ELSEVIER

Journal of Econometrics 82 (1997) 157–192

JOURNAL OF Econometrics

A single-blind controlled competition among tests for nonlinearity and chaos

William A. Barnett[a,*], A. Ronald Gallant[b], Melvin J. Hinich[c], Jochen A. Jungeilges[d], Daniel T. Kaplan[e], Mark J. Jensen[f]

[a] *Department of Economics, Washington University in St. Louis, Campus Box 1208, Brooking Drive, St Louis, M063130, USA*

[b] *University of North Carolina at Chapel Hill*

[c] *University of Texas at Austin*

[d] *University of Osnabrück, Germany*

[e] *Macalester College*

[f] *University of Missouri*

Abstract

Interest has been growing in testing for nonlinearity or chaos in economic data, but much controversy has arisen about the available results. This paper explores the reasons for these empirical difficulties. We designed and ran a single-blind controlled competition among five highly regarded tests for nonlinearity or chaos with ten simulated data series. The data generating mechanisms include linear processes, chaotic recursions, and non-chaotic stochastic processes; and both large and small samples were included in the experiment. The data series were produced in a single blind manner by the competition manager and sent by e-mail, without identifying information, to the experiment participants. Each such participant is an acknowledged expert in one of the tests and has a possible vested interest in producing the best possible results with that one test. The

*Corresponding author.

This research was partially supported by the National Science Foundation in the US (Barnett, Gallant), the Office of Naval Research in the US (Hinich). The Medical Research Council in Canada (Kaplan), and the German Science Foundation in Germany (Jungeilges). Mark Jensen was employed as a research assistant under Barnett's NSF grant SES 9223557. During the design stage of this competition, one of the included tests and one of the competitors were proposed to us by the Santa Fe Institute. We benefited from the comments of William Brock, Charles Manski, Steven Durlauf, and the participants at a Santa Fe Institute workshop on 10–12 January 1997 and a Santa Fe Institute colloquium on 9 January 1997.

PII S0304-4076(97)00081-X

results of this competition provide much surprising information about the power functions of some of the best regarded tests for nonlinearity or noisy chaos. © 1997 Elsevier Science S.A.

Keywords: Competition; Chaos; Nonlinearity; Experiment
JEL classifications: C12; C15; C22; C45

1. Introduction

In this paper, we reveal the results of a single-blind controlled competition, in which we compare the power of five highly regarded tests for nonlinearity or chaos against various alternatives. The data used in this competition was simulated data, produced from five different generating models and two different sample sizes with each of those models. Hence there were ten samples involved in the complete competition. One model, and hence two of the data sets, was purely deterministic (and chaotic). The other four models, and hence eight of the data sets, were stochastic processes, in which the randomness was produced by Monte Carlo methods. One of the stochastic processes was linear, while the other three were nonlinear, but not chaotic. Although the same five generating models were used to produce both sample sizes, the participants in the experiment were not aware of that fact. Hence the participants had no reason to believe that fewer than 10 generating models were used to produce the simulated data.

The data were generated at Washington University in St. Louis and sent by electronic mail to the participants in the experiment. Those participants were provided with no information regarding the nature of the simulated data. Each participant used one test to analyze each of the data series. Throughout the competition, William Barnett and Mark Jensen at Washington University served as the competition managers by generating the data. The competition managers were the only participants having any knowledge of the nature of the data. They did not reveal the generating models to the other participants until the competition was complete and all results from all participants had been received.

Only one of the tests used in this competition (the BDS test) was run at Washington University, and that test is one which is available in a widely used computer program written by W.D. Dechert. We acquired the computer program from William Brock and report the results acquired with that program. The simulated data are available to anyone who might wish to replicate the reported results with the BDS test. The other five tests are more complicated to run and possibly could have been prejudiced in some ways, if the generating model were known to the person running the test. Hence each of those tests was run by a competition participant who was supplied with no prior information

about the generating models. In addition, each of those participants has established expertise in the test that he ran and a possible vested interest in producing the best possible results with the particular test that he ran. In three of those cases, the participant was among the originators of the test, and in the remaining case the participant has produced and used a computer program that is especially well suited to conducting that test.

All five of the tests used in this competition are purported to be useful with noisy data of moderate sample size. The two sample sizes used in this competition were intended to include a sample of small size relative to the capabilities of the tests and a sample of large size. The computational cost of running some of these tests with the large sample was very high. With one of the tests, months of execution time on a workstation were needed to complete the test with each of the five large sample data sets. These computational costs limited to five the number of generating models that reasonably could be used to produce the simulated data in the competition, at least in the large sample case.

In recent years there has been growing interest in testing for both chaotic and nonchaotic nonlinearity in economic data, but much disagreement and controversy have arisen about the available results. For example, Barnett and Chen (1986, 1988a, b) claimed a successful detection of chaos. That conclusion was further confirmed with the same data by DeCoster and Mitchell (1991a, 1994), who also contributed relevant theory in DeCoster and Mitchell (1991b, 1992). However, the finding was subsequently disputed by Ramsey et al. (1990) and Ramsey and Rothman (1994), who also raise questions regarding virtually all of the other published tests of chaos. Various replies have been published, including those of Barnett and Hinich (1992, 1993) and that of DeCoster and Mitchell (1994). Further results relevant to those controversies recently were provided by Serletis (1995). In short, there is little agreement about the existence of chaos or even of nonlinearity in economic data, and some economists continue to insist that linearity remains a good assumption for all economic time series, despite the fact that economic theory provides little support for the assumption of linearity. This paper explores the reasons for these empirical difficulties.

