Studies in Nonlinear Dynamics & Econometrics

Volume 7, Issue 3

2003

Article 5

Determinism in Financial Time Series

Michael Small*

Chi K. Tse[†]

*Hong Kong Polytechnic University

[†]Hong Kong Polytechnic University

Studies in Nonlinear Dynamics & Econometrics is produced by The Berkeley Electronic Press (bepress). <http://www.bepress.com/sn timer>

Copyright ©2003 by the authors.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or otherwise, without the prior written permission of the publisher bepress.

Determinism in Financial Time Series

Abstract

The attractive possibility that financial indices may be chaotic has been the subject of much study. In this paper we address two specific questions: “Masked by stochasticity, do financial data exhibit deterministic nonlinearity?”, and “If so, so what?”. We examine daily returns from three financial indicators: the Dow Jones Industrial Average, the London gold fixings, and the USD-JPY exchange rate. For each data set we apply surrogate data methods and nonlinearity tests to quantify determinism over a wide range of time scales (from 100 to 20,000 days). We find that all three time series are distinct from linear noise or conditional heteroskedastic models and that there therefore exists detectable deterministic nonlinearity that can potentially be exploited for prediction.

The hypothesis that deterministic chaos may underlie apparently random variation in financial indices has been received with enormous interest (Peters 1991, LeBaron 1994, Benhabib 1996, Mandelbrot 1999). However, initial studies have been somewhat inconclusive (Barnett and Serletis 2000) and there have been no reports of exploitable deterministic dynamics in financial time series (Malliaris and Stein 1999). Contradictory interpretations of the evidence are clearly evident (Barnett and Serletis 2000, Malliaris and Stein 1999).

In his excellent review of the early work in this field, LeBaron (1994) argues that any chaotic dynamics that exist in financial time series are probably insignificant compared to the stochastic component and are difficult to detect. In some respects the issue of whether there is chaos in financial time series or not is irrelevant. Some portion of the dynamics observed in financial systems is almost certainly due to stochasticity¹. Therefore, we restrict our attention to whether *in addition* there is significant nonlinear determinism exhibited by these data, and how such determinism may be exploited. Crack and Ledoit (1996) showed that certain nonlinear structures are actually ubiquitous in stock market data, but represent nothing more than discretization due to exchange imposed tick sizes. Conversely, in his careful analysis of nonlinearity in financial indices, Hsieh (1991) found evidence that stock market log-returns are not independent and identically distributed (i.i.d.), and deterministic fluctuation in volatility could be predicted. Hence, if we are to exploit nonlinear determinism in financial time series we should first be able to demonstrate that it exists. If we find that these data do exhibit nonlinear determinism then we must assess how best to exploit this.

Before reviewing significant advances within the literature we will define precisely what we mean by “determinism” and “predictable”. Let $\{x_t\}_t$ denote a scalar time series with $E(x_t) = 0$.

- A time series x_t is *predictable* if

$$E(x_t | x_{t-1}, x_{t-2}, x_{t-3}, \dots) \neq 0.$$

- A time series x_t is *deterministic* if there exists a deterministic (i.e. closed-form, non-stochastic) function f such that

$$x_t = f(x_{t-1}, x_{t-2}, x_{t-3}, \dots).$$

- A time series x_t *contains a deterministic component* if there exist a deterministic f such that

$$x_t = f(x_{t-1}, x_{t-2}, x_{t-3}, \dots, e_{t-1}, e_{t-2}, e_{t-3}, \dots) + e_t$$

where e_t are mean zero random variates (not necessarily i.i.d.).

- Determinism is said to be *non-linear* if f is a non-linear function of its arguments.

For example, we note that according to these definitions a GARCH model *contains* determinism but it is not deterministic. We do not (as Hsieh (1991) did) classify a system that is both chaotic and stochastic as stochastic. Such a system we call stochastic with a non-linear deterministic component. Our choice of nomenclature emphasizes that we are interested in whether there is significant structure in this data that cannot be adequately explained using linear stochastic methods.

In an effort to quantify determinism, predictability and nonlinearity, a variety of nonlinear measures have been applied to a vast range of economic and financial time series. Typical examples include

¹By stochastic we mean either true randomness or high dimensional deterministic events that, given the available data, are in no way predictable.

the estimation of Lyapunov exponents (Schittenkopf, Dorffner, and Dockner 2001); the correlation dimension (Harrison, Yu, Lu, and George 1999, Hsieh 1991); the closely related BDS statistic (Brock, Hsieh, and LeBaron 1991); mutual information and other complexity measures (Darbellay and Wuertz 2000); and nonlinear predictability (Agnon, Golan, and Shearer 1999). The rationale of each of these reports was to apply a measure of nonlinearity or “chaoticity” to the supposed i.i.d. returns (or log-returns). Deviation from the expected statistic values for an i.i.d. process is then taken as a sign of nonlinearity.

However, it is not immediately clear what constitutes *sufficient* deviation. In each case, some statistical fluctuation of the quantity being estimated is to be expected. For example, the correlation dimension of an i.i.d. process should be equal to the embedding dimension² d_e . Although not a signature of chaos, for a deterministic chaotic time series, the correlation dimension is often a non-integer. But for practical purposes it is impossible to differentiate between the integer 1 and the non-integer 1.01. Furthermore, for $d_e \gg 1$ and time series of length $N < \infty$, one observes that the correlation dimension is systematically less than d_e , even for i.i.d. noise. These issues are complicated further by the fact that it is almost certain that financial time series contain a combination of deterministic and stochastic behavior (Schittenkopf, Dorffner, and Dockner 2001).

The problem is that for any nonlinear measure applied to an arbitrary experimental time series one does not usually have access to the expected distribution of statistic values for a noise process. In the case of the correlation dimension, the BDS statistic goes some way towards addressing this limitation (Brock, Hsieh, and LeBaron 1991). However, the BDS statistic is an asymptotic distribution and only provides an estimate for i.i.d. noise. Of course, this can be generalized to more sophisticated hypotheses by first modeling the data and then testing whether the model prediction errors are i.i.d. For example, one could expect that linearly filtering the data and testing the residuals against the hypothesis of i.i.d. noise can be used to test the data for linear noise. However, testing model residuals has two problems. Firstly, one is now testing the hypothesis that a specific model generated the data. The statistical test is now (explicitly) a test of a the particular model that one fitted to the data rather than a test of all such models. Secondly, linear filtering procedures have been shown to either remove deterministic nonlinearity or introduce spurious determinism in a time series (Mees and Judd 1993, Theiler and Eubank 1993). A linear filter is often applied to time series to pre-whiten³ the data. For linear time series so-called “bleaching” of the time series data is known not to be detrimental (Theiler and Eubank 1993). However, Theiler and Eubank (1993) show that linear filtering to “bleach” the time series can mask chaotic dynamics. For this reason they argue that the BDS statistic should *never* be applied to test model residuals. Conversely, Mees and Judd (1993) have observed that nonlinear noise reduction schemes can actually introduce spurious signatures of chaos into otherwise linear noise time series.

In the context of nonlinear dynamical systems theory, the method of surrogate data has been introduced to correct these deficiencies (Theiler, Eubank, Longtin, Galdrikian, and Farmer 1992). The method of surrogate data is clearly based on the earlier bootstrap techniques common in statistical finance. However, surrogate methods go beyond both standard bootstrapping and the BDS statistic, and provide a non-parametric method to directly test against classes of systems *other than* i.i.d. noise.

²The embedding dimension, d_e , is a parameter of the correlation dimension estimation algorithm and will be introduced in section 1.1.

³That is, make the data appear like Gaussian (“white”) noise.

Surrogate data analysis provides a numerical technique to estimate the expected probability distribution of the test statistic observed from a given time series for: (i) i.i.d. noise, (ii) linearly filtered noise, and (iii) a static monotonic nonlinear transformation of linearly filtered noise. Some recent analyses of financial time series have exploited this method to provide more certain results (for example (Harrison, Yu, Lu, and George 1999) and (Kugiumtzis 2001)). However, it has been observed that for non-Gaussian time series the application of these methods can be problematic (Small and Tse 2002a) or lead to incorrect results (Kugiumtzis 2000). But it is widely accepted that financial time series are leptokurtotic and it is therefore this problematic situation that is of most interest. Despite the commonly observed “fat” tails in the probability distribution of financial data the possibility that these data may be generated as a static nonlinear transformation of linearly filtered noise has not been adequately addressed.

In this paper we report the application of tests for nonlinearity to three daily financial time series: the Dow-Jones Industrial Average (DJIA), the US\$-Yen exchange rate (USDJPY), and the London gold fixings (GOLD). We utilize estimates of correlation dimension and nonlinear prediction error.

To overcome some technical problems⁴ (Harrison, Yu, Lu, and George 1999) with the standard Grassberger-Proccacia correlation dimension estimation algorithm (Grassberger and Procaccia 1983a), we utilize a more recent adaptation (Judd 1992) that is robust to noise and short time series. To estimate the expected distribution of statistic values for noise processes, we apply the method of surrogate data (Theiler, Eubank, Longtin, Galdrikian, and Farmer 1992). By examining the rank distribution of the data and rescaling, we avoid problems with these algorithms (Kugiumtzis 2000, Small and Tse 2002a).

For each of the three financial indicators examined we find evidence that the time series are distinct from the three classes of linear stochastic system described above. Hence, we reject the random walk model for financial time series and also two further generalizations of this. Financial time series are not consistent with i.i.d. noise, linearly filtered noise, or a static monotonic nonlinear transformation of linearly filtered noise. Therefore, i.i.d. (including leptokurtotic and Gaussian distributions), ARMA process, or a nonlinear transformation of these do not adequately model this data.

