José A. Scheinkman
University of Chicago

Blake LeBaron
University of Wisconsin—Madison

Nonlinear Dynamics and Stock Returns*

Simple deterministic systems are capable of generating chaotic output that "mimics" the output of stochastic systems. For this reason, algorithms have been developed to distinguish between these two alternatives. These algorithms and related statistical tests are also useful in detecting the presence of nonlinear dependence in time series. In this article we apply these procedures to stock returns and find evidence that indicates the presence of nonlinear dependence on weekly returns from the Center for Research in Security Prices (CRSP) value-weighted index.

I. Introduction

Recently algorithms have been proposed (e.g., Grassberger and Procaccia 1983*a*; Takens 1983*a*, 1983*b*) to try to distinguish between data generated by a deterministic system and data generated by a "random" system. These were motivated by the realization that certain deterministic systems have solutions that "look like" they are generated from a random system. A simple example can be exhibited by considering the function f: $[0,1] \rightarrow [0,1]$ such that $f(x) = 2x$ if $0 \leq x \leq 1/2$, and $f(x) = 2(1 - x)$ if $1/2 < x \leq 1$ (this is usually referred to as a "tent" map). For a given $x_0 \in [0,1]$, let $x_t = f(x_{t-1})$; $t = 1,2, \ldots$. This

* We thank Buz Brock, Ivar Ekeland, Lars Hansen, John Bechhoefer, Mogen Jensen, Narayana Kocherlakota, Jean Michel Lasryl, Albert Libchaber, Albert Madansky, Merton Miller, Jacob Palis, David Ruelle, Don Saari, and Florins Takens for conversations on the subject. David Hsieh provided us with the program that estimates the ARCH model and gave generously of his time to guide us through it. This article contains the results of Scheinkman (1985), which was prepared for the Conference on Nonlinear Dynamics in Paris, 1985. Much of Scheinkman's research was done while he was visiting the CEREMADE at the University of Paris IX and the Instituto de Matemática Pura e Aplicada (IMPA, Rio de Janeiro), which gave generous access to their computer facilities. The National Science Foundation provided support through grants SES-8420930 and INT-841-3966.

(*Journal of Business*, 1989, vol. 62, no. 3)

0021-9398/89/6203-0001$01.50

gives us a "time series" $\{x_t\}_{t=0}^{\infty}$ that depends on x_0. Let

$$r_k = \lim_{m\to\infty} \frac{1}{m} \sum_{t=1}^{m-|k|} x_t x_{t+|k|} - \left(\frac{1}{m} \sum_{t=1}^{m} x_t\right)^2,$$

the autocovariance at lag k. It can be shown that, for almost all $k_0 \in [0,1]$ (i.e., if we consider in [0,1] the uniform density, for all x_0 except for a subset of [0,1] with probability zero), $r_k = 0$ if $k \neq 0$.[1]

These same algorithms have been successfully used to distinguish between random systems and deterministic systems coupled with "small" amounts of noise (e.g., Atten and Caputo 1985). In particular they are potentially useful to detect the presence of nonlinearities. Here we will examine U.S. stock returns data (in fact a comprehensive index) using these techniques. For completeness, in Section II we describe the test as well as make precise the term "deterministic." Though the treatment if self contained, readers are strongly advised to consult Brock (1986) or the extensive survey of Eckman and Ruelle (1985). In Section III we explain the data and procedure. Section IV presents some numerical results on the application of these algorithms to several stock return series. We also present the results of the tests developed by Brock, Dechert, and Scheinkman (1986) (henceforth BDS) that produced a distribution theory for statistics based on the Grassberger-Procaccia-Takens measure of correlation dimension. Since the BDS results are asymptotic, we examined the finite sample distribution of these statistics with a method inspired by bootstrapping. Finally, we examine whether the results could be explained by a particular nonlinear alternative—the ARCH models (see Engle 1982).

The results presented here are of a clearly preliminary nature and some of the difficulties are discussed in Section V. Nonetheless it seems that the algorithms are able to distinguish between the real data and those generated by the linear stochastic difference equations that have been used to explain asset returns. In particular they show the inadequacy of the "random-walk" theory (e.g., Granger and Morgenstern 1963; Fama 1970) that states that returns are independently and identically distributed over time. The results suggest that nonlinearities may play an important role in explaining asset returns. In Section V we also present some conclusions and suggestions for further work and speculate on the meaning of the results to complete markets rational-expectations asset-pricing theories.

It is important to underscore that our primary concern in this article is to use new tests to detect nonlinear departures from random-walk behavior in stock returns. There are, of course, many possible alterna-

1. See Sakai and Tokumaru (1980) where this and other examples are given.

tives to the random-walk model, and we merely select and analyze one alternative hypothesis, the ARCH model, for comparison purposes only. Further, in considering departures from alternatives to the random-walk model, the distribution of the test statistics needs to be ascertained for each particular alternative, and that we have not done.[2]

It is also important to point out that the tests discussed here give no guidance as to the approximate form of nonlinearity that may be present in the data. There are, of course, a large number of stochastic, nonlinear models that have been studied in the statistics literature besides the ARCH model discussed above. Examples include the bilinear models (Granger and Anderson 1978, Subba Rao and Gabr 1980) and the threshold autoregressive models (Tong and Lim 1980). Further, one should also include changes in the variance that are not captured by ARCH. For an example of how to deal with this last point in the context of a different data set, see Scheinkman and LeBaron (1988).

Though the tests developed here can, in principle, be generalized to deal with the estimated residuals from any of these models, we have not done so. We do think this is a necessary next step, but the aim of this article is merely to suggest the application of these new techniques to financial data.

II. Description of the Test

We start by making more precise the notion of a "deterministic explanation." The state of the economy is assumed to be a vector $x \in R^n$ for some n. This state changes over time according to a "law" $f\colon \mathrm{R}^n \rightarrow \mathrm{R}^n$, that is, $x_{t+1} = f(x_t)$. The function f is often assumed to be continuously differentiable or at least Lipschitz. We assume that the "orbit" $\{x_t\}_{t=0}^{\infty}$ is bounded and that enough time has elapsed so that x_0 may be taken as (almost) belonging to the Ω-set generated by x_t, that is,

$$x_0 = \lim_{k \rightarrow \infty} x_{t_k},$$

for some subsequence $t_k \rightarrow \infty$. Unfortunately, we have no clue as to what the relevant components of x or its dimension are. In fact, at this point we only use a single real number $y_t = h(x_t)$ (the observable), where h is continuously differentiable.

Clearly for some functions h (e.g., the constant function) not much can be learned about the evolution of x_t by examining the y_ts. A theorem due to Takens (1983*a*) can be used to show that the situation is much better for "most" h's. For each fixed N, consider

$$\phi_N(x) = \{h(x), h[f(x)], \ldots, h[f^{N-1}(x)]\},$$

2. But see LeBaron (1988).

which maps R^n into R^N. If $x_0 = x$, then $\phi_N(x) = (y_0, y_1, \ldots, y_{N-1})$, that is, the history of the observable for the first N periods. Takens (1983*a*) proves that, if $N \geq 2n + 1$, then, "generically," ϕ_N is one-to-one, and, further, $D\phi_N(x)$ is one-to-one at each x. In other words, $2n + 1$ long histories of the y's are observationally equivalent to the x's. For our purposes this will suffice, though the fact that we do not know how large n is will complicate matters. The following definition is due to Grassberger and Procaccia (1983*a*) and Takens (1983*b*).

DEFINITION 1. The correlation dimension of a set

$$\{x_t\}_{t=0}^{\infty}$$

is $\lim_{\gamma \to 0}[\ln C(\gamma)/\ln \gamma]$, if this limit exists, where $C(\gamma) = \lim_{M \to \infty} C_M(\gamma)$,

$$C_M(\gamma) = \frac{2}{M(M-1)} \sum_{1 \leq i < j \leq M} \theta(\gamma - |x_i - x_j|),$$

and

$$\begin{aligned} \theta(a) &= 0 \quad \text{if } a < 0, \\ \theta(a) &= 1 \quad \text{if } a \geq 0. \end{aligned}$$

REMARK 1. $C(\gamma)$ measures the fraction of the total number of pairs (x_i, x_j) such that the distance between x_i and x_j is no more than γ. Intuition about how the correlation dimension is a measure of "dimension" can be obtained by considering two examples. In the first one (fig. 1), points of the set $\{x_t\}_{t=0}^{\infty}$ are uniformly distributed on a line segment in R^2. In the second one (fig. 2), points are uniformly distributed on a "square" in R^2. It is clear that in the first case for γ small, if we double γ, "most" points (i.e., all but the ones close enough to the boundary) gain twice as many neighbors. While in the second one, "most" gain four times as many neighbors.

REMARK 2. Grassberber and Procaccia (1983*b*) discuss properties of this measure and its relationship with more usual measures of dimension such as the Hausdorff dimension, and so on. They also present some estimates of the dimension of computer-generated solutions to the Henon map, Mackey-Glass equation, and others.