Results may be difficult to find that are consistent across variations in sample size, test method, and aggregation. That possibility was the subject of Barnett et al. (1995), who used five of the most widely used tests for nonlinearity or chaos with various monetary aggregate data series of various sample sizes and acquired results that differed across tests and over sample sizes, as well as over the statistical index number formulas used to aggregate over the same component data. These conclusions applied to the tests for nonchaotic nonlinearity as well as to the tests for chaos.

It is possible that none of these tests completely dominates the other, since some tests may have higher power against certain alternatives than other tests. This competition was designed for the purpose of exploring the relative powers of the five tests used by Barnett et al. (1995) against various alternatives and

to investigate the various possible explanations for the existence of so much controversy regarding the available tests for chaotic and nonchaotic nonlinearity.

2. Data generation

The sample sizes generated consisted of a 'small sample' of size 380 and a 'large sample' of size 2000. The observations were produced with each of the two sample sizes from each of the following five models.

Model I:

Model I is the fully deterministic, chaotic Feigenbaum recursion of the form:

$$y_t = 3.57 y_{t-1}(1 - y_{t-1}),$$

where the initial condition was set at $y_0 = 0.7$.

Model II:

Model II is a GARCH process of the following form:

$$y_t = h_t^{1/2} u_t,$$

where h_t is defined by

$$h_t = 1 + 0.1 y_{t-1}^2 + 0.8 h_{t-1},$$

with $h_0 = 1$ and $y_0 = 0$.

Model III:

Model III is a nonlinear moving average (NLMA) process of the following form:

$$y_t = u_t + 0.8 u_{t-1} u_{t-2}.$$

Model IV:

Model IV is an ARCH process of the following form:

$$y_t = (1 + 0.5 y_{t-1}^2)^{1/2} u_t,$$

with the value of the initial observation set at $y_0 = 0$.

Model V:

Model V is an ARMA model of the form:

$$y_t = 0.8 y_{t-1} + 0.15 y_{t-2} + u_t + 0.3 u_{t-1},$$

with $y_0 = 1$ and $y_1 = 0.7$.

With the four stochastic models, the white noise disturbances, u_t, are sampled independently from a standard normal distribution. Those white noise disturbances were generated using the fast acceptance-region algorithm of Kinderman and Ramage (1976), with the initial seed value set by the clock of the computer at the time the program was run.[1] Of the five generating models, only Model V is linear, only Model I is chaotic, and only Model I is noise free. The simulated data is available online in the Working Paper Archive maintained at Washington University.[2]

In evaluating the results with the tests included in this competition, we need to know what hypotheses are satisfied by design in each of the five cases defined above. The hypotheses that are relevant to the tests used in this competition are: linear process, linear process in the mean relative to an information set, Gaussian process, chaotic, and third order nonlinear dependent process. Those terms are defined in Section 3.2 below. Model V is the only linear process and the only Gaussian process, although models II and IV are linear in the mean.[3] Only Model I is chaotic. Models II, IV, and V are linear in the mean. Models I and III exhibit third order nonlinear dependence.[4]

3. Background

We use five inference methods to test for stochastic nonlinearity or deterministic chaos with the simulated noisy data: the Hinich bispectrum test, the BDS test, the Lyapunov exponent estimator of Nychka et al. (1992), White's test, and Kaplan's test. We chose those tests as a result of their high repute among tests for nonlinearity and chaos.

---

[1] Strictly speaking, computer generated noise is deterministic, but is high dimensional. Hence the tests of nonlinearity and chaos should be viewed as tests for the existence of a low-dimensional nonlinear or chaotic signal below the high-dimensional chaos. In the language of chaotic dynamics, tests for chaos seek to separate intrinsic from extrinsic probability, where the distinction is in terms of the dimension of the determinism of each.

[2] The location of the simulated data in that archive is ewp-data/9510001. A direct link to that location in the archive is provided in paragraph 8 at the following web location: http://wuecon.wustl.edu/~barnett/Papers.html. In addition, code for the competing tests is available online, and links to the location of the code for each test can be found in that same paragraph 8 on the web.

[3] Treating prior observations as the information set and conditioning upon that information set, each of those two processes has zero conditional mean. It is also the case that both of those models are Gaussian in the mean, since their distributions, conditionally upon the past observations, are Gaussian. But we do not include tests for Gaussianity in the mean in this competition.

[4] Models II and IV also would exhibit third order nonlinear dependence, if the innovations were not Gaussian.

