# Nonparametric estimation of the correlation exponent

E. Scott Mayfield
*Department of Economics, Boston College, Chestnut Hill, Massachusetts 02167*

Bruce Mizrach
*Department of Finance, The Wharton School, Philadelphia, Pennsylvania 19104*

(Received 7 February 1991)

The correlation exponent is widely utilized in experimental situations as an estimate of an attractor's fractal dimension. However, in the presence of noise, the slope of the correlation integral may increase gradually, biasing dimension estimates. We propose a nonparametric statistical procedure for distinguishing the attractor from the noise process.

PACS number(s): 05.40.+j, 47.25.Ae

## I. INTRODUCTION

Because of its computational feasibility, the correlation exponent [1,2] is widely utilized in experimental situations as an estimate of an attractor's fractal dimension. However, the observer function is often contaminated by noise, causing the researcher to distinguish between the attractor and noise processes. Since noise will not obscure the structure of the attractor at length scales greater than the noise strength, calculation of the correlation integral over a range of embedding spaces allows the attractor to be differentiated from the noise process [3]. The space-filling nature of the noise process segments the correlation integral, producing a kink that must be identified by the researcher. In practice, since noise gradually increases the slope of the correlation integral, there is not a unique kink point, and dimension estimates can become contaminated. As a result, a variety of estimates may be drawn from the same data, or alternatively, experimenter bias can enter, skewing an estimate toward the researchers prior. We propose a nonparametric procedure for distinguishing the attractor from the noise process.

Noise can be modeled as an infinite-dimensional process, while the deterministic process reconstructed by the attractor is finite. The correlation integral maps a spatial problem into pairs, ordered by scale, where the two models can be distinguished by the integral's slope. Standard techniques [4,5] for distinguishing between the means of two candidate models suffer from the same shortcomings as mentioned above. We implement a nonparametric technique that tests for the stability of regression coefficients over the domain of the correlation integral. This inference requires no prior information other than the sequencing of the data.

## II. CUMULATIVE SUM OF SQUARES TEST (REF. [6])

The correlation integral can be represented as segmented linear models: a model for the attractor and the higher-dimensional model for the noise process. In this section, we discuss recursive estimation of the slope of the correlation integral and introduce a nonparametric procedure for distinguishing the underlying models.

We write the model for the correlation integral as

$$\underline{Y}=\begin{bmatrix}\underline{Y}_1\\ \underline{Y}_2\end{bmatrix}=\begin{bmatrix}\underline{X}_1 & 0\\ 0 & \underline{X}_2\end{bmatrix}\begin{bmatrix}\underline{\beta}_1\\ \underline{\beta}_2\end{bmatrix}+\underline{\mu}\ , \tag{1}$$

where $\underline{Y}$ is a $T$ vector of $y$ coordinates partitioned into the $m\times 1$ and $n\times 1$ vectors $\underline{Y}_1$ and $\underline{Y}_2$, the $\underline{X}_i$ are $m\times 2$ and $n\times 2$ matrices of dependent variables, with 1 in the first column, the $\underline{\beta}_i$ are $2\times 1$ vectors of coefficients with slopes $\beta_{11}$ and $\beta_{21}$, and $\underline{\mu}$ is an independently distributed, mean zero disturbance term with variance $\sigma^2$. For a given $m$ and $n$, there are exact test statistics for the hypothesis $H_0$: $\underline{\beta}_1=\underline{\beta}_2$. Our problem is to identify the lengths of the subsamples.