Further, since ϕ_N is an embedding for $N \geq 2n + 1$, the correlation dimension of the orbit $z_t = \phi_N(x_t)$ is the same as that of x_t, provided N is large. Thus, in principle, one can do without measurements of the unobservable x_t by utilizing, in its place, vectors $z_t = (y_t, \ldots, y_{t+N-1})$ of long enough histories (i.e., $N \geq 2n + 1$) of the observable. The length N of the vector z_t is called the embedding dimension. For each N, we let

$$C_m^N(\gamma) = \frac{2}{m(m-1)} \sum_{1 \leq i < j \leq m} \prod_{k=0}^{N-1} [\theta(\gamma - |y_{i+k} - y_{j+k}|)],$$

where m is chosen such that $m < M - N + 1$.

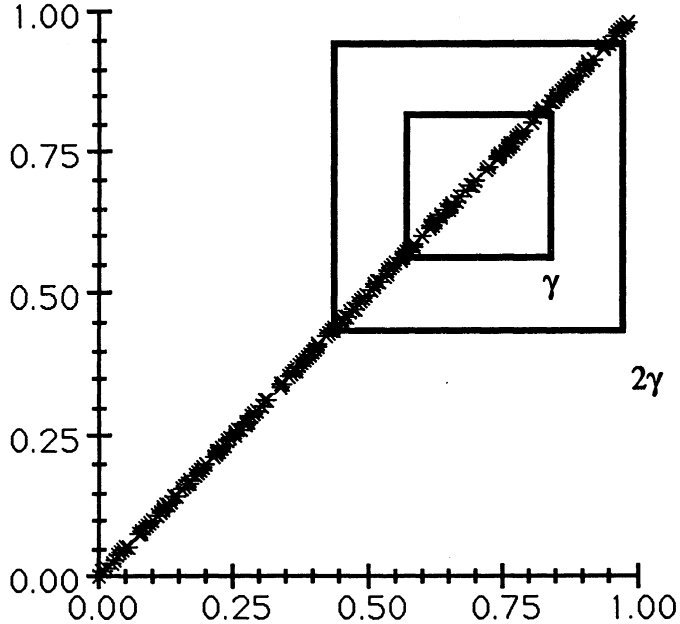

FIG. 1.—Uniform line

If $\{u_t\}_{t=0}^{\infty}$ is the outcome of a sequence of random experiments that are independently and identically distributed (henceforth i.i.d.) and have a nondegenerate density, and if $w_t^N = (u_t, \ldots, u_{t+N-1})$, then the correlation dimension of w_t^N is N. The reader can easily convince himself of this fact by considering the case where each experiment is uniformly distributed on [0,1]. Then w_t^N is uniformly distributed on $[0,1]^N$, and, as in remark 1 above, the dimension of w_t^N is readily seen to be N. The proof goes through for the case of an ergodic Markov process with a nondegenerate stationary density. Brock (1986) establishes an analogous result for other measures of dimension. Thus, the finding of a correlation dimension that does not grow with N is indicative of the existence of a deterministic explanation, although one cannot rule out all random phenomena.

In reality, however, the presence of a finite data set presents serious limitations. In estimating the correlation dimension from the data, one plots $\log[C_M(\gamma)]$ against $\log(\gamma)$, where M is the cardinality of the data set. Clearly this cannot work for γ too small with a finite data set, and thus one must be content to find a linear segment "close to" zero. Takens (1984) proposes a different method of estimating, which when tried yielded similar results. Takens's estimates consist of first assuming that the distance between z_ts are independent and distributed as γ^{-d}, and then estimating d by maximum likelihood. There are some problems with this estimate. First, once the embedding dimension ex-

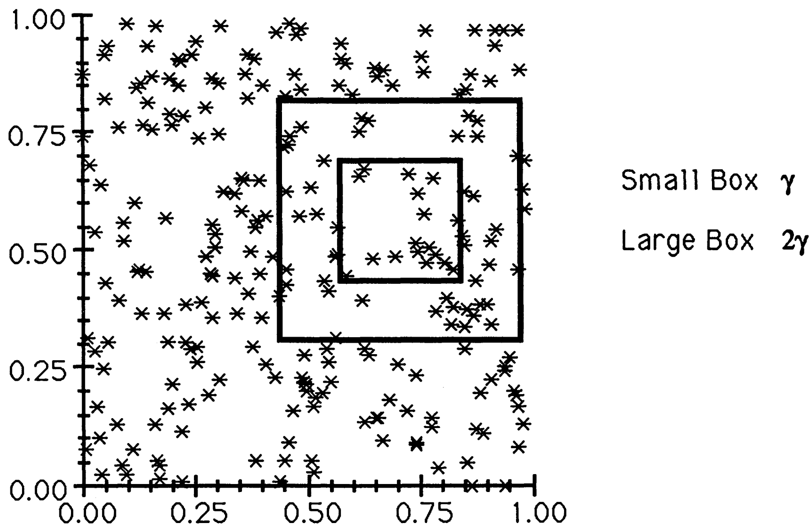

FIG. 2.—Uniform square

ceeds 1, the z_ts are not independent even if the original data is i.i.d. Second, the Takens estimator is an unbiased estimator of $1/d$, and therefore it is an upward biased estimator of d. Nonetheless, this estimate works rather well on numerical examples, and we present it here.

Much work has also been done in the study of deterministic systems "contaminated" by random noise. By looking at deterministic systems where noise distributed uniformly $(-a,a)$ is added to a system of known dimension, many researchers (e.g., Zardecki 1982; Ben-Mizrachi, Grassberger, and Procaccia 1984; and Atten and Caputo 1985) have found that the graph of $\log[C_m^N(\gamma)]$ against $\log(\gamma)$ has the slope of the embedding dimension for $\gamma \leq a$ and the slope of the dimension of the deterministic system above that level. Thus, at a certain scale, one observes behavior as in a random system, while at a larger one, one sees the deterministic motion. Since our data set is much smaller than the ones studied in these references, we performed similar experiments with the number of points equal to the cardinality of our data set and obtained very clearly the same type of results. Takens (1984) also discusses the effects on his measure in the presence of noise. He concludes that it also raises the estimate of the dimension.

Let us write $C^N(\gamma)$ for the quantity defined in definition 1 above when the embedding dimension is N. The quantities $S^{N+1}(\gamma) = C^{N+1}(\gamma)/C^N(\gamma)$ have an interesting interpretation when the *supremum norm* (sup norm) is used. They give you an estimate of the conditional probability that

$$\sup_{0 \leq i \leq N} |y_{t_{1+i}} - y_{t_{2+i}}| \leq \gamma,$$

given that

$$\sup_{0 \le i < N-1} |y_{t_{1+i}} - y_{t_{2+i}}| \le \gamma.$$

This is just the conditional probability that two points are close given that their past N histories are close. For fixed γ, S^N should be independent of N if the y_ts are independent (note that this does *not* assume the existence of a density). On the other hand, if past y_ts help predict future ones, S^N will tend to increase with N (this of course depends on the system not expanding "too much"). Thus the behavior of S^N, as N varies, gives us a measure of departures from independence.

Until recently there was no distribution theory for these types of statistics. This gap has been filled by Brock, Dechert, and Scheinkman (1986) (BDS), who introduced asymptotic distribution theory for some of these types of statistics. We will present two of these statistics for the weekly stock returns data that are closely related to $C^N(\gamma)$ and $S^N(\gamma)$.

Let x_t be a series of length M, and choose an m such that $m \le M - N + 1$. This m is used to shorten the original series so that the embedded series is well defined. Under the assumption that x_t is independent, identically distributed, BDS prove for any $N > 1$, $\gamma > 0$, as $m \to \infty$,

$$\sqrt{m}\{C_m^N(\gamma) - [C_m^1(\gamma)]^N\} \xrightarrow{d} N(0, V_C),$$

$$\sqrt{m}[S_m^N(\gamma) - C_m^1(\gamma)] \xrightarrow{d} N(0, V_S).$$

Then they develop consistent estimators for V_C and V_S. Formulas for these estimators are in the Appendix.

III. Description of Data and Procedure

The initial data set consisted of 5,200 + daily returns (including dividends) on the value-weighted portfolio of the Center for Research in Security Prices at the University of Chicago (CRSP). From this initial data set we constructed a weekly returns series. The latter is, in principle, less "noisy" since the daily returns are sensitive to weekend effects. Most of our results concern the weekly data but we also looked at other series constructed from the original data set.

We decided to compare the correlation dimension, the measure S^N, and the Takens estimate for the data with the ones from computer-generated solutions to possible alternative models. Comparison data sets were created in the following manner: first, returns were regressed on past returns. Then we sampled (with replacement) from the residuals and rebuilt our data set using the estimated linear system and the same initial values as in the real data (from now on we refer to such data as "scrambled data"). Several possible dimensions for this linear model were tried. Note that these "random" data were in fact deterministic since they were generated by computer. However, random number generators mimic randomness well, and in our case the resulting "data" seemed much more random than the initial ones.[3]

For the BDS statistics, we compare statistics measured on residuals of linear models fit to the original series with residuals of linear models fit to the above-mentioned scrambled data. (This process is closely related to Efron [1982, ch. 5].) This careful procedure is followed for two reasons. First, the BDS statistics are sensitive to any deviation from i.i.d., linear or nonlinear, and we want to make sure that linear effects have been removed. Second, fitting a linear model to a time series induces some dependence in the residuals. By comparing our actual data with residuals of fitted models, we can determine if this dependence is affecting our results.[4]

There are, of course, no a priori estimates for the dimension of the state vector. This leads to difficulties in selecting the embedding dimension (i.e., the length of the vector of histories we will want to consider). This situation is made worse by the presence of noise of unknown amplitude. For a given small γ, as the embedding dimension is increased the fraction of pairs within γ falls for two reasons. First, it is harder, in the presence of noise, for two long histories to be within γ of each other. Second, even if noise were absent, with finite data an increase in the embedding dimension by itself lowers the fraction of points within γ. Thus while for infinite data an increase in the embedding dimension after a certain N would not affect the estimates, it certainly does have an effect with the number of points at our disposal. Hence we cannot let the data tell us about the level of noise by choosing larger embedding dimensions.