3.1. The tests

In this competition, we use tests derived for use with noisy data. The Hinich bispectral test is a test in the frequency domain of flatness of the bispectrum. The sampling properties of the test statistic are known, and the approach is based upon conventional time series inference methodology. The test was run by Hinich in Austin, Texas, without knowledge of the models that generated the data. The BDS test is a test for whiteness, which can be used to test for residual nonlinear structure, after linear structure has been removed through prior prewhitening. The test was run by Mark Jensen at Washington University. Although he was aware of the generating models, he used the BDS test program that has been supplied widely on floppy disk by the originators of the BDS test and was programmed by W.D. Dechert. We acquired the program from William Brock. The NEGM (Nychka, Ellner, Gallant, and McCaffrey) test is a non-parametric test for positivity of the maximum Lyapunov exponent.[5] The NEGM test was run by Gallant in North Carolina without knowledge of the models that generated the data. White's test is a test for nonlinearity, and was run by Jochen Jungeilges without knowledge of the models that generated the data. He used his own program, which implements White's test. Kaplan's test can be used to test either for nonlinearity or for more focused special cases of nonlinearity. In this competition that test was used as a test for general nonlinearity. Kaplan's test was run by Kaplan in Quebec without knowledge of the models that generated the data.

By using conventional stochastic process methods for testing for nonlinear dynamics, we largely are limited to tests for general nonlinearity, which is necessary but not sufficient for chaos. There are three particularly well known tests currently in use for testing for nonlinearity: the BDS (Brock et al. (1996)) test, White's neural network test, and the Hinich bispectrum test.[6]

The BDS test provides an important advance in testing for stochastic dependence, and hence the BDS test is a significant new contribution to the field of statistics. But the BDS test does not currently provide a direct test either for nonlinearity or for chaos, since the sampling distribution of the test statistic is not known, either in finite samples or asymptotically, under the null hypothesis

[5] Gencay and Dechert (1992) recently have proposed a test that is similar in some respects to the NEGM test. As a result of that similarity, we did not believe that a comparison between those two tests was a likely place to look for a robustness problem. In addition, we believe that a comparison among such related tests would require a much larger number of replications than we had available with the data used in the current study. From this class of tests, we therefore decided to run only the NEGM test.

[6] As a result of space constraints, our descriptions of the tests are necessarily brief. For a more detailed discussion of those tests, see Barnett et al. (1996a) and Barnett et al. (1996b).

of nonlinearity, linearity, chaos, or the lack of chaos. The asymptotic distribution is known under the null of independence. Hence the hypotheses of nonlinearity and chaos are nested within the alternative hypothesis, which includes both nonwhite linear and nonwhite nonlinear processes.

Nevertheless, it is possible to use the BDS test to test any parametric stochastic process against the remaining alternatives, if the parametric process null has been removed from the data by prefiltering. For example, if all linear possibilities have been removed by fitting an ARIMA model, the BDS test can be used to test the residuals for remaining nonlinear dependence.

Similarly, if all nonchaotic possibilities could be removed by fitting the best possible nonchaotic model, the BDS test could be used to test the residuals for remaining chaotic dependence. But filtering out all possible nonchaotic possibilities with certainty seems to be beyond the state of the art. Hence it is not clear how the BDS test can be used to produce a convincing inference regarding noisy chaos. For a formal definition of noisy chaos, see Nychka et al. (1992).

Filtering out all linear possibilities with certainty is difficult at best, but nevertheless prefiltering by ARIMA fit is often viewed as a reputable means of linear prewhitening, and hence we use the BDS test to test for remaining nonlinear dependence in the residuals of an ARIMA process fitted by the Box–Jenkins approach.[7] There have been a number of other recent attempts to apply the BDS test to nonlinearity testing of filtered data. For one such interesting example, see Scheinkman and LeBaron (1989). Despite our reservations regarding the usefulness of the BDS test in testing for chaos, we do believe that the BDS test produces a viable test of linearity against the omnibus alternative of nonlinearity, when the data is prefiltered by ARIMA fit. We use the BDS test for that purpose.

The Hinich bispectrum approach provides a direct test for nonlinearity as well as a direct test for Gaussianity, since Hinich's approach produces a test statistic having known asymptotic sampling distribution under the null of linearity, as well as another test statistic having known asymptotic sampling distribution under the null of Gaussianity. However, the alternative hypothesis is not as broad as that for the BDS test, as defined in Brock et al. (1996).[8] With the bispectrum test, the alternative hypothesis is all nonlinear processes having nonflat bispectrum. However, there are some nonlinear processes displaying nonflat polyspectra only at the trispectrum or higher order. Hence, the

[7] We used the conditional maximum likelihood routine contained in the RATS computer program package.

[8] Two widely used implementations of that test exist: Dechert's program and LeBaron's program. We used Dechert's program. Both programs are available online, and links to them can be found in paragraph 8 of the following web page: http://wuecon.wustl.edu/ ~ barnett/Papers.html.

bispectrum test has zero power against some forms of nonlinearity. In such cases, the nonlinearity often can be detected by subsequently running the trispectrum test of Dalle Molle and Hinich (1991, 1995) or of Walden and Williams (1993). The sample size requirements of the trispectrum test are formidable. The BDS test, on the other hand, has high power against a vast class of nonlinear alternatives.

