STATISTICAL METHODOLOGY FOR NONPERIODIC CYCLES: FROM THE COVARIANCE TO R/S ANALYSIS

BY BENOIT B. MANDELBROT[1]

This paper begins with a critical and idiosyncratic survey of some of the less known and more dangerous pitfalls of the common statistical methods of time series analysis—namely of the methods using the correlation. It then promotes a new and promising alternative, R/S analysis, and describes a good family of random processes, the fractional noises.

1. INTRODUCTION

This paper is a comparative analysis of a number of statistical techniques for the study of economic fluctuation, of old techniques which I have found—to various degrees—to be lacking, and of a new technique called R/S analysis of which I am enamored today. I had thought of titling this paper "New Methods of Statistical Economics," but unfortunately I used that title in 1963 [10]. Even though the old "new" methods have taken root and spread around,[2] I don't want to confuse matters with the new "new."

Though I followed up [10] and [11] with more work in the same vein, [14], [21]. I have also instigated a different development, in a sense one perpendicular to the older one. The present paper is meant to be a proselytizing introduction to it. The methods on which I was working circa 1960 had concerned primarily the *marginal distribution* of various economic time series, *irrespective of their structure of dependence*. Notably, I concentrated on the fact that among them, especially price series, many are non-Gaussian to the extreme.[3] For this behavior—borrowing

[1] Thanks are due, in chronological order, to Harold A. Thomas of Harvard for showing me Hurst's papers that made me aware of R/S; to James R. Wallis of IBM for stimulating discussions during our study of R/S by computer simulation, and for permission to include a few examples from our papers; to the National Bureau of Economic Research for support of further simulation work on R/S; to Murad Taqqu of Columbia University for assistance in the NBER supported work, for permission to quote his unpublished theorems, and for stimulating discussions; to Hirsh Lewitan of IBM for programming assistance; and to Christopher Sims of the University of Minnesota and the NBER for a most helpful reading of a draft of this paper.

[2] I view my 1963 proselytizing as having been successful in the quantity, and very successful in the quality, of the work it triggered. I was especially fortunate in having Eugene Fama join me, first to help interpret my infinite variance theses to economists [4], then to add his own empirical evidence to mine [5], to work out applications [6] and to train students [2], [28], [31]. Additional notable work on the same lines is to be found in [29], [30]. Compared to Paul Cootner's statement in 1964 that "there can be little doubt that Mandelbrot's theses are [I blush] the most revolutionary development in the theory of speculative prices since Bachelier's initial work (1900)," the casual treatment of the same theses by Richard Roll [31], not as hypothetical but as natural and nearly common-sensical, marks a definite change of attitude.

[3] A little repetition of known facts at this point may help this paper also serve as a tutorial. The main characteristic of the well known "Galton ogive" distribution is that it has a round head and no tails to speak of. By contrast, the daily changes of the logarithm of, say, the spot price of cotton have a distribution with a pointy head, and very long tails, features which express—respectively—that by Gaussian standards very small price changes are much too numerous, and large price changes much too large. Many authors believe that very large price changes should be handled apart, but I don't see why it should be so, and I therefore made it a point to examine the observed distributions as wholes. Stared at intensely, they turn out to look less like the Gaussian than like another classical distribution, Cauchy's, whose density is $(1 + x^2)^{-1}\pi^{-1}$. More precisely, they look like a cross between the Cauchy and the Gaussian. There happens to exist in the literature a whole family of such "hybrids," depending on a parameter α that varies from 1 (Cauchy) to 2 (Gauss). They are ordinarily called stable, and sometimes—in a search for a less overworked term—stable Lévy, and I confess having added to the confusion by calling them stable Paretian or Pareto–Lévy. They turn out to fit price change distributions well.

from the Bible and from hydrology—I have since coined the self explanatory term "Noah Effect" [23].[4] The methods I explore currently concern *the structure of dependence* in various economic time series, especially long run dependence, *irrespective of their marginal distribution*. Notably, economic time series tend to be characterized by the presence of clear-cut but not periodic "cycles" of all conceivable "periods," short, medium, and long, where the latter means "comparable to the length of the total available sample," and where the distinction between "long cycles" and "trends" is very fuzzy. For this sort of behavior I have coined the term "Joseph Effect" [23].

Although cycles have been studied extensively, I see in the literature a glaring gap. While considerable work was invested both in accumulating data and in investigating econometric models that generate "cyclic-looking" artificial time series, efforts to characterize the structure of actual series have been minimal; they have mostly consisted in stressing that few are either strict or hidden periodic, and that even when a clear seasonal is present, removal of the seasonal tends to leave a cyclic remainder. The first step in dealing with such "cyclic behavior" mathematically was taken several decades ago, and has consisted in observing that something "roughly like it" is encountered in oscillatory autoregressive processes. The initial pioneering observation to this effect was naturally based upon intuitive and casual tests, and this was admissible. But for a whole branch of econometrics those same tests provide inacceptably flimsy foundations. Pioneering remarks, due among others to Adelman [1] and Granger [7], have not been followed up. Though non-periodic cyclic behavior is both important and peculiar enough to be viewed as a distinct "phenomenon," the available mathematics (both probability and statistics) had not studied it squarely.

The above remarks set the framework of my current search for the following:

(A) Ways of grasping intuitively the concept of non-periodic "cyclic" long run dependence, contrasting it with the two customary patterns, namely short dependence (Markov character) and periodic variation. The differences between the above kinds of dependence are quite as deep as those in physics between—respectively—liquids, gases, and crystals. Contributing to this comparison, I have participated in extensive computer simulations of a variety of different processes, classical or otherwise. "Eyeball" tests of goodness of fit—as long as their realm of applicability is kept in mind—are peerless.

(B) Alternative methods of statistical testing and estimation that stress such long run dependence, including methods insensitive to non-Gaussian margins. The methods I advocate are range analysis and, more important, R/S analysis.

(C) Simple one variable stochastic processes that look like the data and exhibit the same typical R/S statistical behavior. The main tools I have used are various "fractional noises," which are stationary processes such that their span of dependence is infinite. (Their role is the counterpart of that of the stable Paretian processes I injected into economics around 1960.)

The present paper primarily surveys and elaborates the findings on those points scattered in my earlier papers. I do have some ideas about the causes of the effects to be described, and about their interrelationships (see, e.g. [13], [18]), but

[4] The equivalent of a flood may be of either of two kinds: a price change, either up or down, that some may ride and others will consider catastrophic.

my overall approach below, as it was circa 1960 when I was advancing the infinite variance thesis (and in various investigations outside of economics) is characterized as "phenomenological." The approach plays the same role as phenomenological thermodynamics in the theory of heat (but physics will not be mentioned again). Rather than to leave the data and specific models face to face, I have endeavored to establish "buffers" by identifying "phenomena" that embody something essential in the data and also accept simple mathematical expressions.

The quickest way to make a new viewpoint be appreciated is unfortunately "negative": to show how it affects the use of old techniques. This is why this paper grew to also include much critical exposition of the work of others, I shall list, in turn, some limits to the use of autocorrelation function (ACF) analysis, and of variance time function (VTF) analysis. The former is familiar in economics—the phrase "pitfalls of correlation" is a cliche and perhaps a bore—, while acceptance of the latter is still only a threat. Even in "the best" case, both are known to exhibit a long line of pitfalls—more accute and better known in the ACF case—, and it will be shown that the prevalence of the Noah and Joseph Effect in Nature indicates that both the marginal distributions and the laws of dependence of many actual cases are very far from being "the best" for an application of ACF and VTF. One must beware of the pitfalls of apparent lack of correlation. Then I shall proceed to range (R) analysis, which is a variant—a significantly modified one—of which may be termed "high minus low" (HLF) analysis. Finally, I shall reach range over standard deviation (R/S) analysis. I hope and trust that many readers—especially those who have been sheltered from VTF and highs minus lows—will become convinced by my criticism before they are exhausted by hearing it stated fully, will become impatient with "negative" sections and will skip on to the "positive" sections concerning fractional noises and their simulation, and R/S analysis. Those sections have been written to accommodate such readers.[5]

2. Notation and Definition of Fractional Gaussian Noise

$X(t)$ will designate either a time series (t.s.), i.e. an empirical record as function of time t, or a random function (r.f.) of t. Random variables will be denoted as r.v. Time is assumed integer-valued. As in [15], a star will denote summation, with $X*(t) = \sum_{-\infty}^{t} X(s) - \sum_{-\infty}^{0} X(s)$. Thus,

$$X*(0) = 0 \qquad \text{for } t = 0,$$

$$X*(t) = \sum_{s=1}^{t} X(s) \qquad \text{for } t \geq 1, \text{ and also } (X^2)*(t) = \sum_{s=1}^{t} X^2(s)$$

$$X*(t) = \sum_{s=t+1}^{0} X(s) \qquad \text{for } t \leq -1.$$

[5] Note that two other introductions to R/S already exist in the literature. One is informal—as is the present one—but addressed to the very different problem of water engineering, and written in the vocabulary of hydrology. See Mandelbrot and Wallis [23], [24], [25], [26], [27]. The other is formal and written ex-cathedra, with minimal motivation and extensive mathematical discussion, Mandelbrot [20].

Similarly, two stars will denote repeated summation, as in

$$C**(t) = \sum_{\sigma=1}^{t} \sum_{s=1}^{\sigma} C(s).$$

Abbreviations such as r.f., t.s., ACF will serve both as singular and as plural.

In this paper, a special role will be played by r.f. with independent values, Markov processes, and "discrete fractional Gaussian noises" (dfGn). The latter are the counterpart of the Paretian stable process encountered in studies of infinite variance. The dfGn of exponent H is the special stationary Gaussian r.f. denoted $F_H(t)$, defined as having a population ACF equal to[6]

$$C_H(d) = (1/2)[|d + 1|^{2H} - 2|d|^{2H} + |d - 1|^{2H}].$$

For all values of H, one has $C_H(0) = 1$, as required of a correlation. For $|d| \geq 1$, the cases to consider number three. $C_H(d) = 0$ if $H = 0.5$—in which case the values of $F_{0.5}(t)$ are independent, so it is a white noise. On the other hand, $C_H(d) > 0$ if $H > 0.5$, and $C_H(d) < 0$ if $H < 0.5$. By every one of several criteria to be studied below, we shall see that $F_H(t)$ exhibits very long run dependence whose intensity is measured by $H - 0.5$, where the exponent H lies between 0 and 1. Tables of $C_H(d)$ and of approximations thereto are found in Mandelbrot [17]. To avoid dragging $|d|$ throughout, we shall write the formula for $d > 0$ only.

The reasons why dfGn is needed will become increasingly apparent; for example, Section 7 will illustrate and describe intuitively the behavior of its sample functions. However, it may be good to give without waiting one reason why in [12] I had added this special family to the model maker's kit. Suppose a probabilist is told that a t.s. of price changes looks cyclic but nonperiodic, and is asked for examples of r.f. having such a behavior. His first choice might well be a random walk or Brownian motion (see Feller's Volume 1, page 86 of the third edition). But of course, the economist knows this would not do because what looks in a first approximation like Brownian motion, is the t.s. of price itself, not of its changes. This same approximation says that price changes are like white noise, which is of course the rough model we want to go beyond. Thus, in loose intuitive terms, what the model maker should strive for is a hybrid between white noise itself and its integral, some kind of partial integral or "fractional integral of white noise".