Although this result is consistent with the hypothesis of deterministic nonlinear dynamics, it is not direct evidence of chaos in the financial markets. As a further test of nonlinearity in this data we apply a new surrogate data test (Small, Yu, and Harrison 2001). Unlike previous methods, this new technique has been shown to work well for non-stationary time series (Small and Tse 2002a). This new test generates artificial data that are both consistent with the original data and a noisy local linear model (Small, Yu, and Harrison 2001, Small and Tse 2002a). This is therefore a test of whether the original time series contains any complex nonlinear determinism. In each of the three time series we find sufficient evidence that this is the case. We show that this new test is able to mimic ARCH, GARCH and EGARCH models of financial time series but is unable to successfully mimic the data. Therefore, financial time series contain deterministic structure that may not be modeled using these conditional heteroskedastic processes. Therefore, state dependent linear or weakly nonlinear stochastic processes are not adequate to completely describe the determinism exhibited by financial time series. Significant detail is missing from these models.

⁴These technical issues and their remedies will be discussed in more depth in section 1.1.

To verify this result and in an attempt to model the deterministic dynamics we apply nonlinear radial basis modeling techniques (Judd and Mees 1995, Small and Judd 1998a, Small, Judd, and Mees 2002) to this data both to quantify the determinism and to attempt to predict the near future evolution of the data. We find that in each of the three cases, for time series of almost all lengths, the optimal nonlinear model contains nontrivial deterministic nonlinearity. In particular we find that the structure in the GOLD time series far exceeds that in the other two data sets. Our conclusion is that these data are best modeled as a noise driven deterministic dynamical system either far from equilibrium or undergoing deterministic bifurcation.

In section 1 we will discuss the mathematical techniques utilized in this paper. Section 2 presents the results of our analysis and discusses the implications of these. Section 3 is our conclusion.

1. Mathematical Tests for Chaos

In this section we provide a review of the mathematical techniques employed in this paper. Whilst these tests may be employed to find evidence consistent with *chaos*, it is more correct to say that these are mathematical tests for *nonlinearity*. Nonlinearity is merely a necessary condition for chaos. In section 1.1 we discuss correlation dimension and section 1.2 describes the estimation of nonlinear prediction error. Section 1.3 describes the method of surrogate data and section 1.4 introduces the new pseudo-periodic surrogates proposed by Small, Yu, and Harrison (2001).

1.1. Correlation dimension

Correlation dimension (CD) measures the structural complexity of a time series. Many excellent reviews (for example (Peters 1991)) cover the basic details of estimating correlation dimension. However, the method we employ here is somewhat different.

Let $\{x_t\}_{t=1}^N$ be a scalar time series, consisting of N observations, x_t being the t -th such observation. In the context of this work $x_t = \log \frac{p_{t+1}}{p_t}$ is the difference between the logarithm of the price on successive days. Nonlinear dynamical systems theory requires that the time series x_t is the output of the evolution on a deterministic attractor in d_e -dimensional space (Takens 1981). Taken's embedding theorem (Takens 1981) (later extended by many other authors) gives sufficient conditions under which the underlying dynamics may be reconstructed from $\{x_t\}_{t=1}^N$. The basic principle of Taken's embedding theorem is that (under certain conditions) one may construct a *time delay embedding*

$$\begin{aligned} x_t &\longmapsto (x_t, x_{t-1}, x_{t-2}, \dots, x_{t-d_e+1}) \\ &= z_t \end{aligned} \tag{1.1}$$

such that the embedded points z_t are homeomorphic to the underlying attractor⁵, and their evolution is diffeomorphic to the evolution of that underlying dynamical system⁶. The necessary conditions of this theorem require that x_t is a sufficiently accurate measurement of the original system. This is an

⁵The *attractor* is the abstract mathematical object on which the deterministic dynamic evolution is constrained.

⁶Those adverse to mathematics may substitute “the same as” for both “diffeomorphic to” and “homeomorphic to”.

assumption we are forced to make. Application of the theorem and the time delay embedding then only requires knowledge of d_e ⁷. In general there is no way of knowing d_e prior to the application of equation (1.1). One may either apply (1.1) for increasing values of d_e until consistent results are obtained, or employ the method of false nearest neighbors or one of its many extensions (see for example (Kennel, Brown, and Abarbanel 1992)). For the current application we find that both methods produce equivalent results and that a value of $d_e \geq 6$ is sufficient. Ding, Grebogi, Ott, Sauer, and Yorke (1993) demonstrate that the requirements on d_e for estimating correlation dimension are not as restrictive as one might otherwise expect.

Embedding aside, correlation dimension is a measure of the distribution of the points z_t in the embedding space $\mathbf{R}^{d_e}$. Define the correlation function $C_N(\epsilon)$ by

$$C_N(\epsilon) = \binom{N}{2}^{-1} \sum_{1 \leq i < j \leq N} \Phi(\|z_i - z_j\| < \epsilon) \quad (1.2)$$

where $\Phi(X) = 1$ if and only if X is true. The correlation function measures the fraction of pairs of points z_i and z_j that are closer than ϵ ; it is an *estimate* of the *probability* that two randomly chosen points are closer than ϵ .

The correlation dimension is then defined by:

$$d_c = \lim_{\epsilon \rightarrow 0} \lim_{N \rightarrow \infty} \frac{\log C_N(\epsilon)}{\log \epsilon}. \quad (1.3)$$

The standard method to estimate correlation dimension from (1.3) is described by Grassberger and Procaccia (1983a). The Grassberger-Procaccia algorithm (Grassberger and Procaccia 1983a, Grassberger and Procaccia 1983b) estimates correlation dimension as the asymptotic slope of the log-log plot of correlation integral against ϵ . Unfortunately this algorithm is not particularly robust for finite time series or in the case of noise contamination (see for example the discussion in (Small, Judd, Lowe, and Stick 1999)). Careless application of this algorithm can lead to erroneous results. Several modifications to this basic technique have been suggested in the literature (Judd 1992, Diks 1996). The method we employ in this paper is known as Judd's algorithm (Judd 1992, Judd 1994) which is more robust for short (Judd 1994) or noisy (Galka, Maaß, and Pfister 1998) time series.

Judd's algorithm assumes that the embedded data $\{z_t\}_t$ can be modeled as the (cross) product of a fractal structure (the deterministic part of the system) and Euclidean space (the noise). It follows (Judd 1992) that for some ϵ_0 and for all $\epsilon < \epsilon_0$

$$C_N(\epsilon) \approx \epsilon^{d_c} p(\epsilon) \quad (1.4)$$

where $p(\epsilon)$ is a polynomial correction term to account for noise in the system⁸. Note that, correlation dimension is now no longer a single number d_c but is a function of ϵ_0 : $d_c(\epsilon_0)$. A discussion of the

⁷Oftentimes, a second parameter, the *embedding lag* τ is also required. The embedding lag is associated with characteristic time scales in the data (such as decorrelation time) and is employed in the embedding (1.1) to compensate for the fact that adjacent points may be over-correlation. However, for inter-day recording of financial indicators, a natural choice of embedding lag is $\tau = 1$.

⁸The polynomial $p(\epsilon)$ has order equal to the topological dimension of the attractor. In numerical applications linear or quadratic $p(\epsilon)$ is sufficient.

physical interpretation of this is provided by Small, Judd, Lowe, and Stick (1999). The parameter ϵ_0 may be thought of as a viewing scale: as one examines the features of the attractor more closely or more coarsely, the structural complexity and therefore apparent correlation dimension changes. The concept is somewhat akin to the multi-fractal analysis introduced by Mandelbrot (1999)⁹.

1.2. Nonlinear prediction error

Correlation dimension measures the structural complexity of a time series. Nonlinear prediction error (NLPE) measures certainty with which one may predict future values from the past. Nonlinear prediction error provides an indicator of the average error for prediction of future evolution of the dynamics from past behavior. The value of this indicator will depend crucially on the choice of predictive model. The scheme employed here is that originally suggested by Sugihara and May (1990). Informally, NLPE measures how well one is able to predict the future values of the time series. This is a measure of critical importance for financial time series, but, even in the situation where predictability is low, this measure is still of interest.

The local linear prediction scheme described by Sugihara and May (1990) proceeds in two steps. For some embedding dimension d_e the scalar time series is embedded according to equation (1.1). To predict the successor of x_t one looks for near neighbors of z_t and predicts the evolution of z_t based on the evolution of those neighbors. The most straightforward way to do this is to predict z_{t+1} based on a weighted average of the neighbors of z_t :

$$\langle z_{t+1} \rangle = \sum_{\tau < t} \alpha(\|z_\tau - z_t\|) z_{\tau+1} \quad (1.5)$$

where $\langle z_{t+1} \rangle$ denotes the prediction of z_{t+1} and $\alpha(\cdot)$ is some (possibly nonlinear) weighting function¹⁰. The scalar prediction $\langle x_{t+1} \rangle$ is simply the first component of $\langle z_{t+1} \rangle$. Extensions of this scheme include employing a local linear or polynomial model of the near neighbors z_τ . Another common application of schemes such as this is for estimation of Lyapunov exponents (Wolf, Swift, Swinney, and Vastano 1985). To estimate Lyapunov exponents one must first be able to estimate the system dynamics. Estimating Lyapunov exponents requires additional computation and the estimates obtained are known to be extremely sensitive to noise. For these reasons we chose to only estimate nonlinear prediction error in this paper.