In the next section, we describe the Hinich bispectrum approach, which is related to the Subba Rao and Gabr (1980) approach. It should be observed that Hinich (1996) has a related newer test, which is an analog to the bispectrum test, but in the time domain. Although that newer test may have power against a broader alternative than the frequency domain bispectrum test, Hinich's newer test is not yet as widely known as his popular bispectrum test. As a result, we have not included Hinich's newer test in this competition.

White's test uses neural net methods to test for nonlinearity. A connection exists between the White test, which we use as a candidate for a test of nonlinearity, and the NEGM test for chaos, since the NEGM test uses a neural net as a data model configured as a predictor before testing for chaos with the resulting fitted neural net. Since chaos is a stronger hypothesis than nonlinearity, the connection between the two tests could be useful in sequential testing. In particular, if nonlinearity is rejected with the White test, then there is diminished reason to proceed further with the NEGM test for chaos, since chaos is a strictly nested special case of nonlinearity.

While the BDS, White, and Hinich tests currently are among the best known tests available for testing nonlinearity in noisy data, we believe that there currently is only one well established candidate for a test for chaotic signal in small samples of noisy data. That is the NEGM test.[9] We describe the NEGM test in a later section below.

A new test that examines the evidence for the continuity of dynamical maps has recently been proposed by Kaplan (1993). At present, Kaplan's test has not been subjected to the extensive Monte Carlo comparisons that are available for the NEGM test. The Kaplan test compares a test statistic computed directly from the data with the test statistic produced from surrogate data. In our application of his approach, the surrogate data are produced from linear processes having the same histogram and an almost identical autocorrelation function as the actual data. The null hypothesis is linearity of the dynamics found in the data. However, depending on the manner in which the surrogate data is produced, the method appears relevant to investigating more sharply

[9] The Gencay and Dechert (1992) method mentioned above is among the other promising possibilities, but that test as well as the others have not been subjected to the degree of experimentation that currently is available for the NEGM test with noisy data.

focused forms of complex dynamics. We describe the test briefly in a later section below. For more details, see Kaplan (1993).

Our discussions of each test are rather brief, since those tests are described in greater detail in Barnett et al. (1995, 1996a, b). An exception is the Kaplan test, which is used in this competition in a somewhat different manner than earlier applications. Those differences are described in detail in this paper.

3.2. Definitions

If $\{x(t)\}$ is a zero mean third-order stationary time series, then the mean $\mu_x = \mathrm{E}[x(t)] = 0$, the second-order autocovariance $c_{xx}(m) = \mathrm{E}[x(t+m)x(t)]$, and the third-order autocovariances $c_{xxx}(s, r) = \mathrm{E}[x(t+r)x(t+s)x(t)]$ are independent of t.[10] If $c_{xx}(m) = 0$ for all nonzero m, the series is white noise. We define a pure (also called 'strict' sense) white noise series as a white noise process in which $x(n_1), \ldots, x(n_N)$ are independent random variables for all values of $n_1, \ldots, n_N$. All pure white noise series are white. All white noise series are not pure white noise. However, Gaussian white noise series are necessarily pure white noise series.

In addition to stationarity, whiteness, and pure whiteness, linearity is another often assumed property of a time series. The conventional definition of a linear stochastic process is a linear filter of independent and identically distributed inputs. An ARIMA process is a finite-order linear filter, while a first degree Volterra expansion (with zero higher degree Volterra kernals) is infinite dimensional and spans the space of linear filters.[11] In some definitions of linearity, the innovations are assumed to be white noise martingale differences, since the linear predictor is the best predictor in that case. However, we conform to the more conventional definition requiring independent and identically distributed inputs.

---

[10] See Hinich (1996) for a test of the maintained hypothesis of third-order stationary.

[11] In the literature on chaos, the search for chaos is in reality a search for 'low'-dimensional chaos, since knowing that data has been produced deterministically from high-dimensional chaos is not useful. Similarly the distinction between a high-order linear filter and a nonlinear process is of little use, since the ability to separate the two can disappear in the limit as a linear moving average filter becomes infinite order. Hence in reality, any test of the null of linearity must in reality be interpreted to be a test of 'low'-order linear filter. In this competition, the simulated linear data is produced by a low-order ARMA process. In later research, it could be interesting to generate data from increasingly high-order MA processes to find out how high the order of an MA process must become before some of the tests of linearity would reject linearity. However, it would be difficult to argue in practice that such a rejection would be an 'error', since few statisticians would prefer to estimate a high-order MA process to a sparsely parameterized nonlinear process, especially if the order of the 'true' MA process that generated the data exceeds the sample size. See Bickel and Bühlmann (1996).

A related property of a process is 'linearity in the mean' relative to an information set. Such a process has a conditional mean function that is a linear function of the elements of the information set. For a formal definition of linearity in the mean, see Lee et al. (1993, Section 1). The information set usually contains lagged observations on the process. A process that is not linear in the mean is said to exhibit 'neglected nonlinearity'. A process that is linear is also linear in the mean, but the converse need not be true. Similarly a process is Gaussian in the mean relative to an information set, if the distribution of the process conditionally upon the information set is a Gaussian process.