By unlikely chance, a concept bearing the name in quote marks does happen to have been defined in pure mathematics, by Abel, Holmgren, Riemann and Liouville, their definition being later modified by Herman Weyl. What I have claimed in [12] is that—by an even less likely coincidence—the process yielded by fractional integration of white noise is precisely what is needed for modeling nonperiodic cycles. For Weyl's formula of fractional integration, the reader is referred to [22] or [24]. Note that the order of integration is $H - 0.5$. One can show that a relation familiar in ordinary integration—namely that the integral of order J_1 of an integral of order J_2 of $X(t)$ is an integral of order $J_1 + J_2$ of $X(t)$—extends to fractional orders. This property is needed, when applying Weyl's definition to random functions, to avoid some conceptual complications which would have

[6] Mnemonic device: this is the second finite difference of the function $(1/2)|d|^{2H}$.

marred a direct application. Thus, a good implementation of the fractional integral of a white Gaussian noise is achieved by first taking an integral of order $H + 0.5$ and then taking the sequence of first differences of the result. This yields the dfGn of exponent H as defined above. The fact that for $H > 0.5$ the correlation is positive for all d expresses that—just like classical integration—fractional integration of positive order smoothes out a process and makes it persistent. Fractional integration of negative order—like classical differentiation—roughens it up. As a result, the case $H < 0.5$ is by far the less important one. Its properties will be mentioned only when they both require only a small effort and are illustrative.

3. Two New Pitfalls of Serial Correlation

A t.s. $X(t)$ was given from $t = 1$ to $t = T$, and its sample average $T^{-1}X*(T)$, its sample variance $T^{-1}(X^2)*(T)$, and its sample covariance function (ACF) (or serial correlation, or lagged correlation) $(T-d)^{-1}\sum_{t=1}^{t=T-d} X(t)X(t+d)$ have all been evaluated. General question: What have we learned? Answer: When either the Noah or the Joseph Effect prevails, very little.[7] More specific question: Does it suffice to match these quantities in a model and in reality? Answer: No. Even more specific question: According to Box, Jenkins and Watts [3], it suffices that the above characteristics of the data be matched in a model that is a mixed autoregressive-moving average Gaussian r.f. Is this advice sensible? Answer: Only within sharp limits: ACF analysis is effective primarily in the search for models of those aspects of statistical dependence that refer to near Gaussian r.f. and that concern high frequencies $1/d$ and small lags d, that is, for models of *short run near Gaussian effects*. For example, it is recommended to economists wishing to forecast next year's value of a t.s. known to be Gaussian—as it was to hydrologists wishing to forecast next year's discharge in a Gaussian river. Obviously, Gaussian short run problems are important, but they are not the only ones. Let us examine some arguments which make it likely that, as correlation methods have been applied in the past (for different reasons in different instances), both the short and long run dependence have often been underestimated.

3.1. *Interpretation of the Sample ACF when the Generating Process is Highly Non-Gaussian, Namely Noah Erratic*

Many economic t.s. exhibit the Noah Effect, that is, have extremely long tailed (= leptokurtic) distributions. This is the feature I proposed circa 1960 to model by a class of r.f. characterized by finite mean $EX(t)$ and infinite variance $EX^2(t)$, including notably the Pareto–Lévy (= stable Paretian) r.f. Let me show that *when an ACF analysis program is used blindly for such t.s., the degree of dependence is grossly underrated.* This novel effect is the inverse of another (better known) unfortunate aspect of correlation, that it tends to "uncover" non-existing relations. The difficulty to be described has been—implicitly—widely known, and it could, of course, be avoided by using "common sense," e.g. by eliminating outliers or by transforming the variables before taking the correlation so as to make them near

[7] There would be no improvement if the denominator in the definition of the sample covariance were T^{-1} instead of $(T-d)^{-1}$.

Gaussian, but I think to elaborate on the precise nature of this underrating is illuminating.

As a preliminary, consider the case—especially familiar to computer-oriented scientists—where a sample of 100 unit variance Gaussian r.v. has been "contaminated" through keypunch error with one value of about 100. This has the following effects. The sample variance around the expectation jumps *up* from about 100/100 to about (99 + 10,000)/100. The sample average jumps *up* from about $\pm 10/100$ to about $(100 \pm 10)/100$. The sample variance around the sample average jumps *up* from about 1 to about 100. The sample covariance of lag one, $\sum [X(t)X(t+1)]/99$, jumps *up* from about $\pm\sqrt{99}/99$ to about $(\pm\sqrt{98} \pm 100\sqrt{2})/99$. However, the ACF of lag $d = 1$ changes in the opposite direction, namely jumps *down* from about ± 0.1 to about ± 0.015. ACF for larger lags d also decrease sharply. Many practitioners are familiar with this effect and recognize it as a symptom of keypunch error.

Next, consider T values of an independent r.f. exhibiting the kind of more moderately non-Gaussian distribution characteristic of the Noah Effect, for example a t.s. of the changes of commodity and security prices. In the r.f. with finite mean and infinite variance which I have proposed as model, one has $EX(t)X(t+d) = 0$. More specifically, if $X(t)$ is approximately stable Paretian (= Pareto–Lévy) of exponent α, with $1 < \alpha < 2$, then the orders of magnitude of $\sum X(t)X(t+d)$ and $\sum X^2(t)$ are, respectively, $T^{1/\alpha}$ and $T^{2/\alpha}$ and so the order of magnitude of a correlation of lag $d > 0$ is $T^{-1/\alpha}$. The Gaussian would predict convergence to zero proportionately to $T^{-1/2}$, and by this standard a convergence proportionate to $T^{-1/\alpha}$ should be considered as abnormally rapid. In other words, if statistical correlation analysis is judged by comparison with confidence levels valid for Gaussian r.f., it should not only be expected to call our t.s. independent, but should indicate an absurdly high level of significance. To avoid such anomalies, it would be necessary to establish for each broad family of non-Gaussian r.f. a special set of alternative standards of statistical significance: an obviously unrealistic requirement.

Now pass on to a highly non-Gaussian r.f. with some positive serial dependence. Unless unreasonably severe caution has been exerted, it is tempting to compare its sample ACF with the textbook confidence level, forgetting that the latter is only appropriate for Gaussian r.f. Even though it falls *above* the confidence level that it would have been appropriate to consider, it is quite possible that the ACF of a sample from a non-Gaussian r.f. should fall *below* the textbook confidence level. As a result, the textbook test will affirm that the dependence is *not* significant!

3.2. *Interpretation of the Sample ACF when the Generating Process Exhibits a Very Strong Long Run Dependence, Namely is "Joseph Erratic"*

Let us now consider t.s. that exhibits the peculiar "Joseph" forms of long run dependence. This last notion will be explored in a later section, but we can anticipate by noting that I have proposed [12] to model it by a class of r.f. characterized by an infinite memory. This last feature may be present independently of whether or not the Noah Effect is observed, but we shall for the moment divide the difficulties by assuming the latter Effect is absent, and by examining a Gaussian r.f. such that sample ACF analysis underrates the degree of dependence. It is the dfG noise $F_H(t)$ defined above.

It is a familiar fact that, like any other statistical procedure, model fitting must start with the choice of a criterion that attributes weights to different kinds of error. Unavoidably, this choice of a distance is largely arbitrary, and—depending both upon one's aims and assumptions—different definitions may be recommended. A variety of distances will be considered. Since Gaussian r.f. are fully determined by their ACF, a measure of distance for them only involves the respective ACF, and we shall assume this restriction remains applicable to the near Gaussian phenomena with which we deal at the present.

The most naive distance involves only the mean and the variance. This amounts to neglecting dependence altogether, calling all r.f. independent. In these *Annals*, this viewpoint deserves no further discussion, so we shall assume henceforth that mean and variance are matched by being put equal, respectively, to 0 and 1.

The second most naive distances between two r.f. are increasing functions of the absolute difference between the respective ACF corresponding to the lag $d = 1$. From this viewpoint, every r.f. is Markov, and possibly even independent. For example, define $M_H(t)$ as the Markov Gauss (MG) r.f. of ACF equal to $d \geq 0$ to $[C_H(1)]^d = (2^{2H-1} - 1)^d$. Since the ACF of $F_H(t)$ and $M_H(t)$ have been constructed so as to coincide for both $d = 0$ and $|d| = 1$, the naive conclusion could be that F_H and M_H are indistinguishable.

The notion that every non-independent r.f. is Markovian is surprisingly widespread; but it is, of course, absurd. This shows that the above second most naive distance is inadequate, and that a realistic definition of a distance must also consider the values of ACF for $|d| > 1$. For example, one can take account of these values by taking as distance the maximum difference between two ACF's, as Box and Jenkins [3] do implicitly when they trace "confidence levels."

Although the domain of validity of this last distance is non-negligible, it is limited. For example, compare numerically the above r.f. $M_H(t)$ and $F_H(t)$. Despite the fact that the analytical expressions of their respective ACF $C_H(d)$ and $[C_H(1)]^d$ differ greatly for $d > 1$, the actual numerical differences between the two r.f., for given d, happen to be small. For example, for $H = 0.6$, the difference lies below 0.08. Therefore, the Box Jenkins distance would lead us to conclude that in a first approximation $C_H(d)$ and $[C_H(1)]^d$ differ little, bringing us back to the naive notion that $F_H(t)$ is nearly Markovian. But when a better approximation is desired, one will invoke a higher autoregressive r.f., and what one will find is that the order of this r.f. will behave strangely: it will be rather sharply sample dependent, and it will tend to increase without bound as the sample size increases. This suggests that the lack of difference between $F_H(t)$ and such a process is perhaps an illusion, and the absolute difference of ACF is perhaps a bad distance. This hunch will be confirmed in Section 5, in which we discuss VTF analysis.

4. Concepts of Short (Finite) and Long (Infinite) C-dependence

The practical difficulty we encountered in trying to apply ACF to the dfGn has deep mathematical roots. An ACF is indeed an infinite sequence of numbers, and the definition of a distance for infinite sequences is an order of magnitude harder and more indeterminate than for finite sequence.[8] Some occurrences of mathematical infinity can in practice be dismissed as relevant only to events that will not happen before Doomsday, but the present role of infinity is different. Roughly, a sequence of events can be thought as effectively finite if successive events are qualitatively alike, and conversely as effectively infinite if successive occurrences are qualitatively very different. A first example had been encountered in my Noah Effect studies circa 1960, a second example in the kind of cyclic behavior with which we are now concerned, with ever new kinds of cycles of ever longer "period" appearing as the sample increases. The presence of cycles thus constitutes, in my opinion, prima facie evidence that at least some among economic r.f., though not periodic, have effectively an infinite span. The main alternative description I know assumes economic behavior to be wholly non-stationary. If such were really the case, the possibility of a rational description in economics would be negated. This being the alternative, the rational descriptive economist should consider the assumption of infinite span as optimistic.

4.1. Perhaps the simplest distinction involving infinity explicitly is the dichotomy between convergent and divergent series. An early example, again, occurs in my work on infinite variance. As a second example, consider the covariance of a Gaussian process. Each individual value of $C(d)$ is finite, but consider the series

$$S'(0) = (1/2) \sum_{d=-\infty}^{\infty} C(d) = \left[C(0)/2 + \sum_{d=1}^{\infty} C(d) \right] = C(0)/2 + C*(\infty).$$

It occurs in spectral analysis—whence the notation $S'(0)$ (see below)—and also in the study of physical fluctuations when it is often called Taylor's Eulerian scale. Clearly, the series may either converge, or diverge in the sense that its partial sum from $d = -D$ to $d = D$ tends to infinity with D, or be indefinite in the sense that the partial series has no limit. In addition, series that converge to 0 must be singled out.

For $M_H(t)$, with $C(d) = [C_H(1)]^d$, $S'(0)$ satisfies $0 < S'(0) < \infty$

For $F_H(t)$, with $C(d) = C_H(d)$, $\begin{cases} S'(0) = \infty & \text{when } H > 0.5 \\ S'(0) = 0 & \text{when } H < 0.5 \end{cases}$

For periodic r.f. with fixed period and random amplitude and phase, $C(d)$ is a sine function and $S'(0)$ is indefinite.

