# Correlation integrals of stock returns with volatility clustering*

Cees Diks†

September 21, 2000

**ABSTRACT**

Deviations from scaling in the correlation integrals of the returns of the Center for research in Securities Prices (CRSP) value weighted index are examined using the gaussian kernel correlation integral method. The data are found to be consistent with a, possibly stochastic, low-dimensional process superimposed with additive observational noise. For a class of processes with volatility clustering it is shown analytically that the correlation integrals exhibit this empirically observed behavior. This implies that volatility clustering, apart from leading to scaling behavior in correlation integrals of financial data, can also explain the typical deviations from scaling observed empirically.

In the past decades a wealth of methods have been developed for examining nonlinear dependence in empirical time series. The correlation integral method (Grassberger and Procaccia, 1983) has become a well-known method for characterizing deterministic time series by means of the correlation dimension and correlation entropy. More recently, methods based on correlation integrals and derived quantities were introduced to examine structure in time series that are not deterministic. For example, the BDS test (Brock *et al.*, 1996) can be used to test the null hypothesis that a time series is a realization of an independent, identically distributed (i.i.d.) process. The BDS test enables one to validate a fitted econometric model by testing for the absence of remaining serial dependence in the standardized residuals. Hiemstra and Jones (1994) developed a test for nonlinear Granger causality based on correlation integrals.

The development of the Grassberger-Procaccia method initiated searches for nonlinear determinism and evidence for chaos in nearly all fields in which time series play an important role, including finance. Scheinkman and LeBaron (1989) found evidence for nonlinear dependence in the returns of the value-weighted portfolio of the Center for Research in Security Prices at the University of Chicago (CRSP), and estimated the correlation dimension to be approximately 5.7. Initially it was widely assumed that the behavior of correlation integrals can, formally, be used to distinguish between low dimensional deterministic and stochastic time series. However, since it became clear that, apart from deterministic time series, also

*The author would like to thank Peter Boswijk, William Brock, Kung-Sik Chan, Cars Hommes, Blake LeBaron, Thomas Lux, Bruce Mizrach, Richard Smith and Howell Tong for helpful comments. Earlier versions of this paper were presented at the 1999 Workshop on Financial Statistics, Hong Kong University and the 2000 Applications of Physics in Financial Analysis conference in Liège.

†CeNDEF, Department of Economics, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands, e-mail: diks@fee.uva.nl, URL: http://www.fee.uva.nl/cendef

many stochastic processes can give finite correlation dimension estimates, the strategy of finding evidence for chaos through dimension estimation methods has lost most of its appeal. The possibility of chaos in stock returns is now being studied mainly in economic dynamics. For example, Brock and Hommes (1997, 1998) recently showed that chaos can easily arise for a large number of parameter values in systems with boundedly rational agents that are allowed to change strategies depending on past performance of prediction algorithms. Scheinkman and LeBaron (1989) already noted that finding a correlation dimension which does not increase with the embedding dimension does not rule out all random phenomena. Hsieh (1991) concluded that most of the nonlinear structure found in financial data can be attributed to ARCH structure. On the one hand the results of Scheinkman and LeBaron (1989) indicate that the CRSP returns contain nonlinear dependence incompatible with a random walk, but on the other hand the data were found not to be entirely deterministic. The increase in the slope of the correlation integrals (to be introduced in the next section) for small values of the radius parameter $\epsilon$ suggests the presence of some kind of noise, such as observational noise, dynamic noise, or both.

Here we examine whether the simplest of these types of noise, observational noise, which in the financial time series context is just additional constant volatility, can explain the observed deviations from scaling of the correlation integral. To this end a generalized version of the correlation integral, the gaussian kernel correlation integral, is used (Diks, 1996), which was originally designed to take into account the presence of gaussian observational noise on chaotic time series. By construction, the functional form of these correlation integrals allows for deviations from scaling resulting from observational noise. The idea is to examine whether the deviations from scaling observed in financial time series can be explained by the addition of observational noise to an otherwise self-similar, not necessarily chaotic, process. This would explain the empirically observed fact that they can show approximate scaling with signs of noise.

This paper is organized as follows. Section I describes the gaussian kernel algorithm for estimating the correlation dimension and correlation entropy, together with the noise level, from a time series. In section II we apply the gaussian kernel algorithm to the original CRSP returns as well as a simulated returns time series generated with an EGARCH model estimated from the original returns. In section III we analytically determine the behavior of the correlation integral for a simple stochastic process with volatility clustering. In the limit where the volatility clustering of this process tends to infinity, the correlation dimension is finite. Simulations show that time series generated by this stochastic model indeed have correlation integrals similar to those of chaotic time series. Section IV concludes and provides some directions for future research.

# I. The gaussian kernel algorithm

The correlation integral method (Grassberger and Procaccia, 1983) was developed for characterizing self-similarity in an attractor of a deterministic, possibly chaotic time series. As will be shown later, some stochastic processes also show scaling of the correlation integral similar to the scaling caused by deterministic dynamics. Therefore we will avoid references to deterministic dynamics and attractors, and merely refer to scaling and self-similarity instead.

The gaussian kernel algorithm (Diks, 1996) is a modified version of the Grassberger-Procaccia algorithm which provides a convenient way of characterizing self-similar time series in the presence of observational noise. The modification consists in replacing the familiar kernel function by a gaussian kernel function, which is more convenient for calculating the effect of gaussian observational noise analytically. Before describing the gaussian kernel algorithm in detail, the standard correlation integral method is briefly reviewed.

Let $\{X_n\}_{n=1}^N$ be an observed time series. The $m$-dimensional reconstruction vectors are given by

$$\mathbf{X}_n = (X_n, \ldots, X_{n+m-1}). \tag{1}$$

The term reconstruction vector is derived from the work of Takens (1981) who proved that, for series generated by a deterministic dynamical system, reconstruction vectors for sufficiently large $m$ completely specify the state of the dynamical system, and in this sense provide a reconstruction of the state space of the dynamical system. By definition, if a time series is stationary, its $m$-dimensional reconstruction vectors are distributed according to a well-defined probability measure $\mu_m$, which we will refer to as the reconstruction measure and which can be thought of as the stationary distribution of $m$-histories. The parameter $m$ in the chaos literature is usually referred to as the embedding dimension.

