The statistical properties of dimension calculations using small data sets

Nonlinearity 3 (1990) 155–176. Printed in the UK

The statistical properties of dimension calculations using small data sets

James B Ramsey and Hsiao-Jane Yuan
Department of Economics, Faculty of Arts and Sciences, New York University, 269 Mercer St, 3rd floor, New York 10003, USA

Received 23 May 1988, in final form 15 May 1989
Accepted by J D Farmer

Abstract. The statistical properties of estimates of pointwise dimension and their errors have been investigated for some maps and for some random variables. Substantial bias in the estimates is detected and modelled as a function of the sample size and the embedding dimension. The usual methods for calculating error bars are shown to underestimate the actual error bars by factors of ten and more. Procedures to improve the estimation of dimension in the cases studied are discussed, as are methods to improve the ability to distinguish noise from an attractor when using small data sets. Some idea of small is given when the attractors are similar to the ones studied.

AMS classification scheme number: 58F12, 54F45

1. Introduction

During the past few years, the concept of dimension has progressed from a simple intuitive idea about the fractal structure of sets to a much more interesting, but complex, set of interrelated concepts, as discussed, for example, in [1–5]. Many intriguing new ideas for the measurement of the 'average scaling' of an attractor, such as in [6–9], have been proposed. Also, there is the realisation that in most cases, no single summary number can suffice. Nevertheless, the Grassberger and Procaccia correlation integral procedure [11–14] has endured because of its relative simplicity, intuitive appeal and, most importantly, its successful use in the analysis of experimentally generated data. A useful review of basic dimension concepts, entropy, and their relationships is given in [14].

Simultaneously, interest in the use of dimension concepts has spread far beyond the early physical and chemical experiments wherein data sets with observations numbered in the tens of thousands were common. Recently, researchers in a variety of fields, such as brain research, optics, meteorology, and economics [15–22] have tried to apply the correlation integral approach to dimension to small data sets; indeed, very small, with numbers of observations ranging from less than a thousand to less than two hundred. Even for attractors with fractal dimension less than two, such data sets are miniscule. Consequently, there is considerable concern, even

0951-7715/90/010155 + 22$02.50

skepticism, about the validity of any findings based on such small numbers of data points. Intuitively, measures of scaling measure the relative sparseness of the attractor's points within some n-dimensional subspace, i.e. the attractor does not fill out the space within a given region, but leaves a series of self-similar open subspaces. Consequently, any physical measure of fractal dimension using experimental data cannot easily distinguish the absence of data points due to the structure of the attractor from the absence of points due merely to the sparseness of a few data points scattered in n-dimensional space. Further, the degree of scaling of the correlation integral often varies with the range of scales used. In short, the estimation of fractal dimension depends upon the interaction between sample size, the scaling range, the embedding dimension, and the structure of the attractor itself.

While the downward bias in the estimation of the dimension of random variables has long been recognised, the possibility of an upward bias in the estimate of the dimension of an attractor has not received as much concern. However, if true, and if one is limited to small numbers of observations, distinguishing low-dimensional noisy attractors from random variables is made more problematic; see for example [23].

Even with the very simple examples used in this paper, it is clear that there are many estimation problems that have serious implications for the reliability and the confidence one can place in dimension estimates when data sets are small. The fact that these problems can be ignored with very large data sets is of little comfort. In any event, some idea of what 'large' might mean is needed [24]. Attractors that are more complex than those examined in this paper are most likely to reinforce our conclusions. Our most succinctly stated conclusion is that dimension calculations are very much a large-numbers game.

More elaborately, our main conclusions with respect to the very simple examples that we examined in this paper are:

(i) dimension can be estimated with substantial upward bias for attractors;

(ii) dimension is always estimated with downward bias for random noise;

(iii) the bias effect increases with embedding dimension, but decreases with sample size;

(iv) dimension estimates are normally distributed for sample sizes larger than 1000;

(v) the rate of increase in estimated dimension to embedding dimension for random noise is an increasing function of the distribution's entropy;

(vi) the actual variance of dimension estimates can be as high as 64 times larger than that estimated by the usual least squares approach;

(vii) the variance of the dimension estimate decreases with sample size, but increases very rapidly with embedding dimension.

Practical methods to mitigate these problems are suggested as well as some rough lower bounds on the required sample size to achieve useful estimates.

Section 2 of this paper states the research objectives. The next two sections describe the examples used and the nature of the computer simulation experiments run. The fifth and sixth sections discuss in turn the empirical results and their implications for future research.

2. The research objective

Attention in this paper will be focused entirely on the Grassberger–Procaccia correlation approach alluded to above, which will be formally defined below. There

are two major reasons for this restriction: it is by far the most commonly used procedure; and capacity measures require enormously large numbers of data points for the same level of accuracy and so are impractical for the current objective [10, 25]; a useful pair of references to the earlier literature that precedes Grassberger and Procaccia is [8, 25]. The results obtained for the Grassberger–Procaccia approach should at least be indicative for the results that one would obtain for other measures of dimension or approaches to the calculation of any given dimensional concept. A related paper is that of Abraham *et al* [26], who have also been exploring the potentiality of small data sets for dimension estimates.

If the available experimental data series is represented by an ordered sequence of observations $\{X_t\}$, then for a given embedding dimension d, one can reconstruct the attractor, if it exists, by creating a sequence of d-tuples $\{(x_t, x_{t+\tau}, \ldots, x_{t+(d-1)\tau})\}$; the d-dimensional vectors so created are elements of a d-dimensional Euclidean space in which the attractor is embedded for large enough d. In the following, $\tau = 1$. This is an inessential simplification in this paper, although the value of τ is of crucial importance in actual experimental data [24]. If the series is in fact a realisation of a random variable then, for each d, the 'attractor' is embedded in itself, i.e. the points are space filling for any d-dimensional Euclidean space.

The sample calculation for the correlation integral is given by:

$$C_r^N = N^{-2} \sum_{i,j} \theta(r - |\boldsymbol{x}_i - \boldsymbol{x}_j|) \qquad r > 0 \qquad \boldsymbol{x}_i = (x_i, x_{i+1}, \ldots, x_{i+d-1}) \tag{2.1}$$

and $\theta(\cdot)$ is the Heaviside step function which maps positive arguments onto 1 and non-positive arguments onto 0. $|\ |$ designates some suitable norm, in our case the 'sup' norm. Essentially, $\theta(\cdot)$ counts the number of points within a distance r of each other. The reason for considering C_r^N is that

$$\lim_{\substack{N\to\infty \\ r\to 0^+}} C_r^N \to C \tag{2.2}$$

and $\mathrm{d} \ln C / \mathrm{d} \ln r = D_2$, whenever the derivative is defined [10]; D_2 is a member of a general class of dimensions D_q, $-\infty \leqslant q \leqslant \infty$, defined by:

$$D_q = -\lim_{r\to 0} K_q(r)/\ln r \qquad K_q = (1-q)^{-1} \ln \sum_{i=1}^{N(r)} P_i(r)^{q-1} \tag{2.3}$$

where $P_i(r)$ is the probability of a point of the attractor being within the ith box of a space that has been partitioned into a sequence of boxes of size r and $N(r)$ is the number of such boxes needed to cover the attractor; details are in [1, 11, 12, 14]. In the rest of the paper, D_2 will be designated d_c to stand for correlation dimension.

Correlation dimension is usually estimated from experimental data by a linear regression of the observed values of $\ln C_r^N$ on $\ln r$ over a suitably chosen subinterval of the range of r, $(0, 1)$. The estimated slope coefficient of this regression, hereafter designated $\hat{d}_c$, is the usual estimator of correlation dimension cited in the literature and is the basic variable used in this paper.

The entire focus of this paper is on the statistical properties of $\hat{d}_c$ under a variety of assumptions. Little attention is paid to the all-important issues involving the role of dimension calculations in the context of dynamical analysis.

A problem that is not addressed in this paper is the use of the 'naive estimate' of the probability of a point being in the ith cell that is given by n_i/N, where there are

N observations altogether and n_i observations in the ith cell. Grassberger [27, 28] has shown that for finite N, there is a bias introduced in the calculation of dimension by using the naive estimate of probability, except in the case where q, as defined in equation (1.3), is 2; this is the case that we are examining so that the estimate n_i/N can be used without producing yet another source of bias.

