i \varepsilon_{t-i} \sum_{j=0}^{t-h-1} c_j \varepsilon_{t-h-1-j}\right) \\
&\quad - \mathrm{E}\left(\sum_{i=0}^{t-1} c_i \varepsilon_{t-1-i} \sum_{j=0}^{t-h} c_j \varepsilon_{t-h-j}\right) - \mathrm{E}\left(\sum_{i=0}^{t-1} c_i \varepsilon_{t-1-i} \sum_{j=0}^{t-h-1} c_j \varepsilon_{t-h-1-j}\right) \\
&= \sum_{j=0}^{t-h} c_j c_{j+h} \sigma^2 + \sum_{j=0}^{t-h-1} c_j c_{j+h+1} \sigma^2 - \sum_{j=0}^{t-h} c_j c_{j+h-1} \sigma^2 - \sum_{j=0}^{t-h-1} c_j c_{j+h} \sigma^2 \\
&= \sigma^2 \left[\sum_{j=0}^{t-h-1} c_j \left(c_{j+h+1} - c_{j+h-1}\right) + c_{t-h}\left(c_t - c_{t-1}\right)\right] \\
&= \sigma^2 \left[\sum_{j=0}^{t-h-1} c_j c_{j+h-1} \left(\tfrac{j+h+d}{j+h+1} \tfrac{j+h+d-1}{j+h} - 1\right) + c_{t-h}\left(c_t - c_{t-1}\right)\right] \qquad (*)
\end{aligned}$$

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, $c_j c_{j+h-1}\left(\frac{j+h+d}{j+h+1}\frac{j+h+d-1}{j+h}-1\right) \sim O(j^{2d-3})$, 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 anologously be written as:

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

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

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:

$$\mathrm{Corr}(Y_t Z_t, Y_{t-h} Z_{t-h}) = \frac{\mathrm{Cov}(Y_t, Y_{t-h})\mathrm{Cov}(Z_t, Z_{t-h})}{\sqrt{\mathrm{Var}(Y_t Z_t)}\sqrt{\mathrm{Var}(Y_{t-h} Z_{t-h})}} + \frac{\mathrm{Cov}(Y_t, Z_{t-h})\mathrm{Cov}(Z_t, Y_{t-h})}{\sqrt{\mathrm{Var}(Y_t Z_t)}\sqrt{\mathrm{Var}(Y_{t-h} Z_{t-h})}}$$

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

$$\begin{aligned}\mathrm{Corr}(Y_t Z_t, Y_{t-h} Z_{t-h}) &\xrightarrow{\text{for large } t} \frac{\mathrm{Cov}(Y_t, Y_{t-h})}{\sqrt{\mathrm{Var}(Y_t)}\sqrt{\mathrm{Var}(Y_{t-h})}} \frac{\mathrm{Cov}(Z_t, Z_{t-h})}{\sqrt{\mathrm{Var}(Z_t)}\sqrt{\mathrm{Var}(Z_{t-h})}} \\ &= \mathrm{Corr}(Y_t, Y_{t-h})\end{aligned}$$

Here, the equality sign follows with Proposition 3 (a), using that $Z_t \sim \mathrm{I}(d)$.

As $Y_t \sim \mathrm{I}(d-1)$, $\Delta(X_t^2) = Y_t Z_t \sim \mathrm{LM}(d-1)$ and thus $X_t^2 \sim \mathrm{LM}(d)$.

(b) According to Granger (1988) it suffices to show that $\left(\sum_{j=0}^{t} c_j^3\right)^{1/3} \left(\sum_{j=0}^{t} c_j^2\right)^{-1/2} \xrightarrow{t\to\infty} 0$.

Note that $\left(\sum_{j=0}^{t} c_j^3\right)^{1/3} \underset{\text{for large t}}{\approx} \left(c + \sum_{j=0}^{t} j^{3d-3}\right)^{1/3} \sim \left[O(t^{3d-2})\right]^{1/3} = O(t^{d-2/3})$, and

$\left(\sum_{j=0}^{t} c_j^2\right)^{1/2} \underset{\text{for large t}}{\approx} \left(c + \sum_{j=0}^{t} j^{2d-2}\right)^{1/2} \sim \left[O(t^{2d-1})\right]^{1/2} = O(t^{d-1/2})$. Hence, for $d > ½$,

$\left(\sum_{j=0}^{t} c_j^2\right)^{1/2}$ diverges faster than $\left(\sum_{j=0}^{t} c_j^3\right)^{1/3}$, so that the ratio converges to zero.