These possibilities suggest a tetrachotomy we shall first state, then discuss, then modify and improve.

4.2. *The C-tetrachotomy; Statistical C-dependence*

A tetrachotomy is of course a classification into four categories. In the present instance, they are defined as follows

→ Finite C-dependence = short run C-dependence = *short C-dependence* = *vanishing long C-dependence*. Defined by $0 < S'(0) < \infty$.

[8] Very different comments in the same spirit can be found in papers by Christopher Sims [32], [33].

→ Positive infinite C-dependence = positive long run C-dependence = *positive long C-dependence.* Defined by $S'(0) = \infty$.

→ Negative infinite C-dependence = negative long run C-dependence = *negative long C-dependence.* Defined by $S'(0) = 0$.

→ When the series that defines $S'(0)$ is indefinite, the concept of C-dependence is *inapplicable.*

4.3. *Remark*

The only sound reason for starting with the C-tetrachotomy is that it requires so little preliminary. It has many flaws, some of them real. For example, the process of successive differences of an independent Gaussian r.f. is called negative long dependent. Soon we shall investigate other alternative tetrachotomics—which is why the above definition has included the index C-. The first two, Γ- and R-tetrachotomies, are nearly identical to the C-tetrachotomy, but better, and often the three are indistinguishable. On the other hand, as soon as $EX^2 = \infty$ (strong Noah Effect), all three become meaningless. A final different tetrachotomy, involving R/S dependence (to be introduced in Section 10) will on the contrary apply irrespectively of whether $EX^2 = \infty$ or $EX^2 < \infty$. This, in my opinion, will be a considerable practical and theoretical asset.

The issue of how to define dependence may be further illustrated by analogy with the history of the concept of I.Q. Binet and the Stanford psychologists who followed had only some vague and intuitive ideas of what they wanted to measure and of how to measure it. Before an operational procedure implementing these ideas was selected, many doubtful issues had to be settled more or less arbitrarily. As a result, the claim that the Binet–Stanford Intelligence Quotient "really measures the intelligence" came rapidly to be questioned, and different I.Q.'s came to be considered, each of them measuring a different "kind" of intelligence. All told, "intelligence is what is measured by some I.Q.," where the original I.Q. measures the "Binet–Stanford intelligence." The indeterminacy of the concept of statistical dependence is, luckily, less extreme.

4.4. *Comments on the Spectral Analytic Origin of the Basic Tetrachotomy. The Typical Shape of Economic Spectra*

The above notation $S'(0)$ was selected because when the spectral density of the r.f. is well defined, it is equal to $S'(f) = \sum_{-\infty}^{\infty} C(d)\, e^{-2\pi i f d}$, and so $S'(0)$ is its value for the frequency $f = 0$.[9]

Now let a sample spectral density (s.s.d.) be obtained for a sample of duration T. It will vary little from $f = 0$ up to at least $f = 1/T$, independently of the form of $S'(f)$, and so the three non-trivial sides of the C-tetrachotomy manifest themselves as follows:

For Markov and finite autoregressive processes, and more generally whenever $0 < S'(0) < \infty$, there exists a well determined and intrinsic time scale T^* such

[9] Formula $S'(f)$ implies a decision about notation, because a different definition of the spectral density $S'(f)$ corresponds to each of the several accepted ways of choosing the unit of frequency. I prefer to measure frequency in cycles per unit of time. Those who measure frequency in radians per unit of time are accustomed to writing e^{-ift} (or $e^{-i\omega t}$) instead of $e^{-2\pi i f t}$.

that the population $S'(f)$ is constant and positive for $0 < f < 1/T^*$. Hence, s.s.d. is flat from $f = 0$ up to an upper limit which is $1/T^*$ for large samples ($T \gg T^*$) and $1/T$ for small samples, defined as having a size T satisfying $T \ll T^*$. This upper unit is denoted as "break-over frequency". What we have established is that it possesses a lower bound equal to $1/T^* > 0$.

When $S'(0) = \infty$, on the contrary, T^* is infinite. In other words, s.s.d. *always* is flat between $f = 0$ and $f = 1/T$. That is, as T increases, the break-over frequency tends to 0, and the s.s.d. at $f = 0$ increases without bound. This behavior has been observed by Adelman [1] and Granger [7] in many economic time series, and Granger called it "typical" of economics.

When $S'(0) = 0$, s.s.d. dips down near $f = 0$. This behavior has also been observed. When an economic t.s. exhibits Granger's typical form, then before analysis it is often "prewhitened"—differentiated—to erase the spectral peak at $f = 0$. Such processing often overshoots its aim and replaces the peak by a dip.

Interpretation of sample spectra at low frequencies is known to practitioners as tricky. But the above examples do demonstrate that in the C-tetrachotomy, each of the first three terms has a possible practical application. The fourth term represents r.f. for which $S(f)$ is non-differentiable, for example, exhibits jumps corresponding to pure periodic components. These are beyond our concern in this paper.

4.5. *A Digression on Spectral Analysis and Synthesis*

Spectral *analysis* is not the concern of this paper, but its current popularity implies it may be useful to digress in order to stress the importance of its relations with spectral *synthesis*. Spectral analysis relies on the Euler–Fourier theorem, which says that any well behaved t.s. can be decomposed into a sum of periodic harmonic components. In addition, the original contexts of optics and acoustics include an important converse: each of those periodic components has an independent physical reality, and so the original light and noise could be "synthesized" from meaningful building blocks. The original applications of Fourier methods to economics was similarly spurred by the hope of also discovering periodicities, which might be initially hidden but would turn out after the fact to be real. However, it has turned out that the typical economic t.s. does not exhibit any such periodicity. (Seasonals do not count, because they are hardly hidden to begin with.) As a result, economic t.s. cannot be spectral synthesized, and their spectral analysis is purely a formal technique lacking concrete backing. A taste for it—if it proves worthwhile—would have to be acquired.

By way of contrast, R and R/S analysis—to be discussed below—have an intuitive grounding in "high minus low" analysis of economics, a taste for which seems fairly spontaneous.

5. The Variance Time Function

The population variance time function $V(d)$ of a stationary r.f. $X(t)$, taken around the population expectation, is defined as the second moment of the sum of

d successive values of $X(t)$, namely—assuming $EX = 0$—as

$$
\begin{aligned}
V(d) &= E[X*(t+d) - X*(t)]^2 = \operatorname{Var} X*(d) \\
&= E[X*(d)]^2 = E[X(1) + \cdots + X(d)]^2 \\
&= \sum_{1 \le t', t'' \le t} E[X(t')X(t'')] = dC(0) + 2 \sum_{\sigma=1}^{d-1} \sum_{s=1}^{\sigma} C(s) \\
&= dC(0) + 2C**(d-1).
\end{aligned}
$$

The equality between the first and last lines is known to some engineers as "the formula of G. I. Taylor." Note that the variance of the average of d successive values of a process equals $V(d)/d^2$. The variance $V*(d)$ of the difference between the respective averages of the d past and the d future values of $X(t)$ is $V*(d) = [4V(d) - V(2d)]/d^2$; it measures the error in estimating a future average from the average of a past record of identical length.

In terms of the asymptotic behavior of $V(d)$ for $d \to \infty$, the first three cases of the C-tetrachotomy have the following effects:

In cases of short C-dependence, $V(d)$ is asymptotically proportional to d. The simplest is $V(d) = d$, as encountered for independent reduced Gaussian (IG) r.f. In the MG case—the Markov Gauss process $M_H(t)$ whose ACF is $[C_H(1)]^d$—the function $V(d)$ nearly equals

$$d\{1 + 2C_H(1)[1 - C_H(1)]^{-1}\}.$$

In cases of positive long C-dependence, $V(d)$ grows more rapidly than d.

In cases of negative long C-dependence, $V(d)$ grows more slowly than d.

The simplest behavior $V(d)$ can take in either of the cases of long dependence is $V(d) = d^{2H}$, which in the case of discrete time Gaussian r.f. turns out to mean that $X(t)$ is a dfGn. One has $H > 0$ because $V(d)$ cannot decrease as $d \to \infty$.

5.1. *Comparison Between the V(d) of IG and MG*

The exponents of d in the corresponding $V(d)$ are identical, but the multiplying factors in front are different. As $C_H(1) \to 0$, the factor of MG tends to 1 and MG tends to IG. When $C_H(1)$ is small, MG and IG are near each other from the two viewpoints of the nearness to each other of their respective ACF and VTF.

5.2. *Comparison Between the V(d) of IG and dfGn*

The exponents of d are different. Hence, despite the fact that in a sense (as we have seen) the ACF are close to each other, the processes are very far from identical. In other words: even *though every C(d), when considered singly for large d, is negligible, their accumulation may be very significant.*

5.3. *Digression: The Self Similarity of Fractional Noise*

The simplicity of the relation $V(d) = d^{2H}$ for dfGn had constituted the original motivation for introducing dfGn and also—since $2H$ is a fraction with $H \neq 0.5$—

one of the many equivalent motivations for the term "fractional noise." A consequence of $V(d) = d^{2H}$ is that $X*(t)$ is "statistically self-similar." To define this concept, consider two sample sizes T' and T'' and form the rescaled r.f. $T'^{-H}X*(hT') = X*(hT')/\sqrt{V(T')}$ and $T''^{-H}X*(hT'')$. When h is a multiple of both $1/T'$ and $1/T''$, both rescaled r.f. are defined, and it is easy to see that their distributions are identical. When $T' < T''$, this identity expresses that a portion of the $(1, T'')$ sample is obtained from the whole by geometric similarity, hence the term "statistically self-similar." The complication that h must be a multiple of both $1/T'$ and $1/T''$ is conceptually unimportant; it can be avoided by considering $X(t)$ as the r.f. of the discrete time increments of a r.f. $X*(t)$ defined in continuous time, but we shall not dwell on this matter.

6. Concepts of Short (Finite) and Long (Infinite) Γ-Dependence

6.1. *Correlation Between Long Past and Future Averages*

A second tetrachotomy for r.f. is based upon the behavior of the function $\Gamma(d', d)$, equal to the correlation between the past and future averages $X*(d')/d'$ and $X*(-d)/d$—that is, between $X*(d')$ and $X*(-d)$ themselves. Γ is given by

$$\Gamma(d', d) = \frac{\operatorname{Var} X*(d'+d) - \operatorname{Var} X*(d') - \operatorname{Var} X*(d)}{2[\operatorname{Var} X*(d')]^{1/2}[\operatorname{Var} X*(d)]^{1/2}}$$

$$= \frac{V(d+d') - V(d) - V(d')}{2[V(d)]^{1/2}[V(d')]^{1/2}}.$$

Adding the assumptions that $d' \gg 1$ and $d \gg 1$, we obtain the approximation

$$\Gamma(d', d) \sim \frac{C**(d'+d) - C**(d') - C**(d)}{[d'C(0) + 2C**(d')]^{1/2}[dC(0) + 2C**(d)]^{1/2}}.$$

It is a part of folklore among users of probability theory that, as $d \to \infty$ and $d' \to \infty$, long past and future averages necessarily tend to become independent. However, such is not necessarily the case, as shown by one example: when $X(t)$ is dfGn with $H \neq 0.5$, $\lim_{d\to\infty} \Gamma(dh, d)$ exists, and is nonzero; it has the same sign as $H - 0.5$. To restate this result, we shall need a new tetrachotomy. The term "Γ-dependence" will stand for "correlation-dependence," i.e., non-orthogonality. But when the r.f. is Gaussian, it is well known that non-orthogonality is a synonym for statistical dependence.

Notation: For the sake of simplicity, $\Gamma(dh, d)$ will designate $\Gamma(d', d)$ when d' is the integer closest to dh.