The correlation integral is defined as the cumulative distribution function of the distances between pairs of points drawn independently according to $\mu_m$, i.e.

$$C_m(\epsilon) = \int\int \Theta(\epsilon - \|\mathbf{x} - \mathbf{y}\|)\mu_m(\mathrm{d}\mathbf{x})\mu_m(\mathrm{d}\mathbf{y}), \tag{2}$$

where

$$\Theta(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0. \end{cases} \tag{3}$$

In other words, the correlation integral is the fraction of pairs of points chosen independently according to $\mu_m$ that are closer than $\epsilon$. For computational convenience, the supremum norm is usually taken, that is,

$$\|\mathbf{x} - \mathbf{y}\| = \sup_{i=1,\ldots,m} |x_i - y_i|. \tag{4}$$

(Grassberger and Procaccia, 1983) showed that the correlation integral of deterministic time series behaves as

$$C_m(\epsilon) \sim \epsilon^D \exp(-Km), \tag{5}$$

for small $\epsilon$ and large embedding dimension $m$, where $D$ is the correlation dimension and $K$ is the correlation entropy per time unit. For fixed embedding dimension $m$ the correlation integral exhibits power law behavior in $\epsilon$, $C_m(\epsilon) \sim \epsilon^D$, which is usually referred to as a scaling law in the physics literature. The correlation dimension $D$ and the correlation entropy $K$ are important quantities for characterizing dynamical systems. The correlation dimension $D$ characterizes the geometry of the attractor in terms of its fractal dimension, and can be considered a lower bound for the number of variables needed to model the dynamics of the system. The correlation entropy $K$ characterizes the dynamical behavior of the system. It quantifies the rate at which the distance between two initially nearby states increases under the dynamics. A positive value of $K$ indicates that the system exhibits so-called sensitive dependence on initial conditions, which is often taken as the defining characteristic of chaos.

The correlation integral can be estimated straightforwardly by counting the fraction of distances among the reconstruction vectors $\mathbf{X}_n$ smaller than $\epsilon$:

$$\widehat{C}_m(\epsilon) = \frac{2}{(N-m+1)(N-m)} \sum_{i=2}^{N-m+1} \sum_{j=1}^{i-1} \Theta(\epsilon - \|\mathbf{X}_i - \mathbf{X}_j\|). \tag{6}$$

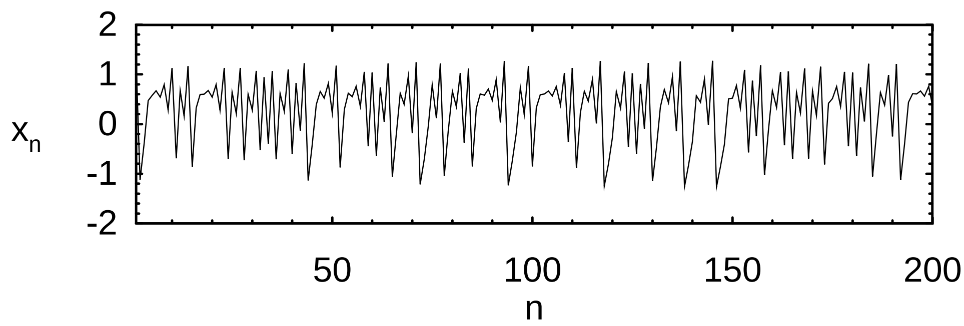


Figure 1: Time series $\{x_n\}_{n=1}^{200}$ generated with the Hénon model, equation (7).

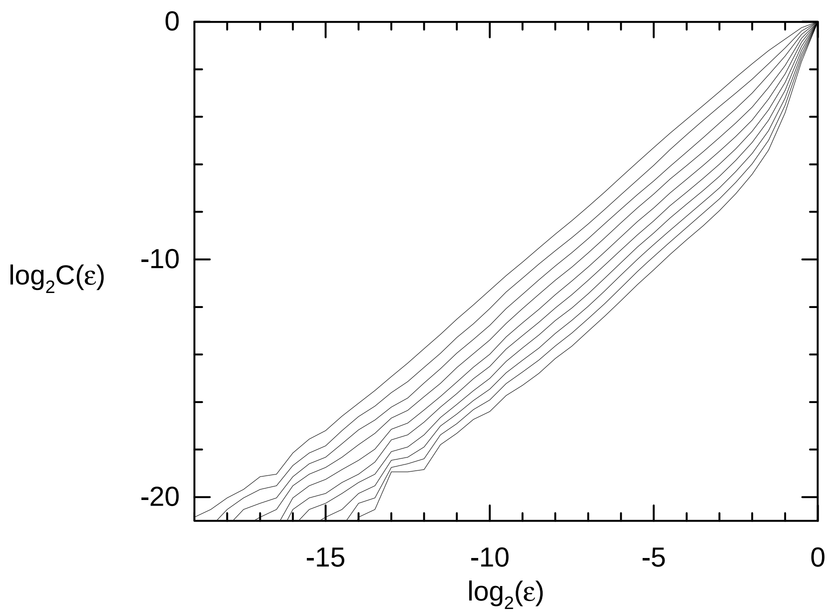


Figure 2: Correlation integrals estimated from the Hénon time series, for embedding dimensions $m = 2, \ldots, 11$ (lower curves correspond to higher embedding dimensions).

As an example we consider Hénon's dynamical system, given by

$$\begin{aligned} x_{n+1} &= 1 - 1.4x_n^2 + y_n \\ y_{n+1} &= 0.3x_n. \end{aligned} \tag{7}$$

We generated a time series consisting of $N = 4000$ consecutive observations of the variable $x_n$, the first 200 of which are shown in Figure 1. Figure 2 shows the estimated correlation integrals of the time series for embedding dimensions $m = 2, \ldots, 11$, on a double logarithmic scale. The scaling behavior is clearly visible, since for small $\epsilon$ the logarithm of $C_m(\epsilon)$ is linear in $\log \epsilon$. For very small values of $\epsilon$ only a few distances are smaller than $\epsilon$ so that statistical fluctuations dominate the estimated correlation integral and deviations from the scaling law become visible. Also, for large values of $\epsilon$ the scaling law breaks down, due to the finiteness of the attractor of the system. In practice, a certain range, called scaling region, of $\epsilon$ values is selected, within which the estimates of $C_m(\epsilon)$ are used in the fit procedure. Figure 3 shows the estimated values of the correlation dimension $D$ and the correlation entropy $K$ obtained by fitting the model of the correlation integral given in equation (5) (the scaling law) for pairs

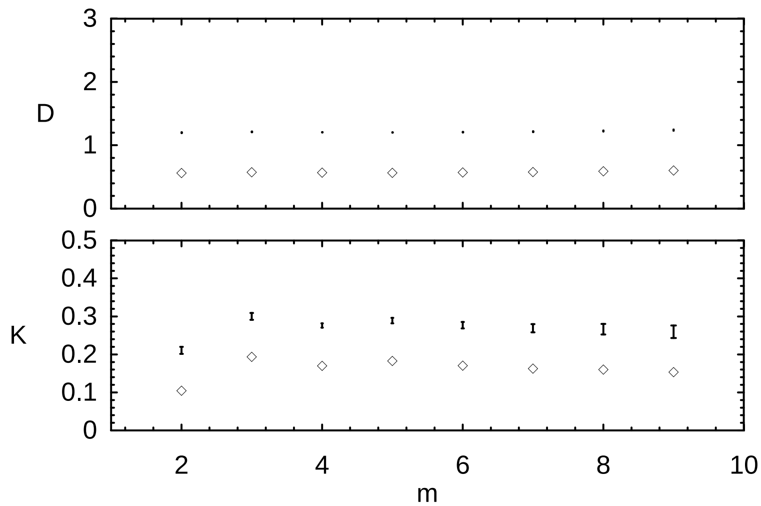


Figure 3: Estimated values of the correlation dimension $D$ and the correlation entropy $K$ for the Hénon time series (the bars indicate two estimated standard errors).

of consecutive embedding dimensions. (We need to fit the model to the estimated values of the correlation integral for at least two different embedding dimensions to estimate the correlation entropy $K$, since $K$ enters the scaling law only via the embedding dimension $m$, as can be seen from equation (5)).