An approach to this problem was a result of our belief that in reality the relevant state vector x_t is of very large (essentially infinite) dimension. By choosing an "arbitrary" noise amplitude γ we are saying that the movements on the data below that scale are "random," and we are

3. Simple congruential random number generators are formed as follows:

$$z_{t+1} = (az_t + c) \bmod 1,$$

for some a, c. Thinking of the random number generator as a mapping on the unit square, it is clear how it "looks" random. The graph of the random number generator would be formed by parallel stripes covering the square. If the fineness of these stripes were small relative to the sizes of γ used, the random number generator would appear to fill the entire space. When γ is small relative to the distance between the stripes, the system will show up as a deterministic system, which it is. With a long enough series, and small enough γ, any random number generator will be detected using these tests, but for the small samples used here the random number generator appears random to these tests.

4. The weekly returns data show only a weak correlation (.09) at lag 3. The numerical results were the same whether we used the filtered or the unfiltered data. The daily index shows a stronger correlation (.23) at lag 1, and this affected the BDS statistics. It did not, however, affect the dimension estimate. For results on BDS statistics on fitted residuals see Brock, Dechert, Scheinkman, and LeBaron (1988).

trying to see whether movements above that scale could be explained by a small number of factors. In practice, this consists of choosing $\underline{\gamma}$ and embedding dimension N such that for $\gamma \in [\underline{\gamma}, \underline{\gamma} + \epsilon]$ for some $\epsilon > 0$, $\log C^N(\gamma) = d\log(\gamma) + k$ (i.e., a line segment of slope d) and that for embedding dimensions $N + i$, $1 \le i \le \bar{i}$, again $\log C^N(\gamma) = d\log(\gamma) + k$, for $\gamma \in [\underline{\gamma}, \underline{\gamma} + \epsilon]$, that is, increasing the length of past histories (the embedding dimension) would not change the correlation dimension of the data at scale $\underline{\gamma}$. We also checked whether Takens's estimates of d stayed constant across dimensions and across γ's close to $\underline{\gamma}$.

While we were able to do this for our weekly data (and less successfully for the daily data), we show that for the scrambled data the estimated d kept increasing with N even after we conceded a large role to noise.

IV. Numerical Results

A. Results on Weekly Returns

Figure 3 presents the estimates $S^N(\gamma)$ of the conditional probabilities for different values of γ corresponding to multiples of the original standard deviation of the data. The value of $S^N(\gamma)$ goes up with N for each γ until $C^N(\gamma)$ gets too small. In figure 4, the estimates $S^N(\gamma)$ for a "typical" scrambled returns series is presented. Note in figure 4 $S^N(\gamma)$ is essentially constant until N gets large and the number of points counted gets small, making the $S^N(\gamma)$ ratio unstable. Figure 5 shows the ratios $S^N(\gamma)/S^1(\gamma)$ (where γ = one-half the standard deviation of the weekly returns) for the weekly returns, and the highest, lowest, and median values of this ratio from a sample of 156 scrambled returns.[5] As expected, the scrambled returns exhibit a ratio clustered around one. However, the data set has the ratio outside the range of the "scrambled" ones. The exact same picture appears for different values of γ. These results indicate that patterns of past returns help predict future ones, though, of course, they do not necessarily establish that one can use past returns to improve the prediction of future *mean* returns.

In figure 6, plots of $\log C^N(\gamma)$ against $\log(\gamma)$ for embedding dimensions 1, 2, 12, and 13 are shown. In this figure, as in all of this type, a plot for a higher dimension lies below one of a lower dimension (this is a result of using the sup norm). Notice that the change is much higher when one goes from dimension 1 to dimension 2 than when we go from 12 to 13. This becomes even clearer when we look at the diagram for $\gamma \ge \underline{\gamma} = .5\sigma$ (σ will refer to the sample standard deviation of each series). The estimated slopes are in table 1.

5. In computing the statistics on $S^{N+1}(\ )/S^N(\ )$, a given scrambled return was dropped from the sample once the number of pairs within γ dropped below 50.

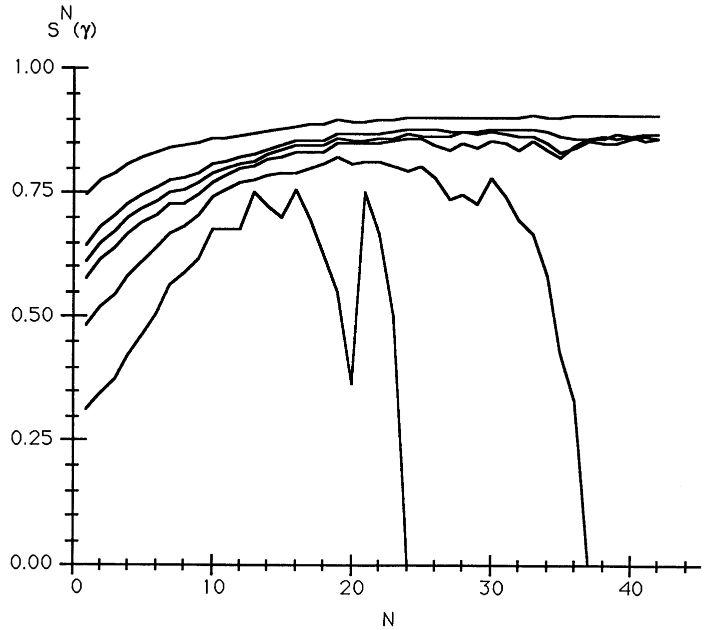

FIG. 3.—Value-weighted weekly conditionals: γ = .5, .8, .9, 1, 1.1, 1.2, 1.5 std. (higher curves correspond to higher γ's).

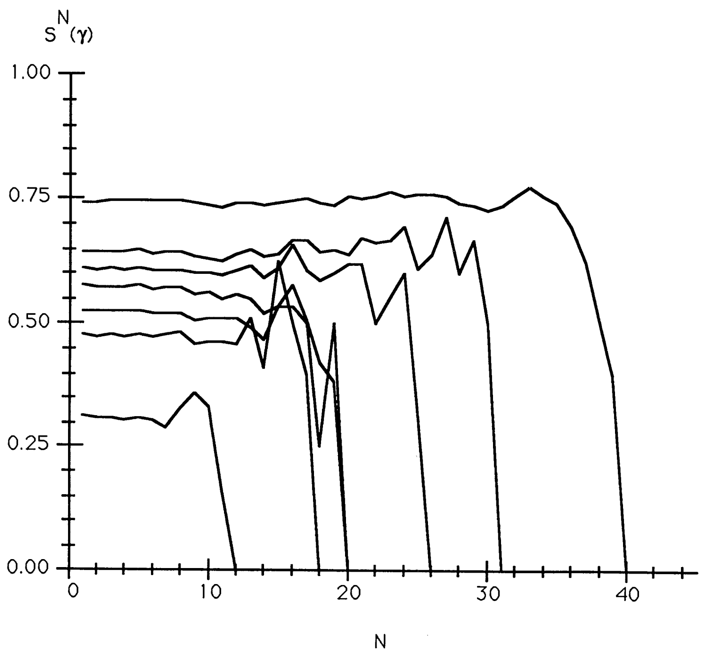

FIG. 4.—Scrambled value-weighted weekly conditionals: γ = .5, .8, .9, 1, 1.1, 1.2, 1.5 std. (higher curves correspond to higher γ's).