6.2. *The Γ-tetrachotomy; Statistical Γ-dependence*

→ Finite Γ-dependence = short run Γ-dependence = *short Γ-dependence* = *vanishing long Γ-dependence*. Defined by $\lim_{d\to\infty} \Gamma(dh, d)$ exists and = 0 for all $h > 0$.

→ Positive infinite Γ-dependence = positive long run Γ-dependence = *positive long Γ-dependence*. Suppose that for all $h > 0$, $\lim_{d\to\infty} \Gamma(dh, h)$ exists and is >0. This defines $X(t)$ as exhibiting positive long Γ-dependence, of a specified intensity equal to the above limit. (For a generalization, see a digression below.)

→ Negative infinite Γ-dependence = negative long run Γ-dependence = *negative long Γ-dependence*. Suppose that for all $h > 0$, $\lim_{d\to\infty} \Gamma(dh, h)$ exists and is < 0. This defines X as exhibiting negative long Γ-dependence, of a specified intensity equal to the above limit. (For a generalization, see a digression below.)

→ In all other cases, the concept of Γ-dependence of specified intensity is inapplicable.

6.3. *Examples: Among Processes for which Successive Averages are Asymptotically, for Large d, Non-independent the simplest are the dfGn*

They have a specified intensity of long Γ-dependence, which is, respectively, > 0, $= 0$, and < 0 when $H > 0.5$, $H = 0$ and $H < 0.5$. As a matter of fact, when h is an integer, $\Gamma(dh, d)$ is independent of d, which is one expression of the self similarity of $X*(t)$. In particular, $\Gamma(d, d) \equiv \Gamma(1, 1) = C_H(1) = 2^{2H-1} - 1$.

6.4. *The Relations Between C and Γ-dependence. Generalization of the Above Examples*

To a large extent—and up to comparatively unimportant exceptions, extra factors and corrections—nearly all short Γ-dependent processes are also short C-dependent, and the only processes that exhibit long Γ-dependence of specified non-vanishing intensity are those sharing the same covariance with dfGn.[10]

Digression: Proof. For the existence of $\lim_{d\to\infty} \Gamma(dh, d)$ for all h, one necessary condition is that the limit exist for $h = 1$. Writing

$$\Gamma(d, d) = [\operatorname{Var} X*(2d)/2 \operatorname{Var} X*(d)] - 1,$$

this necessary condition becomes that $\operatorname{Var} X*(2d)/\operatorname{Var} X*(d)$ must have for $d \to \infty$ a limit to be designated as $\varphi(2)$. Next, set $h = 2$ and rewrite Γ as

$$\Gamma(2d, d) = \frac{\dfrac{\operatorname{Var} X*(3d)}{\operatorname{Var} X*(d)} - \dfrac{\operatorname{Var} X*(2d)}{\operatorname{Var} X*(d)} - 1}{2\left[\dfrac{\operatorname{Var} X*(2d)}{\operatorname{Var} X*(d)}\right]^{1/2}}.$$

When $\operatorname{Var} X*(2d)/\operatorname{Var} X*(d)$ and $\Gamma(2d, d)$ have limits, $\operatorname{Var} X*(3d)/\operatorname{Var} X*(d)$ must tend to a limit. By induction, a necessary condition for the existence of a specified intensity of dependence is that $\lim_{d\to\infty} \operatorname{Var} X*(hd)/\operatorname{Var} X*(d) = \varphi(h)$ should exist for every integer h, and therefore also for every rational h. Obviously, $\varphi(h)$ satisfies for all rational h' and h'' the equation $\varphi(h'h'') = \varphi(h')\varphi(h'')$ which is a

[10] The former condition is enormously less specific than the latter. One is reminded of the conditions required for the central limit theorems with—respectively—a classical Gaussian and a non-classical stable Paretian limit. The former requires that $EX^2 < \infty$, which is not specific at all; the latter requires that for large x, the probability $\Pr(X > x)$ behave near identically for the addends and for the desired limit—again up to comparatively unimportant extra factors—which is a highly specific requirement. This complication is unavoidable.

form of Cauchy's functional equation. When in addition $\varphi(h)$ is assumed continuous (which might or might not be a new independent assumption), the above equation must hold for all real h, and the necessary condition above becomes that $\lim_{d\to\infty} \text{Var } X*(hd)/\text{Var } X*(d) = \varphi(h) = h^{2H}$ with H a constant. Plugging back, $\Gamma(d, d)$ becomes $2^{2H-1} - 1$; this is a correlation only if $H < 1$ which is a second necessary condition. Finally, asymptotic mean square continuity of X requires that $H > 0$, which is a third necessary condition. The combined necessary condition, namely: $\varphi(h) = h^{2H}$ with H a constant between 0 and 1, is obviously also sufficient. H is called the exponent of Γ-dependence.

Now we turn to the three specific sub cases $H = 0.5$, $H > 0.5$ and $H < 0.5$. Short Γ-dependence is equivalent to $H = 0.5$. In particular, short C-dependence implies short Γ-dependence, because when $\lim_{d\to\infty} C*(d)$ is defined, finite and positive, it follows that $\lim_{d\to\infty} C**(d)/d$ is defined, finite and positive, $\lim_{d\to\infty} C**(hd)/C**(d) = h$, and $\lim_{d\to\infty} \Gamma = 0$. The converse is unfortunately not quite true: For example, if $C(d) = 1/d$, then $X(t)$ is long C-dependent but $C**(d) \sim d \log d$ so that $C**(dh)/C**(d) \to 1$, meaning X is short Γ-dependent. Thus, short Γ-dependence is slightly less demanding than short C-dependence.

When $H > 0.5$, it has been shown [35] that $C(d)$ must be of the form $C(d) = d^{2H-2}L(d)$, where $L(d)$ is slowly varying for large d, meaning $L(hd)/L(d) \to 1$. In the dfGn case, $L(d) \to 1$.

When $H < 0.5$, one must have in addition $2^{-1}C(0) + \Sigma_{d=1}^{\infty} C(d) = 0$.

6.5. *Digression: Generalization of Long Γ-dependence to the Case of Unspecified Intensity*

Categories 2 and 3 of the Γ-tetrachotomy can be widened, and category 4 correspondingly narrowed, by allowing long Γ-dependence to be present without having a specified intensity. Suppose that

$$0 \leq \lim\inf_{d\to\infty} \Gamma(dh, h) \leq \lim\sup_{d\to\infty} \Gamma(dh, h)$$

holds for all $h > 0$, with the second inequality replaced by $<$ for "sufficiently many" h (see below). If so, the intensity of dependence is unspecified but dependence is either non-negative (first inequality being ≤ 0), or positive (first inequality being < 0). The obvious parallel definitions hold when

$$\lim\inf_{d\to\infty} \Gamma(dh, h) \leq \lim\sup_{d\to\infty} \Gamma(dh, h) \leq 0.$$

Though the notion of "sufficiently many" above is not yet explored fully, examples where lim inf < lim sup for all h, meaning that the intensity of dependence is unspecified, have indeed been constructed.

7. Simulation of Fractional Noises

Before we continue, it will be useful to acquire an intuitive feeling for the shapes of the sample function of various short and long dependent processes. Such feeling is best obtained by examining pseudo random simulations.[11] The

[11] The following description has in part been previously used in Mandelbrot and Wallis [24], pp. 229 to 232.

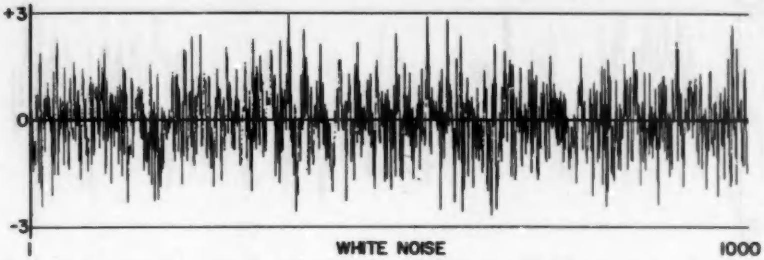


Figure 1 A sample of 1,000 values of white noise, also called sequence of independent Gauss random variables, of zero mean and unit variance, also called fractional noise with $H = 0.5$

first exhibit on Figure 1 is a sample of a process of independent Gaussian r.v. A short sample suffices, because this process is monotonous and featureless. Being analogous to the hum in electronic amplifiers, it is often called a discrete-time white noise. It can also be considered a discrete time fractional Gaussian noise with $H = 0.5$. For the definition of dfGn, see Section 2 above.

Another small sample of dfGn, with $H = 0.1$, is given as Figure 2. It is richer than white noise in high frequency terms, owing to the fact that large positive values tend to be followed by compensating large negative values, but on a graph this is not very apparent.

Friezes 1 to 5 carry successive samples, each containing 1,000 values, of a moderately nonwhite fractional noise with $H = 0.7$. Similarly, Friezes 6 to 10 carry a strongly nonwhite fractional noise, with $H = 0.9$. Whenever $H > 0.5$, a fractional noise is richer than white noise in low frequency terms. Therefore, large positive or negative values tend to persist, and the dependence between successive averages fails to die out. Even on a casual glance at the two Friezes the effect of such low frequency terms is obvious. The reader should compare these artificial series with the natural records with which he is concerned. To be meaningful, in both cases the comparison must involve the same degree of local smoothing of high frequency jitter. If the artificial series feel very different from his natural records, then fractional noises are probably inappropriate for his problems.

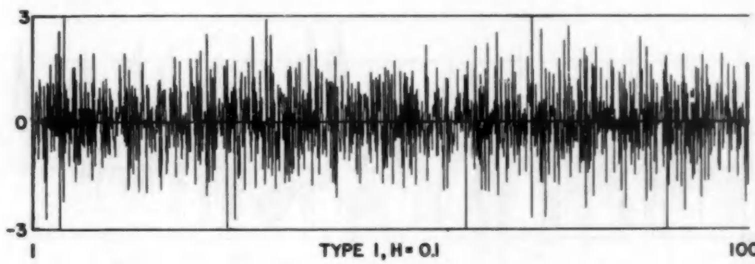
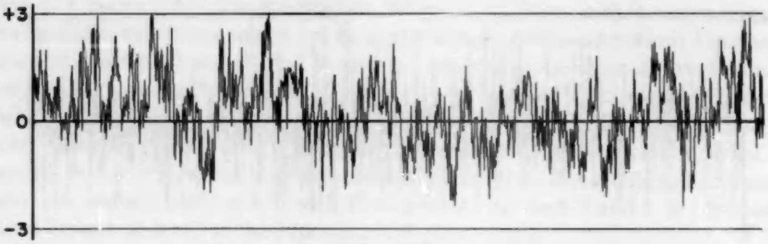


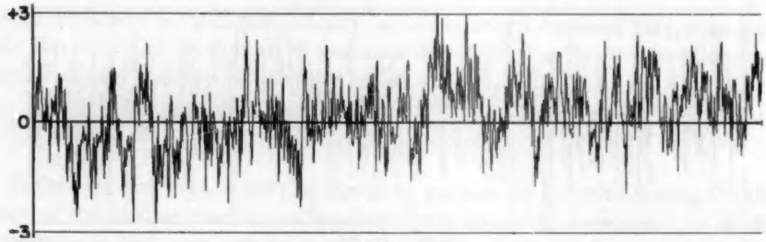
Figure 2 A sample of 1,000 values of a Type 1 fractional noise with $H = 0.1$. The sample was normalized to have zero mean and unit variance



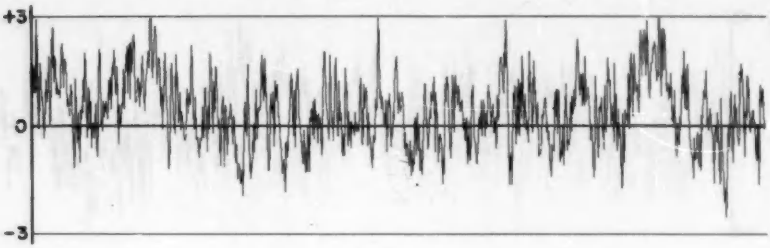
Friezes 1 to 5 A sample of 5,000 values of an approximation to fractional Gaussian noise. In this instance, $H = 0.7$, meaning the strength of long run dependence is moderate but definitely positive (continued on the following pages)

If they feel close, but not quite right, then perhaps another value of H will suffice. If they feel right, then he should proceed to further and more formal statistical tests of fit. Such formal tests are indispensable but should neither be blind nor come first. A statistical test by necessity focuses upon a specific aspect of a process, whereas the eye can often balance and integrate various aspects. Formal test and visual inspection should be combined.