Applying equation (1.5) one has an estimate of the deterministic dynamics of the system. NLPE is computed as the deviation of these predictions from reality

$$\left(\frac{1}{N} \sum_{t=1}^N (\langle x_{t+1} \rangle - x_{t+1})^2 \right)^{\frac{1}{2}}.$$

⁹The main difference is that Mandelbrot's multi-fractal analysis assumes some degree of fractal structure at all length scales; Judd's algorithm does not. Typically, below some scale the dynamics are dominated by noise and are no longer self-similar.

¹⁰Typically, $\alpha(x) = \exp(-x^2)$ or $\alpha(x) = \begin{cases} 1, & x < \ell \\ 0, & \text{otherwise} \end{cases}$.

To normalize data of different magnitude and provide a useful comparison between the time series, nonlinear prediction error p is normalized by the standard deviation of the data σ_s :

$$p = \frac{1}{\sigma_s} \left(\frac{1}{N} \sum_{t=1}^N (\langle x_{t+1} \rangle - x_{t+1})^2 \right)^{\frac{1}{2}}. \quad (1.6)$$

From equation (1.6) it is clear that a value of $p \approx 1$ indicates prediction that is no better than guessing¹¹. If $p < 1$, then there is significant predictable structure in the time series; if $p > 1$, the local linear model (1.5) is performing worse than the naive predictor.

As we will see in section 1.4 such schemes provide potential to mimic a wide range of nonlinear functions, but they have their limitations. More advanced nonlinear modeling schemes are possible and will be described in section 2.3. The purpose of nonlinear prediction error is only to provide a statistical indication of the amount of determinism in a time series. For time series of financial returns we find that the predictability using this method is extremely poor.

1.3. The method of surrogate data

For a random process, one expects the nonlinear predictability to be low. Deterministic processes have a higher predictability. Similarly, a random process has an infinite correlation dimension, and the correlation dimension of a (purely) deterministic one is finite. However, for a given embedding dimension d_e one's estimates of correlation dimension will always be finite. Furthermore, N points of a trajectory of a random walk (embedded in d_e dimensions) will have correlation dimension much lower than d_e . Finally, most real world time series (and certainly financial data) contain both deterministic and stochastic behavior. Therefore, correlation dimension or any other nonlinear measure alone is not sufficient to differentiate between deterministic and stochastic dynamics. The purpose of surrogate data methods is to provide a benchmark — the expected distribution for various stochastic processes — against which a particular correlation dimension estimate may be compared.

Before describing surrogate data analysis it is useful to review the BDS statistic (Brock, Hsieh, and LeBaron 1991). The BDS statistic, a technique popular in the financial literature, may be employed to differentiate between correlation dimension estimates for a given time series, and that expected for i.i.d. noise. In many ways the purpose of surrogate data analysis is very similar to a generalized BDS statistic.

The BDS statistic is an answer to the question “how should the correlation integral behave if this data is i.i.d.?” In this respect, the BDS statistic and surrogate data analysis share a common motivation. From the correlation integral (1.2) one may analytically derive the expected behavior of the BDS statistic for a time series of finite length drawn from an i.i.d. distribution. Any deviation from this expected value indicates potential temporal dependence within the data. The magnitude of deviation required to indicate statistically significant deviation may be calculated for particular (presumed) alternative hypotheses (Brock, Hsieh, and LeBaron 1991). However, the behavior of this statistic for an arbitrary alternative is unknown and certain pathological processes are known to be indistinguishable from i.i.d. noise (Brock, Hsieh, and LeBaron 1991). Furthermore, in its original form the BDS statistic tests only

¹¹For a mean zero time series.

for i.i.d data, acceptable alternatives include: linear noise, nonlinearity, chaos, non-stationarity or persistent processes. The BDS statistic cannot (directly) differentiate between chaos and linearly filtered noise.

It has been suggested that the BDS statistic could be extended to test more general hypotheses by testing model residuals against the hypothesis of white noise. However, the procedure is ill-advised for two reasons. First, if one constrains the data to test for i.i.d. residuals (by linearly filtering or applying nonlinear models) then one implicitly is only testing the validity of that one particular model. In contrast, the surrogate methods we employ here provides a test against the entire class of *all* models. For example, it is a trivial task to build an excessively large model so that the residuals *are* i.i.d. noise. One could then apply the BDS statistic to the residuals and conclude that that model class is consistent with that data. It is not. The only valid conclusion that one could make is that there is no evidence that that particular model is inconsistent with the data. Simulating the null distribution of the BDS statistic for an arbitrary class of model suffers the same weakness. Instead of testing the entire class of (for example) all GARCH models, one is only estimating the null distribution for the specific GARCH processes to simulated. This is a considerably weaker test. Our second objection to employing the BDS statistic in this manner is that it has been shown that “bleaching” chaotic data will mask the chaotic dynamics (Theiler and Eubank 1993).

The principle of surrogate data analysis is the following. For a given experimental time series $\{x_t\}_t$ one estimates an ensemble of n_s surrogate time series $\{s_t^{(i)}\}_t$ for $i = 1, 2, \dots, n_s$. Each time series $\{s_t^{(i)}\}_t$ is a random realization of a dynamical system consistent with some specified hypothesis $\mathcal{H}$, but otherwise similar to $\{x_t\}_t$. To test whether $\{x_t\}_t$ is consistent with $\mathcal{H}$ one then estimates the value of some statistic $u(\cdot)$ for both the data and surrogates. If $u(\{x_t\}_t)$ is atypical of the distribution $u(\{s_t^{(i)}\}_t)$ then the hypothesis $\mathcal{H}$ may be rejected. If not, $\mathcal{H}$ may not be rejected. As is the norm with statistical hypothesis testing, failure to reject $\mathcal{H}$ does not indicate that $\mathcal{H}$ is true, only that one has no evidence to the contrary. For this reason, one typically employs a variety of different test statistics. Conversely, if $u(\{x_t\}_t) < u(\{s_t^{(i)}\}_t)$ for all i then the probability that $\{x_t\}_t$ is consistent with $\mathcal{H}$ is less than $\frac{2}{n_s}$ ¹². Furthermore, if one has good reason to believe that the $u(\{s_t^{(i)}\}_t)$ follow a normal distribution, then the standard Student's t -test may be applied.

The three standard surrogate techniques are known as algorithm 0, 1 and 2 and address the three hypotheses of: $\mathcal{H}_0$ independent and identically distributed noise; $\mathcal{H}_1$ linearly filtered noise; and, $\mathcal{H}_2$ a static monotonic nonlinear transformation of linearly filtered noise¹³.

Surrogate generation with algorithm 0 has been widely used in the financial analysis literature for some time. It is common practice to compare (nonlinear) statistic values of data to the values obtained from a scrambled time series. However, it is not sufficient to compare the data to a single scrambled time series: such a test has no power¹⁴ (Theiler and Prichard 1996). A single surrogate time series provides only a (poor) estimate of the mean of the distribution for $\mathcal{H}$ and no estimate of the variance of that distribution.

¹²This assumes a two-tail test, for a one-tailed test the probability is $\frac{1}{n_s}$.

¹³Algorithm 2 has several problems. A complete discussion of the various pitfalls of this algorithm is given by Schreiber and Schmitz (2000) and Small and Tse (2002a).

¹⁴We use the term “power” in its statistical sense: we mean the probability of a statistical test correctly rejecting a given (false) hypothesis.

Furthermore, we should note that the hypothesis $\mathcal{H}_0$ tested by algorithm 0 surrogates is exactly equivalent to the hypothesis tested by the BDS statistic (Brock, Hsieh, and LeBaron 1991). However, the power of the BDS statistic for arbitrary time series is unknown and may only be estimated for specific alternative hypotheses. The power of surrogate analysis is dependent only on the test statistic.

Notice that the hypothesis addressed by algorithm 2, a static monotonic nonlinear transformation of linearly filtered noise, has no equivalent with the machinery of the bootstrap or the BDS statistic. Therefore, the possibility that financial time series can be modeled as a monotonic nonlinear transformation of linearly filtered noise has not been adequately addressed with these alternative methods.

The power of surrogate data hypothesis testing is dependent on the statistic selected. If the chosen statistic is insensitive to differences between data and surrogates then rejection of $\mathcal{H}$ is unlikely, even when other measures may find a clear distinction. A detailed discussion of test statistics is given by Theiler and Prichard (1996). For the results reported here we select two nonlinear measures as test statistics: correlation dimension and nonlinear prediction. Correlation dimension has been shown to be a suitable test statistic and Judd's algorithm appears to provide an unbiased estimate (Small and Judd 1998b). In particular, it has been shown that using correlation dimension as a test statistic avoids many of the technical issues concerning algorithm 2. Nonlinear prediction is selected as a contrast to correlation dimension. The two statistics measure quite different attributes of the underlying attractor. Furthermore, nonlinear prediction is a measure of particularly great interest for financial time series.

1.4. Pseudo-periodic surrogates

These standard surrogate algorithms are widely known and widely applied. However, each of these algorithms only tests against a linear null hypothesis. The most general class of dynamical systems that may be rejected by these hypotheses is a *static* monotonic nonlinear transformation of linearly filtered noise. Various simple noise models (for example GARCH) are not consistent with this hypothesis. It would be useful to have a more general, nonlinear, surrogate hypothesis test. Schreiber (1998) and Small and Judd (1998b) propose two alternative schemes to generate nonlinear surrogates. However, both methods are computationally intensive and rely on parameter dependent algorithms.

Recently, an alternative (nonlinear) surrogate data test, based on the local linear modeling methodology described in section 1.2, has been described (Small, Yu, and Harrison 2001). This surrogate test was introduced in the context of dynamical systems exhibiting pseudo-periodic¹⁵ behavior. However, the applicability of this method has been suggested to extend to arbitrary time series (Small and Tse 2002a).