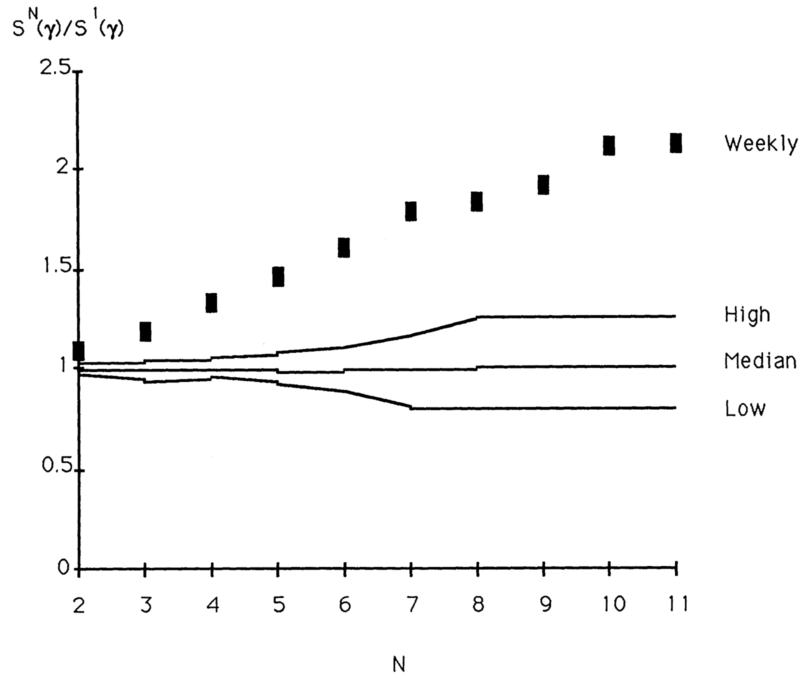

FIG. 5.—Conditional probability ratios: γ = .5 std.

As mentioned in Section II, if a system is subject to noise distributed uniformly on $[-a,a]$, then the dimension of the system should be the estimated slope at $2a$. Hence a choice of a γ implies the admission of a noise term with a standard deviation that equals

$$(\sqrt{3}/6)\gamma.$$

In this case, our choice of γ is such that the standard deviation of the required noise is less than 15% of the original standard deviation of the series. For this γ the estimated slope (5.7) is unchanged for embedding dimensions 12, 13, and 14. It is interesting to note that if, in fact, the state variable x_t were to converge to a set of (fractal) dimension 5.7, we may need to take an embedding dimension of 13 to obtain the correct estimate of d. It is useful to compare these results with the one on figure 7 where the same embedding dimensions are studied for a "typical" scrambled data set. Notice that in such embedding dimensions the scrambled data shows no pairs which are at least as close as γ. This is to be expected due to the "randomness" of the data. Notice that as predicted the estimated slope of the scrambled weekly returns grows with the embedding dimension (see table 1). A search through several

TABLE 1 Dimension Estimates for Weekly Returns

Embedding Dimension	Estimated Slope	Takens Estimate
Weekly returns series ($\gamma = .5\sigma$):		
1	.9	1.0
2	1.7	1.9
12	5.7	6.3
13	5.7	6.3
14	5.7	6.3
Scrambled weekly returns ($\gamma = .8\sigma$):		
1	.6	.9
2	1.2	1.7
12	9.5	10.1
13	10.6	11.2

dimensions and several possible γ's showed similar patterns for the scrambled data.

Although the stable dimension estimates for the unscrambled data are very interesting, we cannot be confident with the estimation of these numbers until their properties are better understood.[6] What is interesting here is the strong difference between the unscrambled and scrambled series, indicating nonlinear dependence in the data.

It is not our purpose to investigate thoroughly the behavior of individual stock returns. However, we did take a random sample of the stocks on the CRSP data base and applied our procedure to their returns. Of these, those of Abbott Laboratories stock were the ones that looked most "random." Figure 8 presents the estimates of $S^N(\gamma)$ for different values of γ corresponding to multiples of the original standard deviation of the data. Figure 9 presents the same estimate for a "typical" scrambled data set. Notice that they hardly differ, indicating the failure of past patterns of returns in predicting future ones. Figure 10 presents the plot of $\log[C^N(\gamma)]$ against $\log(\gamma)$ for embedding dimensions 1, 2, 12, 13. Figure 11 presents the same plots for a "typical" scrambled data set. Table 2 under Abbott Laboratories and scrambled Abbott Laboratories contains the dimension estimates for the series and its scrambled counterpart. Here we fail to distinguish between scrambled returns and the original data. It is natural to expect that single stock returns would be subject to idiosyncratic noise that would disappear in a comprehensive index like the one we use. In such a case one would expect that individual stock returns would look much more "random," as is the case of this example.

6. See Ramsey and Yuan (1987) for some simulations of these types of estimators.

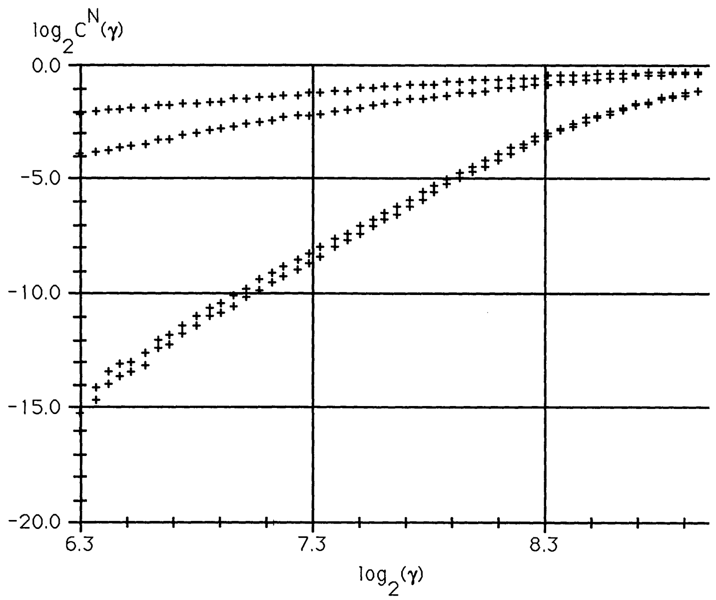

FIG. 6.—Value-weighted weekly returns: $N = 1, 2, 12, 13$ (higher curves correspond to lower N's).

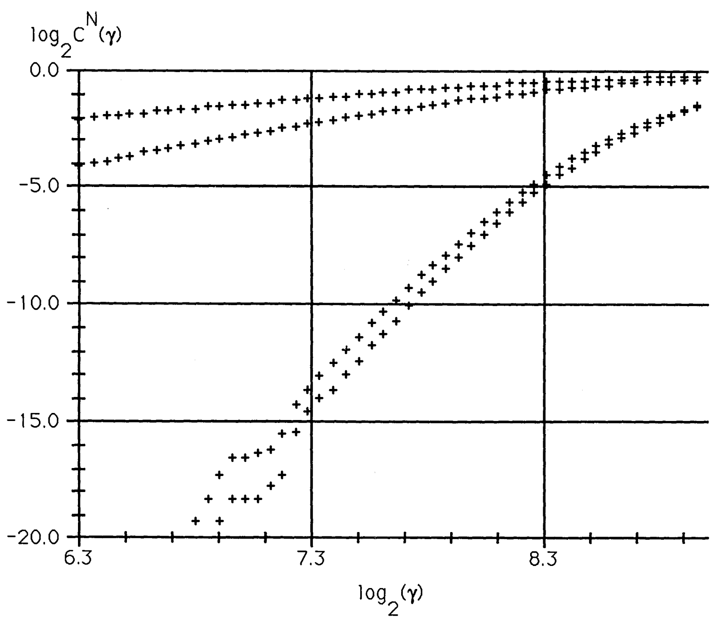

FIG. 7.—Scrambled value-weighted weekly returns: $N = 1, 2, 12, 13$ (higher curves correspond to lower N's).

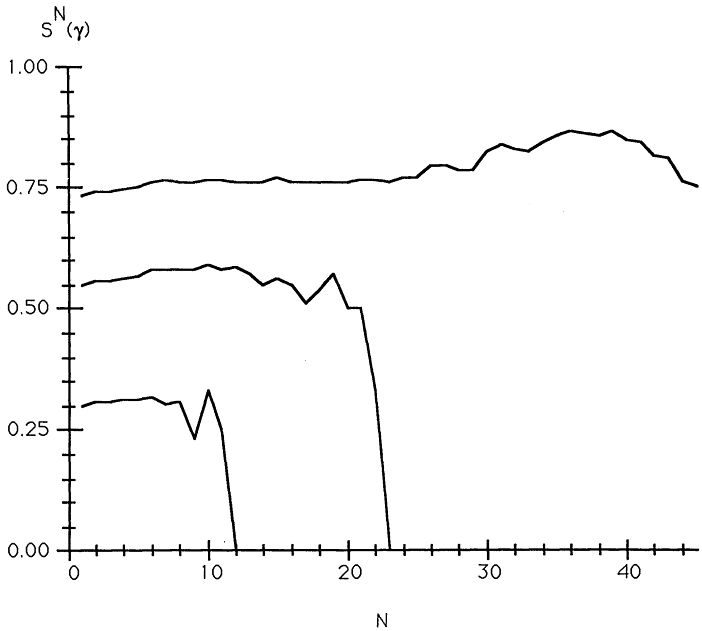

FIG. 8.—Abbott Labs conditionals: γ = .5, 1, 1.5 std. (higher curves correspond to higher γ's).

