Why FARIMA Models are Brittle

D. Veitch[a], A. Gorst-Rasmussen[a,b], A. Gefferth[a,c]

May 23, 2013

[a]*Department of Electrical and Electronic Engineering, The University of Melbourne, Victoria 3010, Australia.*

[b]*Current affiliation: Department of Mathematical Sciences, Aalborg University, 9220 Aalborg East, Denmark.*

[c]*Current affiliation: Innovació i Recerca Industrial i Sostenible, Avda. Carl Friedrich Gauss n^{o}11, 08860 Castelldefels, Spain.*

Email adddresses: *dveitch@unimelb.edu.au, agorstras@gmail.com, andras.gefferth@gmail.com*

Abstract

The FARIMA models, which have long-range-dependence (LRD), are widely used in many areas. Through deriving a precise characterisation of the spectrum and variance time function, we show that this family is very atypical among LRD processes, being extremely close to the fractional Gaussian noise in a precise sense which results in ultra-fast convergence to fGn under rescaling. Furthermore, we show that this closeness property is not robust to additive noise. We argue that the use of FARIMA, and more generally fractionally differenced time series, should be reassessed in some contexts, in particular when convergence rate under rescaling is important and noise is expected.

Keywords: FARIMA, fractionally differenced process, self-similarity, fGn, long-range dependence, Hurst parameter

1 Introduction

Long-Range Dependence (LRD), or long memory, in stationary time series is a phenomenon of great importance [1]. The *Fractional AutoRegressive Integrated Moving Average* (FARIMA) models [2, 3] are very widely used as a class which inherits the flexibility of the ARMA class for short-range dependent modelling, while exhibiting LRD with tunable *Hurst parameter*, the scaling parameter of LRD. They have in particular been widely used to parsimoniously model data sets exhibiting LRD (for example [4]), and more importantly for our purposes here, they have also been employed to make quantitative assessments of the behaviour of stochastic systems in the face of LRD (for example [5]).

A good example is in relation to estimators of the Hurst parameter H. FARIMA models have been used (for example [6, 7, 8]) in order to evaluate the performance of H estimators under circumstances more challenging than that of the canonical *fractional Gaussian Noise* (fGn), in particular to assess small sample size performance using Monte Carlo simulation. In such works, it is an implicit assumption that the FARIMA model class is 'typical' enough so that the results will have relevance to LRD time series more generally.

Of course, no parametric model can be truly typical. However, for a model class to be useful it should be representative for the purposes to which it is commonly put. In this paper, we show that FARIMA time series, and more generally time series whose LRD scaling derives directly from fractional differencing such as the FEXP models [9], are far from typical when it comes to their LRD character, the very quality for which they were first introduced. In a sense we make precise, out of all possible LRD time series, their LRD behaviour is in fact 'as close as possible' to that of fGn. A key technical consequence is ultra-rapid convergence to fGn under the rescaling operation of aggregation. The implications for the role of the family is strong, namely that, in regards to LRD behaviour, *FARIMA offers no meaningful diversity beyond fGn*. A second key consequence is that the addition of additive noise (of almost any kind) pushes a FARIMA process out of the immediate neighbourhood of fGn, changing the convergence rate. In other words FARIMA is structurally unstable in this sense or *brittle*, and may therefore be unsuited for use as a class of LRD time series representing real-world signals.

This work arose out of our prior study of (second-order) self-similarity of stationary time-series [10], which highlighted the benefits of the variance time function (VTF) formulation of the autocovariance structure, over the more commonly used autocovariance function (ACVF) formulation. Using the VTF, questions of process convergence under rescaling to exactly (second-order) self-similar limits can often be more simply stated and studied.

The paper is structured as follows. After Section 2 on background material, Section 3 establishes the main results. It begins by characterising a link between a fractionally differenced process and fGn in the spectral domain. Using it, we prove that related Fourier coefficients in the time domain decay extremely quickly, and then show that as a result the VTFs of the fractionally differenced process and fGn are extremely close. We then explain why this behaviour is so atypical, and how it results in fast convergence to fGn. In Section 4 we explain why fractional processes are not robust to the addition of additive noise, even noise of particularly non-intrusive character. We also provide numerical illustrations of this brittleness, and of the fast convergence to fGn of FARIMA processes. We conclude and discuss possible implications of our findings in Section 5.

A technical document containing a complete set of detailed proofs, as well as an extension of the closeness results to the ACVF, can be found in [11].

2 Background

Let $\{X(t),\ t \in \mathbf{Z}\}$ denote a discrete time second-order stationary stochastic process with mean μ, variance $\mathcal{V} > 0$, and *autocovariance function* (ACVF), $\gamma(k) := E[(X(t) - \mu)(X(t+k) - \mu)]$, $k \in \mathbf{Z}$.

A description of the autovariance structure which is entirely equivalent to γ is the *variance time function* (VTF), defined as $\omega(n) = (\mathbf{I}\gamma)(n) := \sum_{k=0}^{n-1} \sum_{i=-k}^{k} \gamma(i)$ $n = 1,\ 2,\ 3, \ldots$, where $\mathbf{I}$ denotes the double integration operator acting on sequences. Its normalised form, the *correlation time function* (CTF), is just $\phi(n) = \omega(n)/\omega(1) = \omega(n)/\mathcal{V}$. In terms of the original process, $\omega(n)$ is just the variance of the sum $\sum_{t=1}^{n} X(t)$. It is convenient to symmetrically extend ω and ϕ to $\mathbf{Z}$ by setting $\omega(n) := \omega(-n)$ for $n < 0$ and $\omega(0) = 0$.

2.1 LRD, Second-Order Self-Similarity, and Comparing to fGn

There are a number of definitions of long-range dependence, all of which encapsulate the idea of slow decay of correlations over time. Common definitions include power-law tail decay of the ACVF $\gamma(n) \overset{n\to\infty}{\sim} c_\gamma n^{2H-2}$, or power-law divergence of the spectral density at the origin $f(x) \overset{x\to 0}{\sim} c_f |x|^{-(2H-1)}$ for related constants c_γ and c_f.

The well known *fractional Gaussian noise* (fGn) family, parameterised by the *Hurst parameter* $H \in [0, 1]$ and variance $\mathcal{V} > 0$, has $\omega(m) = \omega^*_{H,\mathcal{V}}(m) := \mathcal{V}m^{2H}$ (to lighten notation we sometimes write ω^*_H or simply ω^*). It has long memory if and only if $H \in (1/2, 1]$. We denote the corresponding ACVF and spectral density as γ^*_H and f^*_H respectively. The latter is given by

$$\begin{aligned} f^*_H(x) &= c^*_f\, \pi^{-2}(2\pi)^{2H+1} \sin^2(\pi x) \sum_{j=-\infty}^{\infty} |2\pi j + 2\pi x|^{-(2H+1)} \qquad (1) \\ &\overset{x\to 0}{\sim} c^*_f |x|^{-(2H-1)}, \quad x \in [-1/2, 1/2], \end{aligned}$$

where $c^*_f = \mathcal{V}(2\pi)^{2-2H} C(H) > 0$ and $C(H) = \pi^{-1} H \Gamma(2H) \sin(H\pi)$ (see [12], pp 333-4, but note that the change to normalised frequency multiplies f^*_H by 2π, and c^*_f by $(2\pi)^{2-2H}$).