The pseudo-periodic surrogate (PPS) algorithm is described as follows.

Algorithm 3 (The PPS algorithm) *Construct the usual vector delay embedding $\{z_t\}_{t=1}^{N-d_e}$ from the scalar time series $\{x_t\}_{t=1}^N$ (equation 1.1). Call $\mathcal{A} = \{z_t | t = 1, 2, \dots, N - d_w\}$ the reconstructed attractor.*

1. Choose an initial condition $s_1 \in \mathcal{A}$ at random and let $i = 1$.

¹⁵The meaning of pseudo-periodic should be intuitively clear. To be precise we mean that there exists $\tau > 0$ such that the autocorrelation exhibits a significant peak at τ .

2. Choose a near neighbor $z_j \in \mathcal{A}$ of s_i according to the probability distribution

$$\text{Prob}(z_j = z_t) \propto \exp \frac{-\|z_t - s_i\|}{\rho}$$

where the parameter ρ is the noise radius.

3. Let $s_{i+1} = z_{j+1}$ be the successor to s_i and increment i .
4. If $i < N$ go to step 2.

The surrogate time series is the scalar first component of $\{s_t\}_t$. Denote the hypothesis addressed by this surrogate test as $\mathcal{H}_3$.

The rationale of this algorithm is to provide sufficient nonlinearity (in step 2) to mimic minor non-linearity but not to reproduce the original dynamics exactly. By choosing the level of randomization ρ at an appropriate level we may expect to mimic the large scale dynamics (periodic orbits or trends such as persistence or systematic large deviations) present in the original data, but any fine scale structure (such as deterministic chaos or higher order nonlinearity) will be drowned out by the noise. Of course, the crucial issue is the appropriate selection of noise level. We follow the arguments of Small, Yu, and Harrison (2001) and select the noise level ρ to be such that the number of short segments for which data and surrogate are identical is maximized. Extensive computational studies have shown that this method provides a good choice of ρ (Small, Yu, and Harrison 2001, Small and Tse 2002a).

This algorithm is particularly well suited to financial time series as it has been shown to reproduce qualitative features of non-stationary or state dependent time series well. Rejection of surrogate generated by this algorithm will imply that the data may not be modeled by local linear models, ARMA, or state dependent noise processes (such as GARCH)¹⁶.

Finally, we note that if H_i is the set of all dynamical systems consistent with $\mathcal{H}_i$ then it follows that $H_0 \subset H_1 \subset H_2 \subset H_3$. Therefore these surrogate tests may be applied in a hierarchical manner. One applies increasingly more general hypotheses until the data and surrogates are indistinguishable.

2. Determinism in Financial Indicators

We now turn to real data and apply the methods described in the previous section to three time series: DJIA, GOLD and USDJPY. Section 2.1 describes the data used in this analysis. Section 2.2 presents the results of the surrogate data analysis with correlation dimension and nonlinear prediction error as test statistics. Finally, section 2.3 attempts to justify the results of section 2.2 by modeling the dynamics of these data.

¹⁶PPS models are of the same class as local linear models and therefore capture local linear nonlinearity. ARMA are global linear stochastic models and certainly a subset of that class. The issue of whether GARCH models are consistent with $\mathcal{H}_3$ is more complex. Step 2 of the PPS algorithm introduces state dependent variance: the amount of randomization will depend on how many neighbors are nearby. It is not true that this particular form of state dependent variance will capture the structure of all GARCH models, but a suitable choice of d_e and ρ allows the PPS algorithm to mimic low order G/ARCH processes. Our computational experiments (section 2.2) show that ARCH, GARCH and EGARCH processes do not lead to rejection of $\mathcal{H}_3$.

Figure 1. Financial time series. One thousand inter-day log returns of each of the three time series considered in this paper are plotted (from top to bottom): DJIA, USDJPY, and GOLD.

2.1. Data

We have selected daily values of three financial indicators for analysis in this paper: the Dow-Jones Industrial Average (DJIA), the Yen-US\$ exchange rate (USDJPY), and the London gold fixings (GOLD)¹⁷. In each case the time series represent daily values recorded on trading days up to the date January 14, 2002. USDJPY and GOLD consist of 7800 and 8187 daily returns respectively. The DJIA time series is longer, consisting of 27643 daily values.

Each time series was pre-processed by taking the log-returns of the price data:

$$x_t = \log p_t - \log p_{t-1}. \quad (2.7)$$

We then repeated the following analysis with time series of length 100 to 1000 in increments of 100, and length 1000 to 10000 in increments of 1000. The shortest time series cover approximately 3 – 4 months of data, the longest covers about 40 years. All three data sets are only sampled on week days that are not bank holidays. However, local bank holidays and non-trading days cause some variation

¹⁷The data were obtained from: <http://finance.yahoo.com/> (DJIA), <http://pacific.commerce.ubc.ca/xr/> (USDJPY), and <http://www.kitco.com/> (GOLD).

Figure 2. Typical surrogate time series. For the DJIA($N = 1000$) data set (the top panel of figure 1) we have computed representative surrogates consistent with each of $\mathcal{H}_0$, $\mathcal{H}_1$, $\mathcal{H}_2$ and $\mathcal{H}_3$ (from top to bottom). The PPS surrogate ($\mathcal{H}_3$) appears qualitatively most similar to the data.

between the data sets. Each time series consisted of the most recent values. Representative samples of all three time series are shown in figure 1.

2.2. Surrogate analysis

For each time series (three data sets, considered at 19 different lengths) we computed 50 surrogates from each of the four different surrogate generation algorithms. For each surrogate and original data set we computed correlation dimension (with $d_e = 6$) and nonlinear prediction error (also with $d_e = 6$).

The embedding dimension for the PPS algorithm was chosen to be $d_e = 8^{18}$. Hence for each of 19 data sets we generated 200 surrogates and computed 201 estimates of correlation dimension and nonlinear prediction error. Representative surrogates for one of the data sets are shown in figure 2.

We compared the distribution of statistic values to the value for the data in two ways. First, we rejected the associate null hypothesis if the data and surrogates differed by at least three standard deviations. We found that in all cases the distribution of statistic values for the surrogates were approximately normal and this therefore provided a reasonable estimate of rejection at the 99% confidence level. For correlation dimension $d_c(\epsilon_0)$ we compared estimates for fixed values of ϵ_0 only.

Whilst it may seem a reasonable assertion, this technique does assume that the underlying distribution of test statistic values is Gaussian. To overcome this we compared data and surrogates in a non-parametric way. We rejected the associated hypothesis if the correlation dimension of the data was lower than the surrogates. Nonlinear prediction was deemed to distinguish between data and surrogates if NLPE for the data exceeded that of the surrogates. The NLPE statistic is actually a measure of predictive performance compared to guessing. If nonlinear prediction is less than 1 the prediction does decidedly better than guessing; if prediction is greater than 1 then it is decidedly worse. In none of our analysis did the nonlinear prediction of the data perform significantly *better* than guessing (it often performed worse). Therefore we chose to reject the hypothesis if the nonlinear prediction of the data exceeded that of all 50 surrogates. The probability of this happening by chance (if the data and surrogates came from the same distribution) is $\frac{1}{51}$. Therefore, in these cases we could reject the underlying hypothesis at the 98% confidence level.

In all 228 trials (3 data sets at 19 length scales and 4 hypotheses) we found that non-parametric rejection (at 98% confidence) occurred only if the hypothesis was rejected under the assumption of Gaussianity (at 99% confidence). Table 1 summarizes the results and figure 3 provides an illustrative computation.

For $N \geq 800$ we found that the hypotheses associated with each of $\mathcal{H}_0$, $\mathcal{H}_1$, and $\mathcal{H}_2$ could be rejected. Therefore at the time scale of approximately 3 years, each of DJIA, GOLD and USDJPY is distinct from monotonic nonlinear transformation of linearly filtered noise. This result is not surprising; on this time scale one expects these time series to be non-stationary. This result is only equivalent to the conclusion that volatility has not been constant over the last three years. Certainly, the time series in figure 1 confirm this.

For $N < 800$ we found evidence to reject $\mathcal{H}_1$ but often no evidence to reject $\mathcal{H}_0$ or $\mathcal{H}_2$. With no further analysis this would lead us to conclude that these data are generated by a monotonic nonlinear transformation of linearly filtered noise with non-stationary evident for $N \geq 800$. However, this is not the case. Rejection of $\mathcal{H}_1$ but failure to reject $\mathcal{H}_0$ is indicative of either a false positive or a false negative. The sensitive nature of both CD and NLPE algorithms, and their critical dependence on the quantity of data imply that for short noisy time series, they are simply unable to distinguish data from surrogates. Further calculations of CD with different values of d_e confirm this. In all cases we found evidence to reject $\mathcal{H}_0$ and $\mathcal{H}_2$. The high rate of false negatives (lack of power of the statistical test) is due to the insensitivity of our chosen test statistic for extremely short and noisy time series. That is, for

¹⁸As discussed by Schreiber and Schmitz (2000) one should select parameters of the surrogate algorithm independent of parameters of the test statistic. For the test statistics we selected $d_e = 6$ based on evidence in the literature concerning the dimension of this data (Barnett and Serletis 2000). For PPS generation we found $d_e = 8$ performed best.