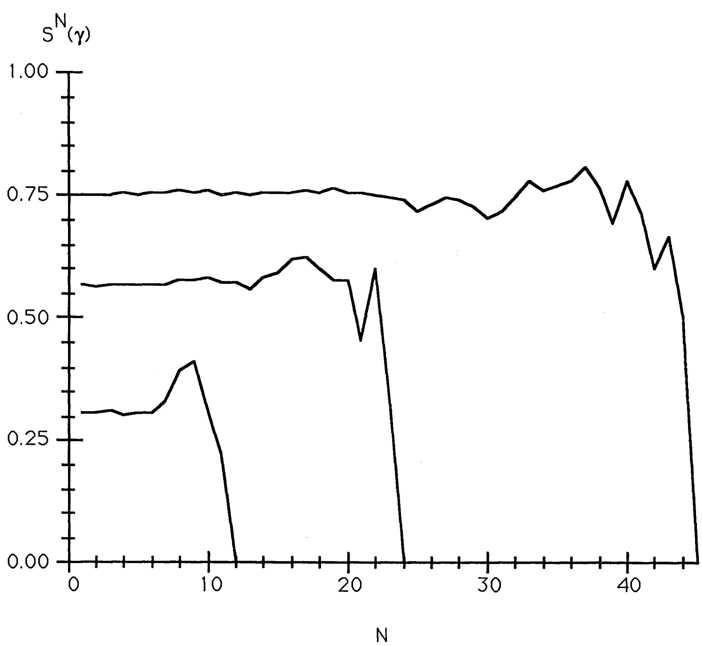

FIG. 9.—Scrambled Abbott Labs conditionals: γ = .5, 1, 1.5 std. (higher curves correspond to higher γ's).

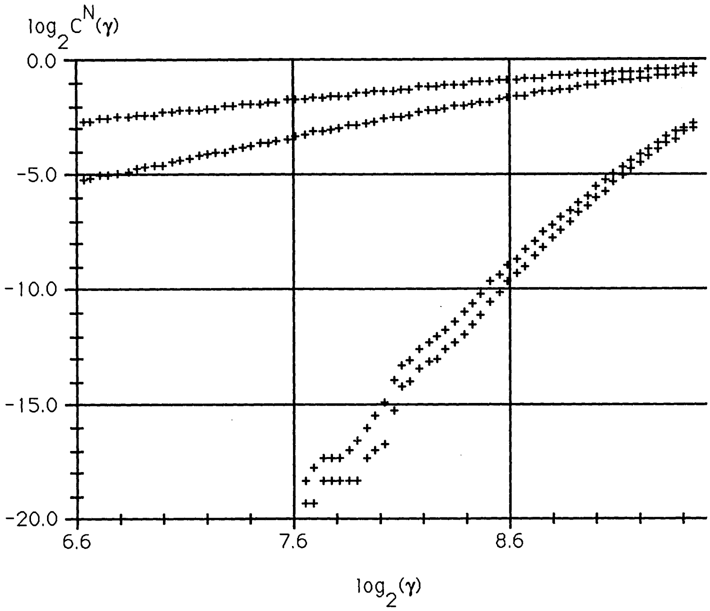

FIG. 10.—Abbott Labs weekly returns: $N = 1, 2, 12, 13$ (higher curves correspond to lower N's).

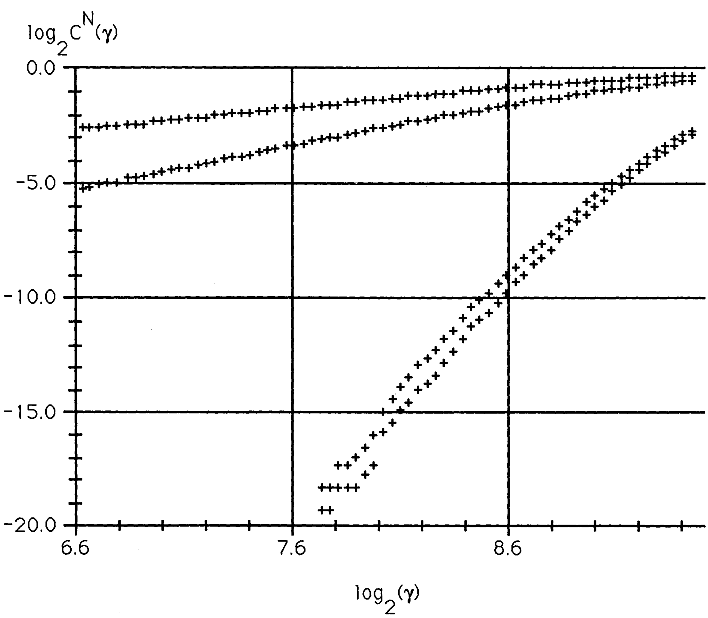

FIG. 11.—Scrambled Abbott Labs weekly returns: $N = 1, 2, 12, 13$ (higher curves correspond to lower N's).

TABLE 2 Dimension Estimates for Abbott Laboratories Stock

Embedding Dimension	Estimated Slope	Takens Estimate
Abbott Laboratories series ($\gamma = \sigma$):		
1	.6	.9
2	1.3	1.7
12	8.0	8.5
13	8.7	9.2
Scrambled Abbott Laboratories series ($\gamma = \sigma$):		
1	.6	.9
2	1.3	1.3
12	8.1	8.6
13	8.8	9.3

TABLE 3 BDS Statistics for Weekly Returns

N	$\zeta_C^N(\gamma)$	$\zeta_S^N(\gamma)$
2	8.2(0)	8.2(0)
3	10.7(0)	10.0(0)
4	15.2(0)	14.3(0)
5	21.2(0)	16.7(0)

NOTE.—Values in parentheses are explained in text.

B. *BDS Statistics*

Table 3 presents results for the statistics

$$\zeta_C^N(\gamma) = \frac{\sqrt{m}\{C_m^N(\gamma) - [C_m^1(\gamma)]^N\}}{\sqrt{V_C}}$$

and

$$\zeta_S^N(\gamma) = \frac{\sqrt{m}[S_m^N(\gamma) - C_m^1(\gamma)]}{\sqrt{V_S}}.$$

Under the assumption of independent, identically distributed returns, these statistics are asymptotically distributed $N(0,1)$.

They are estimated on residuals of the weekly returns series with $m = 1{,}109$, $\gamma = .5\sigma$. The numbers in parentheses represent the number of scrambled comparison runs (out of 100 runs) giving a statistic at least as large in absolute value. Each scrambled comparison run rebuilds the initial estimated linear model and then fits a linear model to this series, estimating the BSD statistics on the residuals of this model. Table 3 clearly shows that the tests are rejecting the hypothesis that the data are i.i.d.

C. Results on Daily Returns

The daily returns series shows strong correlation between consecutive days. Since it is our objective to display the role of nonlinearities, we looked at both the original daily returns and the residuals obtained after a linear regression of returns on past returns. Figure 12 represents the plots of $\log[C^N(\gamma)]$ against $\log \gamma$ for embedding dimensions 1, 2, 19, 20, for the original series, and figure 13 shows the plots for the residual.

Table 4 presents the slope estimates for $\gamma = .9\sigma$ on the residual series. Such a γ implies a readout error of 26% of the standard error of the residual data set. Notice that this is larger than in the case of the weekly data, but this is to be expected. Further, the estimated slope was still somewhat sensitive to the embedding dimension, contrary to what was achieved in the use of weekly returns.

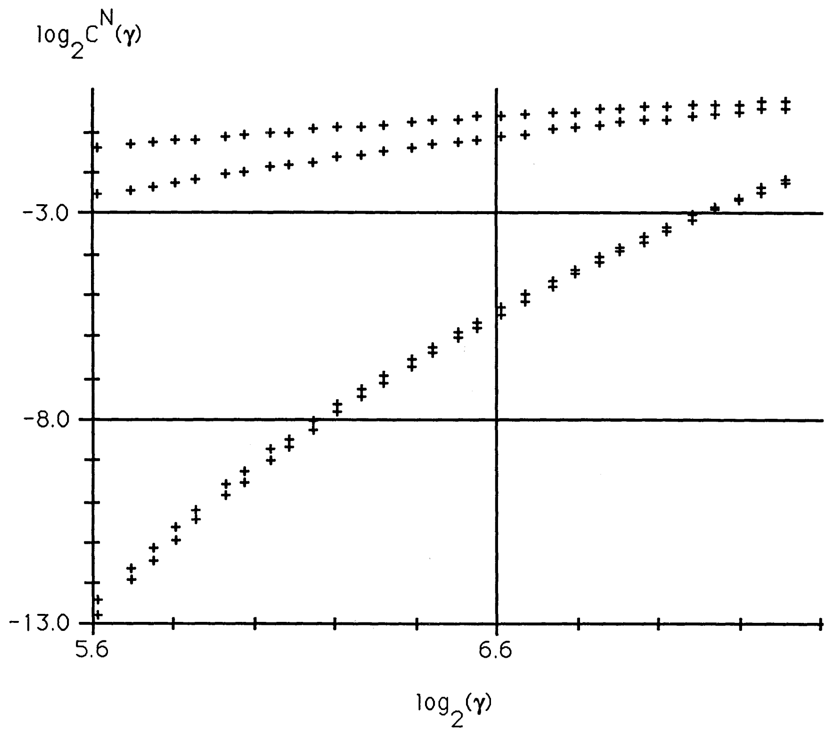

FIG. 12.—Daily value-weighted returns: $N = 1, 2, 19, 20$ (higher curves correspond to lower N's).