The overall objective of this research is to illustrate, amplify and clarify the difficulties inherent in dimension estimation with small data sets; this effort is particularly important in that unless one is already aware of the difficulties to be enumerated, they can easily be missed. Subsidiary outcomes are to indicate a potentially more efficient use of small data sets in estimating dimension, to attempt to clarify what is meant by 'small' at least with respect to the examples studied, and to suggest a potentially useful method for deciding whether one is observing an attractor or not, a surprisingly difficult decision with many small data sets and even for some not so small [23].

The first specific task is to illustrate with a particular set of examples the estimation bias that exists in small data sets; i.e., the expectation of $\hat{d}_c$ is not equal to the theoretical value of the dimension being estimated; the value of the bias is given by $(E(\hat{d}_c) - d_c)$. A related question concerns the consistency of the estimator $\hat{d}_c$ and the rate of convergence: as the sample size increases without limit, does the distribution of $\hat{d}_c$ collapse onto d_c, and how quickly?

Brock *et al* [29] have demonstrated the consistency and asymptotic normality of the estimators $\hat{d}_c$ discussed in this paper. Our experimental results should shed light on the extension of that result to other distributions and attractors as well as document for the cases studied the rate of convergence of the estimator.

We illustrate the trade-off in effects on estimation bias and on the size of error bars between sample size and embedding dimension; higher embedding dimensions require a more than proportional increase in sample size to offset bias and reduce error bars.

The second major objective is to reassess the procedures used to estimate the variance of estimate and consequently reassess the size of the error bars in dimension estimation. The standard procedure substantially underestimates the true value of the variance of estimate, even in the context of the simple examples we have examined. A most useful earlier reference in this respect is Mayer-Kress [30], who examined the effect on the variance of alternative averaging procedures. A related question is whether the assumption of asymptotic normality (Gaussian distribution) for the distribution of $\hat{d}_c$ is tenable with small sample sizes.

A subsidiary task was to demonstrate that different density functions give rise to differences between the rate of increase in estimated dimension and the rate of increase in embedding dimension. This result is of use in clarifying our understanding of dimension estimates of 'noise', especially when such estimates are used as a 'benchmark' of comparison for dimension calculations for hypothesised attractors.

A technique that was suggested by one of the authors some time ago and was implemented independently by Scheinkman and Le Baron [22] to aid in distinguishing attractor results from non-attractor, was to compare the dimension calculations of experimental data with those that have been randomised in time order. If one has data from an attractor that are uncorrelated, the estimated dimensions from the randomised data will increase. However, if the uncorrelated data are in fact random, the estimated dimension will not increase. The usefulness of this idea is examined experimentally in this paper.

Finally, the last set of objectives of this paper is to explore how the results obtained can be used to improve, or at least to modify our confidence in, dimension estimates obtained from small data sets.

3. The examples examined

The examples subjected to the statistical analysis that are to be discussed in section 4 are the Henon map, the quasiperiodic 5-torus, three random distributions: normal, gamma and uniform; and, lastly, randomised orderings of the Henon and the torus.

The Henon example chosen is the now classic

$$x_{n+1} = 1 - 1 \cdot 4x_n^2 + y_n \qquad y_{n+1} = 0 \cdot 3y_n. \tag{3.1}$$

These particular values were chosen to facilitate comparisons with the existing literature. After the first few hundred iterations were performed, a sequence of fifty thousand observations were generated for analysis. As is well known, the currently accepted value for d_c for the Henon map is 1.21 [31].

The 5-torus was defined in the canonical form:

$$\theta_i(t) = \cos(\theta_{0i} + w_i t) \qquad i = 1, 2, \ldots 5 \qquad x_t = \sum_{i=1}^{5} \theta_i(t). \tag{3.2}$$

The particular example used was defined by the parameter values:

$\theta_{0i} = 0$ for all i

$$\omega_1 = \sqrt{2} \qquad \omega_2 = \frac{1+\sqrt{5}}{2} - 1 \qquad \omega_3 = \sqrt{7} \qquad \omega_4 = \sqrt{3} \qquad \omega_5 = \sqrt{13}.$$

When $\theta_{0i} = 0$, a serious practical difficulty occurs. The 'observed' dimension of the 5-torus at finite length scales is greater than 5 and over the range of length scales that we used is approximately 5.2 due to the highly convoluted structure of the torus under these circumstances (Mayer-Kress, personal communication and [32]). The theoretical limit as the value of r goes to zero is still 5; but this fact is not helpful given the practical limitations on the minimum size of r that can be used to estimate dimension. For values of r that are greater than the relevant length scales for the 'surface' of the attractor, the dimension that will be observed will be greater by a 'fractal', because in essence one is measuring 'global' behaviour, not local behaviour. For sufficiently large θ_{0i}, of course, the observable dimension of the 5-torus is 5, because now it is practically possible to observe local behaviour on the attractor with r values that are small relative to the length scales of the convolutions of the attractor. These comments illustrate the great difficulties inherent in both the calculation and interpretation of dimension calculations.

Three different distributions of random variables were generated, fifty thousand observations each. Each set of data was normalised to the unit interval to facilitate the numerical analysis. The three distributions were:

Uniform:	$u(x) = k$	$k \equiv 1$	$x \in [0, 1]$
Normal:	$\phi(x) = \exp(-0.5x^2)/\sqrt{2\pi}$		
Gamma:	$\exp(-x)$	$x > 0.$	

The parameters of the distributions were chosen to contrast varying levels of entropy and skewness across the three distributions. These three distributions were obtained by use of random number generators created by STSC for the Statgraphics system.

The randomised Henon and randomised 5-torus were obtained by a random permutation of the order of the observations generated by each map. These two sets of observations were generated in order to examine the role of dimensional analysis in distinguishing random from chaotic attractors. The latter are spatially ergodic, but sequentially ordered in time. Consequently, randomisation of the order of the observations should produce data sets that scale similarly to random variables as conventionally defined; this approach is restricted to those cases in which the data are not autocorrelated. This idea has been explored by both the authors and by Scheinkman and Le Baron [22]

4. The sampling experiments and some caveats in dimension calculations

Because the experimental procedure followed was the same for each of the seven examples, one need only describe the procedures for one example.

Five sample sizes N were chosen for analysis: 200, 400, 1000, 2000 and 5000 observations. Given a pool of 50 000 observations, this meant that the number of replications (N_j) for each experimental run was 250, 125, 50, 25 and 10, respectively, for sample sizes 200, 400, 1000, 2000 and 5000.

On each replication at each sample size the correlation dimension d_c was estimated by the usual method of regressing $\ln C$ on $\ln r$, where C and r are defined in equation (2.1), but the superscript 'N' has now been deleted. Because each experimental run was repeated N_j times, $j = 1, 2, \ldots, 5$, and to avoid the introduction of operator error caused by the choice of the 'linear region', a single range of r values was chosen to hold over all replications for the map and the quasiperiodic examples. Unfortunately, this was not possible for the random examples. The main problem was a narrowing of the range of r values used as embedding dimension is increased; to some extent the restricted range was baised more towards larger values of r. Efforts were made to avoid undue variation in the range with higher embedding dimensions. The value of $\hat{d}_c$ so calculated is the prime datum for all analysis of the statistical properties of correlation dimension.

One other prime statistic calculated at this stage was the ordinary least squares definition of the variance of the estimate $\hat{d}_c$, hereafter labelled S_0^2. S_0^2 is defined by

$$S_0^2 = (n-2)^{-1}\left(\sum \hat{u}^2\right)(\langle x^2 \rangle)^{-1}$$

where $\hat{u}$ is the vector of residuals from the OLS regression of $\ln C$ on $\ln r$ and the vector x is given by

$$x = \ln r - \langle \ln r \rangle.$$

For each sample size, the first four moments of both $\hat{d}_c$ and S_0 were calculated from the N_j, $j = 1, 2, \ldots, 5$, observations obtained by N_j replications of the

experiment. The actual standardised moments for $\hat{d}_c$ that were used are:

$$\begin{aligned}
m_1'(\hat{d}_c) &= N_j^{-1} \sum_{i=1}^{N_j} \hat{d}_{ci} && \doteq \langle \hat{d}_{ci} \rangle \\
m_2(\hat{d}_c) &= (N_j - 1)^{-1} \sum_{i=1}^{N_j} (\hat{d}_{ci} - m_1')^2 && \doteq \langle (\hat{d}_{ci} - m_1')^2 \rangle \\
m_3^+(\hat{d}_c) &= \frac{1}{m_2^{3/2}} \left((N_j - 1)^{-1} \sum_{i=1}^{N_j} (\hat{d}_{ci} - m_1')^3 \right) && \doteq \langle (\hat{d}_{ci} - m_1')^3 \rangle / m_2^{3/2} \\
m_4^+(\hat{d}_c) &= \frac{1}{m_2^2} \left((N_j - 1)^{-1} \sum_{i=1}^{N_j} (\hat{d}_{ci} - m_1')^4 \right) && \doteq \langle (\hat{d}_{ci} - m_1')^4 \rangle / m_2^2.
\end{aligned} \tag{4.1}$$

The moments of S_0 were calculated similarly, but only the first moment $m_1'(S_0)$ was used in the subsequent analysis. Summary statistics are shown in the appendix.