A perceptually striking characteristic of fractional noises is that their sample functions exhibit an astonishing wealth of features of every kind, including trends and cyclic swings of various frequencies. In some subsamples such swings are rough and far from periodic while other subsamples seem to include absolutely periodic swings. However, the wavelength of the longest among the apparent cycles depends markedly on the total sample size. As one looks at shorter portions of these friezes, shorter cycles become visible. At the other extreme, on plots of these Friezes as strips of 3,000 time units, the impression is unavoidable that cycles of about 1,000 time units are present. Since in the generating mechanisms, there is not built-in periodic structure whatsoever, such cycles must be considered spurious. For example, spectral analysis denies the apparent periodic appearance of fractional noise. On the other hand, they are very real in the sense that something present in human perceptual mechanisms brings most observers to recognize the same cyclic behavior. This makes such cycles useful in describing the past. But they have no



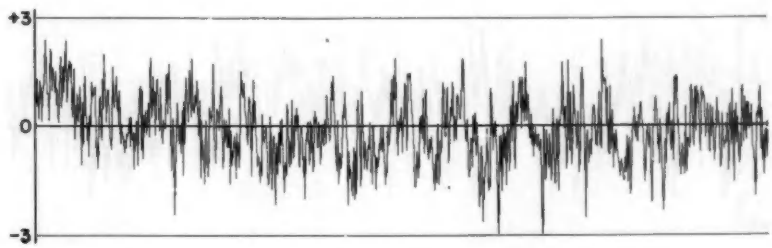
Friezes 6 to 10 A sample of 5,000 values of an approximation to fractional Gaussian noise. In this instance, $H = 0.9$, meaning the strength of long run dependence is high and positive (continued on the following pages)



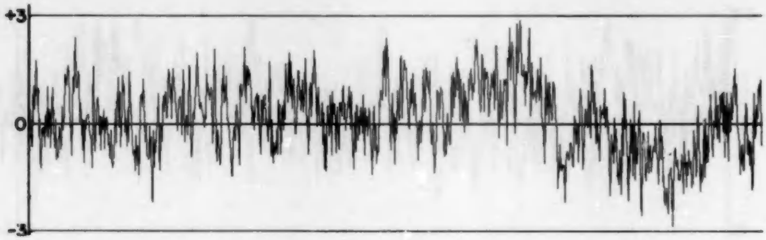
Frieze 2

predictive value for the future. Remarks to the same effect have been made by Keynes [9]. As a rough rule of thumb, unless the sample size is very short, the longest cycles in a record corresponding to $H = 0.7$ have a wavelength equal to one-third of the sample size. The specific value one-third of this rule is only a matter of psychology of perception, but the fact that the ratio of apparent wavelength to sample size is a constant is an aspect of the self-similarity of fractional noise. This ratio depends on H.

A second and even more striking characteristic of fractional noise with $H > 0.5$ is that some periods above or below the theoretical mean, which equals 0 by construction, are extraordinarily long. In fact, portions of these figures are reminiscent of the seven fat and seven lean years in the Biblical story of Joseph son of Jacob. One is tempted to express this perceptual persistence of fractional noise with the help of the ideas of trend and of run. A run of low price changes would be a period when price changes stay below the line; a high run as a period when they stay above the line. However, a careful inspection of samples of fractional noise shows many instances where this concept of run describes its behavior very poorly. Often, one is tempted to call a period a high run although it is interrupted by a very short low run. Should we be pedantic and consider such a sequence as being three runs? Or is it really a single run? Perhaps short runs could be eliminated by a little smoothing? Such a chase after a reasonable definition of runs had to be abandoned because it was found hopeless. Different smoothing procedures (moving averages of various lengths) and definitions of high and low (different crossover levels) were tried. The distribution of the duration of runs was found to depend very



Frieze 7

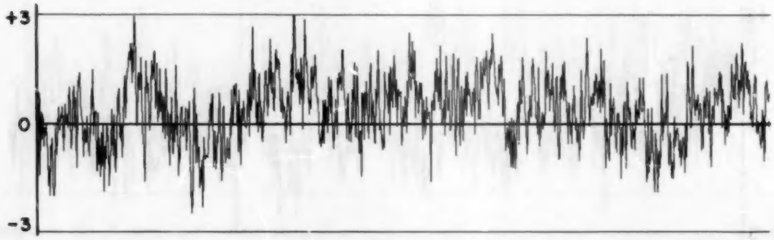


Frieze 3

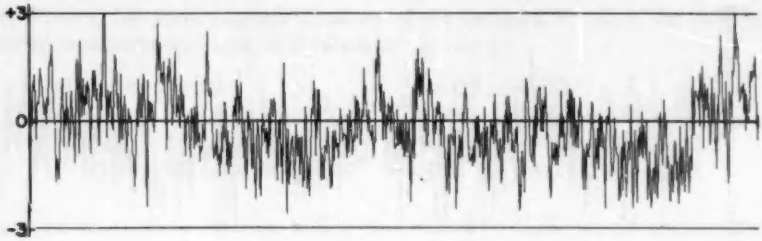
much on otherwise insignificant features of the model, whereas large differences in the relative proportions of low and high frequencies were obscured. Thus, again, we see that runs cannot be a good statistic (which may, incidentally, explain why earlier applications of this statistic had lead to conflicting results).

To complete the comparison between process with zero and infinite C- and Γ-dependence, one should have exhibited some plots of Markov process and other r.f. with finite but non-zero dependence. However, a verbal commentary will suffice, because over any finite time space the behavior of a fractional noise can be mimicked beyond possible detection with the help of an appropriate finitely dependent process. As a matter of fact, that is how I actually perform simulations; see [17]. However, let both a fractional noise and a finite memory mimic be extrapolated to a long time span T, and let their graphs be placed side by side before our eyes to be compared. For that, some compaction by averaging is necessary: for example, each t.s. may be replaced by a sequence of 1,000 values each averaged over a time span of $T/1{,}000$. Then we shall see that, as T varies, the compacted graph of dfGn will exhibit ever new patterns, while the patterns of the Markov mimic will eventually begin to repeat themselves. The art of simulating fractional noises relies on constantly injecting appropriate new low frequency terms in order to prevent such repetition. (Needless to say, this last remark applies only to simulations; the origin of low frequency terms that may be present in actual t.s. is practically unknown, and may be very different.)

We may digress a moment to discuss the scope of various notions of stationarity. The intuitive and mathematical notions agree fully for all comparatively well behaved statistical fluctuations, for example those to which ACF



Frieze 8



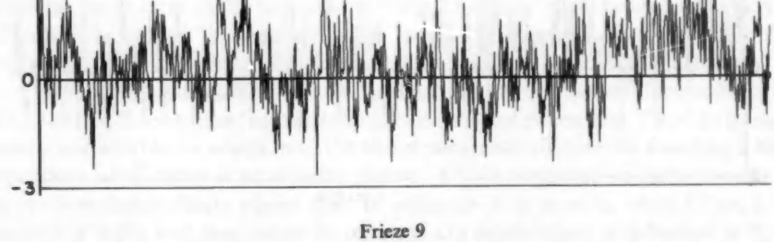
Frieze 4

analysis applies without hitch, and also for fluctuations that are grossly non-stationary—for example, random walks. But in other cases, the two definitions disagree. For example, dfGn is mathematically stationary, but intuitively many view it as non-stationary. It is my belief that when a process is not stationary in the usual intuitive sense, when it is not reducible to such stationarity by differencing, and when no generalization of such stationarity is applicable to it, then a mathematical study of it is impractical. The ambition of [12], in singling out fractional noises and showing their broad practical applicability has been to propose such a generalization.[12]

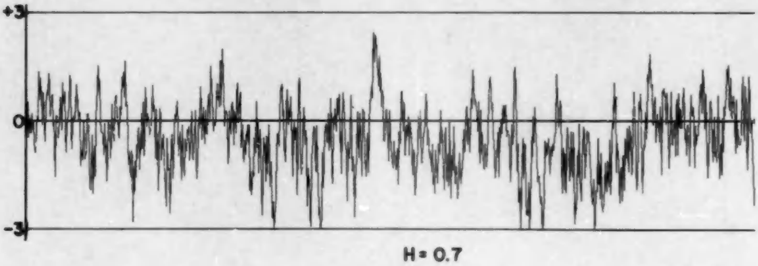
8. Variance Time Analysis

In this paper, fractional noise was first introduced merely to exemplify how ACF analysis can fail; then its VTF behavior was shown to differ from that of white noise, and finally it was shown to be useful. Now we return to statistics, to try to do better than ACF analysis. The fact that the classification of possible behaviors of $V(d)$ is very close to our basic tetrachotomy of finite variance r.f. suggests using $V(d)$ in statistical testing and estimation. The basic idea of VTF analysis is that in order to compare two finite variance r.f. $X'(t)$ and $X''(t)$, it suffices to compare the corresponding VTF $V'(d)$ and $V''(d)$. When V' and V'' are "alike," so are $X'(t)$ and $X''(t)$.

[12] A later achievement was to introduce the concept of generalized conditional stationarity, through which additional r.f. that are not stationary are made manageable, at least in part. But this issue need not be pursued here.



Frieze 9



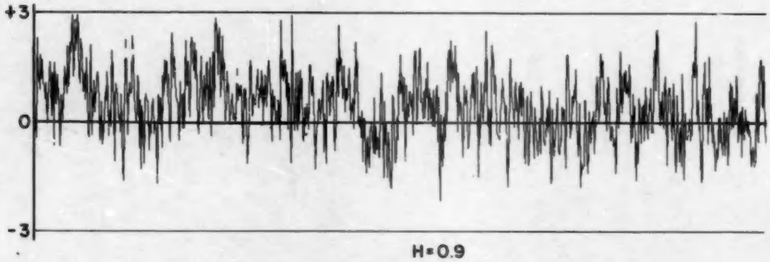
Frieze 5

The "conventional" applications of VTF analysis limit themselves to families of r.f. all of whose members are short C-dependent, so that the purpose of analysis is to estimate the value of $S'(0)$. The present application is different: r.f. with long C-dependence are allowed, and the purpose becomes to identify the asymptotic shape of the VTF, in order to classify $X(t)$ among the alternatives of the basic tetrachotomy.

To ascertain that attempting such classifications makes sense, a first task is to proceed beyond analytic formulas and asymptotics, to numerical values and finite samples. Four tables of typical numerical values follow. Here—as earlier in the paper—MG is the first order Markov r.f. fitted to $C_H(1)$ and IG is the independent Gauss r.f. fitted to $C_H(0) = 1$.