The first step is to construct a series of recursive least squares estimates of the coefficients of the correlation integral, treating the data as if they were generated by a single model. Denote $\underline{Y}_r'=(y_1,y_2,\ldots,y_r)$ and $\underline{X}_r'=(\underline{x}_1,\underline{x}_2,\ldots,\underline{x}_r)$, the first $r$ observations on the dependent and independent variables. The $r$th estimate of the coefficient vector is

$$\hat{\underline{b}}_r=(\underline{X}_r'\underline{X}_r)^{-1}\underline{X}_r'\underline{Y}_r\ . \tag{2}$$

An updated estimate of $\hat{\underline{b}}$ may be obtained recursively from its previous estimate by

$$\hat{\underline{b}}_r=\hat{\underline{b}}_{r-1}+(\underline{X}_r'\underline{X}_r)^{-1}\underline{x}_r(y_r-\underline{x}_r'\hat{\underline{b}}_{r-1})\ . \tag{3}$$

 

The relation (3) is a Kalman filter iteration, where the gain matrix is a time-varying linear function of the one-period-ahead forecast error [7].

The principal advantage of this approach is that parametric assumptions about the densities of the underlying populations are not necessary. The only property of the data we will exploit will be the monotonicity of the squared estimation errors; by converting them to order statistics, we identify nonparametrically $m$ and $n$, the domains of the attractor and the noise process.

Formally, we test the hypothesis $H_0$: $\underline{\beta}_1=\underline{\beta}_2$, using an analysis of the residuals from the recursive estimates proposed by Brown, Durbin, and Evans [6]. Let $v_r$ be the out-of-sample forecast error for the $(r-1)$th estimate, such that

$$v_r=y_r-\underline{x}_r'\hat{\underline{b}}_{r-1}\ . \tag{4}$$

Under $H_0$, $v_r$ has mean zero and variance $\sigma^2 d_r^2$, where $d_r\equiv[1+\underline{x}_r(\underline{X}_{r-1}'\underline{X}_{r-1})^{-1}\underline{x}_r]^{1/2}$. The cumulative standardized prediction error is $w_r\equiv(v_r/d_r)$. The cumulative sum of squares (CSS) is defined as

$$S_r\equiv\sum_{t=3}^{r}(w_t)^2\ . \tag{5}$$

Under the null hypothesis, $S_r$ is equal to the residual sum of squares, yielding a recursive relation for the CSS,

$$S_r=S_{r-1}+w_r^2\ . \tag{6}$$

From (6), the CSS increases monotonically with $r$. After scaling $S_r$ by $S_{m+n}$, we have an ordered sample on the (0,1) interval. Under $H_0$, the statistic

$$s_r\equiv\sum_{t=3}^{r}(w_t)^2\Big/\sum_{t=3}^{m+n}(w_t)^2=S_r/S_{m+n} \tag{7}$$

will have the $\beta$ distribution with mean equal to $(t-2)/(m+n-2)$. Tests of the null hypothesis are based on the symmetric statistic $P=(t-2)/(m+n-2)\pm c_0$, where $c_0$ corresponds to a specific critical value based on calculations by Durbin [8].

This test is robust since distribution of the order statistics is independent of the cumulative distribution function of the model population. If the true data generating process is a single linear model, $s_r$ will not deviate significantly from its mean. Alternatively, if the slope in the noise domain is steeper, $\beta_{12}>\beta_{11}$, then there will be a sequence of large forecast errors. A greater portion of the CSS comes from the noise domain, causing $P$ to cross Durbin's lower boundary.

## III. IMPLEMENTATION

Given an observed time series $\{x_1,x_2,\ldots,x_N\}$, the correlation integral [2,9] is defined as

$$\tilde{C}_d(\epsilon)=\lim_{N\to\infty}(1/N^2)\{\text{number of } (j,k)|[(x_j-x_k)^2+\ldots+(x_{j+d-1}-x_{k+d-1})^2]^{1/2}<\epsilon\} \tag{8}$$

for $d=2,3,\ldots,\infty$. As $\epsilon\to 0$, $\tilde{C}_d(\epsilon)\simeq\epsilon^\nu$, where $\nu$, the correlation exponent, is a lower bound estimate of the Hausdorff dimension [1]. Thus, for small $\epsilon$,

$$\ln_2\tilde{C}_d(\epsilon)=\ln_2 k+\nu\ln_2\epsilon\ , \tag{9}$$

where $k$ is a constant. Since noise will not obscure the fractal structure of the attractor at length scales greater than the noise strength and the $\epsilon$ are ordered,

$$\beta_{11}=\frac{\partial\ln_2\tilde{C}_d(\epsilon)}{\partial\ln_2\epsilon}=\nu\ , \tag{10}$$

and $\beta_{21}$ is an estimate of the dimension of the noise process. In addition, as the embedding space becomes large, the change in $\beta_{10}$ provides a lower bound estimate of Kolmogorov entropy [2,9].