Figure 3. Surrogate data calculation. An example of the surrogate data calculations for the DJIA ($N = 1000$). The top row shows a comparison of correlation dimension estimates for the data (solid line, with stars) and 50 surrogates (contour plot, the outermost isocline denotes the extremum). The bottom row shows the comparison of nonlinear prediction error for data (thick vertical line) and surrogates (histogram). From this diagram one has sufficient evidence to reject $\mathcal{H}_1$ (correlation dimension) and $\mathcal{H}_3$ (both CD and NLPE). Both $\mathcal{H}_0$ and $\mathcal{H}_2$ may not be rejected. Further tests with other statistics indicate that in fact these other statistics may be rejected also. This particular time series is chosen as illustrative more than representative. As table 1 confirms, most other time series led to clearer rejection of the hypotheses $\mathcal{H}_i$ ($i = 0, 1, 2, 3$).

short time series the variance in one's estimates of NLPE and CD becomes greater than the deviation between data and surrogates.

Another possible source of this failure is highlighted by Dolan and Spano (2000). For particularly short time series the randomization algorithms involved in $\mathcal{H}_0$ and $\mathcal{H}_2$ constrain surrogates to consist of special re-orderings of the original data. With short time series the number of reasonable re-orderings is few and the power of the corresponding test is reduced (Dolan and Spano 2000).

We conclude that the evidence provided by linear surrogate tests (algorithm 1) is sufficient to reject the random walk model of financial prices and ARMA models of returns. Consistent rejection across all time scales implies that non-stationarity is not the *sole* source of nonlinearity¹⁹.

¹⁹Here, we assume that the system is stationary for $N \leq 100$. From a dynamical systems perspective this is reasonable, from a financial one it is arguable whether it is reasonable to describe markets as stationary over periods or approximately

Table 1
Surrogate data Hypothesis tests.

Rejection of CD or NLPE with 99% confidence (assuming Gaussian distribution of the surrogate values) is indicated with “CD” and “NP” respectively. Rejection of CD or NLPE with 98% confidence (as the test statistic for the data is an outlier of the distribution) is denoted by “cd” or “np” respectively. This table indicates general rejection of each of the four hypotheses $\mathcal{H}_i$ ($i = 0, 1, 2, 3$). Exceptions are discussed in the text.

N	GOLD				DJIA				USDJPY			
	$\mathcal{H}_0$	$\mathcal{H}_1$	$\mathcal{H}_2$	$\mathcal{H}_3$	$\mathcal{H}_0$	$\mathcal{H}_1$	$\mathcal{H}_2$	$\mathcal{H}_3$	$\mathcal{H}_0$	$\mathcal{H}_1$	$\mathcal{H}_2$	$\mathcal{H}_3$
100	cd	np	–	cd	cd	cd	cd		cd	cd	cd	cd
200	–	–	–	cd	cd	cd+np	cd	cd	–	–	–	cd
300	–	NP	–	cd	–	–	cd	cd				
400	–	cd	–	cd	–	CD	–	cd	–	–	–	cd
500	–	np	–	cd	–	CD	–	cd	–	CD	–	cd
600	cd	NP+cd	CD	cd	cd	CD	–	cd	–	CD	–	cd
700	CD+NP	NP	NP+cd	cd	–	CD	cd	cd	–	CD	–	cd
800	CD+NP	NP	CD+NP		CD	CD+np	–	cd	np	CD	np	cd
900	CD+NP	NP	CD+NP	cd	–	CD	CD	cd	NP	CD+NP	NP	cd
1000	CD+NP	NP	CD+NP	cd	–	CD	–	cd+np	NP	CD+NP	NP	cd
2000	CD+NP	NP	CD+NP	np	CD	CD	CD	cd+np	CD	CD	CD	
3000	CD+NP	NP	CD+NP	np	CD	CD	CD		CD	CD	CD	cd+np
4000	CD+NP	NP	CD+NP		CD+NP	NP	CD+NP	cd+np	CD	CD	CD	np
5000	CD+NP	NP	CD+NP	np	CD+NP	NP	CD+NP	cd	CD	CD	CD	cd+np
6000	CD	NP	CD+NP	np	CD+NP	NP	CD+NP	cd	CD	CD	CD	np
7000	CD	cd	CD	np	CD+NP	NP+cd	CD+NP	cd+np	CD	cd	CD	cd+np
8000	CD	np	CD	CD	CD	NP	CD+NP	cd+np	CD	NP	CD	CD
9000	CD	cd	CD	cd	CD	NP	CD+NP	cd+np	CD	NP	CD+np	CD
10000	CD	cd	CD	cd	CD+NP	NP	CD+NP	CD	CD	NP	CD	CD

The results of the PPS algorithm are clearer, and provide some clarification of the results of the linear tests. In all but six cases the PPS algorithm found sufficient evidence to reject $\mathcal{H}_3$ ²⁰. Since $H_i \subset H_3$ for $i = 0, 1, 2$ we have (as a consequence of rejecting $\mathcal{H}_3$) that $\mathcal{H}_2$ should be rejected. Therefore, we conclude (five of the six exceptions notwithstanding) that these data are not generated by a static transformation of a colored noise process. Rejection of $\mathcal{H}_3$ strengthens this statement further. As PPS data can capture non-stationarity and drift (Small and Tse 2002a) it is capable of modeling locally linear nonlinearities and non-stationarity in the variance (GARCH processes). Therefore, we conclude that these financial data are not a static transformation of *non-stationary* colored noise.

To show that the PPS algorithm can test for ARCH, GARCH and EGARCH type dynamics in the data we apply the algorithm to three test systems²¹. An autoregressive conditional heteroskedastic (ARCH) model makes volatility a function of the state

$$x_t = \sigma_t u_t$$

four months. However, this is more of an issue of definitions of stationarity. If causes of non-stationarity are considered as part of the system (rather than exogenous inputs) then the system is stationary.

²⁰There are six exceptions. Of these six exceptions (see table 1) five occur when $\mathcal{H}_2$ is rejected and therefore linear noise may still be rejected as the origin of this data. This leaves only one exception: for YEN($n = 300$) we are unable to make any conclusion. However, one false negative from 57 trials with two independent test statistics is not unreasonable.

²¹These three systems are given by equations (25), (26) and (27) of (Hsieh 1991).

Figure 4. Testing for conditional heteroskedasticity.. Typical behavior of the surrogate algorithms for ARCH (equation 2.8, top row), GARCH (equation 2.9, middle row), and EGARCH (equation 2.10, bottom row) processes. The calculations illustrated here are for $N = 1000$. Each panel depicts the ensemble of correlation dimension estimates for the surrogates (contour plot) and the value for the test data (solid line, with stars). From this diagram (as with our test calculations for $N = 100, \dots, 10000$) one can reject $\mathcal{H}_0$, $\mathcal{H}_1$, and $\mathcal{H}_2$, but may not reject $\mathcal{H}_3$. This supports our assertion that the PPS algorithm provides a test for conditional heteroskedasticity.

$$\sigma_t^2 = \alpha + \phi x_{t-1}^2. \quad (2.8)$$

For computation we set $\alpha = 1$ and $\phi = 0.5$ (as Hsieh (1991) did). A generalized ARCH (GARCH) process extends equation 2.8 so that the noise level is also a function of the past noise level

$$\sigma_t^2 = \alpha + \phi x_{t-1}^2 + \xi \sigma_{t-1}^2. \quad (2.9)$$

We simulate (2.9) with $\alpha = 1$, $\phi = 0.1$ and $\xi = 0.8$ (Hsieh 1991). Finally, an exponential GARCH (EGARCH) process introduces a static nonlinear transformation of the noise level.

$$\log \sigma_t^2 = \alpha + \phi \left| \frac{x_{t-1}}{\sigma_{t-1}} \right| + 2\xi \log \sigma_{t-1} + \gamma \frac{x_{t-1}}{\sigma_{t-1}} \quad (2.10)$$

($\alpha = 1$, $\phi = 0.1$, $\xi = 0.8$ and $\gamma = 0.1$). For stochastic processes of the forms (2.8), (2.9) and (2.10), figure 4 shows typical results of the application of the PPS algorithm. For data lengths from $N = 100$ to $N = 10000$ we found that linear surrogate algorithms (algorithms 0, 1 and 2) showed clear evidence that this data was not consistent with the null hypothesis (of nonlinear monotonic transformation of linearly filtered noise). However, the PPS algorithm was unable to reject these systems. Therefore we conclude that the financial data we examine here is distinct from ARCH, GARCH and EGARCH processes. Conditional heteroskedasticity is insufficient to explain the structure observed in these financial time series.

Table 2
Rate of false positives observed with PPS algorithm

For each of (2.8), (2.9) and (2.10), 30 realisations of length $N = 100, 200, 300, \dots, 1000$ were generated and the PPS algorithm applied. This table depicts the frequency of false positives (incorrect rejection of the null).

N	100	200	300	400	500	600	700	800	900	10000
ARCH	0	0	0	0	0	0	0	0	1	0
GARCH	1	0	1	0	0	1	0	0	0	0
EGARCH	2	1	1	0	0	0	0	0	0	1

We repeated the computations depicted in figure 4 for multiple realisations of (2.8), (2.9) and (2.10) for different length time series $N = 100, 200, 300, \dots, 1000$. The results of this computation are summarized in table 2. We found that in each case, from 30 realisations of each of these processes, clear failure to reject the null. The highest rate of false positives ($p = 0.1$) occurred for short realisations of EGARCH processes. Hence, this test is robust against false positives.

The local linearity of the PPS data means that certain mildly nonlinear systems may also be rejected as the source of this data. However, we can make no definite statement about how “mild” is mild enough. Instead, we conclude that these data are not a linear stochastic process, but should be modeled as stochastically driven nonlinear dynamical systems.