The values for the embedding dimension, m, ranged from a low of 2 to a high of 25, at least in principle. In some cases, the maximum embedding dimension sustainable by the data was no more than 17 or 18.

An important practical issue involves the appropriate choice of the scaling region r actually used to calculate d_c, [10, 24]. While the theory discusses the properties of C as $r \to 0$, the reality is that the range of r used is far from zero and inevitably increases away from zero as embedding dimension is increased, [24]. Smaller values for r require substantial increases in sample size at any given embedding dimension in order to be able to determine a logarithmic linear relationship between C and r. In fact, the relationship between $\ln C$ and $\ln r$ is only approximately linear over a relatively narrow range of values for r. For large values of r, C saturates at unity so that the regression of $\ln C$ on $\ln r$ is zero. Further, as the value of r declines towards zero even with very large data sets, two complications arise; one is due to the limited precision of the data series and the other is due to the inevitable presence of noise. The former problem sets a practical lower bound on r before C collapses to zero and the latter difficulty offsets the decline in values of C when r reaches the level of the noise scales. An excellent reference is [33].

With limited data sets a further problem occurs in that the plot of $\ln C$ against $\ln r$ for sufficiently high dimension becomes a step function for certain ranges of r values, so that the estimation of slope becomes problematic at best. Ranges of r values which produce reliable slope estimates at lower embedding dimensions will yield apparently very low numbers at higher embedding dimensions. A similar phenomenon is observed as r decreases to very low values with fixed embedding dimension and sample size. This phenomenon is caused by the relative scarcity of data points so that C remains constant for sizable intervals of r values. This problem is particularly difficult with random examples. The existence of a step function may also be due to the fractal structure of the attractor [10].

The practical implications of this are that the choice of r values with respect to which d_c can be usefully calculated is difficult at any given embedding dimension and is doubly so if one is to avoid biases induced by changing the choice of r region as embedding dimension grows; i.e., one needs to avoid moving the r region too close to the saturation level in response to the above-mentioned difficulties. Further, if low scale noise is present, then the appropriate range of r for the calculation of dimension shrinks dramatically [24].

A subtle aspect of the 'choice by eye' of the linear region is that when one is

already aware, or believes one is aware, of the true slope coefficient, there is a natural, albeit unconscious, proclivity to choose the 'right' value. Because the choice of 'linear region' was fixed over all N_j replications in these experiments and because there is considerable variation in slopes from replication to replication, it is to be hoped that the results cited do not contain an implicit experimental bias towards the presumed 'correct' r values. However, this experimental procedure may produce estimates with somewhat larger variances than would be true for an r range chosen for each replication. Unfortunately, this approach precluded our examination of the effects of choice of scaling region on the bias in d_c estimates.

A further difficulty is that the assumption of a linear model even over a limited range of r values is not appropriate. The traditional assumption has been that

$$\ln C = \alpha_0 + \alpha_1 \ln(r) + u \tag{4.2}$$

where u is an additive error term assumed to be distributed normally and independently of r. It is the variance of u that generates the variance in estimating the slope coefficient α_1. This equation is incomplete in some respects and incorrect as stated in other respects.

Equation (4.2) is incomplete because the constant term α_0 is a function of embedding dimension [24, 34]:

$$C \simeq Kr^{D_2} \exp(-m\tau K_2) \tag{4.3}$$

where K is an arbitrary constant, m is the embedding dimension, τ is the delay time, and K_2 is the order-2 Renyi entropy that was defined in equation (2.3). Consequently, a logarithmic linear regression of the theoretical relationship is

$$\ln C \simeq (\ln K - m\tau K_2) + D_2 \ln r. \tag{4.4}$$

For non-zero values of entropy, K_2, α_0 is a decreasing function of m. As is well known, $K_2 = 0$ for ordered (non-chaotic) systems, is infinite in a random system, and is a positive constant for chaotic systems.

The assumptions on the error term, u, are most likely incorrect. The major difficulties concern the assumption that the errors are independently distributed and are independent of r; both are likely to be incorrect. The former because the sampling errors included for a given value of r, will be included for larger values of r. The latter because the standard deviation of the error will vary with the value of r.

With respect to the first difficulty, the autocorrelation of the error terms will mean that the standard definition of the standard deviation of the coefficient estimate is incorrect but that this problem does not, by itself, introduce biases in the estimate of the slope coefficient. Further, we strongly suspect that this definition is incorrect for other reasons to be discussed below.

With respect to the second problem, the short interval over which we calculated the regression of $\ln C$ on $\ln r$, meant that for our purposes and within the confines of our procedures, we could ignore this difficulty.

However, it is very clear from the distribution of sample $\ln C$ against $\ln r$ curves as well as from the results of Mayer-Kress [24], (see in particular his figure 1), that the appropriate regression model is:

$$\ln C = \alpha_0(\epsilon \mid N, m, r) + \alpha_1(\epsilon \mid N, m, r)\ln r + u \tag{4.2'}$$

where the variance of u is, for small N and large m, small relative to the variance of

ϵ, which represents an experimental error term whose distribution depends on sample size, N, embedding dimension, m, and the scaling region, r. Equation (4.2′) is known as a random coefficient regression model; the parameter of interest is the mean of the distribution of $\alpha_1(\epsilon)$, i.e. $E\{\alpha_1(\epsilon)\}$ which is a constant, not dependent on N, m, or r. But what the OLS regression estimates with error 'u' is a single sample value from the distribution of $\alpha_1(\epsilon)$, which does depend on N, m, and r. Equation (4.2′) indicates clearly that the actual variance of any estimate of $E\{\alpha_1(\epsilon)\}$ cannot be estimated from the usual OLS formula using an estimate of the variance of u and that in any event the OLS estimates may be substantially biased. One of the tasks of this paper is to discover the dependence of the distribution of the slope coefficient on sample size and embedding dimension. Some insight into the dependence of $\hat{d}_c$ on r is given in [33].

The reasons for this sampled randomness in the slope coefficients are varied. The fractal structure of some attractors is one problem, [10]. Lacunarity [35] most likely also plays a role. But the more important reasons, especially for simple maps, are the very low sample sizes and high embedding dimensions. From sample realisation to sample realisation different areas of the attractor are sampled at frequencies not equal to those predicted by the underlying invariant distribution. Further, the relative sparseness of the data at high embedding dimensions increases the sensitivity of the slope coefficient estimates to relatively small variations in the sampled spatial distribution of the data. This source of randomness is in addition to that summarised by the error term, u; the prime source of which is errors of observation.

The r values that we chose to generate the original trial plots of $\ln C$ against $\ln r$ varied linearly from a low of 0.002 to a high of 0.4 in steps of 0.002. Out of this total region a small sub-region was selected to be used for the actual calculations in the sampling experiment. The criteria were to have a region that would be suitable for almost all samples at a wide variety of embedding dimensions, that would avoid the curvature induced by saturation, and that would avoid the difficulties met with small data sets for small values of r. The desired goal was to have a single region for the values of r that entered the regressions, but the variation in the plots, especially with the random data, meant that a single r region could not be maintained, although we kept as close to this ideal as we could. Because of these constraints the approximate number of points used in each regression was of the order of about 50–100 for most values of m in almost all examples. But in a few cases at the highest values for m and at the smallest size, the number of observations fell to about 15–20.

5. The empirical results

The outcome of the experiments are summarised in terms of the summary statistics listed in appendix 1, some discussion of the standardised third and fourth moments, and the formulation of two empirically obtained equations, equations (5.1) and (5.2).