A further special case of nonlinearity is third-order nonlinear dependence, which we shall define as a frequency domain concept. We define a process to exhibit third-order nonlinear dependence, if the skewness function in the frequency domain is not flat as a function of frequency pairs. A formal definition of the skewness function is provided below in Eq. (4.2). This form of nonlinearity is called third-order, since the skewness function is a normalization of the Fourier transform of the third-order autocovariances. That Fourier transform is called the bispectrum, and is the third-order polyspectrum.[12]

Many researchers implicitly assume the errors of their models are Gaussian, and test for pure white noise by using the covariance function $c_{xx}(m)$, but ignore the information regarding possible nonlinear relationships which are found in the third-order moments $c_{xxx}(s, r)$. The above discussion suggests the need to test for both nonlinearity and Gaussianity, in addition to testing in the usual manners for whiteness. In addition, unconditional properties need to be distinguished from those that are 'in the mean' and those that are third order.

4. The Hinich bispectral approach

Hinich (1982) argues that the bispectrum in the frequency domain is easier to interpret than the multiplicity of third-order moments $\{c_{xxx}(r, s): s \leqslant r, r = 0, 1, 2, \ldots,\}$ in the time domain. For frequencies f_1 and f_2 in the principal domain

$$\Omega = \{(f_1, f_2): 0 < f_1 < 0.5, f_2 < f_1, 2f_1 + f_2 < 1\},$$

the bispectrum, $B_{xxx}(f_1, f_2)$, is defined by

$$B_{xxx}(f_1, f_2) = \sum_{r=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} c_{xxx}(r, s)\exp[-i2\pi(f_1 r + f_2 s)]. \tag{4.1}$$

[12] As defined in the denominator of Eq. (4.2), the normalization is in terms of a noncausal prewhitening. Hence stochastic processes are compared for third-order nonlinearity after normalization by a linear adjustment that flattens the power spectrum (the second-order polyspectrum).

The bispectrum is the double Fourier transformation of the third-order moments function and is the third-order polyspectrum. The regular power spectrum is the second-order polyspectrum and is a function of only one frequency.

The skewness function $\Gamma(f_1,f_2)$ is defined in terms of the bispectrum as follows:

$$\Gamma^2(f_1,f_2) = |B_{xxx}(f_1,f_2)|^2/S_{xx}(f_1)\,S_{xx}(f_2)\,S_{xx}(f_1+f_2), \tag{4.2}$$

where $S_{xx}(f)$ is the (ordinary power) spectrum of $x(t)$ at frequency f. Since the bispectrum is complex valued, the absolute value (vertical) lines in Eq. (4.2) designate modulus. Brillinger (1965) proves that the skewness function $\Gamma(f_1,f_2)$ is constant over all frequencies $(f_1,f_2)\in\Omega$ if $\{x(t)\}$ is linear; while $\Gamma(f_1,f_2)$ is flat at zero over all frequencies if $\{x(t)\}$ is Gaussian. Linearity and Gaussianity can be tested using a sample estimator of the skewness function $\Gamma(f_1,f_2)$. But observe that those flatness conditions are necessary but not sufficient for general linearity and Gaussianity, respectively. On the other hand, flatness of the skewness function is necessary and sufficient for third-order nonlinear dependence, as defined in Section 3.2.

The Hinich (1982) 'linearity test' tests the null hypothesis that the skewness function is flat, and hence is a test of lack of third-order nonlinear dependence. For details of the test, see Hinich (1982), Hinich and Patterson (1985,1989), and Ashley et al. (1986). In particular, the final transformed test statistic is distributed as a standard normal random variate under the null hypotheses of flat skewness function. When the null is Gaussianity, a related test statistic is denoted by H and is a standard normal random variate under the null.[13] When the null is absence of third-order nonlinear dependence, the test statistic is denoted by Z. In both cases, the distribution of the standard normal is used to produce a one sided test, in which the null is rejected if the test statistic is large.[14]

---

[13] Strictly speaking the test can reject Gaussianity, but cannot accept it, since violation of Gaussianity may not appear at the bispectrum level and may appear only at the level of higher-order polyspectra.

[14] Ashley et al. (1986, p. 174) presented an equivalence theorem which proves that the Hinich bispectral linearity test statistic is invariant to linear filtering of the data, when the parameters of the linear filter are known. An important implication of the theorem is that if $x(t)$ is found to be nonlinear, then the residuals of a linear model of the form $y(t) = f(x(t))$ will also be nonlinear, since the nonlinearity in $x(t)$ will pass through any linear filter, f. The above paper further reported tables on the power of the Hinich linearity test for detecting violations of the linearity and Gaussianity hypotheses for a number of sample sizes. The table indicates substantial power for both tests, even when the sample size is as small as 256.