There are two further significant observations to be made from this data. For CD, rejection of the hypotheses occurred when the correlation dimension of the data was *lower* than that of the data. Therefore, these time series are *more* deterministic than the corresponding surrogates. This implies that some of the apparent stochasticity in these time series is deterministic. This does not imply that the financial markets exhibit deterministic chaos. This does imply that these data exhibit deterministic components that are incorrectly assumed (by the random walk, ARIMA and GARCH models) to be stochastic.

For NLPE one finds that the surrogate and data only differ significantly when the nonlinear predictive power of the data is *worse* than the surrogates. Hence, these time series are *less* predictable than the corresponding stochastic processes. This result may be somewhat surprising, but it is perhaps less shocking than the converse would be. If this data was found to be more predictable than noise then that would indicate that local linear models could be employed for profitable trading (at least trading with a positive expectation). Hence, this result implies that the model used to estimate nonlinear prediction error, a local linear model (1.5), is unable to capture the determinism of this data. This result supports the conclusion of the PPS hypothesis test: the dynamics of this system may not be adequately modeled by local linear nonlinear models.

2.3. Nonlinear modeling and prediction

Surrogate data analysis has demonstrated that these data contain significant determinism that cannot be captured with either linear stochastic or local linear nonlinear models. In this section we employ a popular nonlinear modeling regime to extract that determinism. We employ *minimum description length* (MDL) *radial basis models* (RBM) as described by Judd and Mees (1995). The choice of this particular modeling scheme is based on the personal biases of the first author (Small, Judd, and Mees 2002) and the fact that we have at our disposal a software implementation of this scheme. We make no suggestion *here* that this particular methodology is superior to any other. An alternative, but similar technique favored in the finance literature are the neural network models (Small and Tse 2002b). Lee, White, and Granger (1993) and Lee (2000) apply neural network models as a test for nonlinearity in a manner similar to the surrogate data analysis and radial basis functions discussed in this paper.

Again, we start from embedding the scalar time series according to equation (1.1). We model the dynamics of the map $z_t \mapsto z_{t+1}$ by constructing a function f such that

$$y_{t+1} = f(z_t) + e_t$$

where e_t are expected to be i.i.d. random variates. The function $f(\cdot)$ is of the form

$$f(z_t) = \lambda_0 + \sum_{i=1}^m \lambda_i y_{t-\ell_i} + \sum_{j=1}^n \lambda_{m+j} \phi\left(\frac{\|z_t - c_j\|}{r_j}\right) \quad (2.11)$$

where: $\lambda_i \in \mathbf{R}$ are the *weights*; $\ell_i \in \mathbf{R}$ such that $0 \leq \ell_i < \ell_{i+1} \leq d_w$ are the *lags*; $\phi(x) = \exp \frac{x^2}{2}$ are the *basis functions*; $c_j \in \mathbf{R}^{d_w}$ are the *centers*; and, $r_j \in \mathbf{R}^+$ are the *radii*. Various schemes exist to estimate each of these parameters; we follow the methods described by Judd and Mees (1995) and Small and Judd (1998a). In addition to each of these parameters, the model size $(m+n)$ must also be selected so that the e_t are indeed i.i.d. yet model over-fitting must be avoided. In other words, we wish to capture all the deterministic features of the dynamics, but not guess features for which there is insufficient evidence in the data²².

²²This is a common failing of neural network techniques. The number of parameters in neural network models often approaches or exceeds the amount of data and sufficient “training” will result in an excellent fit to the data. However, because the models are essentially fitted to noise, prediction is abysmal. We address this problem in a separate publication (Small and Tse 2002b).

Figure 5. Description length and in-sample prediction error. Description length and in-sample prediction error $\frac{1}{N} \sum_t e_t$ as a function of model size n are depicted for a typical data set (GOLD($N = 1000$)). In-sample prediction error is monotonically decreasing with model size; description length has a distinct minimum.

We employ minimum description length as described by Rissanen (1989) to estimate the correct model size. Roughly speaking, the description length of a data set is the length of shortest description (*encoding*) that can be employed to reconstruct the entire data set. For random variates, that encoding will simply be the data itself and the description length will be the length of that data set. For data containing some determinism, the most compact encoding of that data will be the description of a model of the deterministic component and the model prediction errors.

A complete derivation of description length is beyond the requirements of this paper. Roughly, description length is given by

$$L(x) = L(x|\Lambda) + L(\Lambda) \quad (2.12)$$

where $L(x|\Lambda)$ is the description length of the data given by the model $\Lambda = (\lambda_0, \lambda_1, \lambda_2, \dots, \lambda_{m+n})$ (note that $L(x|\Lambda)$ is a monotonic decreasing function of model size), and $L(\Lambda)$ is the description of the model (a monotonic increasing function of model size). For some intermediate value of model size there will be a minimum. The description length of the model is given by

$$L(\Lambda) \propto \sum_{i=0}^{m+n} \log \frac{1}{\delta_i}.$$

where δ_i is the cost of specify the i -th parameter λ_i , and $L(x|\Lambda)$ is the negative log likelihood of the model prediction errors under the assumed distribution. A more detailed description of the computation of description length may be found in the references. Figure 5 depicts a representative computation of the behavior of MDL with model size.

Figure 6. Model size. Model size n of the MDL best nonlinear model as a function of length of data is shown. Models for GOLD are larger than those for the DJIA, which in turn are larger than those for USDJPY. Systematic increase in model size with data length N indicates additional history does contribute additional information.

For each data set and for $N \leq 5000$ we built an MDL best nonlinear model according to equation 2.11²³. To ensure that the results were not due to chance the modeling procedure was repeated five times. In each instance the results obtained were almost identical (with the exception of the future predictions, described below). Finally, we repeated this analysis for different starting dates within the data window (i.e. using a sliding window through the historical data) and found consistent results. This suggests that non-stationarity is indeed a problem for this data. Certainly, for the DJIA data one explanation of this non-stationarity may be the trend break in price-earnings ratios after 1982. Using short sliding windows we applied the statistical stationarity test described in (Yu, Lu, and Harrison 1998) and found that this data did indeed exhibit non-stationarity (Yu, Lu, and Harrison 1999).

Figure 6 depicts the optimal model size n (the number of nonlinear terms in the best model) as a function of length of data. With the exception of USDJPY ($N \leq 300$) the MDL best model contained at least one nonlinear term; in most cases the best model contained many nonlinear terms. This supports the conclusion of our surrogate data analysis and demonstrates that there is significant deterministic

²³Building MDL models for $N > 5000$ was beyond the modest computational resources at our disposal.

nonlinearity in these data. One interpretation of this is that the model is fitting non-stationary trends in the data. However, this is not supported by the evidence: even for very short time series ($N < 500$) the number of terms in the nonlinear models of GOLD and DJIA are significant. Furthermore, the increase in model size with N is not strictly monotonic (as one would expect if all the additional information was due to non-stationarity). Part of the observed determinism will, most likely, correspond to non-stationary effects — particularly for USDJPY. However, both GOLD and DJIA contain deterministic nonlinearity that is fitted by the model and not merely an effect of non-stationarity. We have repeated these calculations, and each time we obtain the same results. The models for GOLD are substantially larger (more nonlinear structure) than those for DJIA, which, in turn are larger than those for USDJPY. This is indicative of increasing nonlinear deterministic structure in the corresponding time series.

These models fit the available data and provide the most compact description of that data. The models extract any deterministic structure in the time series and express it in the most compact form. Non-deterministic structure is left as i.i.d. random variates, meaning that the non-trivial models depicted in figure 6 have extracted significant deterministic structure from these data. This does not necessarily mean that these models will *predict* well.

To test the predictive power of these models we applied the models to predict the prices of GOLD, DJIA and USDJPY over the 70 samples subsequent to the end of the data sets (January 14, 2002). Predictions on these 70 samples are honest predictions: they are made without any additional knowledge and are genuine twenty-four hour predictions. We denote these as *out-sample* predictions. In contrast, we call predictions made on the data from which the model was built *in-sample* predictions. Note that, by building a sufficiently large model in-sample prediction error can be made arbitrarily small. But doing this is only over-fitting the data. Applying minimum description length should prevent over-fitting and ensure that in-sample and out-sample predictions are comparable: we should do no better on the seen or unseen data.

Figure 7 depicts in-sample and out-sample prediction error as a function of data length. Whilst there is some variability between data sets, in-sample and out-sample prediction are comparable. From figure 7 it appears that the DJIA is predicting better than the model average and GOLD is predicting substantially worse. In fact, this anomaly is due to the relatively low volatility in the DJIA and relatively high volatility in GOLD over the 70 day period we have tested. The mean volatility (standard deviation of x_t) over this 70 day period for DJIA, GOLD and USDJPY is 0.0131, 0.0029 and 0.0048 respectively. We note that the in-sample prediction performs significantly better than the naive predictor, but over the 70 day test sample, out-sample prediction does not.

We have also examined the number of days that the model correctly predicted only the direction of price movement. The simplest way in which to compare predictive power of this model to the naive predictor is to treat the data as binomial distribution with $p = \frac{1}{2}$. This simple method and several alternatives are discussed by (Diebold and Mariano 1995). We find that the observed data is not significantly different from this hypothesis, and neither are the model predictions. Hence, we conclude that for each data set our predictions did not perform significantly better than random. Although the poor predictive power of this model is to be expected²⁴ we can draw two conclusions from this. We could conclude that the test sample of ten days is either too short, too noisy or not representative of the dynamics. However, we find this conclusion to be somewhat unsatisfactory. This conclusion could be tested by repeating our

²⁴If the model performed substantially better than random then this paper would never have been written.