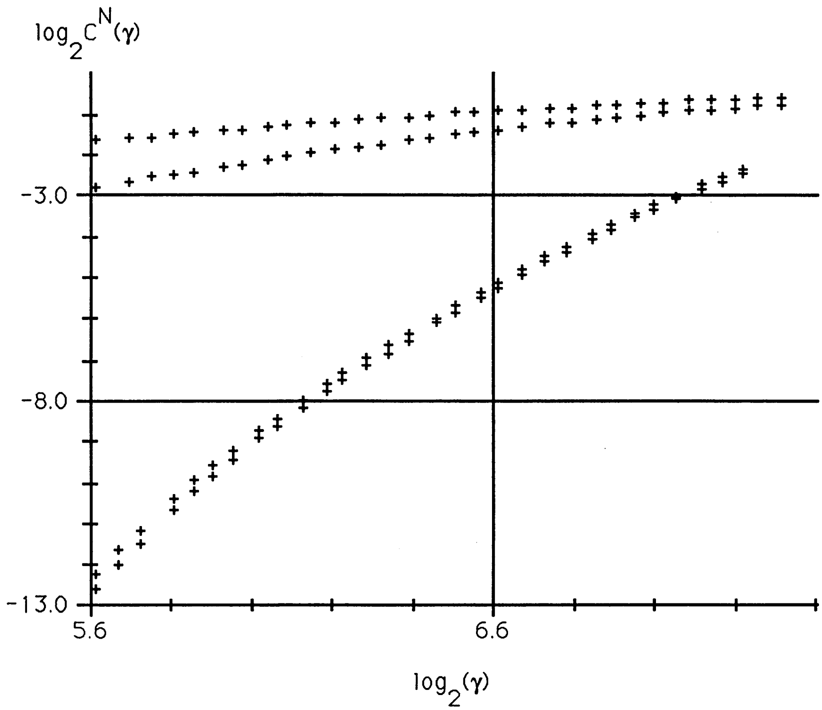

FIG. 13.—Daily value-weighted returns residuals: $N = 1, 2, 19, 20$ (higher curves correspond to lower N's).

TABLE 4 Dimension Estimates for Daily Returns (Residual Series)

Embedding Dimension	Estimated Slope	Takens Estimate
1	.6	.9
2	1.1	1.6
19	5.7	6.4
20	5.9	6.6

NOTE.—$\gamma = 0.9\sigma$ for all dimensions.

REMARK (from Brock 1986). If a system is deterministic and if ϵ_t are residuals generated by a regression of the type

$$y_t = \sum_{i=1}^{I} \alpha_i y_{t-i} + \epsilon_t,$$

then the correlation dimension of ϵ_t is the same as the one generated by y_t. This follows from the fact that, if $y_t = h(x_t)$ and $x_t = f(x_{t-1})$, then $\epsilon_t = g(x_{t-i})$ for an appropriate g. Based on this observation, Brock pro-

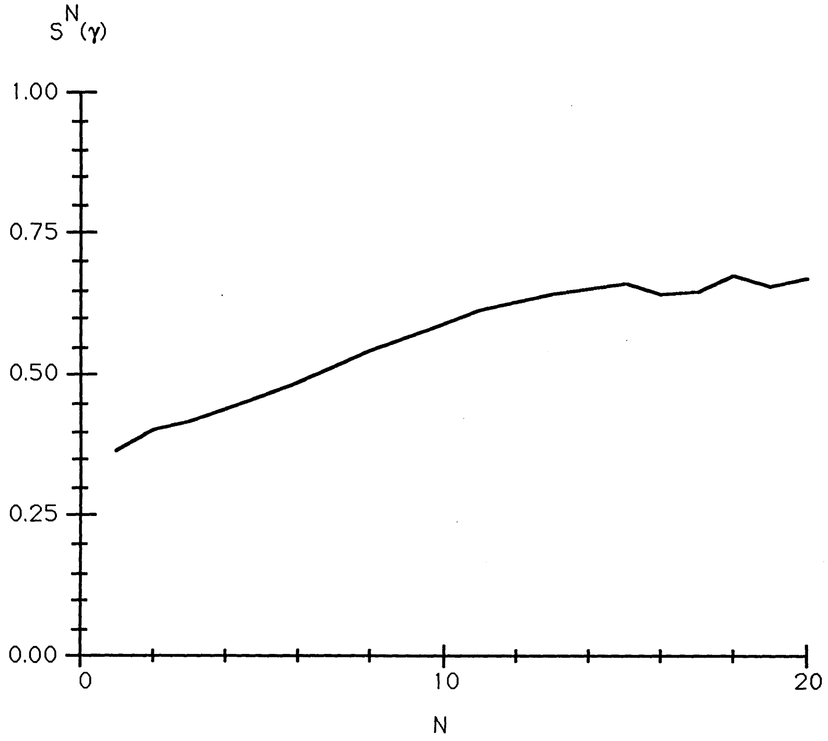

FIG. 14.—Daily residual conditionals: $\gamma = .5$ std.

posed a test for determinism that consisted of regressing the data on past values and looking at the residuals. The estimated dimension should then be the same.[7] Our data pass Brock's test without any difficulty.

Figure 14 plots $S^N(\gamma)$ for the residuals of a regression of returns on past returns with $\gamma = .5\sigma$. Again $S^N(\gamma)$ goes up, indicating that patterns of past residuals help predict future ones.

D. Comparison with ARCH Models

The ARCH models (see Engle 1982) are a nonlinear stochastic alternative that has been fit to economic time series. A version of such models is given by the following points: (1) The distribution of y_t conditional on past y_ts is normal with mean $\bar{y}_t$ and variance $\bar{v}_t$; (2) $\bar{y}_t = \bar{y} + \alpha\bar{y}_{t-1} + \beta \bar{v}_t$; (3) $\bar{v}_t = \bar{v} + \sum_{i=1}^{I} \delta_i(y_{t-i} - \bar{y}_{t-i})^2$. We estimated (by maximum likelihood) a model following points 1, 2, and 3 on our weekly data set.

7. It should be noted that Brock's test covers the purely deterministic case. If one allows noise to enter as in $y_t = h(x_t) + \mu_t$, and $x_t = f(x_{t-1})$, where μ_t is independent over time and μ_t is uniformly distributed in $[-a,a]$, then $\epsilon_t = y_t - \alpha y_{t-1}$ will satisfy $\epsilon_t = g(x_{t-1}) + \mu_t - \alpha\mu_{t-1}$, and thus ϵ_t will be subject to a "larger" error.

Using the Akaike information criteria, we selected an optimal $I = 11$. We then looked at the residuals

$$\epsilon_t = (y_t - \bar{y}_t)/\sqrt{\bar{v}_t},$$

which according to the model should be distributed as $N(0,1)$. We also looked at "scrambled" residuals and to data generated by using the ARCH model with the estimated coefficients.

REMARKS. Following Brock's remark discussed above we can see that, if $y_t = h(x_t)$ where $x_{t+1} = f(x_t)$, then $\epsilon_t = g(z_t)$ where $z_t = F(z_{t-1})$. It suffices to choose $z_t = (x_{t-I}, \bar{y}_{t-1}, \ldots, \bar{y}_{t-I})$. But, when I is large (as it is in the best fit), the increase on the estimated dimension will be quite large, and with the number of data points we are dealing with it will be hard to distinguish them from random residuals. Further, even if we look at the residuals of an ARCH regression with a low I (say $I = 1$), if errors are present (i.e., $y_t = h[x_t] + \mu_t$), the error of the residuals will be changed and, in particular, will no longer be i.i.d.

Figure 15 gives the estimates of the conditional probability of a typi-

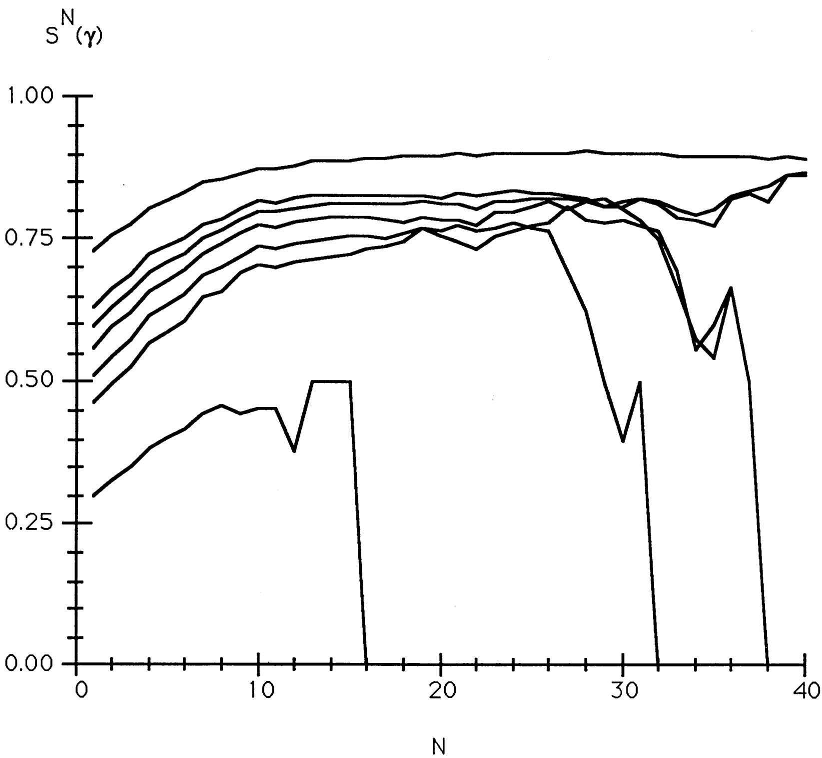

FIG. 15.—Generated ARCH (11) conditionals: γ = .5, .8, .9, 1, 1.1, 1.2, 1.5 std. (higher curves correspond to higher γ's).

cal generated ARCH run with our maximum-likelihood estimates at the optimal I ($I = 11$). Notice that the curve for $\gamma = .5\sigma$ shows that the ARCH-generated data exhibits the "dependence" one expects. A comparison with figure 2, however, shows that the original data set exhibits a much stronger dependence than the generated ARCH data.