5. The BDS test

The details of the BDS test (Brock et al., 1996) are well known in the literature. The test uses the correlation function (also called the correlation integral) as the test statistic. This choice is in contrast to the Grassberger–Procaccia test, which uses the correlation dimension. The correlation function is needed in deriving the correlation dimension, but the two are not the same.[15]

While correlation dimension is potentially very useful in testing for chaos, the sampling properties of the Grassberger–Procaccia correlation dimension are unknown. The BDS test uses the correlation function (not the correlation dimension) as the test statistic. The asymptotic distribution of the correlation function is known under the null hypothesis of pure whiteness. As a result, the BDS test can be used to produce a formal statistical test of whiteness against general dependence. However, the sampling distribution of the BDS test statistic is not known under the nulls of chaos, nonlinearity, or linearity We are left with the uncomfortable choice between the correlation dimension, which produces a direct test for chaos, but only when no stochastic shocks exist within the model, or the correlation function, which does have known sampling properties when there are stochastic shocks within the model, but only under a different null hypothesis (i.e., pure whiteness).

Nevertheless, the BDS test can be used to produce indirect evidence about nonlinearity. In particular, an ARIMA process can be fitted to the data in an attempt to remove linear structure. The BDS test then can be used to determine whether there is evidence of remaining dependence in the data. If all linear dependence has already been removed, then any remaining dependence must be nonlinear.[16] We use the Box–Jenkins approach to fit an ARIMA (i, j, k) model to

---

[15] The correlation dimension's value has a direct connection with the Hausdorff dimension of the attractor. Hence the correlation dimension, in principle, has a direct connection with chaos. In particular, low fractional Hausdorff dimension is the result sought by those looking for useful chaos. The determinism in high-dimensional chaos cannot be modeled without large numbers of variables, and in the limit, infinite-dimensional chaos is noise.

[16] In principle, there are some difficulties with this approach. The Box–Jenkins estimate of the ARIMA process may not succeed in removing all forms of linear dependence. In addition, the sampling distribution of the BDS test statistic is affected by the nonzero variances of the coefficient estimators in the ARIMA process. Although exact analytical results are not available on the effects of these problems on the test statistic, a large and growing body of Monte Carlo results has cast much light on implications of these matters for the use of the test. In particular, the power of the test depends upon the setting of the embedding dimension, the metric bound, and the time delay within the test statistic, and the Monte Carlo results provide useful information on the settings that maximize power. See, e.g., Brock et al. (1991) and Hsieh and LeBaron (1991). In addition, Hsieh and LeBaron (1991) have found that the effect of the nonzero variances of the coefficient estimators in the ARIMA process is small in models for low order ARMA's for samples of 500 or more with modest settings of the embedding dimension. Furthermore, by bootstrapping BDS under the null, these

the data.[17] In every case, the Box–Jenkins approach resulted in setting $j = 0$ (so the fit was ARMA). The BDS test statistic asymptotically becomes a standard normal Z statistic, under the null of pure whiteness. The null of pure whiteness is rejected, if the test statistic is large. By convention with a Z statistic, 'large' means larger than 2 or perhaps 3.[18]

The test has two free variables, the embedding dimension m and the metric bound ε, which can be set at various levels to check for robustness.[19] The need to choose the values of ε and m can be a complication in using the BDS test. We adopt the approach used by advocates of the test. In particular, we set ε equal to the standard deviation of the data.[20] At our chosen setting for ε, we produce the BDS test statistic for all settings of embedding dimension from 2 to 8, in the hope that the same inference will be produced at each of those embedding dimensions. Fortunately in our large sample cases, the inference was robust to the setting of m within the 2 to 8 range.[21]

(footnote 16 continued).
problems can be mitigated somewhat. This bootstrapping can be done using LeBaron's software written in C-source code that will run in a UNIX environment. That code is available at the web location provided in footnote 8 above. One further can do convergence experiments of bootstrap for BDS along the lines of LeBaron's experiments on page 1754 in Brock et al. (1992).

[17] Here i is the order of the AR (i) autoregressive part, k is the order of the MA(k) moving average part, and j is the number of times that the data is differenced before fitting the moving average.

[18] Strictly speaking, the definition of 'large' should depend upon sample size, with rejection requiring higher values of the test statistic for larger sample sizes. In our experiments, clear rejections occurred with extremely high values of the test statistic, and clear acceptances occurred with very low values of the test statistic. As a result, we viewed conclusions with the BDS test to be ambiguous, when the test statistic was close to the conventional critical values of the test, or when the inference depended upon embedding dimension.

[19] In addition, there is a free parameter in the correlation function, and that free parameter must be set at one fixed value. That parameter is the time delay used in embedding the univariate observations into a multivariate phase space. In this case, a finite choice for that parameter must be made in either the Grassberger–Procaccia test or the BDS test. In the BDS test, the convention is to set the time delay equal to one, so that m successive observations are stacked, without skipping any intervening observations, in producing the embedded phase space vectors.

[20] Through Monte Carlo studies, Hsieh and LeBaron (1988) found that the power and size of the test is maximized when ε is selected to be between 1/2 and 1.5 times our choice. Hence our choice is in the center of that region. We further investigated variations of the setting throughout that range. Our inferences were not changed at either the upper or lower bound of the region. Lower settings for ε, including the square of the standard deviation, produced results evidencing domination of the test by noise in the data. In particular, the test statistic became a strong function of embedding dimension and varied between very positive and very negative values as m was increased at fixed ε.