Equations (5.1) and (5.2), given below, were obtained by empirical fits to the experimental data. However, the choice of functional form was guided by two criteria. First, the asymptotic properties should be in accord with known results. Second, we tried to propose a single class of function to fit all the data sets in order to facilitate comparisons across different examples; i.e., to the extent that the

Table 1. Regression results to fitting

$$\ln(k + d_c) = \gamma_1 + \gamma_2 N^{\gamma_3} + \gamma_4 N^{\gamma_5}[\exp(\gamma_6/m^{\gamma_7}) - 1]$$

to several examples†.

Examples (preset coefficients)	γ	Coefficient estimates	Estimated student-t ratios	R^2 value/asymptotic $\bar{d}_c$ estimate
Henon	1	0.802	105.3	$R^2 = 0.9992$
($k = 1$,	2	0.235	4.8	Asymptotic $\bar{d}_c = 1.23$
$\gamma_3 = -1$)	4	0.264	5.3	Per cent error = 1.7
	5	−0.930	−36.0	89 degrees of freedom (DOF)
	6	−3.012	−17.4	
	7	2.104	5.3	
5-torus	1	1.868	394.7	$R^2 = 0.99995$
($k = 1$,	2	0.016	18.2	Asymptotic $\bar{d}_c = 5.47$
$\gamma_3 = -1$)	4	1.187	93.6	Per cent error = 4.5 (for a
	5	−0.023	−7.0	hypothesised d_c value of 5.2)
	6	−0.156	−38.8	89 DOF
	7	1.864	56.2	
Normal	2	4.246	4.8	$R^2 = 0.9993$
($k = 0$,	3	0.016	4.5	90 DOF
$\gamma_1 = 0$,	4	5.290	2.1	
$\gamma_5 = 0$)	6	−0.540	−7.7	
	7	0.465	2.3	
Gamma	2	3.266	3.7	$R^2 = 0.998$
($k = 0$,	3	0.031	3.5	90 DOF
$\gamma_1 = 0$,	4	5.418	0.7	
$\gamma_5 = 0$)	6	−0.305	−1.1	
	7	0.520	1.3	
Uniform	2	5.994	4.6	$R^2 = 0.9998$
($k = 0$,	3	0.0004	0.6	81 DOF
$\gamma_1 = 0$,	4	6.606	3.6	
$\gamma_5 = 0$)	6	−0.145	−1.6	
	7	0.391	3.4	
Randomised	2	4.968	2.6	$R^2 = 0.9993$
Henon	3	0.006	2.3	83 DOF
($k = 0$,	4	7.565	1.0	
$\gamma_1 = 0$,	6	−0.478	−2.3	
$\gamma_5 = 0$)	7	0.367	1.3	
Randomised	2	3.809	15.6	$R^2 = 0.9998$
5-Torus	3	0.007	8.6	83 DOF
($k = 0$,	4	4.767	5.5	
$\gamma_1 = 0$,	6	−0.435	−15.0	
$\gamma_5 = 0$)	7	0.566	6.4	

† N is in thousands; m is in tens; R^2 = (regression sum of squares about mean)/(the error sum of squares). See text for an explanation; 5.2 is the 'observable' dimension.

chosen class of function provides a useful approximation to the bias effects of the estimation procedure, cross-example comparisons can easily be made by comparing coefficient values. To the extent that for relatively simple attractors, the proposed equations provide a useful approximation, some insight is gained into the likely bias problems to be encountered with other attractors and different random variables.

In the discussion to follow, $E\{\hat{d}_c \mid M, N, m, r\}$, the conditional expectation of the random variable $\hat{d}_c$ on the example, M, sample size N, embedding dimension m, and scale r, is written as $\bar{d}_c \cdot m_1'(\hat{d}_c)$, defined in equation (4.1), is a consistent estimator of $\bar{d}_c$.

Equations (5.1) and (5.2) are respectively the conditional means of $\hat{d}_c$ and of $\sqrt{m_2(\hat{d}_c)}$, defined in equation (4.1), with respect to example type, sample size and embedding dimension. The dependence of the means of $\hat{d}_c$ and of $\sqrt{m_2(\hat{d}_c)}$ on variations in the scaling region could not be evaluated from these experiments.

5.1. Dependence of $\hat{d}_c$ on N and m

A general empirical example that seems to provide a reasonable fit for the dependence of mean d_c estimates on embedding dimension and sample size is

$$\ln(k + \bar{d}_c) = \gamma_1 + \gamma_2 N^{\gamma_3} + \gamma_4 N^{\gamma_5}[\exp(\gamma_6/m^{\gamma_7}) - 1.0] \tag{5.1}$$

where $\bar{d}_c$ is the mean of $\hat{d}_c$. The empirical results are summarised in table 1 and in the appendix.

This equation is not as parameter extensive as it would appear. The parameter k equals 0 or 1; its purpose is to facilitate the numerical approximation for very low-dimensional attractors and to allow γ_1, the asymptotic limit, to be strictly positive for attractors. k could easily have been set to 0 for all data sets in this study with little serious impact on the overall results. The right-hand side of equation (5.1) can be considered in two parts:

$$\gamma_1 + \gamma_2 N^{\gamma_3} \tag{5.1'a}$$

$$\gamma_4 N^{\gamma_5}[\exp(\gamma_6/m^{\gamma_7}) - 1.0]. \tag{5.1'b}$$

The expression $\gamma_1 + \gamma_2 N^{\gamma_3}$ indicates the main effect of small sample size on the expected value of the estimator $\hat{d}_c$. For random variables that scale monotonically in m, $\gamma_1 = 0$ and both γ_2 and γ_3 are positive; i.e., the limit towards which the term (5.1'b) is approaching is given by $\gamma_2 N^{\gamma_3}$; the larger N, the larger the limiting value for $\bar{d}_c$ expressed as a function of m.

If, however, one has an attractor, then $\gamma_1 > 0$ and $\gamma_3 < 0$, so that as $N \to \infty$ the small sample bias provided by the term $\gamma_2 N^{\gamma_3}$ goes to zero; one might guess that γ_3 is -1.0 for attractors with low dimension. The asymptote as both N and $m \to \infty$, but such that $\lim_{N\to\infty} m/N \to 0$, is given by γ_1:

$$\lim_{\substack{N\to\infty \\ m\to\infty}} \ln(1 + \bar{d}_c) = \gamma_1 \qquad \text{or} \qquad \lim_{\substack{N\to\infty \\ m\to\infty}} \bar{d}_c = e^{\gamma_1} - 1.$$

The second part of the expression shown in (5.1'b) models the bias effect due to the embedding dimension. For both random and attractor generated data, one expects γ_6 to be negative, but γ_7 to be positive so that the approach to the limit is a sigmoid shape. The factor $\gamma_4 N^{\gamma_5}$ modifies the m bias effect. γ_5 seems to be negative for attractors and zero for random variables. Thus, for low-dimensional attractors,

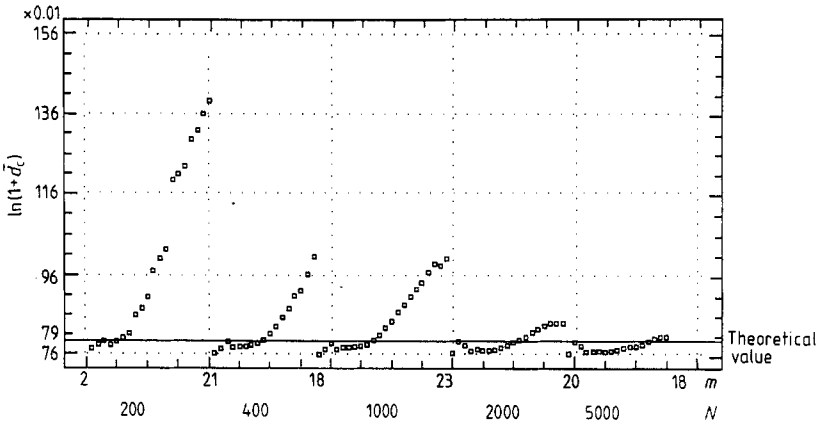

Figure 1. Plot of $\ln(1 + \bar{d}_c)$ against N and m for the Henon map, equation (3.1). This set of five plots shows the graph of $\ln(1 + \bar{d}_c)$, the mean $\bar{d}_c$ estimate, against sample size N and embedding dimension m. The correct value for $\ln(1 + \bar{d}_c)$ is 0.79. The substantial bias that occurs with even modest levels of m is striking. In this case, a sample size of 5000 is beginning to yield reasonable estimates. These data points are fitted by equation (5.1).

the effect of a downward bias on $\bar{d}_c$ due to small values of m declines as N increases. γ_4 depends on the units of measurement chosen for m and the relative weight of the m effect to sample size, N. See figures 1, 3, and 4; the bias effect is striking.