Figure 7. In-sample and out-sample prediction error. In-sample (solid line) and out-sample (dashed) prediction as a function of model size is shown. Out of sample prediction error (true predictions) performance is independent of model size (and performs no better than guessing) indicating that additional history does not contribute to improved prediction. In-sample and out-sample predictions perform comparably, indicating that the modeling algorithm is not over fitting the data.

analysis for very many alternative test dates. It is likely that this lengthy analysis would only confirm that our chosen test date is not particularly atypical. Instead, we conclude that the deterministic dynamics are swamped by stochastic²⁵ behavior making prediction on the deterministic dynamics impossible. Furthermore, we note that the out-sample prediction performance does not improve with the length of time series used for the model. This supports the conclusion that stochasticity is dominant.

While figure 6 provides evidence that there is deterministic nonlinearity in this data, figure 7 shows the poor predictive power of these models. This provides good evidence that while these models do capture deterministic features in the data, these features cannot be profitably used for prediction, or, these features are relatively rare and therefore absent during the test period selected in this analysis. We find that the most reasonable conclusion is that these time series are best described as a nonlinear noise driven dynamical system far from equilibrium *undergoing bifurcation or change in system dynamics*. Such changes in system dynamics would be modeled well using these nonlinear modeling techniques but would not be able to be exploited for prediction. Models based on nonlinear dynamical systems

²⁵That is, apparently random behavior or extremely high dimensional dynamics.

Figure 8. Correlation of model size n and rejection of $\mathcal{H}_2$. Model size n is plotted as a function of significance of the rejection of $\mathcal{H}_2$ for all data. The DJIA is depicted as diamonds, USDJPY as stars, and GOLD as circles. A positive correlation is apparent, indicating that model size is an indicator of nonlinearity.

theory are not capable of extrapolation (to unseen dynamics), they only interpolate between embedded points (equation (1.1)).

We have observed that GOLD exhibited substantially larger nonlinear models than the other time series and we interpreted this as evidence of increased nonlinearity in the time series. Table 1 also shows that rejection of the surrogate hypothesis tests was most clear for GOLD. For each surrogate data test we computed the deviation of the data from the surrogates:

$$\text{deviation} = \frac{u - \mu_s}{\sigma_s} \quad (2.13)$$

where: u is the test statistic value for the data; μ_s is the mean value for the surrogates; and, σ_s is the standard deviation of the surrogates. From this we obtain a numerical value for how different the data and surrogates are²⁶. Figure 8 plots the deviation between data and surrogates for $\mathcal{H}_2$ tested with CD against the nonlinear model size. Figure 8 exhibits a definite trend (correlation coefficient of 0.78). This is empirical evidence that surrogate data hypothesis testing (and in particular, algorithm 2) provides a definite indicator of nonlinear determinism. When the same calculation was repeated with the PPS

²⁶We do not make any claims here about Gaussianity or employ this quantity in a Student's t -test.

algorithm we found only a very weak trend (correlation coefficient of 0.17). This is to be expected as the null hypothesis $\mathcal{H}_3$ and surrogates generated by the PPS algorithm incorporate nonlinearity. Hence, rejection of $\mathcal{H}_3$ is a test of more than just weak nonlinearity that may be mimicked with local linear models. Therefore, these data exhibit nontrivial deterministic nonlinearity. However, the models described in this section are unable to fully capture this determinism.

These computational models, like most others, are based upon certain assumptions about the distribution of errors. For fitting purposes we assume that the errors are mean zero; to compute the cost function (2.12) we assume that they are i.i.d. Gaussian. If we assume that all market participants behave rationally and equally, then the mean value theorem implies that this should be the case. However, there is a growing body of empirical evidence (Malliaris and Stein 1999) that not all players in the market are equally informed. Therefore, they do not all play by the same rules and it is not necessarily the case that the sum of their behavior will be normal. Ample evidence in the literature, including the results of this paper support this view. Hence if models are to perform better, one must first ensure that the assumptions concerning the distribution of noise, often implicit in the model, reflect reality.

For nonlinear or chaotic systems it has been shown that even in fairly generous circumstances false assumptions about the distribution of errors can prove dangerous (McSharry and Smith 1999). In the case of market price data we have no strong evidence to assume more than that the errors have mean zero (any non-zero expectation can be incorporated in the model itself). Therefore, to apply complicated modeling procedures one must first ensure that the models assume *only* that the expectation of the errors is zero, or else restrict one's attention to a market in which all players are equally knowledgeable.

Black (1986) showed that noise in financial time series is the cause of inefficiency, and therefore potential profitability. Noise is an intrinsic part of the dynamic behavior of financial data and cannot be neglected (Black 1986). Equally, one should not assume that all noise is i.i.d. Gaussian.

3. Conclusion

We have provided empirical evidence that the random walk model is wrong (or at the very least, highly improbable). We have compared the measured statistic values for financial time series and noisy simulations. We find that the simulations, representative of the noise process most like the observed data, are clearly distinct from it. From this we have concluded that the time series of inter-day returns do not behave as i.i.d. noise. Therefore, we reject the random walk model as a likely origin of this data. Furthermore, we find that this data is inconsistent with colored noise, or a monotonic nonlinear transformation of colored noise. Therefore, a static monotonic nonlinear transformation is insufficient to explain the fat tails observed for financial time series.

This implies that many common models of financial process are inappropriate. Because the inter-day returns do not behave as i.i.d. noise, returns may not be considered as either i.i.d. Gaussian or leptokurtotic. Because they are not colored noise, they cannot be modeled as a linear stochastic process (autoregressive and/or moving average models are wrong). Because they are not a monotonic nonlinear transformation of linearly filtered noise, the many extensions of ARMA models are also inappropriate. Finally, rejection of the PPS test for local nonlinearity shows that even simple forms of dynamic nonlinearity, including ARCH, GARCH and EGARCH models, are insufficient to model market dynamics.

Therefore conditional heteroskedasticity is insufficient to explain the dynamics observed in financial time series.

We conclude that the system that generated these time series is a nonlinear dynamical system driven by noise that is either not in equilibrium or undergoing bifurcation/non-stationarity. Our evidence supports the view that, despite this non-stationarity there is a characteristic nonlinear deterministic structure that persists through the length of these recordings (approximately 40 years). Inter-day returns of financial time series must be modeled as a nonlinear dynamical system, albeit driven by noise. Agnon, Golan, and Shearer (1999) argue that a deterministic nonlinear system could act as a source or filter and produce complex nonlinear stochastic behavior consistent with that observed in financial data. Our results support this view, and provide the beginnings of a method to better understand that nonlinear system.

One can argue that, while there is no evidence against nonlinear determinism or even chaos in financial markets (see for example McCauley (2000)) there is strong evidence against the random walk model (e.g. (Mandelbrot 1999)). McCauley (2000) argued that market indicators do not follow a random walk (a conclusion that our empirical results support) and that the system is best considered as a complex stochastic dynamical system, far from equilibrium. Such a system could easily admit chaotic dynamics — in addition to the stochastic (or very high dimensional) component.

Some of the previous analysis of financial time series with methods of nonlinear dynamical systems theory have not reached results as unambiguous as those presented here and the literature contains many apparently contradictory reports (Barnett and Serletis 2000, Malliaris and Stein 1999). The reason for this is that the techniques of nonlinear dynamical systems theory do not necessarily provide results that are as unequivocal as first suggested (Peters 1991). The field of dynamical systems theory is still very young and results obtained with these techniques should be interpreted carefully. Estimates of dynamic invariants (correlation dimension, Lyapunov exponents, entropy and the like) rarely provide reliable error bars and so cannot be applied without checking for simpler, linear stochastic explanation of the data. Surrogate data analysis must be applied to test the data and ensure that false evidence of chaos is not reported (Kugiumtzis 2001). However, surrogate analysis brings its own pitfalls: misapplied surrogate analysis can lead to both false positive (Schreiber and Schmitz 1996) and false negative (Theiler and Prichard 1996) results. Finally, the PPS algorithm introduced in (Small, Yu, and Harrison 2001) has been applied in this paper and provides a clear rejection of the linear hypothesis when (for $N < 800$) algorithm 2 surrogates suggest the opposite. Without this additional check our results would have been more ambiguous.

We assert that the problem with attempting to identify *chaos* in financial time series is that the whole issue is potentially ill-conceived. Competing analyses could provide evidence that a particular time series is consistent with a noisy chaotic system or a purely stochastic one. The level of noise in this data far surpasses that found in most physical or even physiological systems (see for example (Small, Judd, Lowe, and Stick 1999)). Therefore, the only sensible question to ask is whether linear models are sufficient to describe this data, or if nonlinear models are required. We find that linear models only provide an approximation to the underlying dynamics and that deviation from that approximation is empirically observable.

Our analysis also suggests a method to quantify the level of nonlinear determinism. We showed that the deviation between data and algorithm 2 surrogates provided a good indicator of the size of the best

nonlinear model. In the three time series we examined we found nonlinear determinism to be greatest in GOLD. That there is nonlinearity in financial time series should not be particularly surprising. Hsieh (1991) examined financial time series from a dynamical systems perspective and found that while ARCH was insufficient, *certain* conditional heteroskedasticity (GARCH type models) could explain the observed nonlinearity. For the data examined in this paper we have found significant differences between GARCH models and the observed dynamics. However, we have not explicitly considered predictability in the volatility of the time series. This is already an area of active interest and shall continue to receive close examination.