[21] Hsieh and LeBaron (1991) have found that type I error is large with the BDS test, when the sample size is not adequately large, since the nonzero standard error of the ARIMA coefficient estimators biases the BDS test. By their criterion, our small sample size of 380 observations is barely

6. The Lyapunov exponent test

A method of testing for chaos is to compute the dominant Lyapunov exponent. Testing for a positive value for that exponent for a bounded system is equivalent to testing for the sensitivity to initial conditions property of chaos. Hence, testing for positivity of that exponent produces a direct test for chaos.

Algorithms for estimating that exponent fall into two classes: the Jacobian method (see, e.g., Ellner et al. (1991)) and the direct method. In the past, such computations were applied deterministically. In physics experiments with very large sample sizes and no stochastic shocks internal to the system, noise in the data could be filtered out (see, e.g., Smith (1992)) and the Lyapunov exponent computed by one of the two approaches. Recently an estimator became available which is applicable with more modest sample sizes and with systems containing internal stochastic shocks. The approach is presented and explored with simulated data and biological data by Nychka et al. (1992). The approach proceeds as follows.

Consider the nonlinear autoregressive model of the form

$$x_t = f(x_{t-L}, x_{t-2L}, \cdots, x_{t-dL}) + e_t \tag{6.1}$$

for $1 \leqslant t \leqslant N$, where L is the time delay parameter, d is the length of the autoregression, and $\{x_t\}$ are real valued.[22] Here f is a smooth, unknown function, and $\{e_t\}$ is a sequence of independent random variables with zero mean and unknown constant variance. While (6.1) itself is an unlikely data generating model, Takens' theorem (Eckmann and Ruelle, 1985) for dynamical systems states that in this class of nonlinear autoregressions there exists at least one model that can track any deterministic chaotic solution on an attractor with

(footnote 21 continued).

adequate. Hence, to avoid rejecting a true null hypothesis, we should refrain from rejecting the null unless the test statistic is very large. As mentioned above, our experiment produced unusually extreme values of the test statistic in many cases. As a result, our clear rejections corresponded to extremely low tail areas (P-values), and our clear acceptances corresponded to extremely high tail areas. We viewed as ambiguous the cases that did not correspond with such decisive tail areas.

Brock, Hsieh, and LeBaron have found that the asymptotic properties of the BDS test deteriorate, when the embedding dimension increases to more than 3 at sample sizes comparable to ours. Although we report results with embedding dimensions varying from 2 to 8, the results with embedding dimensions of 2 or 3 should be given the most serious consideration. But again, we acquired inferences that were robust to variation of embedding dimension from 2 to 8 in the large sample cases, so that the issue regarding deteriorating asymptotic properties with large embedding dimensions did not arise.

[22] The procedure that follows is phase space reconstruction in lag coordinates based upon Takens Theorem. This procedure is standard in this literature. Regarding its use and implications, see Broomhead et al. (1992).

finite dimension, and any such model having that property can be used to compute the Liapunov exponent of the unknown true data generating process. The proof of this Takens representation result in the stochastic case can be found in Casdagli et al. (1991). Nychka et al. (1992) fit f nonparametrically using either a spline or a neural net. They then compute the Liaponuv exponent from the fitted function, f, using the Jacobian approach.

Based upon the findings of Nychka, Ellner, Gallant and McCaffrey (hereafter NEGM), Gallant used the neural net approach. As in their study, he used the feed-forward single hidden layer networks with a single output. The neural net approach to nonlinear regression has a selection parameter, q, which equals the number of units in the hidden layer of the neural net. Hence, in addition to the coefficients θ of the neural net, there are three parameters that must be selected in the NEGM approach: q, L, and d.

As appropriate values of d, L, and q are unknown, they must be estimated. Nychka et al. (1992) recommend selecting that value of the triple (d, L, q) that minimizes the Bayesian BIC criterion (Schwarz, 1978) jointly in (d, L, q, θ), where θ is the vector of other parameters of the fitted neural net.[23] In the more recent version of the NEGM approach, the closely related GCV (generalized cross validation) criterion is minimized. In this competition, the GCV criterion, rather than the BIC criterion, is used. The estimate of the dominant Lyapunov exponent then is compute from gradient method along the fitted neural net.[24] For further details of the implimentation of the test used in this competition, see Barnett et al. (1995, 1996a, b).[25]

Although the standard error of the Lyapunov exponent estimate $\hat{\lambda}$ is not known, NEGM display plots that are informative about precision. One plot illustrates the sensitivity of the estimate of λ to variations in the initial conditions used in estimating the coefficients, θ, of the neural net and to variations in (L, d). We shall refer to that plot as the 'NEGM sensitivity plot'. The other plot illustrates the effect on the estimate of λ of variations in (L, d) and also indicates

[23] It should be observed that throughout, we use BIC to stand for 'Bayesian information criterion', and not to Akaiki's nonBayesian 'information criterion B', which in some other papers is called BIC to distinguish it from Akaiki's 'information criterion A'.

[24] For more details regarding the neural net fit, see McCaffrey et al. (1992). For proof of the consistency of the nonparametric function estimator by neural net, see Gallant and White (1988, 1991).