Next we describe the gaussian kernel algorithm. The idea is to modify the scaling relation in such a way that it also describes the effect of gaussian observational noise. For the usual definition of the correlation integral this approach is analytically difficult. Upon using a slightly modified definition of the correlation integral this problem can be avoided. Note that the correlation integral as given in equation (2) can be written as the expected value of the kernel function, $C_m(\epsilon) = E\{\Theta(1 - \|\mathbf{X} - \mathbf{Y}\|/\epsilon)\}$, where $\mathbf{X}$ and $\mathbf{Y}$ are two vectors drawn independently according to the reconstruction measure $\mu_m$. A more general version of the correlation integral is given by

$$C_m(\epsilon) = E\left\{\kappa(\|\mathbf{X} - \mathbf{Y}\|/\epsilon)\right\}, \tag{8}$$

where $\kappa$ is some kernel function. Throughout we assume that $\kappa(0) = 1$, and that $\kappa$ is non-increasing on $[0, \infty)$ and satisfies $\lim_{x\to\infty} K(x) = 0$. In order to obtain the usual scaling relation (see equation (5)) of the correlation integral for deterministic time series in the absence of noise, some further conditions need to be imposed on the kernel function $\kappa$. We will not go into detail here, but refer the interested reader to Ghez and Vaienti (1992). We choose the gaussian kernel function

$$\kappa(x) = \exp(-x^2/4) \tag{9}$$

together with the Euclidean norm, i.e. $\|\mathbf{x} - \mathbf{y}\|^2 = \sum_{i=1}^{m}(x_i - y_i)^2$. The correlation integral thus obtained will be referred to as the gaussian kernel correlation integral. Ghez and Vaienti (1992) used a gaussian kernel function for the estimation of dimension and entropy from a noise-free time series. Diks (1996) showed that this kernel is convenient when the observed time series is superimposed with independent normally distributed observational noise, that is, when instead of observing a deterministic time series, one observes

$$X_n = Y_n + \epsilon_n, \qquad \epsilon_n \overset{\text{i.i.d.}}{\sim} N(0, \sigma_{\text{n}}^2) \tag{10}$$

where $\{Y_n\}$ is a finite dimensional deterministic time series, and $\sigma_{\rm n}^2$ is the variance of the noise.

After rescaling of the time series to unit variance, the gaussian kernel correlation integral can be shown to behave as

$$C_m(\epsilon) \simeq \psi \epsilon^D \left(\frac{\epsilon}{\sqrt{\epsilon^2+\sigma^2}}\right)^{m-D} m^{-D/2} \mathrm{e}^{-Km} \tag{11}$$

for small $\epsilon$ and large embedding dimension $m$. Here $\sigma^2$ denotes the normalized noise variance after rescaling the time series to unit variance, i.e.

$$\sigma^2 = \frac{\sigma_{\rm n}^2}{\sqrt{\sigma_{\rm n}^2+\sigma_{\rm s}^2}} \tag{12}$$

where $\sigma_{\rm s}^2$ is the standard deviation of the noise-free time series $\{Y_n\}$. The correlation dimension $D$ and correlation entropy $K$ are those of the underlying deterministic time series $\{Y_n\}$. The term between braces accounts for the presence of noise. The factor $m^{-D/2}$ merely occurs as a result of the use of the Euclidean norm. It can be readily verified that, apart from this factor, equation (11) reduces to the usual scaling law given in equation (5) which holds in the absence of noise ($\sigma = 0$). Estimates of the correlation dimension $D$, the correlation entropy $K$ and the noise level $\sigma$ for embedding dimension $m$ are obtained by fitting the modified scaling relation, equation (11), for two consecutive values of the embedding dimension simultaneously. We use a weighted least squares nonlinear fit procedure (Levenberg Marquardt, see e.g. Press *et al.*, 1992).

It was shown in Diks (1996) using computer generated time series that the gaussian kernel algorithm gives good dimension and entropy estimates for noise levels as high as 20% ($\sigma = 0.2$). Also, the method was found to be robust with respect to the details of the distribution of the noise. For example, if uniformly distributed noise is superimposed on the time series rather than normally distributed noise, the noise variance is underestimated slightly but one still obtains good dimension and entropy estimates.

## II. Application to CRSP returns

We applied the gaussian kernel algorithm to the CRSP daily returns time series (shown in Figure 4) based on the 6345 closing prices in the period from July 3rd 1962 to September 30th 1987. Scheinkman and LeBaron (1989) already observed the typical signs of noise in the behavior of the correlation integral; the local slopes $\mathrm{d}\log C_m(\epsilon)/\mathrm{d}\log\epsilon$ of the correlation integrals are approximately constant for moderate values of $\epsilon$, but increase for small $\epsilon$. They partly accounted for this by estimating the dimension as the local slope for values of $\epsilon$ which are large enough to avoid most of the overestimation of the correlation dimension due to noise.

Figure 5 shows the estimates of $D$, $K$ and $\sigma$ obtained with the gaussian kernel correlation integral method for increasing embedding dimensions. An upper bandwidth of $\epsilon_{\rm u} = 0.5$ (after rescaling the time series to unit variance) was used in the fit. The estimated correlation dimension saturates at a value of approximately 4.5, the correlation entropy saturates at 0.06 nats/day (the unit nats refers to the use of the natural logarithm), and the noise level $\sigma$ converges to about 0.24. Note that this noise level is just below the maximum noise level allowed by Scheinkman and LeBaron (1989) (they estimated the correlation dimension from the local slopes $\mathrm{d}\log C_m(\epsilon)/\mathrm{d}\log\epsilon$ for values of $\epsilon$ sufficiently large to allow for a noise level