In this paper we compare against fGn with $H \in (1/2, 1]$ as it plays a special role among among LRD processes; that of being a family of *second-order self-similar* time series[1]. To understand how this comparison can be made, we must define self-similarity and related notions.

Self-similarity relates to invariance with respect to a rescaling operation. In the present context, the time rescaling is provided by *aggregation*. For a fixed $m \geq 1$, the *aggregation of level* m of the original process X is the process $X^{(m)}$ defined as

$$X^{(m)}(t) := \frac{1}{m} \sum_{j=m(t-1)+1}^{mt} X(j).$$

The γ, ω, ϕ functions and the variance of the m-aggregated process will be denoted by $\gamma^{(m)}$, $\omega^{(m)}$, $\phi^{(m)}$ and $\mathcal{V}^{(m)}$ respectively. It is not difficult to show [10] that

$$\omega^{(m)}(n) = \frac{\omega(mn)}{m^2}, \qquad \mathcal{V}^{(m)} = \frac{\omega(m)}{m^2}. \tag{2}$$

To seek invariance, the time rescaling must be accompanied by a compensating amplitude rescaling. This is performed naturally by dividing by $\mathcal{V}^{(m)}$, which amounts to examining the effect of aggregation on the correlation structure. Combining the time and amplitude rescalings yields the correlation renormalisation

$$\phi^{(m)}(n) = \frac{\phi(mn)}{\phi(m)} = \frac{\omega(mn)}{\omega(m)}. \tag{3}$$

We can now define second-order self-similarity as the fixed points of this operator.

Definition 1. A process is second-order self-similar iff $\phi^{(m)} = \phi$, for all $m = 1, 2, 3, \ldots$.

It is easy to see that fGn, which has $\phi(m) = \phi^*_H(m) := m^{2H}$, satisfies this definition for all $H \in [0, 1]$.

Given a fixed point ϕ^*_H, we define its *domain of attraction* (DoA) to be those time series which converge to it pointwise under the action of (3). This definition is very general, in particular it includes processes whose VTFs have divergent slowly varying prefactors. It provides a natural way to define LRD which subsumes and generalises most other definitions including those above [10], namely: *a time series is long-range dependent if and only if it is in the domain of attraction of ϕ^*_H for some $H \in (0.5, 1]$.*

With the above definitions the DoA are revealed as the natural way to partition the space of all LRD processes, namely into equivalence classes of processes corresponding to the different fGn fixed points. Since all processes within a DoA converge to the same fixed point, their asymptotic structure can be meaningfully compared both against each other and to the fixed point itself. Alternatively if two processes were in different DoAs then they cannot be close asymptotically as they would converge to different processes.

Section 3.2 provides a precise characterisation of the closeness of a fractionally differenced process to its corresponding fixed point, and its associated fast convergence under renormalization. More generally, within a given DoA, one can further partition processes according to some measure of distance from the common fixed point. Section 3.3 establishes such a notion.

2.2 Fractionally Differenced Processes and FARIMA

Let B denote the backshift operator. The fractional differencing operator of order $d > -1$ is given by

$$(1-B)^d := \sum_{j=0}^{\infty} \frac{\Gamma(j-d)}{\Gamma(-d)\Gamma(j+1)} B^j.$$

Let $\{Y(t), t \in \mathbf{Z}\}$ be a second-order stationary stochastic process. Assuming $H \in (0, 1)$ the process

$$X := (1-B)^{-(H-1/2)} Y$$

[1] Until recently, fGn was considered to be the only such family. A second (and final) family was discovered recently [10].

is called a fractionally differenced process, driven by Y, with differencing parameter $H - 1/2$.

If h is the spectral density of Y then X has spectral density ([13], Thm. 4.10.1)

$$f_H(x) = h(x)\left|1 - \mathrm{e}^{2\pi i x}\right|^{-(2H-1)} = h(x)|2\sin \pi x|^{-(2H-1)} \quad x \in [-1/2, 1/2]. \tag{4}$$

In this paper we assume that Y is short-range dependent, and in particular that h satisfies:

- $h(x) > 0$ and is continuous for all $x \in [-1/2, 1/2]$ (and is therefore bounded);
- h is three times continuously differentiable on $(-1/2, 1/2)$.

Under such conditions, the ACVF of X exists and satisfies $\gamma_H(n) \sim c_\gamma n^{2H-2}$ for some constant c_γ ([13], Thm. 13.2.2). Hence, when $H \in (1/2, 1)$ the process X is LRD with Hurst parameter H. We denote the VTP of such a process as ω_H, and continuing from (4),

$$\begin{aligned} f_H(x) = f_{(H,h)}(x) &= c_f(2\pi)^{2H-1}\frac{h(x)}{h(0)}|2\sin \pi x|^{-(2H-1)} \\ &\overset{x\to 0}{\sim} c_f|x|^{-(2H-1)}, \quad x \in [-1/2, 1/2], \end{aligned} \tag{5}$$

where $c_f = (2\pi)^{1-2H}h(0) > 0$.

An important example of a fractionally differenced process is the FARIMA class [2] where h is the spectral density of a causal invertible ARMA model. This family includes the ARMA family as the special case $H = 1/2$. Another class is the class of FEXP-models (e.g. [14, 9, 15]) which comes from taking the logarithm of h to be a trigonometric polynomial, i.e. $\log h(x) = \theta_1 \cos x + \theta_2 \cos(2x) + \cdots + \theta_{q-1}\cos((q-1)x)$ for real coefficients. Both FARIMA and FEXP models are widely used in statistical applications since, in addition to exhibiting LRD, they both enable modelling of arbitrary short-range correlation structures.

3 Fractionally Differenced Processes are Unusual LRD Processes

This section establishes our main results, detailed characterisations of the closeness of the asymptotic covariance structure of a fractionally differenced process to that of fGn.

Our approach can be described as follows. We begin in the spectral domain where the relationship between the processes can be simply stated through a function g by defining

$$f_H(x) = f_H^*(x)g(x). \tag{6}$$

The simple closed form of the spectra (1) and (5) allow g to be explicitly written. We study the properties of g, obtaining a characterisation of the closeness of the processes in the spectral domain (Theorem 1). This leads to a convolution formulation $\gamma_H = \gamma_H^* \star G$ in the time domain, where G is the Fourier Series of g, and thereby to a similar relationship for the VTFs, where the fast decay of the Fourier coefficients can be used to characterise the closeness (Theorem 2). The VTF result then allows the closeness within the DoA and the convergence speed to be easily established (Theorem 3).