(c) follows directly from (a) and (b). ■

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 an 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 paramter $d > ½$ (see Proposition 3). As a consequence, we have to consider first differences in the proof of Proposition 4. First differences of a non-linear 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.

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

**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

Table 2 illustrates the findings of the previous two propositions. It contains the average estimated long-memory parameter of some polynomial transformations of three non-stationary $\tilde{\mathrm{I}}(d)$ processes with $d = 0.6$, 0.8 and 1. The first line of each cell shows the long-memory parameter estimated from the "original" transformed series, while the second line displays the average parameter estimated from the rescaled transformed series as discussed in the previous section. The numbers in brackets 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 $\tilde{\mathrm{I}}(d)$ 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 $m = [T^{0.8}]$.

| 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) |

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

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 ($0 < d < ½$) 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. Indeed, the results of Proposition 1 do not hold for the exponential transformation, which can be written as $\exp(x) = \exp(0.5)\sum_{j=0}^{\infty}\frac{H_j(x)}{\sqrt{j!}}$ (see Cramér, 1946, p. 133). With Lemma 1 and $g_j = 1/\sqrt{j!}$ we then obtain:

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

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 (½ < $d$ < 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 $d$ > ½, 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 $d = 0.8$ 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 correlogram of 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 $d$ > ½, the long-memory of the output is less than that of the input, but still larger than ½. 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 $d$ > ½ – 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).

## References


BARRETT, J. F. AND D. G. LAMPARD (1955): An expansion for some second-order probability distributions and its application to noise problems, *IRE Transactions – Information Theory*, 10-15.

BOHRNSTEDT, GEORGE W. AND ARTHUR S. GOLDBERGER (1969): On the exact covariance of products of random variables, *Journal of the American Statistical Association*, 64, 1439-1442.

BERAN, JAN (1994): *Statistics for Long-Memory Processes*, Chapman & Hall, New York.

CRAMÉR, HARALD (1946): *Mathematical Methods of Statistics*, Princeton University Press.

GEWEKE, JOHN AND SUSAN PORTER-HUDAK (1983): "The estimation and application of long memory time series models", *Journal of Time Series Analysis*, 4, 221-238

GOURIEROUX, CHRISTIAN AND JOANN JASIAK (1999): Nonlinear persistence and copersistence, mimeo.

GRADSHTEYN, I. S. AND I. M. RYZHIK (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. AND JEFF HALLMAN (1988): The algebra of I(1), Finance and Economics Discussion Series 45, Board of Governors of the Federal Reserve System

GRANGER, C. W. J. AND JEFF HALLMAN (1991): Nonlinear transformations of integrated time series, *Journal of Time Series Analysis*, 12, 207-224

GRANGER, C. W. J. AND P. NEWBOLD (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, CLIFFORD M., ROHIT DEO AND JULIA BRODSKY (1998): "The mean squared error of Geweke and Porter-Hudak's estimator of the memory parameter of a long-memory time series", *Journal of Time Series Analysis*, 19, 19-46

JOHNSON, NORMAN I. AND SAMUEL KOTZ (1970): *Distributions in Statistics: Continuous Univariate Distributions 1*, Houghton Mifflin, Boston

JOHNSON, NORMAN I. AND SAMUEL KOTZ (1972): *Distributions in Statistics: Continuous Multivariate Distributions*, John Wiley & Sons, New York

KENDALL, MAURICE, ALAN STUART AND J. KEITH ORD (1987): *Kendall's Advanced Theory of Statistics*, Vol. 1 (Distribution Theory), Oxford University Press, New York

TAQQU, MURAD S. (1979): Convergence of integrated processes of arbitrary Hermite rank, *Probability Theory and Related Fields*, 50, 53-83