The functional class, defined up to parameter values in equation (5.1), is sufficiently general so as to be able to describe the bias scaling effects of all the alternative examples considered in this paper and would seem to be sufficiently general as to provide at least useful guidance for other, and perhaps less simple, attractors. Empirical use of this function with small-sample data from economics seems to support this conjecture. It is to be hoped that while the coefficient values of equation (5.1) are clearly example specific, the form of the equation is invariant to a wide class of attractors and a wide class of distribution functions.

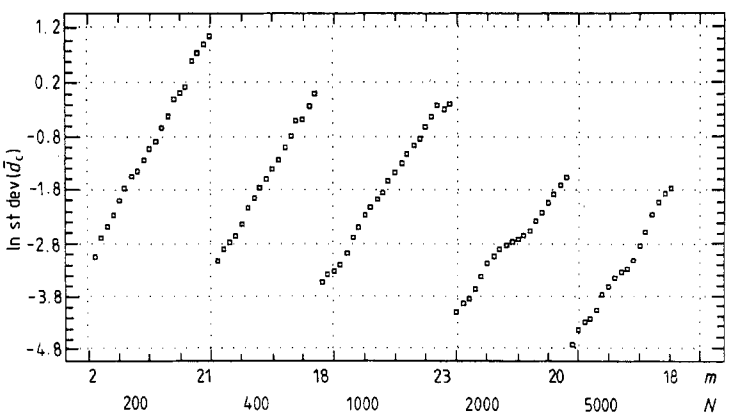

Figure 2. Plot of ln (standard deviation of d_c) against N and m for the Henon map, equation (3.1). This set of five plots shows the effect of N and m on the actual standard deviation of the estimate $\hat{d}_c$. Small increases in m more than offset substantial increases in N. These data points are fitted by equation (5.2).

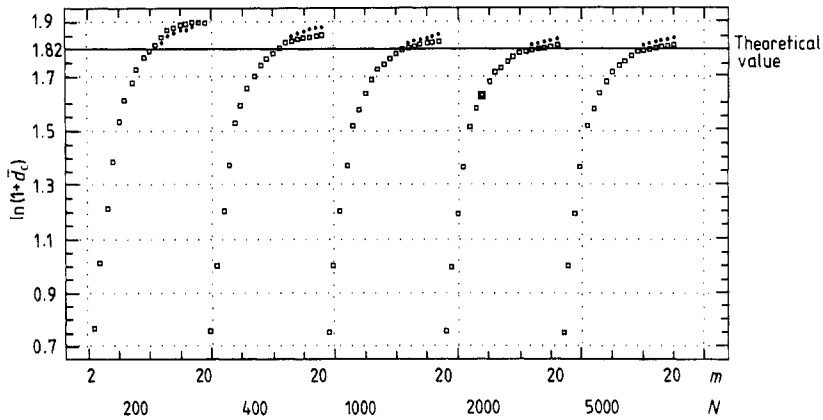

Figure 3. Regression of $\ln(1 + \bar{d}_c)$ against N and m for (five-dimensional) tori, equation (3.2). This set of five plots shows the relationship between $\ln(1 + \bar{d}_c)$ and N, m. The value of $\bar{d}_c$ for a 5-torus in the canonical form shown in equation (3.2) is approximately 5.2 (Mayer-Kress, personal communication), so that $\ln(1 + \bar{d}_c) \doteq 1.82$. There is substantial downward bias for low m leading to modest upward bias for high m. Reasonably useful results seem to be attainable at sample size as low as 1000, but only by approximating the asymptote involved in equation (5.1).

The full circles represent fitted values and the open squares represent observed values.

For the two attractors, the asymptotic estimates of d_c are remarkably close, given that the maximum sample size was only 5000.

Our results confirm the conjecture that different probability distributions scale with embedding dimension at different rates. For each sample size and hence overall sample sizes the average estimated dimensions at each embedding dimension are largest for the uniform distribution, least for the gamma, and intermediate for the normal distribution. The observed decrease across distributions in the rate of increase in dimension as a function of embedding dimension is matched by the

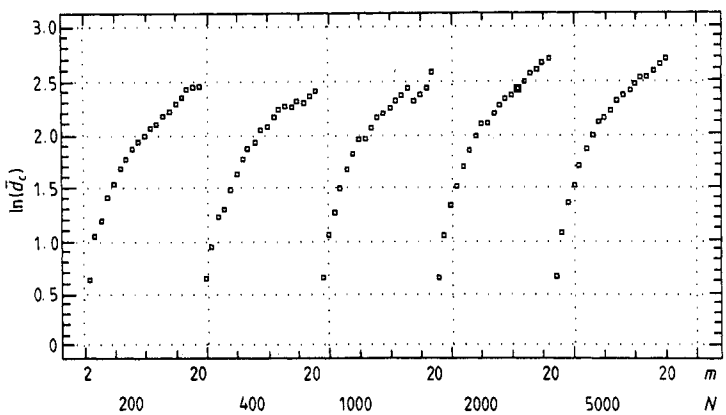

Figure 4. Plot of $\ln(\bar{d}_c)$ against N and m for normal (0, 1) distribution. This set of five plots shows the relationship between $\ln(\bar{d}_c)$ and N, m, when the underlying data are generated by a sequence of independent $N(0, 1)$ random variables. Ln (20), the maximum $m = 2.995$, so the substantial downward bias is clear.

increase in kurtosis, as measured by the standardised fourth moment (fourth moment divided by the second moment squared); the standardised fourth moments are 1.8, 3.0, and 9.0 for the uniform, normal, and gamma distributions respectively.

An alternative way of expressing the matter is in terms of the entropy of the distributions, except that an allowance must be made for the different ranges for the support of the distributions. If we compare entropy measures for the three distributions after rescaling so that the effective support for each distribution is over $[0, 1]$, the entropy measures are 0, -2.1 and -38.0 respectively for the uniform, normal, and gamma distributions. Thus, estimated dimension for random variables seems to be an increasing function of the entropy of the density function, when entropy is evaluated over a similar range of support for the distribution. The rate of increase in estimated dimension to increases in embedding dimension is greatest for the uniform distribution, less for the normal, and least for the gamma, distribution.

5.2. Relationship of standard deviation to N and m

With respect to the relationship between the standard deviation and sample size and embedding dimension, an attempt was made to propose a class of functions that would provide, up to estimable parameter values, a reasonably general example for a wide variety of simple attractors and different distributions of noise.

The basic empirical relationship between the estimated standard deviation $\sqrt{m_2(\hat{d}_c)}$, hereafter written as Std, sample size N and embedding dimension m is

$$\ln(\text{Std}) = \alpha_1 + \alpha_2 \ln N + \alpha_3 \ln m + \alpha_4 m/N. \tag{5.2}$$

Essentially, this equation indicates that

$$\text{Std} \propto N^{\alpha_2} m^{\alpha_3}.$$

The last term in (m/N) is a small-sample-size correction to the effect of m on the standard deviation. We expected that Std would scale for most of the examples as $O(1/\sqrt{N})$; this was confirmed in the empirical results. The parameter values of this equation are example specific; i.e. the values of the parameters will differ with the type of attractor or type of distribution function for random variables. The regression results for the basic example are shown in table 2; and figure 2 is instructive.

As one might suspect *a priori,* the standard deviation increases with embedding dimension and decreases with sample size. The equation enables one to obtain a rough estimate of the trade-off between sample size and embedding dimension to maintain a fixed standard deviation. One aspect worthy of further attention is that the effect of sample size on standard deviation, the value of α_2, is roughly twice as strong for the randomised Henon and 5-torus than for any other example, except for the 5-torus itself. For the remaining examples, standard deviation declines according to the approximate square root of sample size. The increase in standard deviation due to m varies from a low of the square root of m to an enormous power of 16. While the larger coefficients cannot really be believed, notwithstanding the observed student-t ratios (the ratio of the coefficient estimate to its estimated standard error), the order of magnitude certainly reflects the relative difficulty in trying to determine error bars for estimates of dimension. The uniform distribution was particularly troublesome in this respect.