Proofs are left to the appendix, and to focus on the key points some are either omitted or abbreviated. Full details can be found in [11].

3.1 Closeness of the Spectrum

The following spectral closeness result, proved in the appendix, is the foundation of our VTF results.

For $\alpha \geq 0$ let Λ_α denote the normed space of uniformly α-Hölder continuous functions, and V denote the linear space of functions of bounded variation, each on $[-1/2, 1/2]$. Each is closed under pointwise multiplication, addition, and reciprocation for functions bounded away from zero. By *smooth function* we mean one in C^∞.

Theorem 1. *Assume that* $H \in [1/2, 1)$ *and define* $g(x) := f_H(x)/f_H^*(x)$, $x \neq 0$ *and* $g(0) := \lim_{x\to 0} g(x)$. *Then* $g(0) = c_f/c_f^* = h(0)/(2\pi\mathcal{V}C(H))$ *and* g *satisfies the following over* $[-1/2, 1/2]$*:*

(i) *g is even, continuous, positive, bounded, and L^p, $p > 0$;*
(ii) *g is twice differentiable, and smooth away from $x = 0$;*
(iii) *$g'' \in \Lambda_{2H-1} \cap \mathrm{V}$, but $g'' \notin \Lambda_{\beta'}$ for $\beta' > 2H - 1$;*
(iv) *g admits a Fourier series with coefficients $\{G_j\}$ such that $\sum_{j=-\infty}^{\infty} j^2 |G_j| < \infty$. In particular $\sum_{j=-\infty}^{\infty} |G_j| < \infty$ and $\sum_{j=-\infty}^{\infty} j^\alpha |G_j| < \infty$ for $1 < \alpha < 2$.*

Note that for an arbitrary LRD process, only boundedness of g at the origin would be automatic. In contrast, as shown in Figure 1 the g for a fractionally differenced process is a very well behaved function.

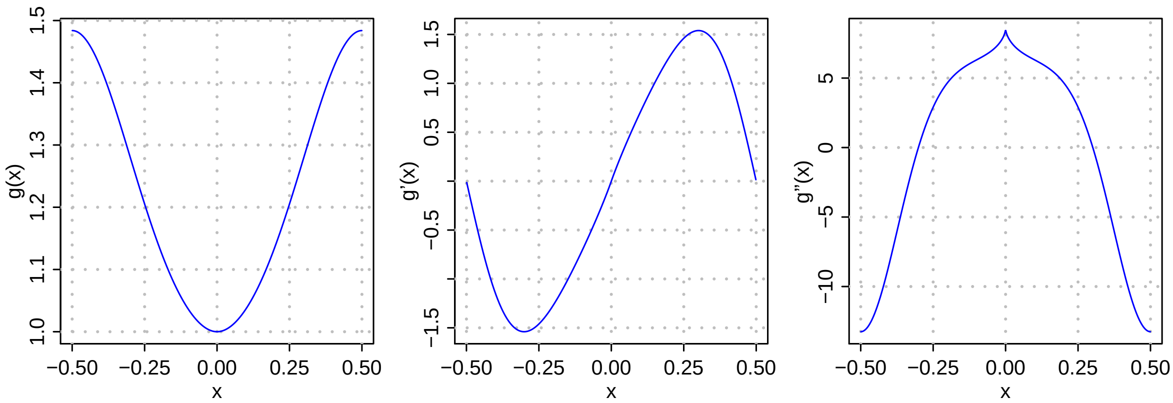

Figure 1: The function $g(x) = f_H(x)/f_H^*(x)$ and its first two derivatives in the canonical case of a pure fractionally differenced process (FARIMA with trivial ARMA components or 'FARIMA0d0') with $H = 0.8$. Here we have set $c_f = c_f^*$, so $g(0) = 1$.

3.2 Closeness of the VTF

It is straightforward to confirm that, thanks to the nice behaviour of g and G detailed in Theorem 1, the spectral relationship $f_H(x) = f_H^*(x) g(x)$ translates as expected.

Lemma 1. *The auto-covariance functions γ_H and γ_H^* are related through the convolution $\gamma_H = \gamma_H^* \star G$.*

Since $\omega_H = \mathbf{I}\gamma_H$, it is tempting to seek a relationship of the form $\omega_H = G \star \omega_H^*$ through taking the 'double integral' of $\gamma_H = G \star \gamma_H^*$. The following lemma provides a sufficient condition for the existence of such a convolution, as well as some of its important properties which will be crucial in what follows.

Lemma 2. *Assume $1 < \alpha < 2$ and let $a = \{|n|^\alpha, n \in \mathbf{Z}\}$. Let b be a symmetric sequence satisfying $\sum_{j=1}^{\infty} j^\alpha |b_j| < \infty$. Then $S_b = \sum_{j=-\infty}^{\infty} b_j$ and the symmetric sequence $c = a \star b$ exist, and $(c_n - S_b a_n) \overset{n\to\infty}{\to} 0$.*

The proof is given in the appendix.

Corollary 1. *The convolution $G \star \omega_H^*$ exists for $H \in (1/2, 1)$.*

Proof. Set $b = G$ in Lemma 2. The condition on b holds since $\sum_{j=1}^{\infty} j^\alpha |G_j| < \sum_{j=1}^{\infty} j^2 |G_j|$ which is finite from Theorem 1. The result then follows immediately by identifying α with $2H$ and a with ω_H^*. □

The following lemma, proved in the appendix, shows that, if existence is granted, taking the 'double integral' of a convolution is straightforward, provided a double counting issue at the origin is allowed for.

Lemma 3. *Let a, b be symmetric sequences and assume that $c := a \star b$ exists. Then $\mathbf{I}c$ exists, and if $(\mathbf{I}a) \star b$ exists, then $\mathbf{I}c = (\mathbf{I}a) \star b - ((\mathbf{I}a) \star b)_0$.*

We are now able to prove our main result on the VTF.

Theorem 2. *Let ω_H denote the VTF of a fractionally differenced process with $H \in (1/2, 1)$ and c_f chosen equal to c_f^*. Then*

$$\omega_H(n) = \omega_H^*(n) + D + o(1)$$

where $D = -2\sum_{j=1}^{\infty} j^\alpha |G_j| < 0$ is a constant.