Since the correlation integral may become distorted at large values of $\epsilon$, sequential application of the CSS test over the remaining $n+1$ observations allows these distorted segments, as well as the noise scale, to be detected. The correlation exponent is identified by the segment of the correlation integral whose estimated slope becomes invariant to embedding.

## IV. APPLICATION

To demonstrate our procedure, we study the Mackey-Glass delay differential equation [10], parametrized as follows:

$$\dot{x}(t)=\frac{ax(t-\tau)}{1+[x(t-\tau)]^{10}}-bx(t),\quad a=0.2,\ b=0.1\ . \tag{11}$$

Following the appendix to [1], we approximate Eq. (11) by a set of 2400 difference equations and generate time series of the form $\{x(t+i\tau), i=1,\ldots,25.000\}$. We estimate correlation exponents for series with delay times

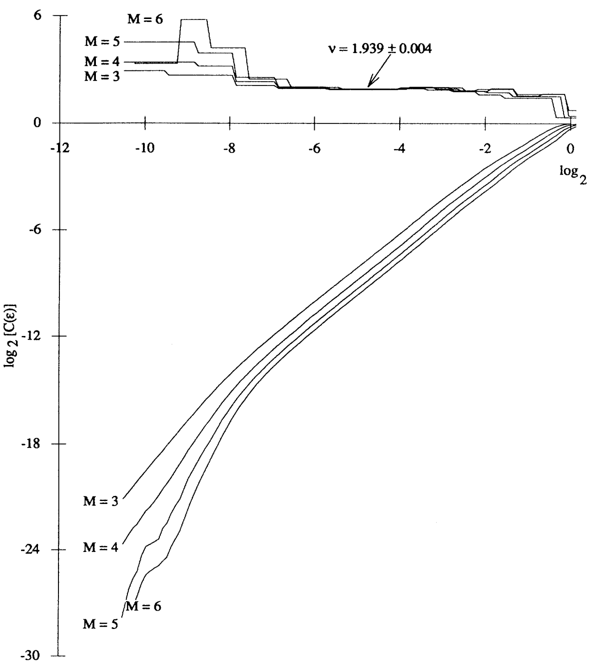


FIG. 1. Correlation integral and estimated slope coefficients (Mackey-Glass attractor, $\tau=17$).

$\tau=17$, 23, 30, and 100. Correlation integrals are calculated using the binning technique described in [1] with the bin width set such that there are approximately 100 points over the domain of each correlation integral.

To highlight our procedure's ability to distinguish the attractor domain from the noise domain, we add a uniformly distributed error term with range $\pm 0.4\times 10^{-2}$ to the time series corresponding to delay time $\tau=17$. Figure 1 shows the calculated correlation integrals and estimated slope coefficients for embedding dimensions 3, 4, 5, and 6. The space-filling nature of the noise process segments the correlation integral; specifically, increases in embedding dimension identify the kink in the correlation integral. At $\epsilon=2^{-7}$, estimated slope coefficients begin to increase with embedding and, for $\epsilon$ less than $2^{-8}$, the estimated slope coefficients clearly vary with embedding dimension. Above this length scale, the estimated correlation exponent $\nu=1.939$ is identified by the common scaling region in the domain $(-5.3, -4.0)$.