Figure 16 exhibits the plots of log $C^N(\gamma)$ against log(γ) for dimensions $N = 1, 2, 12$, and 13. As is expected from "random" series, one cannot estimate the dimension at γ as low as the γ used for the weekly data. Under the entry Simulated ARCH, table 5 presents the slope estimates for $\gamma = .8\sigma$. Contrary to what happens with the weekly data, the estimates are also quite sensitive to your choice of γ. Thus the generated ARCH models look much more random.

The ARCH residuals taken from $I = 11$ lead to high estimated dimension, but this is to be expected since, in principle, one could be adding 11 dimensions to the original dimension of the weekly data set. To show that this resulted from the transformation of the data, we reestimated the model for $I = 1$. Our slope estimates for this series at $\gamma = .6\sigma$, and the scrambled version of this series at $\gamma = .8\sigma$, are in table

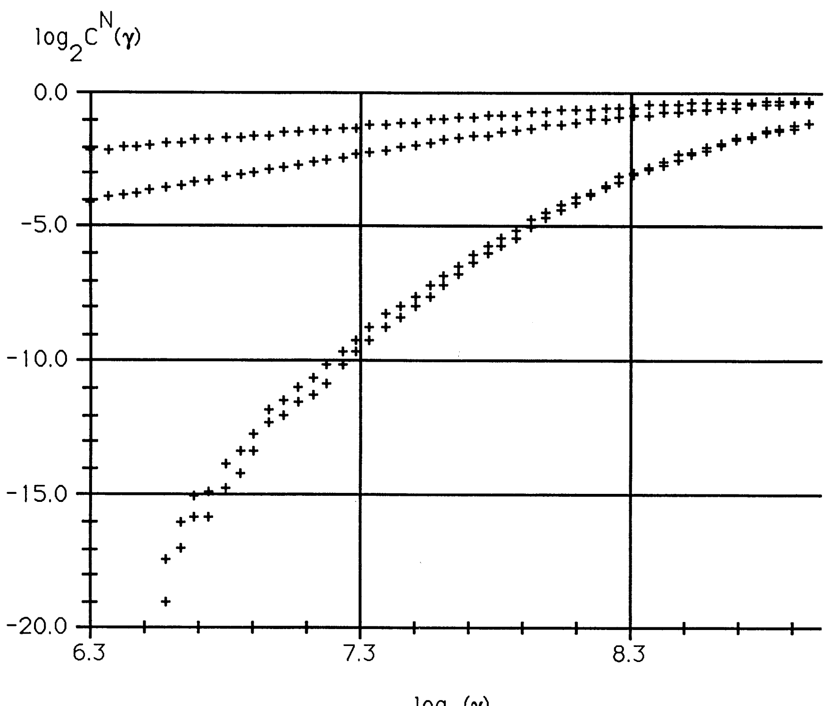

FIG. 16.—Generated ARCH (11): $N = 1, 2, 12, 13$ (higher curves correspond to lower N's).

TABLE 5 Dimension Estimates for ARCH Simulation and Residuals

Embedding Dimension	Estimated Slope	Takens Estimate
Simulated ARCH (γ = .8σ):		
12	8.3	9.3
13	8.4	9.7
14	7.6	10.0
ARCH residuals (γ = .8σ):		
1	.7	...
2	1.5	...
12	5.8	...
13	6.0	...
Scrambled ARCH residuals (γ = .8σ):		
1	.7	...
2	1.3	...
12	8.6	...
13	9.1	...

5 under ARCH residuals. Notice that the scrambled residuals look much more random than the original residuals.

V. Conclusion

In the early 1970s the geometric random walk commanded great respect as a description of asset pricing.[8] Now it has been brought into question by several different studies. Calendar anomalies have been found at several frequencies, indicating larger expected returns during certain periods.[9] LeBaron (1988) has tested some of these possibilities and found that they are not the cause of the results seen here. The behavior of the statistics $S^N(\gamma)$ seems to leave no doubt that past weekly returns help predict future ones even though they are uncorrelated. Further, it seems that a substantial part of the variation on weekly returns is coming from nonlinearities as opposed to randomness. Or, more moderately, the data are not incompatible with a theory where some of the variation would come from nonlinearities as opposed to randomness and are not compatible with a theory that predicts that the returns are generated by i.i.d. random variables.

8. In his well-known 1970 paper, Eugene Fama states that, "indeed, at least for price changes or returns covering a day or longer, there isn't much evidence against the 'fair game' model's more ambitious offspring, the random walk."

9. Anomalies have been detected at annual (Rozeff and Kinney 1976), monthly (Ariel 1987), and weekly (French 1980) frequencies.

The equilibrium asset-pricing theories that followed the work of Lucas (1978) assert that asset prices p_t satisfy

$$p_t = h(x_t), \tag{5.1}$$

and

$$x_t = f(x_{t-1}, \mu_t). \tag{5.2}$$

Here x_t is a vector of "state variables" (typically capital stocks), and μ_t a vector of random variables. If model (5.1)/(5.2) were true and h and f nonlinear, part of the movements on p_t would be caused by the nonlinearities. Though special cases of these asset-pricing theories generate linear (or log-linear) equations, nonlinearities appear in general, and they become even more important once one abandons the complete market framework as in Scheinkman and Weiss (1986). It should be emphasized that most of the mathematics discussed above does not exactly apply to (5.1) and (5.2) but to

$$p_t = h(x_t) + \mu_t, \tag{5.3}$$

and

$$x_t = f(x_{t-1}). \tag{5.4}$$

Here μ_t is an additive measurement error on the price series rather than a shock to the "state variable." Hence the task of showing that the data is compatible with (5.1) to (5.2) (in a nonlinear version) is not complete.[10]

More important, we hope to have convinced the reader that the techniques used here can be useful in testing the "whiteness" of observations and of residuals of proposed models. We presented two tests. The first involves estimates of the conditional probabilities that two N-length "histories" of the variable y_t are close to each other given that the first $N-1$ elements of the "histories" are close enough. The second compares the behavior of the plots of log $C^N(\gamma)$ versus log γ for the proposed series versus the "scrambled" series. This last one can be done in two different ways when testing a model: we can either look at the scrambled residuals, or we can look at generated data obtained from the model and the scrambled residuals.

Using these tests, we examined a series of weekly returns, one of daily returns and a proposed ARCH model. We also compared, as suggested by Brock, the behavior of log $C^N(\gamma)$ versus log(γ) for the

10. We did perform numerical experiments. We let $f(x,0) = 2x$ of $0 \le x \le 1/2$, and $f(x,0) = 2(1 - x)$ if $1/2 < x \le 1$, and $f(x,1) = (1/\lambda)x$ if $0 \le x \le \lambda$, and $f(x,1) = [1/(1 - \lambda)](1 - x)$ if $\lambda < x \le 1$, where $\lambda \in (0,1)$. We studied the output of $x_{t-1} = f(x_t,\mu_t)$, where $\mu_t \in \{0, 1\}$ and μ_t was i.i.d. with prob$\{\mu_t = 1\} = 1/2$. For λ close to $1/2$, the output looked for our test as if it was the outcome of $x_{t+1} = f(x_t,0) + \nu_t$, provided we chose a low embedding dimension.

original data set and the residuals of the proposed models. The results point towards the presence of nonlinearities.

If one adopts a strictly "rational expectations" view, that is, that agents know the price-formation mechanism, even though economists do not, the nonlinearities may well be the result of a "law" as in (5.1) and (5.2). In this case, actual randomness of the returns on the CRSP portfolios studied here is much smaller than what it seems from a linear point of view. As several researchers (see Mehra and Prescott 1985) have pointed out, the risk premia on securities seem to be too large relative to the variability of their returns, unless one assumes absurdly high risk aversion on the part of agents or abandons the hypothesis of complete markets (Scheinkman and Weiss 1986). If, in fact, most variability is "predictable," then even larger risk aversion is required in a complete-markets framework. One should note, however, that the same "sensitive dependence to initial conditions" that makes the trajectory of some deterministic nonlinear systems appear random—and presumably increases the apparent volatility of some nonlinear systems subject to random shocks—also makes the tasks of forecasting future values and of understanding the law of motion of the system extremely difficult. Hence it seems unlikely that even the most rational agents in such a case could come close to understanding the law of motion. A less strict view of rationality may admit that agents are constantly "learning" about the true law. As agents learn more characteristics of the returns process, arbitrage will change the law itself. This of course means that, unless one can model "knowledge" with a low dimensional state variable, the description in (5.1) and (5.2) is incorrect, and it is unlikely that the data are generated by any fixed law (subject perhaps to small noise). In this case it would be hard to interpret the results presented here.