Proof. Since each of $\gamma_H^* \star G$ and $\omega_H^* \star G$ exist, Lemma 3 applies upon identifying $a = \gamma_H^*$, $b = G$ and $c = \gamma_H$ and states that $\omega_H = \omega_H^* \star G - \{\omega_H \star G\}(0)$. From Lemma 2 with $b = G$, $S_G = \sum_{j=-\infty}^{\infty} G_j < \infty$ exists. By introducing the term $S_G \omega_H^*$ we obtain

$$\begin{aligned} \omega_H &= S_G\omega_H^* + \left(\omega_H^* \star G - S_G\omega_H^*\right) - \{\omega_H \star G\}(0) \\ &= S_G\omega_H^* + o(1) - 2\sum_{j=1}^{\infty} j^\alpha |G_j| \end{aligned}$$

by the final part of Lemma 2. Since $S_G = g(0) = c_f/c_f^* = 1$, the result follows. □

The key property underlying this result is $\omega_H^* \star G - S_G\omega_H^* \overset{n\to\infty}{\to} 0$, which shows that G is 'compact' enough to act as an aggregate multiplier S_G asympotically. This is analogous to the role the covariance sum $S_\gamma := \sum_{k=-\infty}^{\infty} \gamma(k)$ plays in the asymptotic variance of aggregated short-range dependence processes [10].

3.3 Atypicality and Speed of Convergence

Theorem 2 showed that the VTF of a fractionally differenced process is asymptotically equal to the VTF of its fGn fixed point up to an additive constant. This makes fractionally differenced processes highly atypical among LRD processes. We show this first for the VTF itself, and then for the speed of convergence of the CTF to the fixed point.

Without loss of generality, the VTF of any time series in the domain of attraction of a given fGn can be expressed as

$$\omega_H(n) = \omega_H^*(n) + \omega_d(n) \tag{7}$$

where ω_d represents the distance of the VTF from its limiting fGn counterpart. By definition, $\omega_d(n) = o(n^{2H})$, but otherwise the growth rate of ω_d is not constrained, implying that there is considerable variety within the domain of attraction.

One way of characterising the size of the difference $\omega_d(n)$ is to use *regular variation* [16, 10]. A regularly varying function $f(n)$ of index β and integer argument $n \in \mathbf{N}^+$ satisfies $\lim_{k\to\infty} f(kn)/f(k) = n^\beta$, $\beta \in \mathbf{R}$. Assume without loss of generality that ω_d is upper bounded by a regularly varying function of index $\beta \in [0, 2H]$, that is

$$\omega_d(n) = O(s(n)n^\beta), \tag{8}$$

where s is a *slowly varying* function (that is regularly varying with index 0), and β is the infimum of indices for which (8) holds. A notion of *closeness* of the process to the limiting fGn can then be defined in terms of β, where the smaller the index, the closer the process.

According to this scheme, Theorem 2 states that fractionally differenced processes belong in the closest layer of the hierarchy, corresponding to $\beta = 0$. Furthermore, the theorem shows that $s(n)$ (which could in general diverge, for example $s(n) \overset{n\to\infty}{\sim} \log(n)$) tends to a constant. Thus, the VTF of a fractionally differenced process lies in a very tight neighbourhood indeed of the VTF of its limiting fixed point. Far from being typical LRD processes, they deviate only in very subtle ways from fGn in terms of their large lag behaviour.

From (3), there is a direct relationship between closeness in the above sense and speed of convergence of the CTF to its fixed point under aggregation.

Theorem 3. *Let ϕ_H denote the CTF of a fractionally differenced process in the domain of attraction of ϕ^*_H with $H \in (1/2, 1)$. Then*

$$\phi_H^{(m)}(n) = \phi^*_H(n) + D(1 - n^{2H})m^{-2H} + o(m^{-2H}) = \phi^*_H(n) + O(m^{-2H})$$

where D is the constant from Theorem 2.

Proof. The result follows from substituting $\omega_H(n) = \omega^*_H(n) + D + o(1)$ from Theorem 2 in (3) and using $(1+x)^{-1} = 1 - x + O(x^2)$. □

Beginning from (7), it holds generally for LRD processes in the DoA of ϕ^*_H that $\phi_H^{(m)}(n) = \phi^*_H(n) + O(s(m)m^{-2H+\beta})$. It follows that fractionally differenced processes, for which $\beta = 0$ and $s(m)$ is identically equal to a constant, converge faster to the fixed point compared to all other processes in the DoA.

4 Fractional Processes are Brittle

As pointed out at the end of Section 3, fractionally differenced processes converge 'almost immediately' to their fGn fixed point compared to other processes in the domain of attraction, and this is true in terms of each of the (normalised) VTF, ACVF and spectrum. In this section we point out and illustrate a key consequence of this fact, namely the *brittleness* of fractionally differenced models.

4.1 Brittleness

Imperfections in physical measurement are often treated through the concept of a random observation noise. A common choice is that of additive independent Gaussian noise, either white or coloured. In the present context, this corresponds to adding to the original VTF (or ACVF, or spectrum) the VTF (respectively ACVF, spectrum) of a *Short-Range Dependent* (SRD) noise process, that is a noise whose own fGn fixed point has $H' = 1/2$.

More generally, we can categorize possible noise time series according to the Hurst parameter H' of their own fixed points. Time series with $H' \in (1/2, 1]$, $H' = 1/2$, $H' \in (0, 1/2)$ are known as LRD, SRD and *Constrained Short-Range Dependent* (CSRD) processes respectively (see [10]). The final special case of $H' = 0$ corresponds to the first difference of white noise and has the property that $\lim_{n\to\infty} \phi_0(n) = 1$.

Because the quantity D from Theorem 2 is a constant, it is evident that we can essentially think of a fractionally differenced process as an fGn to which a $H' = 0$ process has been added. Adding a SRD noise will therefore

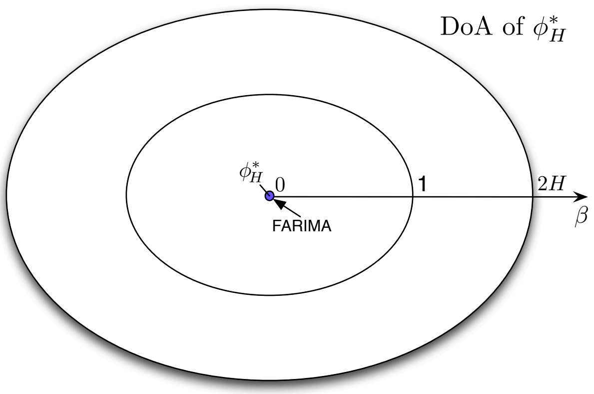

Figure 2: A schematic representation of the DoA of the fixed point ϕ^*_H, and the closeness as measured by β. Fractionally differenced processes, shown as the central disk, in fact correspond to a subset of the innermost 'circle' $\beta = 0$.

change the asymptotic behaviour, because its $H'=1/2$ asymptotics is 'stronger' than the $H'=0$ asymptotics. In terms of the hierarchy within the DoA described by the index β from (8), whereas the original process lies at the centre with $\beta=0$, the SRD-perturbed process will lie considerably further out, with $\beta=1$ (see Figure 2). A similar observation can be made if we instead add a LRD noise with $H' \in (1/2, H)$ (resulting in $\beta \in (1, 2H)$), or even a CSRD process with $H' > 0$ (resulting in $\beta \in (0,1)$). In other works, the 'error' process separating differenced processes from fGn is so special, corresponding to the extreme case of $H'=0$ resulting in $\beta = 2H' = 0$, that (additive) noise of any other kind will alter its character.