In addition to noise, entropy may cause significant variation in the slope of the correlation integral over its domain. If entropy is positive, correlation integrals may shift downward as embedding is increased. Consequently, correlation integrals may be distorted at large values of $\epsilon$. The following exercise demonstrates this. For a delay time $\tau=23$, with no added noise, Fig. 2(a) plots $\Delta \log_2 C(\epsilon)/\Delta \log_2 \epsilon$ over the domain of the correlation integral and Fig. 2(b) shows the corresponding estimated slope coefficients. Even though there is significant variation in the estimated slope coefficients, a common scaling region is identified in the domain $(-6.4, -5.3)$. Table I presents the estimated correlation exponents and the identified domains for all series. In each case, the estimate is consistent with that found in [1].

TABLE I. Estimated correlation exponent Mackey-Glass attractor.

| Embedding dimension | Estimated correlation exponent | Domain ($\log_2$) |
|---|---|---|
| | ($\tau=17$) | |
| 3 | 1.901±0.003 | −5.6 to −4.0 |
| 4 | 1.924±0.004 | −5.6 to −3.8 |
| 5 | 1.935±0.005 | −5.6 to −4.0 |
| 6 | 1.939±0.004 | −5.3 to −4.0 |
| | ($\tau=23$) | |
| 4 | 2.484±0.014 | −6.5 to −5.4 |
| 5 | 2.474±0.008 | −6.5 to −5.4 |
| 6 | 2.482±0.010 | −6.4 to −5.3 |
| 7 | 2.496±0.015 | −6.4 to −5.3 |
| | ($\tau=30$) | |
| 4 | 2.775±0.014 | −6.4 to −5.3 |
| 5 | 2.841±0.010 | −6.4 to −5.3 |
| 6 | 2.869±0.016 | −6.4 to −5.3 |
| 7 | 2.897±0.018 | −6.2 to −5.1 |
| | ($\tau=100$) | |
| 10 | 4.887±0.084 | −0.8 to −0.2 |
| 12 | 4.963±0.208 | −0.2 to −0.7 |
| 14 | 6.864±0.008 | −0.6 to −0.1 |
| 15 | 6.993±0.130 | −0.3 to −0.4 |

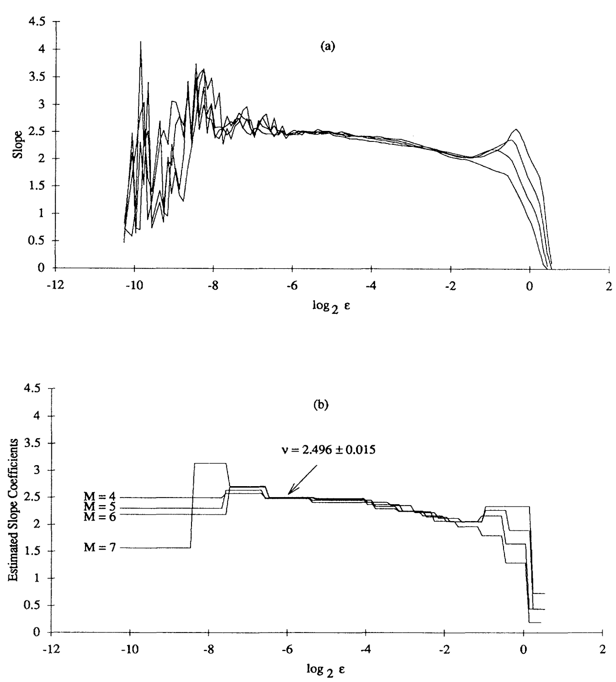


FIG. 2. (a) Slope of correlation integral and (b) estimated correlation exponent (Mackey-Glass attractor, $\tau=23$).

As a final example, our procedure is tested using experimental data. We study Couette-Taylor flow data with Reynolds number $R=12.9R_c$, where $R_c$ is the Reynolds number where Taylor vortex flow appears. From [11], the data are characterized by chaotic flow, the onset of which occurs at $R/R_c=11.7\pm 0.2$. In order to test our procedure's ability to identify a common scaling region using noisy data, we use a sample of data containing a significant amount of noise.