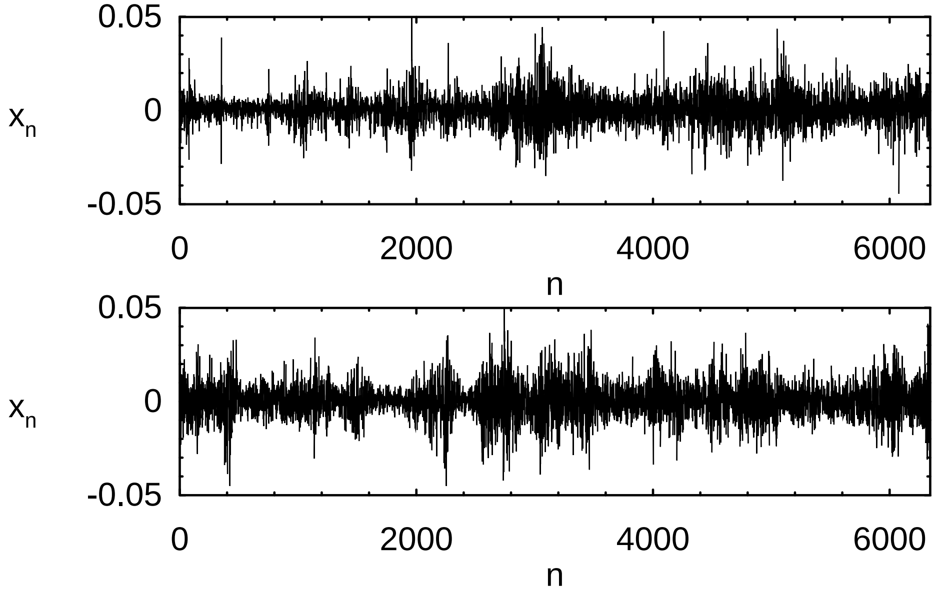


Figure 4: CRSP returns (upper panel) and returns time series generated using an estimated EGARCH model (lower panel).

of 26% of the standard error of the time series). The fact that the estimated correlation dimension is smaller than the estimate (5.7) of Scheinkman end LeBaron can be explained by the fact that observational noise increases the slope $\mathrm{d}\log\log C_m(\epsilon)/\mathrm{d}\log\epsilon$. This leads to a positive bias in estimates of $D$ when the effects of noise are not taken into account.

Our results at first sight appear to confirm the hypothesis that the CRSP returns are a realization of a chaotic process with observational noise. However, before concluding that the CRSP returns consist of low-dimensional chaos with observational noise, we should at least check whether the low dimension estimates are the result of chaos or might alternatively have been generated by some stochastic econometric model. In order to examine this, we fitted an EGARCH model to the data, and used the estimated coefficients to generate a time series of the same length as the original time series. The motivation for choosing an EGARCH model rather than a GARCH model is that we found a highly significant negative correlation between the returns and the volatility, suggesting the presence of a leverage effect: the market response (in terms of volatility) to negative shocks is larger than to positive shocks. The EGARCH model (Nelson, 1991) accounts for this effect. The autocovariance structure was accounted for by an ARMA(2,1) model. The estimated model is

$$x_n = \underset{(0.0001)}{0.0004} - \underset{(0.16)}{0.43}\, x_{n-1} + \underset{(0.04)}{0.12}\, x_{n-2} + z_n + \underset{(0.16)}{0.66}\, z_{n-1}, \qquad (13)$$

where $z_n \sim N(0, \sigma_n^2)$. The conditional variance $\sigma_n^2$ is modeled by

$$\log(\sigma_n^2) = -\underset{(0.03)}{0.21} + \underset{(0.002)}{0.989}\, \log(\sigma_{n-1}^2) + \underset{(0.01)}{0.14} \left|\frac{z_{n-1}}{\sigma_{n-1}}\right| - \underset{(0.01)}{0.07}\, \frac{z_{n-1}}{\sigma_{n-1}}. \qquad (14)$$

In order to test the validity of the EGARCH model we applied the BDS test to the standardized residuals (embedding dimension $m = 3$, $\epsilon = 0.5\sigma$). The $p$-value obtained by constructing

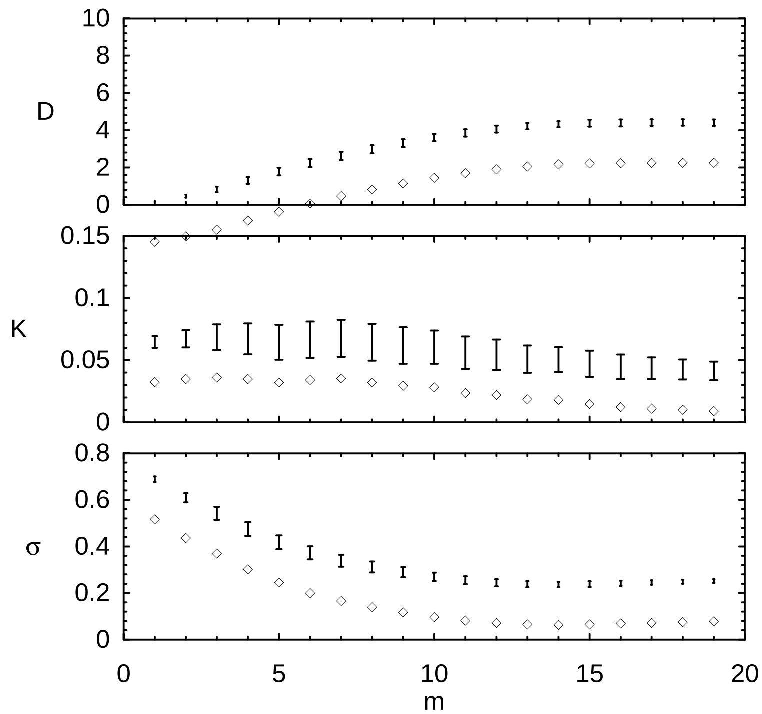


Figure 5: Estimated correlation dimension, correlation entropy and noise level for the CRSP returns ($\epsilon_{\mathrm{u}} = 0.5$).

a (conditional) null distribution from 1000 consecutive random permutations of the elements of the time series, is 0.72 so that we have found no evidence of structure in the residuals.

We applied the gaussian kernel algorithm again to a time series generated by the fitted model. Figure 4 (lower panel) shows the simulated returns time series. Figure 6 shows the resulting estimates of the correlation dimension, entropy and noise level. In contrast with what one might expect for a time series generated by a stochastic process, the resulting estimates are qualitatively similar to those obtained from the original returns time series; the estimated correlation dimension saturates at a value around 5 and the correlation entropy at 0.05 nats/day. The estimated noise level converges to a value of about 25% ($\sigma = 0.25$). Since these values are close to the estimates for the CRSP returns data, this leads us to the conjecture that, somehow, volatility clustering might be responsible for the observed low dimension of the time series.

Figure 7 shows the fitted model parameters for the standardized residuals of the fitted EGARCH model. Clearly, there is no sign of saturation of the parameters with increasing embedding dimension $m$, as expected, since the standardized residuals passed the BDS test for independence.