Since the addition of even trace amounts of noise of diverse kinds will change the asymptotics, pushing the process further from its fGn limit and therefore slowing its convergence rate to it under aggregration, fractional differencing models are 'brittle' or non-robust in this sense. Properties of systems driven by such processes may therefore differ qualitatively from properties of the same system once noise is added.

4.2 Numerical Illustrations

In this section we illustrate the brittle nature of fractionally differenced processes through high accuracy numerical evaluation of the VTF of FARIMA time series, both with and without additive noise.

Three different examples will be considered, two with SRD-noise and one with LRD-noise. More precisely, the perturbed processes are $Z_i(t) = X_i(t) + \sqrt{0.1}Y_i(t)$ for $i = 1, 2, 3$, where

1) X_1: unit variance $\mathrm{FARIMA}(0, 0.3, 0)$;
Y_1: unit variance Gaussian white noise,

2) X_2: unit variance $\mathrm{FARIMA}(1, 0.3, 1)$ with ARMA parameters $(\phi_1, \theta_1) = (0.3, 0.7)$;
Y_2: unit variance $\mathrm{ARMA}(1, 1)$ process also with ARMA parameters $(\phi_1, \theta_1) = (0.3, 0.7)$,

3) X_3: unit variance $\mathrm{FARIMA}(0, 0.3, 0)$;
Y_3: unit variance $\mathrm{FARIMA}(0, 0.2, 0)$.

In each case, the original process X_i and the perturbed process Z_i share a common fGn fixed point, but have unequal variances. It may seem unfair to compare results for processes with different variances, however the opposite is true. In fact, if the variances of Z_i and X_i were chosen equal, this would mean that $c_f \neq c_f^*$, and so their fGn limits would be different, rendering meaningful comparison impossible. To see this more directly, from the definitions in Section 2.1 it is clear that adding a perturbation corresponding to a smaller H value does not alter the fixed point. On the other hand the variance must increase when an independent noise is added.

For each example $i = 1, 2, 3$, we calculate the VTF of Z_i and X_i and normalise them by dividing by their common fGn limit ω_H^*. Closeness to fGn can therefore be evaluated by looking to see how the normalised VTF deviates from 1 for each lag. Maple version 13 was used to numerically evaluate the variance time functions to a high degree of precision.

Figure 3 displays the normalised VTFs for lags 1-10 for aggregation levels $m = 1$, 10, and 100, with one example per column. The graphs clearly demonstrate that even a small departure from FARIMA takes the process much further away from its corresponding fGn. Indeed, after an aggregation of level 100, in each case the VTF of the original process is visually indistinguishable from its fGn limits compared to their perturbed versions.

Note that the second column in the figure gives an example where before aggregation ($m = 1$) the perturbed process was in fact *closer* to the fixed point over the first few lags, where most of the obvious autocovariance lies. Under aggregation however, this quickly reverses as the different asymptotic behaviours of the original and perturbed processes manifest and become dominant at all lags.

5 Discussion

Focusing on the variance-time function, we have shown that fractionally differenced processes have an asymptotic autocovariance structure which is extremely close to that of the fractional Gaussian noise, more

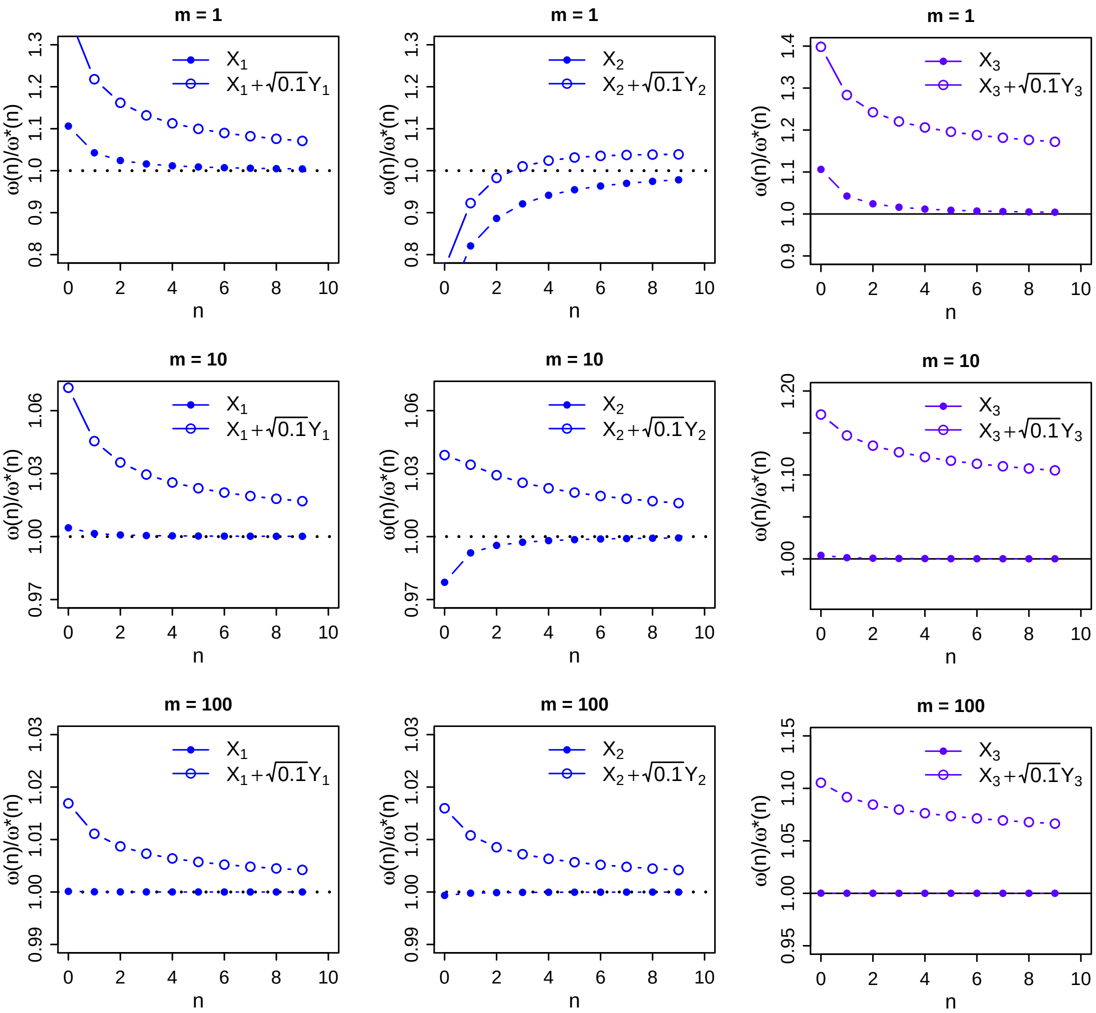

Figure 3: Ratios of VTFs of original FARIMA and perturbed processes to their fGn limit, both originally and under aggregation, one column per example. The solid circles denote unperturbed FARIMA; the hollow circles the perturbed ones. It is seen that the VTFs for unperturbed FARIMA converge much faster than their perturbed counterparts.

specifically, to that of the fGn fixed point to which the given process will tend under aggregation based renormalisation.