Table 2. Variation in Std ($\hat{d}_c$) as a function of N and m:

$$\ln(\text{Std}) = \alpha_1 + \alpha_2 \ln N + \alpha_3 \ln m + \alpha_4 m/N\dagger.$$

Example	Coefficient index	Coefficient estimate	Student-t ratio
Henon			
($\bar{R}^2 = 0.94$)	1	−2.272	−36.892
	2	−0.533	−10.164
	3	1.295	20.131
	4	0.117	4.737
5-torus			
($\bar{R}^2 = 0.99$)	1	−3.661	−160.796
	2	−0.608	−33.728
	3	0.464	19.997
	4	0.139	14.926
Normal			
($\bar{R}^2 = 0.97$)	1	−1.322	−26.397
	2	−0.731	−18.431
	3	1.891	36.973
Uniform	1	−0.659	−6.658
($\bar{R}^2 = 0.96$)	2	−0.589	−7.984
	3	2.33	26.583
	4	0.18	3.636
Gamma	1	−0.714	−21.838
($\bar{R}^2 = 0.98$)	2	−0.53	−20.48
	3	1.713	51.321
	4	−0.071	−5.316
Randomised Henon	1	−0.065	−1.013
($R^2 = 0.96$)	2	−0.799	−15.708
	3	2.558	38.983
	4	−0.117	−4.453
Randomised torus	1	−0.994	−15.58
($\bar{R}^2 = 0.93$)	2	−0.683	−13.532
	3	1.616	24.808
		−0.076	−2.897

† $\bar{R}^2$ = regression sum of squares adjusted for degrees of freedom. N is measured in thousands; m is measured in tens.

Table 3. Estimated relationship between S_0^2 and $m_2(\hat{d}_c)$†.

Example	Slope coefficient	$(\text{Slope})^{-1}$	Student-t ratio for slope	R^2
Henon	0.016	62.5	76.6	0.99
Normal	0.045	22.2	33.8	0.96
Gamma	0.006	166.7	14.8	0.84
Uniform	0.216	4.6	59.7	0.99
Randomised Henon	0.051	19.6	32.8	0.96
Randomised tori	0.027	37.0	17.5	0.88

† Regression: $S_0^2 = \delta_0 + \delta_1 m_2(\hat{d}_c)$.

Table 3 summarises the approximately linear relationship between the ordinary least squares estimate of variance, S_0^2, and $m_2(\hat{d}_c)$; the constant term was in all cases insignificantly different from zero. Insofar as S_0^2 is an estimate of $m_2(\hat{d}_c)$, the slope coefficient in a linear regression of S_0^2 on $m_2(\hat{d}_c)$ indicates that S_0^2 is proportional to $m_2(\hat{d}_c)$. But, as has long been suspected, the actual variances are much larger than the OLS estimates would indicate. For example, the actual variance of the d_c estimates for the Henon map is approximately 63 times greater than would be indicated by the OLS estimate, S_0^2; alternatively expressed, the standard deviation as simulated in these experiments is approximately eight times as large as is indicated by the usual OLS estimate, S_0. The least difference between the simulated variance and the usual estimate occurs with the uniform distribution where the actual variance is only about four times the OLS estimate.

5.3. The standardised third and fourth moments

The standaridised third moment $m_3^+(\hat{d}_c)$ seems to demonstrate no systematic variation as a function of N or m. For most examples its values are less than a fraction of a standard deviation, so that a reasonable hypothesis is that $m_3^+(d_c)$ is uniformly zero for all examples; i.e. the distribution of d_c estimates is symmetric about its biased mean value.

There is only very weak evidence that there is any systematic relationship between $m_4^+(\hat{d}_c)$ and N on m. The overall mean values for $m_4^+(\hat{d}_c)$ for each example differ only very insignificantly from 3.0, the value taken by the normal distribution itself. It would appear that there is some increase in $m_4^+(\hat{d}_c)$ for very small sample sizes, but that for moderate levels, say above 2000, $m_4^+(\hat{d}_c)$ is usefully regarded as being 3.0.

Consequently, for moderate sample sizes, the assumption of normality for the distribution of d_c estimates about their biased means is a useful approximation.

6. Conclusions and implications for further research

From the evidence above with respect only to the examples examined, it is reasonable to assume that for modest sample sizes (1000 to 2000) and above, the assumption of normality for the standardised distribution of d_c estimates is a useful approximation. More precisely, symmetry of the standardised distribution is a robust result across all examples. Leptokursis (extremely peaked distribution) is a problem only with very small sample sizes.

On the all-important issue of the bias of the estimates, a first conclusion is that the bias is very sensitive to both sample size and embedding dimension. For all examples d_c estimates are an increasing function of embedding dimension for given sample size, although for random examples the bias is always negative and is increasing relatively as m increases. For maps, the bias is large even for intermediate embedding dimension. The bias is negative for very small m, but is positive for moderate sizes of m and only declines for those large values N which are sufficient to allow the correction term in m to approach zero. As shown in figure 1, for example, the biased estimate can be twice to three times the size of the true value for small sample sizes and moderate embedding dimension. Even for a sample as large as 2000 the estimate can be 50% greater than the true value at high embedding

dimensions. Consequently some care is needed in interpreting estimates of d_c used with small N and moderate to large sizes of m.

The usual OLS estimates of the standard deviations of d_c estimates are only small fractions of the actual standard deviations, even though they are positively and significantly correlated with the actual standard deviations. The true error bars are much larger than the usual procedure indicates. However, while an increase in sample size reduces standard deviations roughly according to the square root of sample size, standard deviation increases on average in embedding dimension at a much faster rate; i.e. modest increases in m more than offset substantial increases in N, as is demonstrated in table 2 and illustrated in figure 2.

The above empirical results also support the idea that the slope coefficient in the regression of $\ln C$ on $\ln r$ is a random coefficient, the expected value of which is approximated by equation (5.1). Consequently, some idea of the actual variance of the d_c estimate can be obtained from a single experimental trial only if the researcher has some idea of the relationship between S_0^2 and the actual variance as documented, for example, in table 2.

In order to use equation (5.2) to estimate the relationship between the actual variances and N and m requires multiple estimates of d_c at each sample size. Consequently, the only current way to get an accurate assessment of the variance of estimate of d_c is to use multiple samples at each sample size, as was done in the paper. The multiple samples are used to estimate equation (5.2). From this estimated equation the variance of the estimate for $\hat{d}_c$ can be estimated by substituting into (5.2) the values of N and m that were used to obtain the given estimate of d_c. A suggested minimal sample size for estimating moderately useful error bars for low-dimensional attractors by this method is about 35 000 with a maximum sub-sample size of 5000.

A more encouraging conclusion is that equation (5.1) potentially provides the basis for more efficient d_c estimates, i.e. lower mean squared error, for any given sample size. Given that one is estimating a finite-dimensional attractor, then the estimation of γ_1 provides an estimate of d_c that allows explicitly for the impact of bias due to embedding dimension. With the assumption of near normality already justified, useful but realistic error bars could be produced with relatively modest amounts of data, by allowing for the dependence of standard deviation on N and m as is indicated by equation (5.2) and the results in table 2. This is achieved by sub-sampling from the total available sample and estimating d_c with the sub-samples. The resulting statistics can be used to estimate equation (5.1). For attractors the estimate of γ_1 provides the required asymptotic estimate of the dimension. Reasonable estimates should be obtainable for low-dimensional attractors at a sample size of about 7000 and above.

For example, equation 5.1 was estimated for the 5-torus based on sample sizes from 200 to 5000. The results were used to predict the effect of embedding dimension on d_c estimates when the sample size is 13 000; a number dictated by the constraints of our computer system. The fitted line between predicted d_c values at each m and the estimated values has a zero intercept, a slope of 0.999 and an R^2 fit of 0.997. At $N = 13\,000$, the asymptotic estimate for d_c is 5.15.

The final issue of great concern to researchers in many disciplines, especially outside physics, is that of distinguishing chaotic attractors from random phenomena. Of more interest to physicists is the problem of detecting when noise scales begin to intrude on attractor d_c estimates. The results in this paper are of some interest in this regard.