In the next section we will construct a theoretical model which allows us to understand how stochastic time series can show signs of chaos.

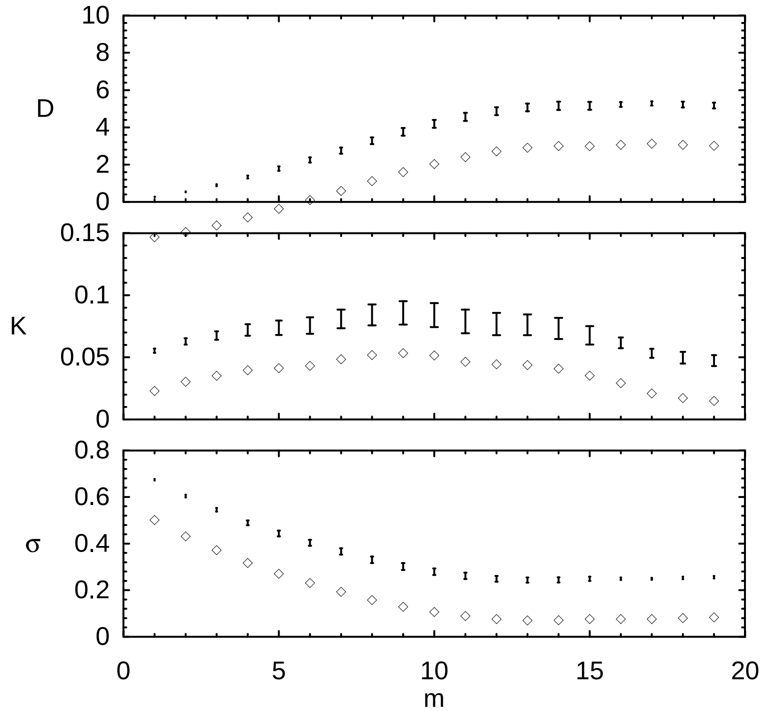


Figure 6: The estimated correlation dimension, correlation entropy and noise level for a artificial time series generated with the estimated EGARCH model for the CRSP data ($\epsilon_{\mathrm{u}} = 0.5$).

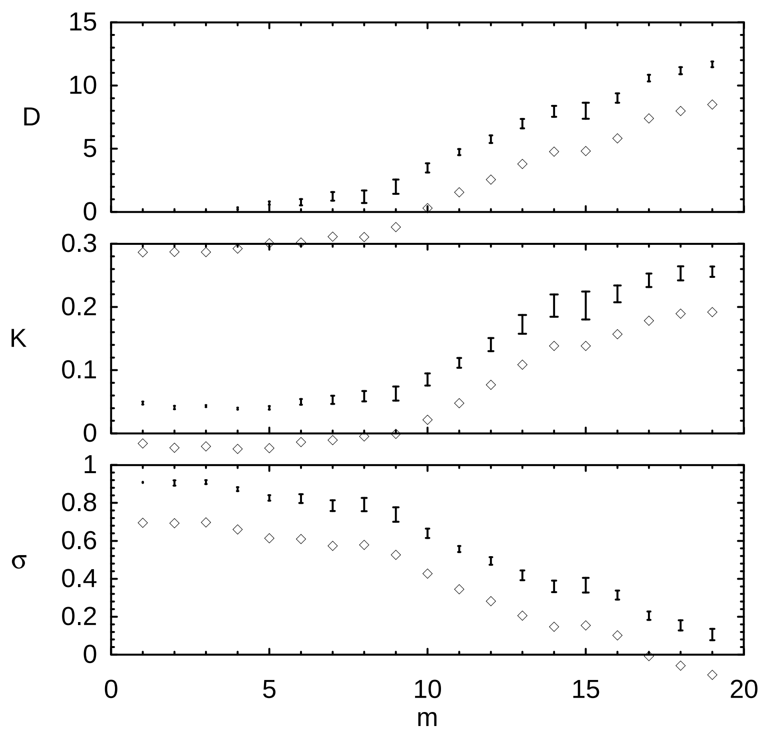


Figure 7: Estimated correlation dimension, correlation entropy and noise level for the standardized EGARCH residuals ($\epsilon_{\mathrm{u}} = 0.5$).

# III. Volatility clustering and correlation integrals

## A. A simple process with volatility clustering

In this section we examine the scaling properties of the gaussian kernel correlation integral for time series with volatility clustering. Before examining the volatility clustering model in detail, let us first develop some intuition by considering an example of a stationary stochastic process with a finite correlation dimension. Consider the i.i.d. process

$$x_t = \begin{cases} 0 & \text{with probability } p \\ 1 & \text{with probability } 1-p. \end{cases} \quad (15)$$

For each embedding dimension $m$, with probability $p^m$, a reconstruction vector $(X_{n-m+1}, X_{n-m+2}, \ldots, X_n)$ is equal to the null vector, $\mathbf{0}$. Therefore, a fraction of at least $p^{2m}$ of the distances is equal to zero, and the correlation integral, which is nothing but the cumulative distribution function of distances, does not approach zero for small $\epsilon$ but instead converges to a finite positive value. As a result the the slope $\mathrm{d}\log C_m(\epsilon)/rmd\log\epsilon$ tends to zero for small $\epsilon$ so that the correlation dimension of this process is zero.

In the above example the finite value of the correlation integral is the result of the distributional properties of the i.i.d. process. The process can take only a finite number of values and hence has to be of dimension zero. It is also possible to construct stochastic processes with a finite correlation dimension with a continuous marginal distribution. As long as the process is sufficiently close to some value, zero say, for sufficiently long time intervals, the time series has a finite correlation dimension. This mechanism is responsible for the scaling of the correlation integral for our volatility clustering model. Since the scaling properties of the correlation integral for more familiar models of volatility clustering such as an ARCH or GARCH models are difficult to derive analytically we choose to use a more simple model for volatility clustering.

Consider the following simple process with volatility clustering:

$$X_n \sim N(0, \sigma_n^2), \quad (16)$$

where

$$\sigma_n^2 = \begin{cases} S_n & \text{with probability } p \\ \sigma_{n-1}^2 & \text{with probability } 1-p, \end{cases} \quad (17)$$

where the $S_n$ are drawn independently from a probability distribution on $[0, \infty)$, with probability density function (pdf), $f(x)$, say. Because at each point in time, there is only a small probability $p$ with which the process switches to a regime with a different volatility, for small $p$ the volatility is likely to stay the same for long periods of time. For time series of length $N$ the expected number of different regimes becomes $Np$, and the expected regime duration $1/p$ for large $N$. Note that, since $f$ is the pdf of the variance, the resulting process has finite variance if and only if

$$\int_0^\infty x f(x)\,\mathrm{d}x < \infty. \quad (18)$$

A limit of process (17) in which $N \to \infty$ and $p \to 0$, such that $Np \to \infty$, will be referred to as a volatility persistence limit. In this limit one obtains the following properties: (i) Persistence of volatility: the probability $P(\sigma_{n+1}^2 = \sigma_n^2)$ tends to one, and (ii) independence:

$\sigma_n^2$ and $\sigma_l^2$, with $m$ and $l$ selected independently according to the uniform distribution on $1, 2, \ldots, N$, become independent.