We showed that the natural class of processes against which this behaviour should be compared are those in the domain of attraction of the fGn fixed point limit. Using regular variation to provide a measure of distance from this fixed point within the DoA, we were able to quantify the nature of this 'closeness' precisely, and to confirm that the fractionally differenced class are indeed highly unusual in this regard, resulting in very fast convergence to fGn under renomalisation. We then used this fact to point out that the fractionally differenced process class is brittle, that is, non robust to the presence of noise. In particular we showed that the addition of arbitrarily small amounts of independent noise, not only Gaussian white noise but also noise sequences which are much gentler in a precise sense, changes the asymptotic covariance structure qualitatively.

The assessment of the impact of the brittleness of fractionally differenced models is beyond the scope of this work, as it will depend intimately on each particular application as well as the nature of the noise in question. However, we argue that conclusions based on the perception that FARIMA and related models represent 'typical' LRD behaviour may need to be reassessed, in particular in contexts where noise is important to consider. To give an example of a possible impact in the noiseless case, we conclude by expanding upon the comments given in the introduction on statistical estimation.

The closeness of a process to its fGn fixed point in functional terms is directly related to the speed of convergence of that process to the fixed point under aggregation. One application where this fact carries direct implications is the performance of statistical estimators for the Hurst parameter H. Fundamentally, semi-parametric estimators of scaling parameters such as H are based on underlying estimates made at a set of 'aggregations' at different levels, that is at multiple scales [9, 17, 18, 6]. The sophistication of particular estimators notwithstanding, this is true regardless of whether they are based in the spectral, time, or wavelet domains, though the technical details vary considerably. In the time domain using time domain aggregation the link is of course direct, and reduces to looking at the asymptotically power-law nature of $\mathcal{V}^{(m)} = \omega^{(m)}(0)$ as a function of m in some form. This is precisely where fractionally differenced processes are at a real advantage, as this quantity converges extremely quickly to that of the fGn fixed point, whose ideal power-law behaviour $\mathcal{V}^{(m)} = \mathcal{V}m^{2H}$ allows H to be easily recovered. As a result, estimator performance evaluated through the use of fractionally differenced models would be superior to that for LRD processes more generally. Note that we are not recommending that H estimation be performed directly in the time domain by regressing $\hat{\mathcal{V}}^{(m)}$ on m, indeed we have argued the opposite [18]. Our point is that the extreme closeness of such models to fGn must ultimately manifest in simpler asymptotic behaviour which will, in general, translate to improved estimation. Indeed, in the spectral domain, the importance of the degree of smoothness at the origin for the ultimate limits on estimator performance has already been noted [19]. Note that the above observations in no way put into question the findings of prior work on estimation of fractional processes in noise.

6 Appendix

Proof of Theorem 1

Proof. First, since $f_H(x) \overset{x\to 0}{\sim} c_f^*|x|^{-(2H-1)}$ and $f_H^*(x) \overset{x\to 0}{\sim} c_f|x|^{-(2H-1)}$, $g(0) := \lim_{x\to 0} g(x) = c_f/c_f^*$.

The proof of (i) is straightforward. To prove the smoothness properties (ii) and (iii), we first establish those of $\tilde{g}$ defined as

$$\tilde{g}(x) := \frac{c_f \pi^{2H+1}}{c_f^* h(0)} \cdot \frac{h(x)}{g(x)} \tag{9}$$

$$= \left|\frac{\sin(\pi x)}{\pi x}\right|^{2H+1} + |\sin(\pi x)|^{2H+1} \sum_{\substack{j=-\infty \\ j\neq 0}}^{\infty} |\pi j + \pi x|^{-(2H+1)} \tag{10}$$

$$:= |a(x)|^{2H+1} + |b(x)|^{2H+1} c(x). \tag{11}$$

It is not difficult to show that each of a, b and c are smooth, and so $\tilde{g}$ is smooth everywhere except at the origin where its smoothness is controlled by that of $|b|^{2H+1}$, which we now study.

Let $\beta = 2H - 1$. Since b is smooth and $\beta \in (0, 1)$, $|b|^{\beta+2}$ is twice differentiable at the origin. The smoothness of its second derivative is controlled by $(b')^2|b|^\beta$, which, since $b \in \Lambda_1$ and $x \mapsto |x|^\beta$ is in Λ_β, is also in Λ_β by the multiplicative and compositional closure properties of Λ_β. It follows that $\tilde{g}''$ exists and is in Λ_β. Since however $x \mapsto |x|^\beta$ is not in $\Lambda_{\beta'}$ for any $\beta' > \beta$, and moreover $b(x) \overset{x\to 0}{\sim} \pi x$ and $b'(0) \neq 0$, $\tilde{g}''$ is not in $\Lambda_{\beta'}$ for any $\beta' > \beta$.

Since smooth functions are in V, by similar arguments using the closure properties of V, we have $\tilde{g}'' \in$ V if $|b|^\beta \in$ V. The latter holds since it is easy to see that $|b|^\beta$ is monotone (with total variation 2).

We have shown that $\tilde{g}''$ exists and is in $\Lambda_{2H-1} \cap$V, but not in $\Lambda_{\beta'}$ for any $\beta' > 2H-1$. We now prove the same for g using (9). It suffices to consider $1/\tilde{g}$ since h''' exists. Since $\tilde{g}$ is bounded away from zero, (ii) follows since $(1/\tilde{g})'' = 2(\tilde{g}')^2/\tilde{g}^3 - \tilde{g}''/\tilde{g}^2$ clearly exists, and is smooth away from the origin. Now consider (iii). It follows from the last expression and the fact that $\tilde{g} > 0$ that $(1/\tilde{g})''$ and hence g'' are in V and Λ_{2H-1} by applying the respective closure properties. Finally, since $1/\tilde{g}^2(0) \neq 0$, the smoothness of $(1/\tilde{g})''$ is controlled by that of $\tilde{g}''$ and so $(1/\tilde{g})'' \notin \Lambda_{\beta'}$ for any $\beta' > 2H - 1$. This completes the proof of (iii).