$H = 0.6; d = 50$		$H = 0.6; d = 100$	
dfGn:	$V(d) \sim 100$	dfGn:	$V(d) \sim 250$
MG:	$V(d) \sim 80$	MG:	$V(d) \sim 160$
IG:	$V(d) \sim 50$	IG:	$V(d) \sim 100$
$H = 0.7: d = 50$		$H = 0.7; d = 100$	
dfGn:	$V(d) \sim 240$	dfGn:	$V(d) \sim 630$
MG:	$V(d) \sim 97$	MG:	$V(d) \sim 194$
IG:	$V(d) \sim 50$	IG:	$V(d) \sim 100.$

The comparison of the three values in each table makes it obvious that the respective r.f. are very different. For $H = 0.6$, MG lies halfway between IG and dfGn. For $H = 0.7$, MG lies nearer to IG.



Frieze 10

We next list some numerical values of the variance $V*(d)$ of the difference between successive averages of d values of various r.f.

$H = 0.6; d = 50$		$H = 0.7; d = 50$	
dfGn:	$V*(d) = 250/2{,}500$	dfGn:	$V*(d) = 330/2{,}500$
MG:	$V*(d) = 160/2{,}500$	MG:	$V*(d) = 200/2{,}500$
IG:	$V*(d) = 100/2{,}500$	IG:	$V*(d) = 100/2{,}500.$

These values confirm that the MG process fitted to $C_H(1)$ would underestimate the long run variability grossly.

8.1. *The Estimation Bias of VTF in the Case of r.f. Having Long Dependence and Finite Variance*

The definition of $V(d)$ involves EX. When this moment is indeed known, and the origin of X is so selected that $EX = 0$, it seems reasonable to estimate $V(d)$ by

$$V_1(d, 1, T) = \frac{1}{T-d} \sum_{t=1}^{T-d} [X*(t+d) - X*(t)]^2.$$

When, however, EX is unknown, its value is estimated efficiently by $T^{-1}X*(T)$ so it seems reasonable to estimate $V(d)$ by

$$V_2(d, 1, T) = \frac{1}{T-d} \sum_{t=1}^{T-d} [X*(t+d) - X*(t) - (d/T)X*(T)]^2.$$

The above V_1 satisfies $EV_1 = V(d)$, so it is an unbiased estimator. When $X(t)$ is short dependent, V_2 has a bias, but it is small. But when $X(t)$ is long dependent, the bias is large and depends on both T and d. The question of whether or not a biased procedure can be useful has been discussed at length in more usual contests of statistics. After a period during which statisticians had sought to avoid bias, they came to accept it, as long as it decreases the variance, and as long as bias corrections are available. In the present case, however, the aim is not to estimate the specific parameters of a r.v. or r.f., but the whole ill defined "shape" of a function $V(d)$ of d. We must proceed through the simultaneous estimation of the separate values of $V(d)$ for different d, and use judgments based on visual inspection, at least as preliminaries. In this context, a bias that depends only on d would be correctable, but when the bias depends on both d and T in inextricable combination, one cannot correct it by eye. In order to correct it by algorithm one needs the exact form of the bias as a numerical function of two variables, which is difficult to store in a computer. If other methods happen to exist, for which the bias is larger but easier to handle, they may be preferable even if less efficient.

The preceding discussion may serve as advance advertisement for the statistics $R(t, d)$ and R/S, for which the bias is large but does not depend on T and so is more readily correctable. In addition to the above pragmatic reasons for avoiding a bias dependent on T, there is an esthetic reason: $V(d)$ is supposed to take account of those dependence effects whose span of influence is at most d, while $EV_2(d, 1, T)$ involves $X*(\Gamma)$, and thus mixes in other effects whose span of influence is T.

Further, observe that many statistics, such as sample moments, covariances and Fourier coefficients, can be constructed in two steps. The first step is independent of the sample size and consists in constructing from the original $X(t)$ a transformed function $\Phi[X(t)]$, such as $X^k(t)$, $X(t)X(t+s)$, $X(t)\sin(2\pi ft)$. The second step consists in taking the average of the transformed function over the available sample. Statistics of this "transform-then-average" form have many advantages. For example, they may be computed sequentially: as the sample size is increased from T' to T'', one increases the number of sample values of $\Phi[X(t)]$ to be evaluated and averaged, but the values of $\Phi[X(t)]$ already computed from the initial sample of size T' need not be recomputed. The sample correlations and V_2 are examples of statistics that are not obtained by transforming then averaging. But the sample averages of $R(t, d)$, $S(t, d)$ and $R(t, d)/S(t, d)$—to be defined below—are computable sequentially.

8.2. *Inapplicability of VTF Analysis when $EX^2 = \infty$. Alternative Statistics*

When $EX^2 = \infty$ and $EV(d) = \infty$, one encounters a difficulty very different from a bias and more severe. We noted that, on the population level, the basic tetrachotomy becomes meaningless. On the sample level, it follows that $EV_2(d, 1, T) = \infty$ and hence the sample values of $V_2(d, 1, T)$ are wildly scattered.

This phenomenon is well known to statisticians in the context of sample averages of such r.v. as Cauchy's, and it is recommended when averages misbehave to replace them by medians. For example, an easy alternative to V, one having acceptable fluctuations, would be obtained by considering

$$V_{\text{med}}(d, 1, T) = \text{median of } [X*(t+d) - X*(t) - (d/T)X*(T)]^2,$$

an expression whose expectation is finite.

Nevertheless, V_{median} is a poor measure of dependence, because, in addition to being biased, it is extremely non-robust. When the r.f. $X(t)$—with $EX^2 = \infty$—are independent, the function that relates EV_{med} to d already depends upon the distribution of the $X(t)$. This implies that when dependence is added, the effects on EV_{med} of the marginal distribution and of the law of statistical dependence of $X(t)$ are mixed up inextricably. Such mixture is not only undesirable, as was the case with biases, but unacceptable. In summary, empirical values of VTF—e.g., those of Young [36]—are almost impossible to interpret. In the section after next, we shall see how the statistic R/S—which avoids biases depending on T—also avoids mixing the effects of margins with the effects of dependence.

9. Range Analysis

In the case of a price t.s., to be viewed as a $X*(t)$, economists are very familiar with a statistic we shall call "high minus low function," which is defined as

$$\text{HLF} = (\max - \min)_{1 \le u \le d} [X*(t+u) - X*(t)].$$

Here as below we shall use the notation

$$\max_{1 \le u \le d} Z(u) - \min_{1 \le u \le d} Z(u) = (\max - \min)_{1 \le u \le d} Z(u).$$

The HLF statistic is mostly useful when the process of differences $X(t) = X*(t) - X*(t-1)$ is stationary. Also, since the value $X(t) = 0$ of X plays a central role here, HLF is useable in assessing the degree of statistical dependence in $X(t)$ only when $EX(t)$ is known, with $EX(t) = 0$. In that case, one takes the average of HLF over all values of t between 1 and $T - d$, and follows its variation with d. There is no bias.

When EX is unkown, it is tempting to modify the definition of HLF by analogy to the definition of the sample VTF, by considering the expression

$$R*(t, d, 1, T) = (\max - \min)_{1 \le u \le d} [X*(t+u) - X*(t) - (u/T)X*(T)],$$

and its average

$$R_A^*(d, 1, T) = (T - d)^{-1} \sum_{t=1}^{t=T-d} R*(t, d, 1, T).$$

However—for reasons that will transpire shortly—I think it better to base range analysis on a different algorithm, a variant of the HLF and of R^* (Figure 3). Evaluate

$$R(t, d) = (\max - \min)_{1 \le u \le d} \{X*(t+u) - X*(t) - (u/d)[X*(t+d) - X*(t)]\},$$

and form the estimator

$$R_A(d, 1, T) = (T - d)^{-1} \sum_{t=1}^{t=T-d} R(t, d).$$

Observe that for each subsample $(t+1, t+u)$ we use a different estimator of EX. As a result, the functional form of the bias is improved. While the bias of HLF, like the bias of VTF, depended on both d and T, the bias of R_A depends only upon d.[13] Price to pay for the above advantage: there is a loss of efficiency in the estimation of EX, and the value of the bias for fixed T is increased.

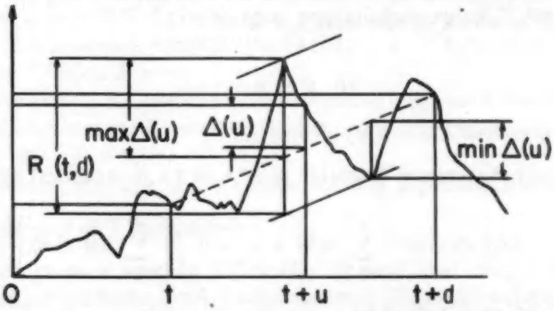


Figure 3 Construction of the sample range $R(t, s)$ (reproduced from [29]). Since empirical records are necessarily taken in discrete time, the function $X(u)$ should have been drawn as a series of points, but it was drawn as a line for the sake of clarity. As marked, $\Delta(u) = X*(t+u) - X*(t) - (u/s)[X*(+s) - X*(t)]$, and $R(t, s) = \max_{0 \le u \le s} \Delta(u) - \min_{0 \le u \le s} \Delta(u)$. In the original application in hydrology [8], $X(t)$ stands for the discharge of a river during year t. In this case, save for corrections due to discrete time, $R(t, d)$ is the volume of a water reservoir that regularizes optimally the flow of water between time $t + 1$ and $t + d$, that is, of a reservoir that produces a uniform outflow, ends at time $t + d$ exactly as full as it started at time t, never overflows and never dries up. The use of $R(t, d)$ in reservoir design originated with Rippl in 1885.

[13] One might have tried, likewise, to apply to VTF the same transformation that leads from HLF to $R(t, d)$. However, this option is closed because it would yield values identically equal to 0.

The definition of $R(t, d)$, contrary to the definition of VTF, does not involve second moments, and as a result, $ER(t, d) < \infty$ only requires $EX < \infty$, while $EV < \infty$ required $EX^2 < \infty$. Nevertheless, as was the case for VTF, $R(t, d)$ must be examined separately for the two cases of r.f. with a finite variance, especially Gaussian, and of sharply non-Gaussian r.f.

Gaussian examples. In the Gaussian cases, and especially for IG, MG and dfGn, the behavior of $R(t, d)$ with varying d can be shown to be as follows:

IG and MG and most other cases of short dependence: $R(t, d)$ is asymptotically proportional to $\sqrt{d}$ and hence to $\sqrt{V(d)}$.

Most other cases of positive long C-dependence: $R(t, d)/\sqrt{d} \to \infty$.

Most other cases of negative long C-dependence: $R(t, d)/\sqrt{d} \to 0$.

dfGn: $R(t, d)$ is asymptotically proportional to d^H and, again, to $\sqrt{V(d)}$.

The behavior of $R(t, d)$ lends itself easily to defining a tetrachotomy of R-dependence, parallel to, but a bit different from, C- and Γ-dependence. However, we need not dwell on it because in Gaussian case, the *asymptotic effectiveness of the sample* VTF *and range is about the same.* The latter requires longer computer runs, a feature that not everybody considers an asset; also, it has the advantage that its bias is independent of T. But the precise payoffs between bias, efficiency and simplicity have not been explored; they deserve the statisticians' attention, but in this paper we have other purposes in mind.

Sharply non-Gaussian cases when $EX^2 = \infty$. Since VTF is now meaningless, the comparison can only run between R and some alternative to VTF, such as V_{median}. Again, $R(t, d)$ has the asset of having better bias properties, but both share the major defect of being extremely non-robust. Therefore, $R(t, d)$, as VTF, is to be avoided (see Figure 8).

Conclusion. R may be of possible use in preference to VTF in the Gaussian case, but it mainly deserves attention as an intermediary step in the progression towards R/S, to which we can finally proceed.

10. R/S Analysis

10.1. *Definitions and Elementary Properties*

The squared standard deviation of a t.s. or r.f. X is defined by

$$S^2(t, d) = d^{-1} \sum_{u=1}^{d} X^2(t + u) - d^{-2}\left[\sum_{u=1}^{d} X(t + u)\right]^2,$$

and the statistic Q is defined by

$$Q(t, d) = \frac{R(t, d)}{S(t, d)}.$$

This last notation is intended to emphasize that the numerator $R(t, d)$ and the denominator $S(t, d)$ are merged into a new entity.