The idea behind the volatility persistence limit is to increase the duration of the time intervals with constant volatility. At the same time we want the number of intervals with different volatility to tend to infinity. If the switching probability goes to zero, the difference $X_n - X_l$ for $|n-l|$ sufficiently large, given $\sigma_n^2$ and $\sigma_l^2$, becomes $m$-variate normal with independent components, each with variance $\sigma_n^2 + \sigma_l^2$. This independence of the components is convenient in the calculations. Formally, the process has a finite correlation dimension only in the volatility persistence limit. However, as our simulations will show, one can obtain low dimension estimates also for finite time series.

In Appendix A it is shown that, if the probability density function $f(x)$ of the volatility behaves as $x^{\alpha-1}$ ($\alpha > 0$) for small $x$, the correlation integral satisfies

$$C_m(\epsilon) \sim \epsilon^{4\alpha}, \tag{19}$$

up to leading order in $\epsilon$, for $m \geq 4\alpha$, which implies that the correlation dimension $D$ is related to $\alpha$ through

$$D = 4\alpha. \tag{20}$$

## B. Results on simulated time series

We generated a time series with the volatility clustering model with $p = 0.05$ of the same length, $N = 6345$, as the CRSP time series. For the probability density function of the volatility we used

$$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ \frac{2}{\sqrt{2\pi}} e^{-x^2/2} & \text{for } x \leq 0. \end{cases} \tag{21}$$

Since $\lim_{x\to 0} f(x)$ is finite, this density function corresponds to the case $\alpha = 1$. Therefore the corresponding correlation dimension $D$ is equal to 4, and we also expect the estimated correlation dimension to be about 4.

Figure 8 (upper panel) shows a realization of the volatility clustering process with $p = 0.02$. The time series is rescaled to have variance equal to the sample variance of the CRSP returns. Figure 9 shows the estimated correlation dimension, entropy and noise level for this series. The estimated dimension converges to a value of about 4 around embedding dimension $m = 10$ and the estimated correlation entropy converges to a finite positive value. The noise level, also converges to a finite value near 0.15 for increasing embedding dimensions, which suggests that it is slightly overestimated.

Figure 8 (lower panel) shows a realization of the same process, but now with normally distributed observational noise with $\sigma = 0.20$ added. Again the time series is rescaled to have variance equal to the sample variance of the CRSP returns time series. The estimated correlation dimension, correlation entropy and noise level for this time series are shown in figure 10. Again, one can observe convergence of the estimates. In this case the dimension is underestimated slightly, and again the estimated noise level is larger than the true noise level.

Qualitatively, the fits for these model generate model generated time series are very similar to those obtained using the original CRSP returns data and the time series generated by the fitted EGARCH model.

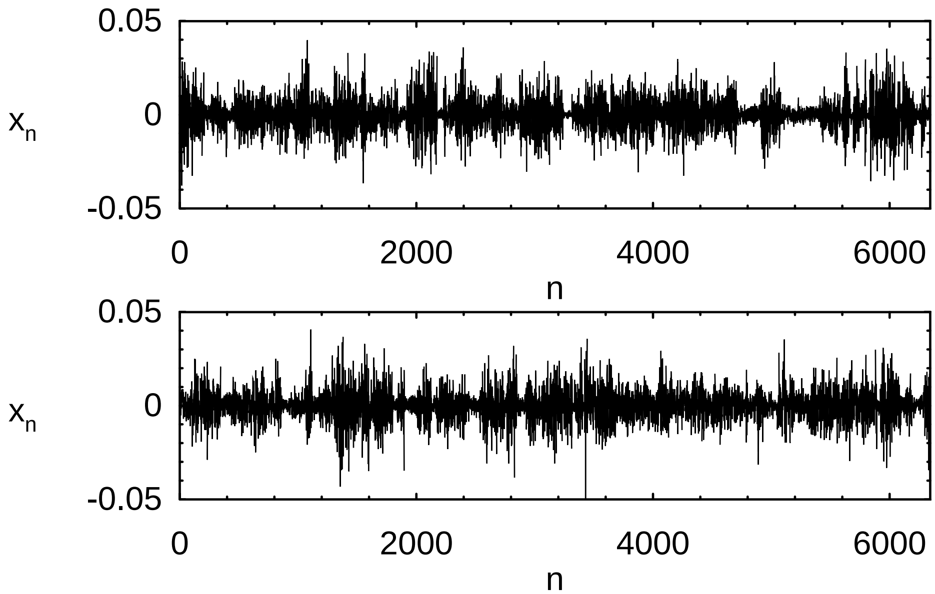


Figure 8: Two time series generated independently with the volatility clustering model ($p = 0.02$) without observational noise (upper panel) and with observational noise, $\sigma = 0.20$ (lower panel).

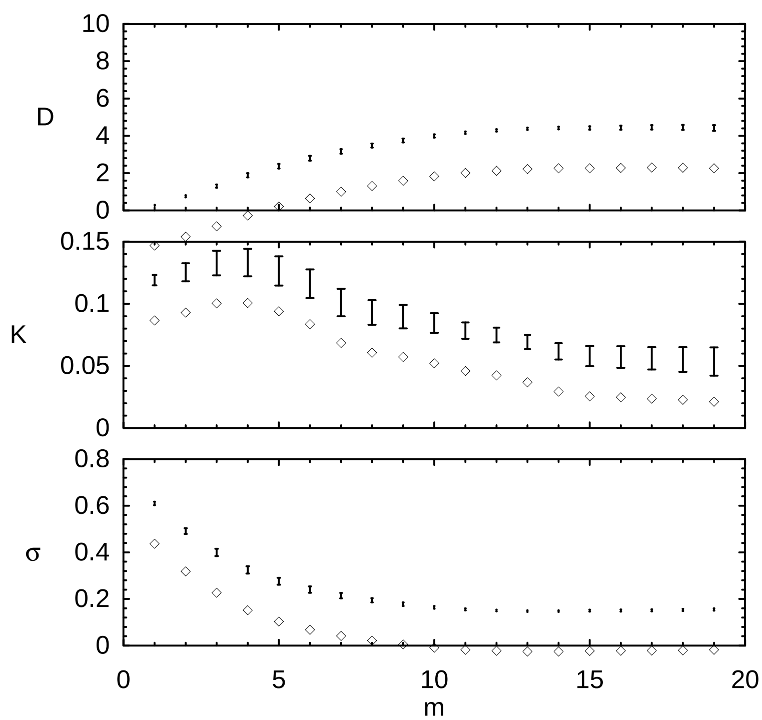


Figure 9: Estimated correlation dimension, correlation entropy and noise level for the time series generated by the volatility model ($p = 0.02$, $\sigma = 0$, $\epsilon_{\rm u} = 0.5$).