We now prove (iv). Since each of g, g', and g'' are continuous and bounded, the Fourier series for each exists and are related by term by term differentiation ([20], Thm. 15.19). In particular $g(x) = \sum_{j=-\infty}^{\infty} G_j e^{2\pi ijx}$, and we can write $g''(x) = -4\pi^2 \sum_{j=-\infty}^{\infty} j^2 G_j e^{2\pi ijx}$. Now Zygmund [21], Thm. VI.3.6 states that the Fourier Series of a function in $\Lambda_\beta \cap \mathrm{V}$ for some $\beta > 0$ converges absolutely. This applies to g'' and proves that $\sum_{j=-\infty}^{\infty} j^2|G_j| < \infty$ as claimed. □

Lemma 2. *Assume* $1 < \alpha < 2$ *and let* $a = \{|n|^\alpha : n \in \mathbf{Z}\}$. *Let* b *be a symmetric sequence satisfying* $\sum_{j=1}^{\infty} j^\alpha |b_j| < \infty$. *Then* $S_b = \sum_{j=-\infty}^{\infty} b_j$ *and the symmetric sequence* $c = a \star b$ *exist, and* $(c_n - S_b a_n) \overset{n\to\infty}{\to} 0$.

Proof. Since $\alpha > 1$, $\sum_{j=-\infty}^{\infty} |b_j| \le |b_0| + 2\sum_{j=1}^{\infty} j^\alpha |b_j| < \infty$, so b is absolutely summable and hence summable. Clearly $c_0 = \sum_{j=-\infty}^{\infty} |-j|^\alpha b_j$ exists by the assumptions on b, and similarly one can show that $|c_n| < \infty$ for $n > 0$. Since both a and b are symmetric, c_n also exists for $n < 0$, and so c exists and is symmetric.
For the last part, since $c_n - S_b a_n$ is symmetric in n we assume $n \ge 0$ and rewrite it as

$$\sum_{j=-\infty}^{\infty} |n-j|^\alpha b_j - n^\alpha \sum_{j=-\infty}^{\infty} b_j = n^\alpha b_0 + \sum_{j=1}^{\infty} (n+j)^\alpha b_j + \sum_{j=1}^{\infty} |n-j|^\alpha b_j - n^\alpha \sum_{j=-\infty}^{\infty} b_j = \sum_{j=1}^{\infty} T_n^j b_j$$

where $T_n^j := |n-j|^\alpha + (n+j)^\alpha - 2n^\alpha$, $n \ge 0$, $j > 0$. Noticing that $T_n^j = f_\alpha(n,j)$ from Lemma A below, we have that $T_n^j < T_0^j = 2j^\alpha$ for each fixed j, and so

$$|c_n - S_b a_n| \le \sum_{j=1}^{N} |T_n^j||b_j| + \sum_{j=N+1}^{\infty} |T_n^j||b_j| < \sum_{j=1}^{N} |T_n^j||b_j| + 2\sum_{j=N+1}^{\infty} j^\alpha |b_j|.$$

Now given any $\varepsilon > 0$, a $N(\varepsilon) > 1$ can be found such that $\sum_{j=N+1}^{\infty} j^\alpha |b_j| < \varepsilon/4$. Next, since $T_n^j \overset{n\to\infty}{\to} 0$ for any fixed j (Lemma A below), there exists an $n_0(N)$ such that $\sum_{j=1}^{N} |T_n^j||b_j| < \varepsilon/2$ when $n \ge n_0$. It follows that $|c_n - S_b a_n| < \varepsilon$ for $n \ge n_0$ and so $(c_n - S_b a_n) \overset{n\to\infty}{\to} 0$. □

Lemma A *Assume* $1 < \alpha < 2$ *and define* $f_\alpha(x,y) := |x-y|^\alpha + (x+y)^\alpha - 2x^\alpha$ *for* $x \ge 0$, $y > 0$. *For each* y, $f_\alpha(\,\cdot\,, y)$ *is positive, strictly decreasing, and* $\lim_{x\to\infty} f_\alpha(x,y) = 0$.

Proof. Fix $y > 0$. We split the domain of $f_\alpha(\,\cdot\,, y)$ into two cases.
Let $x \ge y$. It follows that $f'_\alpha(\,\cdot\,, y) = \alpha f_{\alpha-1}(\,\cdot\,, y)$. Define $g(x) = x^\alpha$. Since $g'(x) = \alpha x^{\alpha-1}$ is strictly concave, $(x-y)^{\alpha-1} + (x+y)^{\alpha-1} < 2x^{\alpha-1}$ and so $f'_\alpha(\,\cdot\,, y) < 0$ and $f_\alpha(\,\cdot\,, y)$ is strictly decreasing. To prove $\lim_{x\to\infty} f_\alpha(x,y) = 0$, we apply the mean value theorem twice to g, and then once to g', to obtain:

$$\begin{aligned} f_\alpha(x,y) &= \big((x+y)^\alpha - x^\alpha\big) - \big(x^\alpha - (x-y)^\alpha\big) && (12)\\ &< \alpha y\big((x+y)^{\alpha-1} - (x-y)^{\alpha-1}\big) && (13)\\ &< 2\alpha(\alpha-1)y^2(x-y)^{\alpha-2} && (14) \end{aligned}$$

(since g' is strictly increasing and g'' strictly decreasing), which tends to zero as $x \to \infty$.
Let $x < y$. In this case, the derivative with respect to x yields

$$\begin{aligned} f'_\alpha(x,y) &= \alpha\big((x+y)^{\alpha-1} - (y-x)^{\alpha-1} - 2x^{\alpha-1}\big) && (15)\\ &< \alpha\big((x+y)^{\alpha-1} - (y-x)^{\alpha-1} - (2x)^{\alpha-1}\big) && (16)\\ &= \alpha\big(h_x(y) - h_x(x)\big) && (17) \end{aligned}$$

where $h_x(y) = (x+y)^{\alpha-1} - (y-x)^{\alpha-1}$. Since the derivative of h_x is negative for $x > 0$, h_x is strictly decreasing. It follows that $f'_\alpha(\cdot, y) < 0$ and so $f_\alpha(\,\cdot\,, y)$ is likewise strictly decreasing.
Finally, since $f_\alpha(x,y)$ is decreasing for all $x \ge 0$ and tends to zero, it is positive. □

Lemma 3. *Let a, b be symmetric sequences and assume that $c := a \star b$ exists. Then $\mathbf{I}c$ exists, and if $(\mathbf{I}a) \star b$ exists, then $\mathbf{I}c = (\mathbf{I}a) \star b - ((\mathbf{I}a) \star b)_0$.*