The source of the nonlinearity observed in this data lies in one of the erroneous assumptions underlying the semi-strong form of the efficient market hypothesis and its corollary: the random walk model. The effect of individuals on the market is not linear. Two investors will not have twice the effect of a single investor. Mob behavior is commonly observed in financial circles.

According to Black (1986) economic research proceeds on the basis of learned debated rather than empirical evidence. The empirical evidence presented here contradicts the standard linear stochastic view of market dynamics. This view is therefore insufficient to describe all the information available from these time series. We are not suggesting that the standard linear techniques of financial analysis are worthless and should be rejected. We are only suggesting that they may be augmented by more exact models. The nonlinear arbitrage pricing model described by Bansal, Hsieh, and Viswanathan (1993) is an example of one potential application of nonlinear techniques to improve standard linear methods²⁷.

Colophon

This work was supported by Hong Kong Research Grants Council and Hong Kong Polytechnic University Research Grants (Nos. A-PE46 and G-YW55). Source code for the algorithms used in this paper are available from the first author (ensmall@polyu.edu.hk). Postal address: Michael Small, Department of Electronic and Information Engineering, Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong.

References

- Agnon, Yehuda, Amos Golan, and Matthew Shearer, 1999, Nonparametric, nonlinear, short-term forecasting: theory and evidence for nonlinearities in the commodity markets, *Economics Letters* 65, 293–299.
- Bansal, Ravi, David A. Hsieh, and S. Viswanathan, 1993, A new approach to international arbitrage pricing, *J Finance* 48, 1719–1747.
- Barnett, William A., and Apostolos Serletis, 2000, Martingales, nonlinearity, and chaos, *Journal of Economic Dynamics and Control* 24, 703–724.
- Benhabib, Jess, 1996, On cycles and chaos in economics, *Studies in Nonlinear Dynamics and Econometrics* 1, 1–2 Editorial.

²⁷However, we note that in general global polynomial approximations such as those employed by Bansal, Hsieh, and Viswanathan (1993) do not perform well (Small, Judd, and Mees 2002).

- Black, Fischer, 1986, Noise, *J Finance* 41, 529–543.
- Brock, William A., David A. Hsieh, and Blake LeBaron, 1991, *Nonlinear dynamics, chaos and instability*. (The MIT Press Cambridge, Massachusetts).
- Crack, Timothy Falcon, and Olivier Ledit, 1996, Robust structure without predictability: The "Compass Rose" pattern of the stock market, *J Finance* 51, 751–762.
- Darbellay, Georges A., and Diethelm Wuertz, 2000, The entropy as a tool for analysing statistical dependences in financial time series, *Physica A* 287, 429–439.
- Diebold, Francis X., and Roberto S. Mariano, 1995, Comparing predictive accuracy, *Journal of Business and Economic Statistics* 13, 253–263.
- Diks, C., 1996, Estimating invariants of noisy attractors, *Physical Review E* 53, R4263–R4266.
- Ding, Mingzhou, Ceslo Grebogi, Edward Ott, Tim Sauer, and James A. Yorke, 1993, Plateau onset for correlation dimension: when does it occur?, *Physical Review Letters* 70, 3872–3875.
- Dolan, Kevin T., and Mark L. Spano, 2000, Surrogate for nonlinear time series analysis, *Physical Review E* 64, 046128.
- Galka, A., T. Maaß, and G. Pfister, 1998, Estimating the dimension of high-dimensional attractors: A comparison between two algorithms, *Physica D* 121, 237–251.
- Grassberger, Peter, and Itamar Procaccia, 1983a, Characterization of strange attractors, *Physical Review Letters* 50, 346–349.
- Grassberger, Peter, and Itamar Procaccia, 1983b, Measuring the strangeness of strange attractors, *Physica D* 9, 189–208.
- Harrison, Robert G., Dejin Yu, Weiping Lu, and Donald George, 1999, Non-linear noise reduction and detecting chaos: some evidence from the S&P Composite Price Index, *Mathematics and Computers in Simulation* 48, 497–502.
- Hsieh, David A., 1991, Chaos and Nonlinear Dynamics: Applications to Financial Markets, *J Finance* 46, 1839–1877.
- Judd, Kevin, 1992, An improved estimator of dimension and some comments on providing confidence intervals, *Physica D* 56, 216–228.
- Judd, Kevin, 1994, Estimating dimension from small samples, *Physica D* 71, 421–429.
- Judd, Kevin, and Alistair Mees, 1995, On selecting models for nonlinear time series, *Physica D* 82, 426–444.
- Kennel, Matthew B., Reggie Brown, and Henry D. I. Abarbanel, 1992, Determining embedding dimension for phase-space reconstruction using a geometric construction, *Physical Review A* 45, 3403–3411.
- Kugiumtzis, D., 2000, Surrogate data test for nonlinearity including nonmonotonic transforms, *Physical Review E* 62, R25–R28.
- Kugiumtzis, D., 2001, On the reliability of the surrogate data test for nonlinearity in the analysis of noisy time series, *International Journal of Bifurcation and Chaos* 11, 1881–1896.
- LeBaron, Blake, 1994, Chaos and nonlinear forecastability in economics and finance, *Phil Trans R Soc Lond A* 348, 397–404.
- Lee, Tae-Hwe, 2000, Neural network test and nonparametric kernel test for neglected nonlinearity in regression models, *Studies in Nonlinear Dynamics and Econometrics* 4, 169–182.
- Lee, Tae-Hwy, Halbert White, and Clive W.J. Granger, 1993, Testing for neglected nonlinearity in time series models, *Journal of Econometrics* 56, 269–290.

- Malliaris, A.G., and Jerome L. Stein, 1999, Methodological issues in asset pricing: Random walk or chaotic dynamics, *Journal of Banking & Finance* 23, 1605–1635.
- Mandelbrot, Benoit B., 1999, A multifractal walk down wall street, *Scientific American* pp. 70–73.
- McCauley, Joesph L., 2000, The futility of utility: how market dynamics marginalize Adam Smith, *Physica A* 285, 506–538.
- McSharry, Patrick E., and Leonard A. Smith, 1999, Better nonlinear models from noisy data: Attractors with maximum likelihood, *Physical Review Letters* 83, 4285–4288.
- Mees, Alistair I., and Kevin Judd, 1993, Dangers of geometric filtering, *Physica D* 68, 427–436.
- Peters, Edgar E., 1991, *Chaos and order in the capital markets*. (John Wiley & Sons, Inc. New York).
- Rissanen, Jorma, 1989, *Stochastic complexity in statistical inquiry*. (World Scientific Singapore).
- Schittenkopf, Christian, Georg Dorffner, and Engelbert J. Dockner, 2001, On nonlinear, stochastic dynamics in economic and financial time series, *Studies in Nonlinear Dynamics and Econometrics* 4, 101–121.
- Schreiber, Thomas, 1998, Constrained randomization of time series, *Physical Review Letters* 80, 2105–2108.
- Schreiber, Thomas, and Andreas Schmitz, 1996, Improved surrogate data for nonlinearity tests, *Physical Review Letters* 77, 635–638.
- Schreiber, Thomas, and Andreas Schmitz, 2000, Surrogate time series, *Physica D* 142, 346–382.
- Small, Michael, and Kevin Judd, 1998a, Comparison of new nonlinear modelling techniques with applications to infant respiration, *Physica D* 117, 283–298.
- Small, Michael, and Kevin Judd, 1998b, Correlation dimension: A pivotal statistic for non-constrained realizations of composite hypotheses in surrogate data analysis, *Physica D* 120, 386–400.
- Small, Michael, Kevin Judd, Madeleine Lowe, and Stephen Stick, 1999, Is breathing in infants chaotic? Dimension estimates for respiratory patterns during quiet sleep, *Journal of Applied Physiology* 86, 359–376.
- Small, Michael, Kevin Judd, and Alistair Mees, 2002, Modeling continuous processes from data, *Physical Review E* 65, 046704.
- Small, Michael, and C.K. Tse, 2002a, Applying the method of surrogate data to cyclic time series, *Physica D* 164, 187–201.
- Small, Michael, and C.K. Tse, 2002b, Minimum description length neural networks for time series prediction., *Physical Review E* 66, 066701 Reprinted in *Virtual Journal of Biological Physics Research* 4 (2002).
- Small, Michael, Dejin Yu, and Robert G. Harrison, 2001, A surrogate test for pseudo-periodic time series data, *Physical Review Letters* 87, 188101.
- Sugihara, G., and R.M. May, 1990, Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series, *Nature* 344, 734–741.
- Takens, Floris, 1981, Detecting strange attractors in turbulence, *Lecture Notes in Mathematics* 898, 366–381.
- Theiler, James, and Stephen Eubank, 1993, Don't bleach chaotic data, *Chaos* 3, 771–783.
- Theiler, James, Stephen Eubank, Andre Longtin, Bryan Galdrikian, and J. Doyne Farmer, 1992, Testing for nonlinearity in time series: The method of surrogate data, *Physica D* 58, 77–94.
- Theiler, James, and Dean Prichard, 1996, Constrained-realization Monte-Carlo method for hypothesis testing, *Physica D* 94, 221–235.
- Wolf, Alan, Jack. B. Swift, Harry L. Swinney, and John A. Vastano, 1985, Determining Lyapunov exponents from a time series, *Physica D* 16, 285–317.

- Yu, Dejin, Weiping Lu, and Robert G. Harrison, 1998, Space time-index plots for probing dynamical nonstationarity, *Physics Letters A* 250, 323–327.
- Yu, Dejin, Weiping Lu, and Robert G. Harrison, 1999, Detecting dynamical nonstationarity in time series data, *Chaos* 9, 865–870.