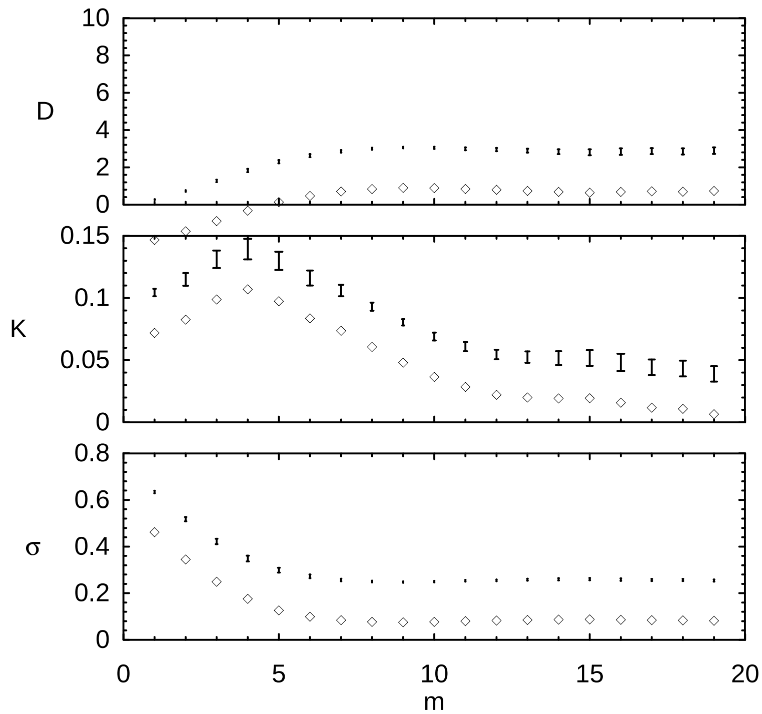


Figure 10: Estimated correlation dimension, correlation entropy and noise level for the time series generated by the volatility model ($p = 0.02$, $\sigma = 0.2$, $\epsilon_{\rm u} = 0.5$).

## IV. Summary and Discussion

Our empirical results suggest that the behavior of the correlation integrals of the CRSP returns can be described well by a modified scaling law which was designed for describing the effect of observational noise on a finite dimensional time series, whether chaotic or not. This approach gives a summary of the data in terms of the correlation dimension, the correlation entropy, and the noise level. A volatility clustering process is presented which generates data for which the correlation integral behaves as prescribed by this model. This provides two possible interpretations of the estimated parameters of the model for the correlation integral, depending on whether: (i) the time series at hand was generated by a chaotic process with noise or (ii) by a stochastic process with volatility clustering. Application of the BDS test to the standardized residuals of a fitted EGARCH model provides no evidence against the the EGARCH specification.

Shifting from the chaotic framework to the volatility clustering framework both the correlation dimension and the estimated noise level obtain a potentially useful new econometric interpretation. The correlation dimension is, up to a constant factor, equal to the exponent describing the probability distribution of the volatility near the smallest volatilities assumed by the process. The estimated noise level provides a measure for "background volatility." It determines the minimum conditional volatility of the process, that is, the minimum uncertainty about tomorrow's stock return. Estimates of these parameters might display useful patterns across assets and across time, and are a potentially useful "information summary" of the underlying stochastic process. As of yet the econometric interpretation of the estimated correlation entropy is not clear to us.

One should keep in mind that the two different frameworks mentioned earlier, determinism with noise, or volatility clustering, are merely two extremes of a broad spectrum. Currently the development of deterministic models with volatility clustering is a growing area of research. It is therefore possible that these models require even more possible interpretations of the estimated model parameters. This is left for future research. However, after fitting and validating a stochastic model with volatility clustering, such as an ARCH or GARCH model, one has a strong case for interpreting the estimated parameters in the stochastic (volatility clustering) framework.

# A Behavior of the correlation integral in the volatility persistence limit

After rewriting the expectation on the left hand side of equation (8) one obtains

$$C_m(\epsilon) = \int\int \kappa\left(\frac{\|\mathbf{x}-\mathbf{y}\|}{\epsilon}\right) \mathrm{d}\mu_m(\mathbf{x})\mathrm{d}\mu_m(\mathbf{y}) \tag{22}$$

where $\mu_m$ is the $m$-dimensional reconstruction measure, i.e. the probability measure associated with the $m$-dimensional reconstruction vectors as defined in equation (1). This gives

$$C_m(\epsilon) = \int \kappa\left(\frac{\|\mathbf{z}\|}{\epsilon}\right) \rho_{\mathbf{Z}}(\mathbf{z})\mathrm{d}\mathbf{z}, \tag{23}$$

where $\rho_{\mathbf{Z}}(\mathbf{z})$ is the probability density function of $\mathbf{Z} = \mathbf{X} - \mathbf{Y}$, with $\mathbf{X}$ and $\mathbf{Y}$ independent random vectors, distributed according to $\mu_m$.

The distribution of $\mathbf{X}$ and $\mathbf{Y}$ is a continuous mixture of normal distributions, with weights determined by the density $f(\cdot)$ of the variances. Both $\mathbf{X}$ and $\mathbf{Y}$ have pdf

$$\int_0^\infty \frac{e^{-\|\mathbf{x}\|^2/(2s)}}{(2\pi s)^{m/2}} f(s)\mathrm{d}s. \tag{24}$$

We can factorize $\mathbf{X}$ as $\mathbf{X} = \sigma_{\mathbf{X}}\tilde{\mathbf{X}}$ and where $\sigma_{\mathbf{X}}$ and $\tilde{\mathbf{X}}$ are independent, $\sigma^2_{\mathbf{X}}$ having pdf $f(\cdot)$, and $\tilde{\mathbf{X}}$ being distributed according to $\mathcal{N}(0, I_m)$. A similar factorization holds for $\mathbf{Y}$. Clearly, given $\sigma^2_{\mathbf{X}} = s$ and $\sigma^2_{\mathbf{Y}} = t$, $\mathbf{X} - \mathbf{Y}$ is distributed normally, with variance $s + t$. This leads to

$$\rho_{\mathbf{Z}}(\mathbf{z}) = \int_0^\infty \int_0^\infty \frac{e^{-\|\mathbf{z}\|^2/(2s+2t)}}{(2\pi(s+t))^{m/2}} f(s)f(t)\mathrm{d}s\mathrm{d}t, \tag{25}$$

which, upon introducing the variable $b = s + t$, can be written as

$$\rho_{\mathbf{Z}}(\mathbf{z}) = \int_0^\infty \frac{e^{-\|\mathbf{z}\|^2/(2b)}}{(2\pi b)^{m/2}} \nu(b)\mathrm{d}b \tag{26}$$

where $\nu$ is the probability density function of $(\sigma^2_{\mathbf{x}} + \sigma^2_{\mathbf{y}})$:

$$\nu(b) = \int_0^b f(b-s)f(s)\mathrm{d}s. \tag{27}$$

Upon substituting this into equation (23) one obtains

$$C_m(\epsilon) = \int_0^\infty \left\{\int \kappa\left(\frac{\|\mathbf{z}\|}{\epsilon}\right) \frac{e^{-\mathbf{z}^2/(2b)}}{(2\pi b)^{m/2}}\mathrm{d}\mathbf{z}\right\} \nu(b)\mathrm{d}b. \tag{28}$$