Proof. Since $(\mathbf{I}c)_n$ is a finite sum of elements of c, it exists for each n. Now

$$(\mathbf{I}c)_n = \sum_{k=0}^{n-1} \sum_{i=-k}^{k} \sum_{j=-\infty}^{\infty} a_j b_{i-j}$$

can be rewritten as $(\mathbf{I}c)_n = \sum_{j=-\infty}^{\infty} a_j H_n(j)$ where $H_n(j) := \sum_{k=0}^{n-1} \sum_{i=-k}^{k} b_{i-j}$, since a finite sum of convergent series is convergent. Since $(\mathbf{I}a)_{j-1} - 2(\mathbf{I}a)_j + (\mathbf{I}a)_{j+1} = a_{-j} + a_j = 2a_j$, we have

$$\begin{aligned}
(\mathbf{I}c)_n = \sum_{j=-\infty}^{\infty} a_j H_n(j) &= \frac{1}{2} \sum_{j=-\infty}^{\infty} \big((\mathbf{I}a)_{j-1} - 2(\mathbf{I}a)_j + (\mathbf{I}a)_{j+1}\big) H_n(j) && (18)\\
&= \frac{1}{2} \left(\sum_{j=-\infty}^{\infty} (\mathbf{I}a)_{j-1} H_n(j) - 2 \sum_{j=-\infty}^{\infty} (\mathbf{I}a)_j \, H_n(j) + \sum_{j=-\infty}^{\infty} (\mathbf{I}a)_{j+1} H_n(j) \right) && (19)\\
&= \frac{1}{2} \sum_{j=-\infty}^{\infty} (\mathbf{I}a)_j \big(H_n(j+1) - 2H_n(j) + H_n(j-1)\big), && (20)
\end{aligned}$$

Step (19) is justified since each of the sums is convergent, because each can be written as a finite sum of series of the form $\sum_{j=-\infty}^{\infty} (\mathbf{I}a)_j \, b_{m-j}$ for some m, and this is just $((\mathbf{I}a) \star b)_m$ which exists by assumption. Now

$$\begin{aligned}
& H_n(j+1) - 2H_n(j) + H_n(j-1) = \big(H_n(j-1) - H_n(j)\big) - \big(H_n(j) - H_n(j+1)\big) && (21)\\
&= \sum_{k=0}^{n-1} \left(\Big(\sum_{i=-k-j+1}^{k-j+1} b_i - \sum_{i=-k-j}^{k-j} b_i \Big) - \Big(\sum_{i=-k-j}^{k-j} b_i - \sum_{i=-k-j-1}^{k-j-1} b_i \Big) \right) && (22)\\
&= \sum_{k=0}^{n-1} \Big(\big(b_{k-j+1} - b_{-k-j}\big) - \big(b_{k-j} - b_{-k-j-1}\big) \Big) && (23)\\
&= (b_{n-j} - b_{-j}) - (b_{-j} - b_{-n-j}) = b_{n-j} + b_{-n-j} - 2b_{-j}. && (24)
\end{aligned}$$

The result then follows by substitution into (20), using the existence of $(\mathbf{I}a) \star b$ to justify splitting the sum, and finally by the symmetry of $(\mathbf{I}a)$ and b. □

References

[1] Taqqu, M. (2002). Fractional Brownian Motion and Long-Range Dependence. In P. Doukhan, G. Oppenheim, and M. S. Taqqu (Eds.), *Theory and Applications of Long-Range Dependence*, pp. 6–38. Birkhäuser.

[2] Hosking, J. (1981). Fractional differencing. *Biometrika 68*, 165–176.

[3] Granger, C. and R. Joyeux (1980). An introduction to long-memory times series models and fractional differencing. *J. Time Series Analysis 1*, 15–29.

[4] Ilow, J. (2000). Forecasting network traffic using FARIMA models with heavy tailed innovations. In *ICASSP '00: Proceedings of the Acoustics, Speech, and Signal Processing, 2000. on IEEE International Conference*, Washington, DC, USA, pp. 3814–3817. IEEE Computer Society.

[5] Barbe, P. and W. McCormick (2010, June). An extension of a logarithmic form of Cramér's ruin theorem to some FARIMA and related processes. *Stochastic Processes and their Applications 120*(6), 801–828.

[6] Taqqu, M., V. Teverovsky, and W. Willinger (1995). Estimators for long-range dependence: an empirical study. *Fractals 3*(4), 785–798. Reprinted in *Fractal Geometry and Analysis*, C.J.G. Evertsz, H-O Peitgen and R.F. Voss, editors. World Scientific Publishing Co., Singapore, 1996.

[7] Taqqu, M. and V. Teverovsky (1997). Robustness of whittle-type estimators for time series with long-range dependence. *Stochastic Models 13*, 323–357.

[8] Abry, P., P. Flandrin, M. S. Taqqu, and D. Veitch (2003). Self-similarity and long-range dependence through the wavelet lens. In P. Doukhan, G. Oppenheim, and M. Taqqu (Eds.), *Theory and Applications of Long-Range Dependence*, pp. 527–556. Birkhäuser.

[9] Robinson (1994). Semiparametric analysis of long-memory time series. *Annals of Statistics 22*, 515–539.

[10] Gefferth, A., D. Veitch, I. Maricza, S. Molnár, and I. Ruzsa (2003, June). The Nature of Discrete Second-Order Self-Similarity. *Advances in Applied Probability 35*(2), 395–416.

[11] Gorst-Rasmussen, A., D. Veitch, and A. Gefferth (2012, Mar. 29). Why FARIMA Models are Brittle. *arXiv:1203.6140v1 [stat.AP].*

[12] Samorodnitsky, G. and M. Taqqu (1994). *Stable Non-Gaussian Random Processes*. Chapman and Hall.

[13] Brockwell, P. and R. Davis (1991). *Time Series: Theory and Methods* (second ed.). Springer.

[14] Bloomfield, P. (1973). An exponential model for the spectrum of a scalar time series. *Biometrika 60*(2), 217–226.

[15] Beran, J. (1993). Fitting long memory processes by generalized regression. *Biometrika 80*, 817–822.

[16] Bingham, N., C. Goldie, and J. Teugels (1987). *Regular Variation*. Cambridge University Press, Cambridge England.

[17] Beran, J. (1994). *Statistics for Long-Memory Processes*. Chapman and Hall, New York.

[18] Abry, P., D. Veitch, and P. Flandrin (1998, May). Long-range dependence: revisiting aggregation with wavelets. *Journal of Time Series Analysis (Bernoulli Society) 19*(3), 253–266.

[19] Giraitis, L., P. Robinson, and A. Samarov (1997). Rate optimal semiparametric estimation of the memory parameter of the Gaussian time series with long range dependence. *J. Time Series Analysis 18*, 49–61.

[20] Champeney, D. (1990). *A handbook of Fourier Theorems*. Cambridge University Press.

[21] Zygmund, A. (2002). *Trigonometric Series, Vol 1* (third ed.). Cambridge University Press.