The depressing result is that for some ranges of values for m, d_c estimates for attractors are biased upwards, but random dimension estimates for random variables are biased downwards. Worse is the fact that for fixed, albeit relatively small, values of N, increases in m lead to increases in d_c estimates for both attractors and for random variables. Such results are not encouraging for distinguishing between attractors and non-attractors.

However, equation (5.1) potentially provides a parametric test of the hypothesis as to whether a set of data is from an attractor or from a model generating random variables. Specifically, with random data $\gamma_1 \equiv 0$ and $\gamma_3 > 0$, but with attractors $\gamma_1 > 0$ and $\gamma_3 \leqslant 0$. Noise scale levels are detected by the shift in values for γ_1, γ_3 as the scale for r is moved through the attractor region into the noise region.

A second approach is to compare the d_c estimates obtained through equation (5.1) on a supposed attractor and the d_c estimates for the same data with randomised order. The former data series obey $\gamma_1 > 0$ and $\gamma_3 \leqslant 0$, the latter $\gamma_1 \equiv 0$ and $\gamma_3 > 0$. In short, randomised attractor generated series behave like random variables, but the randomisation of the order of independent random variables makes no difference at all. Consequently, a shift in the estimates of γ_1 and γ_3 provide an indication that the original data were from an attractor. See figure 5 and compare it with figures 1 and 4.

Finally, our results indicate that for random variables the rate at which the mean of $\hat{d}_c$ estimates increases with increasing embedding dimension seems to be a monotonic function of the distribution's entropy; the higher the entropy, the faster the estimated d_c increases with embedding dimension; but entropy comparisons must be made on the basis of equivalent ranges of support for the distribution, if, as is usual, the observed data are rescaled to a chosen finite range of values.

The results cited in this paper are restricted to the limited examples that we have examined. However, the intent underlying our approach in formulating and estimating equations (5.1) and (5.2) was to provide a general class of functions that would be useful in a wide variety of situations to model the bias and size of error bars to be encountered in experimental situations.

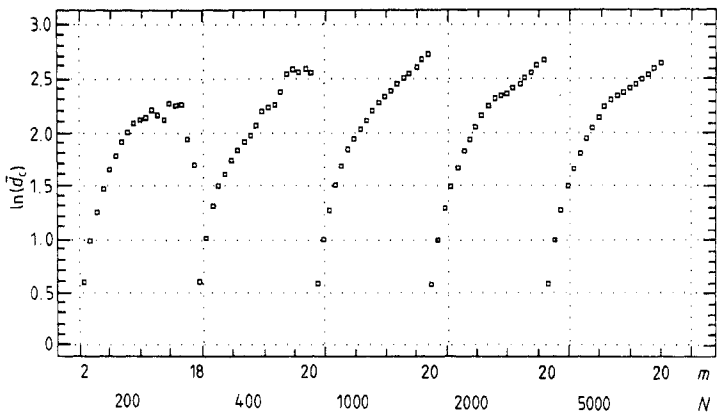

Figure 5. Plot of $\ln(\bar{d}_c)$ on N and m for the randomised Henon map, equation (3.1). This set of five plots shows the relationship between $\bar{d}_c$ and N, m when the underlying data were generated by a Henon map (equation (3.1)) whose time sequence of points were then randomly permuted. Compare this figure with figure 4 for the normal distribution and figure 1 for the non-permuted Henon data points.

Our results should be regarded as very conservative, although tentative. They are conservative in that more complex attractors with fractal structure are likely to exhibit these estimation difficulties to a far greater extent.

Now that the problems of estimation bias and the gross under-calculation of error bars using the conventional procedures have been recognised, the forewarned researcher can take steps to allow for these difficulties. Nevertheless, considerable efforts are still needed to evaluate the statistical properties of d_c estimators in more complex situations.

Acknowledgments

We are grateful for financial and technical support from the C V Starr Center for Applied Economics and to a grant from NYU from a Fund to Develop University/Industry Linkages. The readers, as well as we, owe the following people thanks for their contribution to whatever clarity this paper has achieved; John Lowenstein, Alan Wolf, and especially Bruce Stewart, not to mention several anonymous referees.

References

[1] Badii R and Politi A 1985 Statistical description of chaotic attractors: The dimension function *J. Stat. Phys.* **40** 725

[2] Halsey T C, Jensen M H, Kadanoff L P, Procaccia I and Shraiman B I 1986 Fractal measures and their singularities: The characterization of strange sets *Phys. Rev. A* **33** 1141

[3] Grebogi C, Ott E, Pelikan S and Yorke J A 1984 Strange attractors that are not chaotic *Physica* **13D** 261

[4] Umberger D K, Mayer-Kress G and Jen E 1986 Hausdorff dimensions for sets with broken scaling symmetry *Dimensions and Entropies in Chaotic Systems Quantification of Complex Behavior* ed G Mayer-Kress (Berlin: Springer) pp 42–53

[5] Farmer J D 1986 Scaling in fat fractals *Dimensions and Entropies in Chaotic Systems, Quantification of Complex Behavior* ed G Mayer-Kress (Berlin: Springer) pp 54–60

[6] Termonia Y and Alexandrowicz Z 1983 Fractal dimension of strange attractors from radius versus size of arbitrary clusters *Phys. Rev. Lett.* **51** 1265

[7] Osborne A R, Provenzale A and Bergamasco L 1986 Multivariate scaling portraits of the Lorenz attractor *Phys. Rev. Lett.* submitted

[8] Somorjai R L 1986 Methods for estimating the intrinsic dimensionality of high-dimensional point sets *Dimensions and Entropies in Chaotic Systems, Quantification of Complex Behavior* ed G Mayer-Kress (Berlin: Springer) p 137

[9] Grassberger P 1985 Generalizations of the Hausdorff dimension of fractal measures *Phys. Lett.* **107A** 101

[10] Guckenheimer J 1984 Dimension estimates for attractors *Contemp. Math.* **28** 357

[11] Grassberger P and Procaccia I 1983 Characterization of strange attractors *Phys. Rev. Lett.* **50** 346

[12] Grassberger P and Procaccia I 1983 Measuring the strangeness of strange attractors *Physica* **9D** 189

[13] Ben-Mizrachi A, Procaccia I and Grassberger P 1984 Characterization of experimental (noisy) strange attractors *Phys. Rev. A* **29** 975

[14] Farmer J D, Ott E and Yorke J A 1983 The dimension of chaotic attractors *Physica* **7D** 153

[15] Abraham N B, Albano A M, Das B, Mello T, Tarroja M F H Tufillaro N and Gioggia R S 1986 Definitions of chaos and measuring its characteristics ed Chrostowski J and Abraham N B *SPIE Proc.* **667** to appear

[16] Carter P H, Cawley R, Licht A L, Yorke J A and Melnik M S Dimension measurements from cloud radiance *Dimensions and Entropies in Chaotic Systems, Quantification of Complex Behavior* ed G Mayer-Kress (Berlin: Springer) pp 215–221