Notice that upon rescaling both the process standard deviation by multiplying $\sqrt{b}$ with some factor, and the bandwidth $\epsilon$ with the same factor, the integral in curly braces must remain unchanged. Indeed it can be readily verified that the inner integral is invariant under the scale transformation

$$\begin{aligned} \epsilon &\mapsto \gamma\epsilon \\ b &\mapsto \gamma^2 b \end{aligned} \tag{29}$$

whence this integral is a function of $\epsilon/\sqrt{b}$ only. For convenience we denote this integral by $g_m$, i.e.

$$g_m(\epsilon/\sqrt{b}) = \int \kappa\left(\frac{\|\mathbf{z}\|}{\epsilon}\right) \frac{e^{-\|\mathbf{z}\|^2/(2b)}}{(2\pi b)^{m/2}} \mathrm{d}\mathbf{z}. \tag{30}$$

After rescaling by a factor $\gamma = \sqrt{b}$ one obtains

$$g_m(\epsilon/\sqrt{b}) = \int \kappa\left(\frac{\sqrt{b}\|\mathbf{z}\|}{\epsilon}\right) \frac{e^{-\|\mathbf{z}\|^2/(2)}}{(2\pi)^{m/2}} \mathrm{d}\mathbf{z}. \tag{31}$$

We can interpret $g_m(x)$ as the expectation of the kernel with respect to a multivariate normal distribution as follows:

$$g_m(x) = E\{\kappa(\|\mathbf{Z}\|/x)\} \qquad \text{with } \mathbf{Z} \sim \mathcal{N}(0, I_m). \tag{32}$$

Both the standard kernel and the gaussian kernel factorize in the multivariate case, that is, $\kappa(\|\mathbf{Z}\|/x) = \kappa(\|Z_1\|/x)\ldots\kappa(\|Z_m\|/x)$, provided the appropriate norms are used: the $L_\infty$ and the $L_2$ norm for the standard kernel and the gaussian kernel respectively. As a result, $g_m$ factorizes also:

$$g_m(x) = (g_1(x))^m. \tag{33}$$

In fact, $g_1(x)$ is the contribution to the correlation integral of a univariate distance, distributed according to $N(0,1)$:

$$g_1(x) = E\{\kappa(|Z|/x)\} \qquad \text{with } Z \sim N(0,1). \tag{34}$$

Substitution of equation (31) into equation (28) gives

$$C_m(\epsilon) = \int_0^\infty g_m\left(\frac{\epsilon}{\sqrt{b}}\right) \nu(b)\mathrm{d}b. \tag{35}$$

The function $g_m(x)$ is the correlation integral with the kernel function $\kappa$, of a multivariate gaussian random variable with covariance matrix $I_m$. Therefore, it has the following properties: $g_m(x)$ is non-decreasing, for small $x$, $g_m(x) \sim x^m$, and $g_m(x)$ approaches 1 for large $x$.

The behavior of $C_m(r)$ for small $\epsilon$ is dominated by the contributions from small $b$ in equation (35). The behavior of $\nu(b)$ for small $b$ is determined by the behavior of $f(x)$ for small $x$. Assuming $f(x) \sim x^{\alpha-1}$ for small $x$, for $b > 0$ and $\alpha > 0$, equation (27) then leads to the behavior $\nu(b) \sim b^{2\alpha-1}$ for small $b$. If $m \geq 4\alpha$ we can substitute $u = \epsilon/\sqrt{b}$ into equation (35), which gives

$$C_m(\epsilon) = 2\int_0^\infty g_m(u) \left(\frac{\epsilon^2}{u^2}\right)^{2\alpha-1} u^{-3}\epsilon^2 \mathrm{d}u \tag{36}$$

for small $\epsilon$. This gives

$$C_m(\epsilon) \sim \epsilon^{4\alpha} \tag{37}$$

for small $\epsilon$, so that the correlation dimension $D$ is related to the exponent $\alpha$ through

$$D = 4\alpha. \tag{38}$$

## References


Brock, W. A., Dechert, W. D., Scheinkman, J. A. and B, LeBaron (1996). A test for independence based on the correlation dimension. *Econ. Rev.*, **15,** number 3, 197–235.

Brock, W. A. and Hommes, C. H. (1997). A rational route to randomness. *Econometrica*, **65,** 1059–1095.

Brock, W. A. and Hommes, C. H. (1998). Heterogeneous beliefs and routes to chaos in a simple asset pricing model. *JEDC*, **22,** 1235–1274.

Diks, C. (1996). Estimating invariants of noisy attractors. *Phys. Rev.* E, **53,** R4263–R4266.

Ghez, J. M. and Vaienti, S. (1992). Integrated wavelets on fractal sets: I. The correlation dimension. *Nonlinearity*, **5,** 777–790.

Grassberger, P. and Procaccia, I. (1983). Measuring the strangeness of strange attractors. *Physica* D, **9,** 189–208.

Hiemstra, C. and Jones, J. D. (1994). Testing for linear and nonlinear Granger causality in the stock price-volume relation. *Journal of Finance*, **49,** number 5, 1639–1664.

Hsieh, D. A. (1991). Chaos and nonlinear dynamics: Application to financial markets. *Journal of Finance*, **46,** number 5, 1839–1877.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: a new approach. *Econometrica*, **59,** 347–370.

Press, W. H., Teukolski, S. T., Vetterling, W. T. and Flannery, B. P. (1992). *Numerical Recipes in C: the art of scientific computing*, 2nd edn. Cambridge, United Kingdom: Cambridge University Press.

Scheinkman, J. A. and LeBaron, B. (1989). Nonlinear dynamics and stock returns. *J. Business*, **62,** number 3, 311–337.

Takens, F. (1981). Detecting strange attractors in turbulence. In *Dynamical Systems and Turbulence, Warwick 1980 (Lecture Notes in Mathematics)* (eds D. A. Rand and L.-S. Young), volume 898, pp. 366–381. Springer, Berlin.