Appendix

This appendix presents the formulas for the estimators of the asymptotic variance of the BDS statistics used in this article. Brock, Dechert, and Scheinkman (1986) present more general formulas for a wide range of these types of statistics.

Given that x_t is independent and identically distributed, BDS prove for any $N > 1$, $\gamma > 0$, as $m \to \infty$,

$$\sqrt{m}\{C_m^N(\gamma) - [C_m^1(\gamma)]^N\} \xrightarrow{d} N(0, V_C),$$

and

$$\sqrt{m}[S_m^N(\gamma) - C_m^1(\gamma)] \xrightarrow{d} N(0, V_S).$$

Now we need consistent estimators for V_C and V_S. Let

$$C(\gamma) = E[\theta(\gamma - |x_i - x_j|)],$$

$$K(\gamma) = E[\theta(\gamma - |x_i - x_j|)\,\theta(\gamma - |x_j - x_k|)],$$

$$v(n) = 4[K^n - (2n - 1)C^{2n} + 2\sum_{j=1}^{n-1} K^{n-j}C^{2j}],$$

and

$$cv(n_1,n_2) = 4\{K^{n_1} + K^{n_2} + 2K^{n_1}C^{n_2-n_1} - (1 + 2n_1)C^{2n_1} - (1 + 2n_2)(C^{2n_2} + 2C^{n_1+n_2}) + 2\left(\sum_{j=1}^{n_1} K^{n_1-j}(C^{2j} + C^{n_2-n_1+2j}) + K^{\min(n_1,n_2-j)} \times C^{n_1+n_2-2\min(n_1,n_2-j)} + K^{n_2-j}\,C^{2j}\right) + 2\sum_{j=n_1+1,}^{n_2-1} K^{n_2-j}\,C^{2j} + [K^{\min(n_1,n_2-j)}][C^{n_1+n_2-2\min(n_1,n_2-j)}]\}.$$

Let

$$\Sigma^C_{11} = v(1),$$

$$\Sigma^C_{22} = v(N),$$

and

$$\Sigma^C_{12} = \Sigma^C_{21} = \tfrac{1}{2}[cv(1,N) - v(1) - v(N)];$$

then

$$V_c = (-NC^{N-1}, 1)^T\Sigma^C(-NC^{N-1}, 1).$$

Further, let

$$\Sigma^S_{11} = v(1),$$

$$\Sigma^S_{22} = v(N - 1),$$

$$\Sigma^S_{33} = v(N),$$

$$\Sigma^S_{12} = \Sigma^S_{21} = \tfrac{1}{2}[cv(1,N - 1) - v(1) - v(N - 1)],$$

$$\Sigma^S_{13} = \Sigma^S_{31} = \tfrac{1}{2}[cv(1,N) - v(1) - v(N)],$$

and

$$\Sigma^S_{23} = \Sigma^S_{23} = \tfrac{1}{2}[cv(N - 1,N) - v(N - 1) - v(N)];$$

then

$$V_S = (-1, -C^{2-N}, C^{1-N})^T\Sigma^S(-1, -C^{2-N}, C^{1-N}).$$

In computing V_C and V_S, any consistent estimator for $C(\gamma)$ and $K(\gamma)$ may be used. We choose to use the U-statistic estimators for each. For $C(\gamma)$ this is just

$C^1_m(\gamma)$ already introduced. For $K(\gamma)$, we use

$$K_m(\gamma) = \frac{6}{n(n-1)(n-2)} \sum_{1 \le i < j < k \le m} h_\gamma(x_i, x_j, x_k),$$

where

$$h_\gamma(x,y,z) = \tfrac{1}{3}[\theta(\gamma - |x-y|)\theta(\gamma - |x-z|)\theta(\gamma - |z-y|)].$$

References

Atten, P., and Caputo, J. C. 1985. Estimation experimentale de dimension d'attracteurs et d'entropie. In M. Cosnard and C. Mira (eds.), *Traitement numerique des attracteurs etranges: Conference Proceedings*. Grenoble: University of Grenoble.

Ariel, R. A. 1987. A monthly effect in stock returns. *Journal of Financial Economics* 18 (March): 161–74.

Ben-Mizrachi, A.; Grassberger, P.; and Procaccia, I. 1984. Characterization of experimental (noisy) strange attractors. *Physical Review*, ser. A, 29 (February): 975.

Brock, W. A. 1986. Distinguishing random and deterministic systems: Abridged version. *Journal of Economic Theory* 40 (October): 168–95.

Brock, W. A.; Dechert, W. D.; and Scheinkman, J. 1986. A test for independence based on the correlation dimension. Manuscript. Madison: University of Wisconsin—Madison; and Chicago: University of Chicago.

Brock, W. A.; Dechert, W. D.; Scheinkman, J. A.; and LeBaron, B. 1988. A test for independence based upon the correlation dimension. Unpublished manuscript. Madison: University of Wisconsin.

Eckmann, J. P., and Ruelle, D. 1985. Ergodic theory of chaos and strange attractors. *Review of Modern Physics* 57, no. 3 (July): 617–56.

Efron, B. 1982. *The Jackknife, the Bootstrap, and Other Resampling Plans*. Philadelphia: Society for Industrial and Applied Mathematics.

Engle, R. F. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflations. *Econometrica* 50 (July): 987–1007.

Fama, E. 1970. Efficient capital markets: Review of theory and empirical work. *Journal of Finance* 25 (May): 383–417.

French, K. 1980. Stock returns and the weekend effect. *Journal of Financial Economics* 8 (March): 55–70.

Granger, C. W. J., and Anderson, A. P. 1978. *An Introduction to Bilinear Time Series Models*. Göttingen: Vandenboech & Ruprecht.

Granger, C. W. J., and Morgenstern, O. 1963. Spectral analysis of New York stock market prices. *Kyklos* 16:1–27.

Grassberger, P., and Procaccia, I. 1983*a*. Characterization of strange attractors. *Physics Review Letters* 50 (January): 316.

Grassberger, P., and Procaccia, I. 1983*b*. Measuring the strangeness of strange attractors. *Physical Review*, ser. D, 9 (October): 189.

LeBaron, B. 1988. Nonlinear puzzles in stock returns. Manuscript. Chicago: University of Chicago.

Lucas, R. 1978. Asset prices in an exchange economy. *Econometrica* 46 (November): 1429–45.

Mehra, R., and Prescott, E. 1985. The equity premium: A puzzle. *Journal of Monetary Economics* 15 (March): 145–61.

Ramsey, J. B., and Yuan, H. 1987. The statistical properties of dimension calculations using small data sets. Manuscript. New York: New York University.

Rozeff, M. S., and Kinney, W. 1976. Capital market seasonality: The case of stock returns. *Journal of Financial Economics* 3 (October): 379–402.

Sakai, H., and Tokumaru, H. 1980. Autocorrelations of a certain chaos. *IEEE Transactions on Acoustics, Speech and Signal Processing* 1, no. 5 (October): 588–90.

Scheinkman, J. A. 1985. Distinguishing deterministic from random systems: An examination of stock returns. Manuscript prepared for the Conference on Nonlinear Dynamics, Centre de Recherché de Mathematiques de la Decision (CEREMADE) at the University of Paris IX.

Scheinkman, J. A., and LeBaron, B. 1988. Nonlinear dynamics and GNP data. In W. Barnett, J. Geweke, and K. Shell (eds.), *Economic Complexity: Chaos, Sunspots, Bubbles, and Nonlinearity*. Cambridge: Cambridge University Press.

Scheinkman, J. A., and Weiss, L. 1986. Borrowing constraints and aggregate economic activity. *Econometrica* 54 (January): 23–45.

Subba Rao, T., and Gabr, M. *An Introduction to Bispectral Analysis and Bilinear Time Series Models: Springer-Verlag Lecture Notes in Statistics,* vol. 24. New York: Springer-Verlag.

Takens, F. 1983*a*. Distinguishing deterministic and random systems. In G. Borenblatt, G. Iooss, and D. Joseph (eds.), *Nonlinear Dynamics and Turbulence*. Boston: Pitman Advanced.

Takens, F. 1983*b*. Invariants related to dimension and entropy. *Proceedings of the Thirteenth Coloquio Brasileiro de Matemática*. Rio de Janeiro: Instituto de Matemática Pura e Apicada.

Takens, F. 1984. On the numerical determination of the dimension of an attractor. Manuscript. Gronigen: University of Gronigen, Department of Mathematics.

Tong, H., and Lim, K. S. 1980. Threshold autoregression, limit cycles and cyclical data. *Journal of the Royal Statistical Society* 42B:245–92.

Zardecki, A. 1982. Noisy Ikeda attractor. *Physics Letters*, ser. A, 90, no. 6 (July): 274–77.