Historically, R/S appears to have been first considered in a study of Nile River discharges due to the celebrated hydrologist Harold Edwin Hurst [8] nicknamed Abu Nil (the Father of the Nile). Irrespective of Hurst's original reasons for considering Q, this statistic has turned out to have remarkably good properties, as we shall see.[14]

The normalizing denominator S vanishes if and only if all the $X(t + u)(0 < u \leq d)$ are identical, in which case one also has $R(t, d) = 0$ so that Q takes the indeterminate form $Q = 0/0$. In particular, the indeterminate form 0/0 is always encountered when $d = 1$. Otherwise, namely whenever it is not of the form 0/0, Q is positive and finite and in fact it can be shown that it lies between two effectively attainable limits, equal, respectively, to about 1 and about $d/2$. The lower limit is exactly 1 when d is even, and is $\sqrt{d/(d - 1)}$ when d is odd; for large d, the latter is barely above 1. The upper limit is exactly $d/2$ when d is even, and is $\sqrt{(d^2 - 1)/2}$ when d is odd; for large d, the latter is barely below $d/2$. The upper and lower limits coincide when $d = 2$, in which case $Q(t, d)$ is identically 1. When d is smallish, $Q(t, d)$ only depends on the rules of short run dependence in $X(t)$, but its dependence upon such rules is limited and so in the study of short run dependence the statistic R/S is not at all effective. Quite on the contrary, for large d the dependence of $Q(t, d)$ on the rules of long run dependence in $X(t)$ is very strong and very apparent, and so in the study of long run dependence the statistic R/S is very valuable. The good properties announced as belonging to it consist in the fact that the dependence of $Q(t, d)$ upon d can be made the basis of a new and especially convenient tetrachotomy of statistical dependence. One can adopt either of several variants; it is too early to be sure which is the best, so we shall state and then discuss the simplest.

10.2. *The R/S Tetrachotomy; Statistical R/S Dependence*

→ Finite R/S dependence = short run R/S dependence = *short R/S dependence = vanishing long R/S dependence.* Defined by this: $\lim_{d\to\infty} d^{-0.5}EQ(t,d)$ exists and is positive and finite.

→ Positive infinite R/S dependence = positive long run R/S dependence = *positive long R/S dependence.* Exemplified by $\lim_{d\to\infty} d^{-0.5}EQ(t, d) = \infty$. Defined by $\limsup_{d\to\infty} d^{-0.5}EQ(t, d) = \infty$.

→ Negative infinite R/S dependence = negative long run R/S dependence = *negative long R/S dependence.* Exemplified by $\lim_{d\to\infty} d^{-0.5}EQ(t, d) = 0$. Defined by $\liminf_{d\to\infty} d^{-0.5}EQ(t, d) = 0$.

→ In all other cases, the above variant for R/S dependence is *inapplicable.*

10.3. *The Exponent of R/S Dependence*

When there exists a specific $H(0 \leq H \leq 1)$ such that $\lim_{d\to\infty} d^{-H}EQ(t, d)$ exists and is positive and finite, the R/S dependence of $X(t)$ is said to have a specified intensity, of which H is called the exponent. Thus, the exponents of short, positive long, and negative long R/S dependence, respectively, equal 0.5, lie between 0.5 and 1, and lie between 0 and 0.5.

10.4. *The Value $H = 1$ and the Application of R/S Analysis to Non-stationary r.f.*

In the rest of this paper, $X(t)$ is assumed stationary, and it could not be otherwise because the covariance and the VTF would cease to depend on d alone.

[14] Steiger [34] took, independently, one step towards R/S, but he failed to identify any of the remarkable properties to be described below.

But R/S is another matter; for example, let $X(t)$ be Brownian motion or any other explosive sum of the values of a stationary r.f. Because of the subtraction of sample means throughout, one may check that the distributions of $R(t, d)$ and $S(t, d)$ only depend upon d, just as when $X(t)$ is stationary. The corresponding behavior of $Q(t, d)$ arouses curiosity. A more pragmatic reason for studying it is because one never knows when $X(t)$ is, or is not, stationary—especially when its values are neither printed nor graphed but coded on tape. One must learn to recognize after the fact when a sample $Q(t, d)$ came from a non-stationary r.f.

As it happens, when $X(t)$ is non-stationary but $Q(t, d)$ is in distribution independent of t, then $H = 1$. As a result, when the observed H differs from 1 by a small amount that may be due solely to sample variation, this finding may mean either that $X(t)$ is stationary with a very strong long run dependence, or that $X(t)$ is non-stationary with almost any kind of dependence. To discriminate between those possibilities, one must R/S analyze both $X(t)$ and the sequence of differences of $X(t)$.

10.5. *Transients, the Middle Run, and the Effective Exponent of R/S Dependence*

The fact that the above tetrachotomy involves asymptotics is both its main point and a weakness. It is not at all excluded that for middle sized values of d, $Q(t, d)$ should behave "as if" $d^{-H}EQ(t, d)$ tended to a limit, with $H \neq 0.5$, while eventually it turns out that X is short R/S dependent after all. Such is for example the case when $X(t)$ has a finite but large memory. It may well happen that the research with which one is involved has a finite horizon $d_{\max}$, and that for all d between 5 (say) and $d_{\max}$, $d^{-H}EQ(t, d)$ is near constant. From this limited viewpoint, H is, "effectively," the exponent of R/S dependence of $X(t)$. From the really long run viewpoint, this value of H is no more than a special transient, but this does not make the effective H any less important.

10.6. *Relationships between the R/S and Earlier Tetrachotomies*

The Figures and their captions are an integral part of the discussion that follows. To motivate the construction of the Figures, note that—given a sample of T values of $X(t)$—an exhaustive output of the R/S analysis is constituted by the $(T - 2)(T - 1)/2$ values of Q corresponding to the non-trivial values $d > 2$. Thus, the Q transformation expands the number of values involved. This insures that the output is redundant, and indeed the values of $Q(t, d)$ corresponding to the neighboring values of t and d are strongly interdependent. Pending the development (eagerly awaited) of more efficient methods, I simply select for each variable an appropriately spaced subgrid. The interdependence between Q for different t depends on their difference, so I take uniformly spaced values of t, ordinarily 15 altogether. The interdependence over different d—on the contrary—depends on their ratio, so I take logarithmically spaced values of d, ordinarily 10 per decade, and in graphical methods I use for abscissa $\log d$. Also, I start with $d = 10$ because we know that for small d, $Q(t, d)$ is mostly independent of X. Finally, I plot the values of Q as a collection of T functions $\log Q(t, d)$ of the variable $\log d$, parameterized by t. Successive points plotted are linked to form broken lines.

Such plots make it possible to check whether the output functions fluctuate in a straight "street", and if they do, to estimate H as a slope. In other cases, I plot $\log Q(t, d) - 0.5 \log d$. Such plots make it easier to test the extent to which the street differs from horizontally. "Estimation" and "test" in the previous sentences were originally mostly done by eye, but I am currently proceeding to various algorithms.

First check of consistency: the Gaussian cases. It can be shown that short C-dependent Gaussian r.f. are also short R/S dependent. See Figures 4 and 5 for the case when $X(t)$ is independent and white. In this case $\lim_{d\to\infty} d^{-0.5} EQ(t, d)$ was shown by W. Feller to be about 1.25. When $X(t)$ is non-independent, $\lim_{d\to\infty} d^{-0.5} EQ(t, d) = 1.25 \lim_{d\to\infty} d^{-1} \operatorname{Var} X * (d)$.

On the other hand, when $X(t)$ is a fractional Gaussian noise, it is long R/S dependent of exponent H, and so its dependence is positive when $H > 0.5$ and negative when $H < 0.5$. Figure 9.

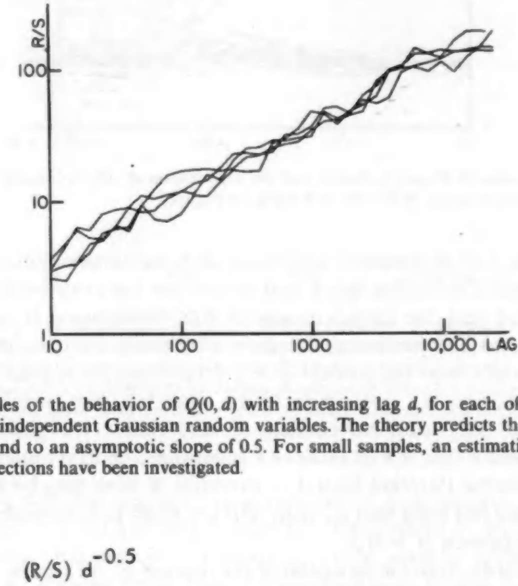


Figure 4 Examples of the behavior of $Q(0, d)$ with increasing lag d, for each of four independent samples of 30,000 independent Gaussian random variables. The theory predicts the trend line of this diagram should tend to an asymptotic slope of 0.5. For small samples, an estimation bias is present. The resulting corrections have been investigated

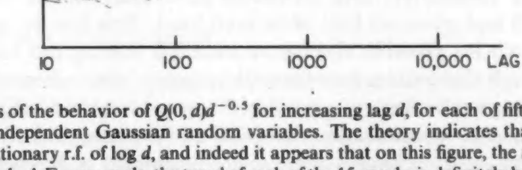


Figure 5 Examples of the behavior of $Q(0, d)d^{-0.5}$ for increasing lag d, for each of fifteen independent samples of 30,000 independent Gaussian random variables. The theory indicates that $Q(0, d)d^{-0.5}$ is asymptotically a stationary r.f. of $\log d$, and indeed it appears that on this figure, the asymptotic stage has already been reached. For example, the trend of each of the 15 graphs is definitely horizontal and the fluctuations around this trend are small

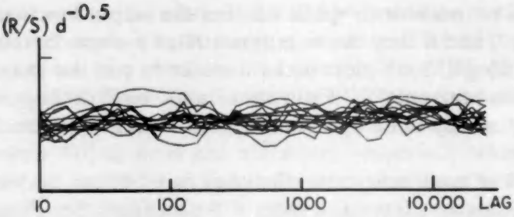


Figure 6 The same as Figure 5, except that the distribution of $X(t)$ is lognormal, $\log_{10} X(t)$ being reduced Gaussian. We see that R/S testing for independence is blind to the extremely non-Gaussian character of the lognormal process. The respective kurtosis of the various samples ranged around 1000

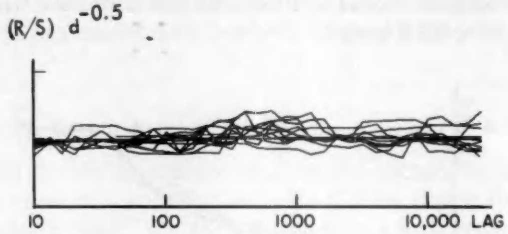


Figure 7 The same as Figure 5, except that the distribution of $X(t)$ is Cauchy. To appreciate the contrast between the behavior of R/S and of R itself, see Figure 8

Second check of consistency: non-Gaussian finite variance cases. The situation is a bit complicated in its fine detail, and has not yet been explored fully, but it has been established that the tetrachotomy of R/S dependence is roughly parallel to those of C- and Γ-dependence, themselves known to be parallel to each other.

Check of usefulness: the concept of R/S dependence for regular independent r.f. with infinite variance. We now arrive to the basic fact that justifies defining R/S dependence. Let $X(t)$ be an independent r.f. such that its marginal distribution is regular, in the sense that $X*(t)$ satisfies a generalized central limit theorem with a non-Gaussian stable Paretian limit. For example, X itself may be stable Paretian. The basic finding has been that all such $X(t)$ are short R/S dependent, that is, R/S dependent of exponent $H = 0.5$.