We reconstruct the attractor using a delay equal to 0.25 mean orbital time, corresponding to the first

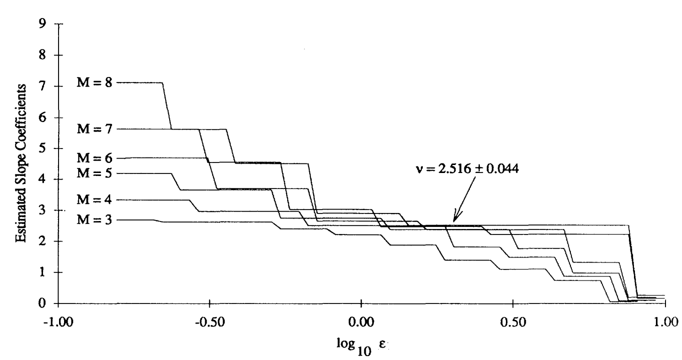


FIG. 3. Estimated correlation exponent (Couette-Taylor flow data).

TABLE II. Estimated correlation exponent Couette-Taylor flow data.

| Embedding dimension | Estimated correlation exponent | Domain ($\log_{10}$) |
|---|---|---|
| 3 | $1.196\pm0.048$ | 0.43 to 0.28 |
| 4 | $1.617\pm0.059$ | 0.46 to 0.31 |
| 5 | $2.202\pm0.074$ | 0.49 to 0.10 |
| 6 | $2.539\pm0.027$ | 0.49 to 0.10 |
| 7 | $2.469\pm0.014$ | 0.40 to 0.07 |
| 8 | $2.516\pm0.044$ | 0.16 to 0.88 |

minimum of the mutual information function. From a datafile of 32 768 observations, we sample approximately 25 points per orbit.

Figure 3 shows the estimated slope coefficients for embedding dimensions $M=3$, 4, 5, 6, 7, and 8. The noise domain is clearly identified for $\epsilon$ less than $10^{-0.15}$. Above this noise scale, the estimated slope coefficients are invariant to embedding for $M\geq 6$. The estimated correlation exponent, $\nu=2.516\pm0.044$ is identified by a common scaling region in the domain (0.07,0.40). Table II presents the estimated slope coefficients, over this domain, for each embedding. These estimates are consistent with those reported in [11].

## V. SUMMARY

The correlation integral has become a standard method for estimating the dimension of strange attractors using experimental data. Noise partitions the integral into two segments that must be identified by the experimenter. We develop a nonparametric statistical procedure for identifying the attractor and noise segments. Using this procedure, we produce estimates for the Mackey-Glass delay differential equation and Couette-Taylor flow data that are consistent with the existing literature.

## ACKNOWLEDGMENTS

We are grateful to Eric Kostelich for supplying the Couette-Taylor flow data and to the Cambridge Scientific Center for providing time on their 3090 vector processor.

[1] P. Grassberger and I. Procaccia, Physica D **9**, 189 (1983).
[2] P. Grassberger and I. Procaccia, Physica D **13**, 34 (1984).
[3] A. Ben Mizrachi, I. Procaccia, and P. Grassberger, Phys. Rev. A **29**, 975 (1984).
[4] C. R. Rao, *Advanced Statistical Methods in Biometric Research* (Wiley, New York, 1952).
[5] G. Chow, Econometrica **29**, 591 (1960).
[6] R. Brown, J. Durbin, and J. Evans, J. R. Stat. Soc. B **37**, 149 (1975).
[7] L. Ljung and T. Soderstrom, *Theory and Practice of Recursive Identification* (MIT Press, Cambridge, MA, 1986).
[8] J. Durbin, *Biometrika* **56**, 1 (1969).
[9] P. Grassberger and I. Procaccia, Phys. Rev. A **28**, 2591 (1983).
[10] M. Mackey, and L. Glass, *Science* **197**, 287 (1977).
[11] A. Brandstater and H. Swinney, Phys. Rev. A **35**, 2207 (1987).