[17] Babloyantz A 1986 Evidence of chaotic dynamics of brain activity during the sleep cycle *Dimensions and Entropies in Chaotic Systems Quantification of Complex Behavior* ed G Mayer-Kress (Berlin: Springer) p 241
[18] Lovejoy S and Schertzer D 1986 Scale invariance, symmetries, fractals, and stochastic simulations of atmospheric phenomena *Bull. Am. Met. Soc.* **67**
[19] Barnett W and Chen P 1988 The aggregation-theoretic monetary aggregates are chaotic and have strange attractors *Dynamic Economic Modelling, Proc. 3rd Int. Symp. on Economic Theory and Econometrics* ed W Barrett, E Berndt and H White (Cambridge, MA: Cambridge University Press)
[20] Brock W A and Sayers C L 1986 Is the business cycle characterized by deterministic chaos? *Journal of Monetary Economics* **22** 71–90
[21] Sayers C L 1986 Work stoppages: Exploring the nonlinear dynamics *Preprint* University of Wisconsin-Madison
[22] Scheinkman J A and Le Baron B 1986 Nonlinear dynamics and stock returns *Journal of Business* in press
[23] Brandstater A, Swinney H L and Chapman G T 1986 Characterizing turbulent channel flow *Dimensions and Entropies in Chaotic Systems* ed G Mayer-Kress (Berlin: Springer) p 150
[24] Caputo J G, Malraison B and Atten P 1986 Determination of attractor dimension and entropy for various flows: An experimentalist's viewpoint *Dimensions and Entropies in Chaotic Systems Quantification of Complex Behavior* ed G Mayer-Kress (Berlin: Springer) p 180
[25] Greenside H S, Wolf A, Swift J and Pignataro T 1982 Impracticality of a box-counting algorithm for calculating the dimensionality of strange attractors *Phys. Rev.* A **25** 3453
[26] Abraham N B, Albano A M, Das B, De Guzman G, Yong S, Gioggia R S, Puccioni G P and Tredicce J R 1986 Calculating the dimension of attractors from small data sets *Phys. Lett.* **114A** 217
[27] Grassberger P 1985 Generalizations of the Hausdorff dimension of fractal measures *Phys. Lett.* **107A** 101–5
[28] Grassberger P 1988 Finite sample corrections to entropy and dimension estimates *Phys. Lett.* A **128** 369–73
[29] Brock W A Dechert W D and Scheinkman J A 1986 A test for independence based on the correlation dimension *Working Paper* Department of Economics, University of Wisconsin Madison, Wisconsin
[30] Holzfuss J and Mayer-Kress G 1986 An approach to error-estimation in the application of dimension algorithms, *Dimensions and Entropies in Chaotic Systems, Quantification of Complex Behavior* ed G Mayer-Kress (Berlin: Springer) p 114
[31] Hentschel H C E and Procaccia I 1983 The infinite numbers of dimensions of probabilistic fractals and strange attractors *Physica* **8D** 435
[32] Mayer-Kress G 1987 Application of dimension algorithms to experimental chaos *Directions in Chaos* ed Hao Bai-lin (Singapore: World Scientific) pp. 122–47
[33] Brandstater A and Swinney H L 1986 A strange attractor in weakly turbulent Couette-Taylor flow *Preprint* University of Texas
[34] Grassberger P and Procaccia I 1983 Estimation of the Kolmogorov entropy from a chaotic signal *Phys. Rev.* A **28** 2591
[35] Smith L A, Fournier J D and Spiegel E A 1986 Lacunarity and intermittency in fluid turbulence *Phys. Lett.* **114A** 465

Appendix

Table A1. Summary statistics on $\hat{d}_c$†.

Example		Sample size					
		All N	200	400	1000	2000	5000
Henon map	mean, m_1	1.388	1.811	1.313	1.357	1.222	1.188
	std, $\sqrt{m_2}$	0.394	0.658	0.177	0.203	0.063	0.03
	minimum	1.14	1.17	1.145	1.14	1.15	1.15
	maximum	3.035	3.035	1.734	1.73	1.327	1.241
	range	1.895	1.865	0.589	0.59	0.177	0.091
5-torus	mean, m_1	4.213	4.441	4.256	4.156	4.106	4.104
	std, $\sqrt{m_2}$	1.279	1.431	1.32	1.268	1.237	1.237
	minimum	1.214	1.142	1.134	1.127	1.125	1.124
	maximum	5.662	5.662	5.357	5.216	5.107	5.097
	range	4.538	4.52	4.223	4.089	3.982	3.973
Uniform	mean, m_1	8.603	7.337	8.109	8.842	9.114	9.291
[0, 1]	std, $\sqrt{m_2}$	4.181	3.535	3.894	4.461	4.409	4.503
	minimum	1.881	1.923	1.889	1.89	1.903	1.881
	maximum	16.97	12.99	13.81	16.91	16.26	16.97
	range	15.089	11.067	11.911	15.02	14.357	15.089
Normal	mean, m_1	7.912	7.22	7.046	7.852	8.663	8.772
(0, 1)	std, $\sqrt{m_2}$	3.382	2.997	2.883	3.155	3.867	3.856
	minimum	1.883	1.891	1.884	1.899	1.883	1.904
	maximum	14.63	11.55	11.03	13.02	14.59	14.63
	range	12.747	9.659	9.146	11.121	12.707	12.726
Gamma	mean, m_1	6.329	4.82	5.904	6.727	6.953	7.239
($\alpha = 1$,	std, $\sqrt{m_2}$	2.548	1.362	2.2	2.692	2.8	2.909
$\beta = 1$)	minimum	1.63	1.68	1.63	1.714	1.654	1.723
	maximum	11.06	6.355	8.537	10.63	10.31	11.06
	range	9.43	4.675	6.907	8.916	8.656	9.337
Randomised	mean, m_1	8.23	6.791	8.186	8.822	8.704	8.646
Henon map	std, $\sqrt{m_2}$	3.648	2.412	3.787	4.104	3.872	3.809
	minimum	1.8	1.805	1.822	1.8	1.802	1.802
	maximum	15.49	9.64	13.36	15.49	14.78	14.28
	range	13.69	7.835	11.538	13.69	12.978	12.478
Randomised	mean, m_1	8.355	7.561	7.928	8.577	8.912	8.794
5-torus	std, $\sqrt{m_2}$	3.384	2.783	3.11	3.513	3.823	3.739
	minimum	1.916	1.945	1.919	1.916	1.918	1.926
	maximum	14.2	10.87	12.06	12.82	14.2	14.09
	range	12.284	8.945	10.141	10.904	12.282	12.164

† $m_1' = \langle \hat{d}_c \rangle$, $m_2 \doteq \langle (\hat{d}_c - m_1')^2 \rangle$.

Table A2. Summary statistics on S_0†.

		Sample size					
Example		All N	200	400	1000	2000	5000
Henon map	mean, m_1	0.304	0.776	0.301	0.27	0.072	0.053
	std, $\sqrt{m_2}$	0.502	0.872	0.277	0.253	0.051	0.049
	minimum	0.008	0.047	0.043	0.028	0.016	0.008
	maximum	2.885	2.885	0.972	0.783	0.192	0.166
	range	2.876	2.84	0.93	0.755	0.176	0.158
5-torus	mean, m_1	0.061	0.167	0.081	0.028	0.018	0.01
	std, $\sqrt{m_2}$	0.073	0.088	0.045	0.01	0.005	0.003
	minimum	0.005	0.042	0.028	0.011	0.008	0.005
	maximum	0.319	0.319	0.169	0.044	0.023	0.013
	range	0.314	0.276	0.141	0.033	0.016	0.009
Uniform	mean, m_1	1.878	3.839	3.203	1.583	1.091	0.383
[0, 1]	std, $\sqrt{m_2}$	3.788	6.13	5.297	6.645	1.768	0.426
	minimum	0.004	0.074	0.044	0.019	0.01	0.004
	maximum	22.419	22.419	17.152	10.01	6.762	1.464
	range	22.414	22.345	17.108	9.991	6.752	1.46
Normal	mean, m_1	0.478	1.048	0.597	0.321	0.305	0.118
(0, 1)	std, $\sqrt{m_2}$	0.582	0.922	0.445	0.26	0.302	0.101
	minimum	0.002	0.052	0.037	0.01	0.009	0.002
	maximum	3.093	3.093	1.464	1.024	0.983	0.308
	range	3.091	3.041	1.427	1.013	0.974	0.306
Gamma	mean, m_1	0.611	0.981	0.721	0.626	0.43	0.298
($\alpha = 1$,	std, $\sqrt{m_2}$	0.493	0.608	0.464	0.501	0.3	0.218
$\beta = 1$)	minimum	0.011	0.07	0.057	0.031	0.028	0.011
	maximum	2.074	2.07	1.361	1.666	0.9	0.672
	range	2.062	2.0	1.304	1.635	0.872	0.66
Randomised	mean, m_1	1.662	2.748	2.434	1.722	0.973	0.432
Henon map	std, $\sqrt{m_2}$	2.049	2.6	2.773	1.654	0.786	0.408
	minimum	0.004	0.063	0.047	0.02	0.011	0.004
	maximum	8.951	7.906	8.951	5.022	2.768	1.416
	range	8.947	7.844	8.904	5.0	2.757	1.412
Randomised	mean, m_1	0.487	0.957	0.566	0.426	0.351	0.136
5-torus	std, $\sqrt{m_2}$	0.452	0.651	0.373	0.239	0.216	0.09
	minimum	0.004	0.051	0.034	0.026	0.016	0.004
	maximum	2.187	2.187	1.246	0.82	0.767	0.296
	range	2.183	2.136	1.212	0.795	0.752	0.292

† $m_1' = \langle S_0 \rangle$, $m_2 \doteq \langle (S_0 - m_1')^2 \rangle$.