In other words, *from the viewpoint of the value of H, all regular independent r.f. are indistinguishable and equivalent*. R/S has the extraordinary ability of separating the long run dependence properties of X from its marginal distribution properties. See Figures 6 and 7, and contrast the latter with Figure 8.

Check of singularity: alternatives to R/S. The reasons that made Hurst introduce R/S had given no hint of its usefulness. This feature spurred me to an extensive search for possible alternative statistics sharing the above property of robustness. Such alternatives were found, but none of them seems clearly preferable and all are more cumbersome to calculate. Observe to start with that all the classical statistics, such as covariance, spectrum, VTF and the range, involve one characteristic that is very sensitive to dependence. The originality of R/S is that while it also involves such a statistic (as its numerator) it also involves (as its denominator)

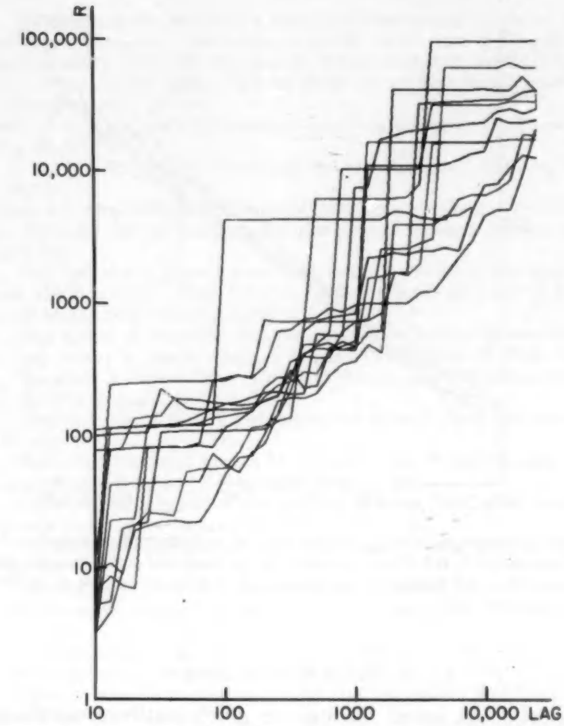


Figure 8 Examples of the behavior of $R(0, d)$ for each of a number of independent samples of 30,000 independent Cauchy r.v.'s. As predicted by the theory, the slope of the trend line is 1, and, more important, fluctuations around the trend line are enormous. The fact that the ratio $Q(0, d)d^{-0.5}$ is well behaved, as shown on Figure 7, turns by contrast to be the more remarkable. What it expresses is that the fluctuation of $S(0, d)d^{-0.5}$ and those of $R(0, d)$ are matched perfectly. The ratio R/S can be considered "self-standardized"

a second, very different statistic, one completely independent of the rule of dependence, that is, invariant with respect to permutations of the quantities $X(t)$. R/S demonstrates that such a combination can be used to achieve robustness.

Figures 4 to 9 show better than a long discussion would the nature of the behavior of $Q(t, d)$. Note that the precise value of $\lim_{d \to \infty} d^{-0.5} EQ(t, d)$ is not robust, i.e., it depends on $X(t)$, but it only varies between 1.25, applicable in the Gaussian case, and 1.

10.7. *Alternative R/S Tetrachotomies*

In the R/S tetrachotomy, as stated, all non-regular independent r.f. fall into the fourth category. Intuitively, however, they belong in the first category, and indeed one may define R/S dependence to bring them back there. However, the non-regular cases are of slight importance here, and to complicate everything just for their sake would be unwarranted.

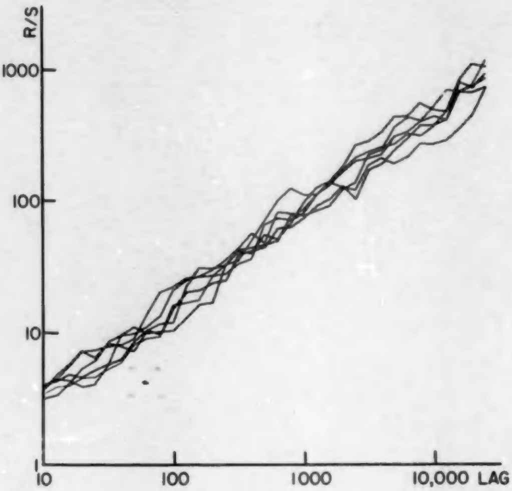


Figure 9 Examples of the behavior of $Q(0, d)$ for each of six independent samples of a fractional Gaussian noise of exponent $H = 0.7$. The theory predicts the trend line of this diagram should tend to an asymptotic slope of $H = 0.7$. Indeed, H can be estimated from the sample (a controllable small sample bias is again present)

11. A Word in Conclusion

The reader who has lasted through so much statistics should have been rewarded by at least a little economics. More specifically, (A) by a little analysis of actual data from economics, to check to what extent the non-periodic cycles exhibited by those data are indeed alike to those of fractional noise, (B) by some economic theory to explain whatever conclusion is reached in (A). I have accumulated such statistical analysis by the cord, and such economic analysis by the gross, but I can only say that I hope to have both ready for publication very soon.

IBM Thomas J. Watson Research Center, and
National Bureau of Economic Research

References

[1] Adelman, I., "Long Cycles: Fact or Artefact?" *American Economic Review*, LV (June, 1956), 444–463.

[2] Blume, M., "The Assessment of Portfolio Performance: An Application of Portfolio Theory," University of Chicago Dissertation, 1968. Revised as "Portfolio Theory: a Step Towards its Practical Application"; *Journal of Business*, XL (April, 1970), 152–173.

[3] Box, G. E. P. and G. M. Jenkins, *Time Series Analysis, Forecasting and Control*, (San Francisco: Holden Day 1970).

[4] Fama, E. F., "Mandelbrot and the Stable Paretian Hypothesis," *Journal of Business*, XXXVI (October, 1963), 420–429.

[5] ———, "The Behavior of Stock-Market Prices," *Journal of Business*, XXXVIII (January, 1965), 34–105. First issued in 1964 as a University of Chicago Dissertation entitled "The Distribution of the Daily First Differences of Stock Prices: A Test of Mandelbrot's Stable Paretian Hypothesis."

[6] ———, "Efficient Capital Markets: a Review of Theory and Empirical Work," in M. D. Intrilligator, ed., *Frontiers of Quantitative Economics* (Amsterdam and London: North Holland Publishing Company, 1971), 309–361. Reprinted from *Journal of Finance*, XXV (1970), 383–417.
[7] Granger, C. W. J., "The Typical Spectral Shape of An Economic Variable," *Econometrica*, XXXIV (January, 1966), 150–161.
[8] Hurst, H. E., R. P. Black, and Y. M. Simaika, *Long-Term Storage, an Experimental Study* (London: Constable, 1965).
[9] Keynes, J. M., "On the Method of Statistical Business Research," *Economic Journal*, L (March, 1970), 154–156.
[10] Mandelbrot, B., "New Methods in Statistical Economics," *Journal of Political Economy*, LXXI (October, 1963), 421–440, or *Bulletin de l'Institut International de Statistique*, Ottawa session, XL (1964), 699–720.
[11] ———, "The Variation of Certain Speculative prices," *Journal of Business*, XXXVI (October, 1963), 394–419. Reprinted in P. H. Cootner, ed., *The Random Character of Stock Market Prices* (Cambridge Mass.: MIT Press, 1964), 297–337.
[12] ———, "Une Classe de Processus Stochastiques Homothetiques a Soi; Application a la Loi Climatologique de H. E. Hurst," *Comptes Rendus* CCLX (March 22, 1965), 3274–3277.
[13] ———, "Forecasts of Future Prices, Unbiased Markets and 'Martingale' Models," *Journal of Business*, XXXIX (January, 1966), 242–255.
[14] ———, "The Variation of Some Other Speculative Prices," *Journal of Business*, XL (October, 1967), 393–413.
[15] ———, "Long-Run Linearity, Locally Gaussian Process, *H*-Spectra, and Infinite Variances," *International Economic Review*, X (February, 1969), 82–111.
[16] ———, "Statistical Dependence in Prices and Interest Rates," read at the Second World Congress of the Econometric Society (September, 1970).
[17] ———, "A Fast Fractional Gaussian Noise Generator," *Water Resources Research*, VII (June, 1971), 543–553.
[18] ———, "When Can Price be Arbitraged Efficiently? A Limit to the Validity of the Random Walk and Martingale Models," *Review of Economics and Statistics*, LIII (August, 1971), 225–236.
[19] ———, "Comments on 'Application of Linear Random Models to Four Annual Streamflow Series," *Water Resources Research*, VII (October, 1971), 1360–1362.
[20] ———, "Mathematical Foundations of the R/S Analysis of Long-Run Nonperiodic Dependence in Random Functions."
[21] ——— and H. M. Taylor, "On the Distribution of Stock Price Differences," *Operations Research*, XV (November–December, 1967), 1057–1062.
[22] ——— and J. W. Van Ness, "Fractional Brownian Motions, Fractional Noises and Applications," *SIAM Review*, X (October, 1968), 422–437.
[23] ——— and J. R. Wallis, "Noah, Joseph and Operational Hydrology," *Water Resources Research*, IV (October, 1968), 909–918.
[24] ——— and J. R. Wallis, "Computer Experiments with Fractional Gaussian Noises," *Water Resources Research*, V (February, 1969), 228–267.
[25] ——— and J. R. Wallis, "Some Long-Run Properties of Geophysical Records," *Water Resources Research*, V (April, 1969), 321–340.
[26] ——— and J. R. Wallis, "Robustness of the Rescaled Range and tee Measurement of Long-Run Statistical Dependence," *Water Resources Research*, V (October, 1969), 967–988.
[27] ——— and J. R. Wallis, "Operational Hydrology Using Self Similar Processes," in John Lawrence, ed., *Proceedings of the Fifth International Conference on Operations Research* (Venice, 1969) (London: Tavistock Press).
[28] Officer, R., "An Examination of the Time Series Behavior of the Market Factor of the New York Stock Exchange," University of Chicago Dissertation, 1971.
[29] Renwick, F. B., "Theory of Investment Behavior and Empirical Analysis of Stock Market Price Relatives," *Management Science*, XV (September, 1968), 57–71.
[30] ———, "Economic Growth and Distributions of Changes of Stock Market Prices," *Industrial Management Review*, IX (Spring, 1968), 39–67.
[31] Roll, R., *Behavior of Interest Rates; the Application of the Efficient Market Model to U.S. Treasury Bills* (New York: Basic Books, 1970). First issued in 1968 as a University of Chicago Dissertation, entitled "The Efficient Market Model Applied to U.S. Treasury Bill Rates."
[32] Sims, C., "Distributed Lag Estimation When the Parameter Space is Explicitly Infinite Dimensional," *Annals of Mathematical Statistics*, XLII (October, 1971), 1622–1636.
[33] ———, "The Role of Approximate Prior Restrictions in Distributed Lag Estimation," *Journal of the American Statistical Association*, LXVII (March, 1972), 169–175.

[34] Steiger, W., "A Test of Nonrandomness in Stock Price Changes," P. H. Cootner, ed., *The Random Character of Stock Market Prices* (Cambridge, Mass.: MIT Press, 1964), 253–261.
[35] Taqqu, M., "Weak Convergence to Fractional Brownian Motion." Columbia University Dissertation in Mathematical Statistics, 1972.
[36] Young, W. E., "Random walk of stock prices; a test of the variance–time function," *Econometrica*, XXXIX (September, 1971), 797–812.