[25] In principle, it should be possible to produce a standard error for the Lyapunov exponent point estimate, perhaps by bootstrapping. But when noise is large, the properties of such a bootstrapped standard error are not known, and there has not yet been any published research on the computation of a standard error for the Nychka, Ellner, Gallant, and McCaffrey Lyapunov exponent estimate. Hence we report only the point estimates of the dominant Lyapunov exponents, as computed by Gallant in this competition.

the precision of the point estimate of (L, d). We find the NEGM sensitivity plot to be especially useful, and hence we supply only that plot, both for our large and small samples, in the cases in which evidence of chaos was found with the NEGM test. That plot is an indication of the sensitivity of $\hat{\lambda}$ to variations in θ about the least squares estimate at various settings of (L, d).

The procedure for producing the NEGM sensitivity plot is the following. For each setting of (L, d), where $L = 1, 2, \ldots, 5$ and $d = 1, 2, \ldots, 6$, the value of q that minimizes GCV conditionally upon (L, d) is found. Let $\hat{q}(L, d)$ be that value. The estimation of θ proceeded by first narrowing down the estimates of that vector to 20 possibilities, through a nested optimization procedure. The one among the 20 that minimized least squares then was selected as the optimum estimate. In the NEGM sensitivity plot, box plots are displayed indicating the range of values of the estimated dominant Lyapunov exponent at each setting of (L, d), with q set at $\hat{q}(L, d)$. The range within the box was acquired at each such setting of $(L, d, \hat{q})$ by varying θ over the 20 possibilities for θ attained through the nested iteration.[26]

The scatter within any such box illustrates the numerical stability of recovering a similar estimate of λ, when only the starting values of θ are varied. Moving between boxes indicates the sensitivity of the estimate of λ to variations in (L, d).

7. The White test

In White's test, the time series is fitted by a single hidden-layer feed-forward neural network, which is used to determine whether any nonlinear structure remains in the residuals of an AR process fitted to the same time series. Recent simulation studies have produced evidence that White's test against nonlinearity, based upon that model of the process, has power against a variety of nonlinear processes. The null hypothesis for the test is linearity in the mean (relative to the information set of lagged observations). All results using White's test were obtained using an implementation of White's test, programmed and applied in this competition by Jochen Jungeilges.

The test procedure applied is essentially due to Halbert White, who proposed his neural network test in White (1989a, b). Efforts to study the operational

[26] A box plot is a graphical display of the center and spread of a set of points and the deviant points within the set. The shaded box indicates the interquartile range (IQR) of the data. The lower limit of that shaded box is the 25th percentile, and the upper limit is the 75th percentile. The (white) horizontal line within the box is located at the median. The whiskers that extend from the top and bottom of the shaded box are the dotted lines capped by brackets at each end. The whiskers extend to either the extreme values of the data or to $1.5 \times$ IQR from the center of the shaded box, whichever is less. The horizontal (black) lines mark deviant points that fall outside the limits of the whiskers.

characteristics of this test against nonlinearity in the mean were undertaken by Lee et al. (1993) and Jungeilges (1996). These studies demonstrate that the test has appropriate size as well as power against various types of nonlinearity in the mean. Details of the algorithm used are given in Jungeilges (1996).

The rationale for White's test can be summarized as follows: under the hypothesis of linearity in the mean, the residuals obtained by applying a linear filter to the process should not be correlated with any measurable function of the history of the process. White's test uses a fitted neural net to produce the measurable function of the process's history and an AR process as the linear filter. White's method then tests the hypothesis that the fitted function does not correlate with the residuals of the AR process. The resulting test statistic has an asymptotic chi squared distribution under the null of linearity in the mean. See Lee et al. (1993, Section 2) for a presentation of the test statistic's formula and computation method.

The formal test is conditional upon the choice of a direction, and in White's method the direction in which the test looks for nonlinearity is chosen at random.[27] See, e.g., White (1989a) and Kuan and White (1991) for details. In White (1989b), it is pointed out that under certain assumptions the parameters of the network do not have to be estimated. White argues that a procedure involving regression and the extraction of principal components leads to an asymptotically equivalent test procedure. See White (1989b), Lee et al. (1993), and Jungeilges (1993).

The order of the AR process is chosen by a conventional selection criterion. For each series in this competition, Jungeilges chose the order which minimized the Schwarzian Bayesian Information Criterion (BIC). This criterion provides asymptotically unbiased order estimates. In Jungeilges (1996), it is demonstrated that choosing the order of the AR process via the BIC criterion may improve the power of White's test against nonlinear chaotic data generating process relative to the power of the version of the test involving alternative selection criteria.

8. The Kaplan test

We begin our discussion of the Kaplan test by reviewing its origins in the chaos literature, although the test is used in this competition as a test of linear stochastic process against general nonlinearity, whether or not noisy or chaotic. In the case of chaos, a time series plot of the output of a chaotic system may be very difficult to distinguish visually from a stochastic process. However, it is well

[27] For a related procedure, see Bierens (1990).

known that plots of the solution paths in phase space (x_{t+1} plotted against x_t and lagged values of x_t) often reveal deterministic structure that was not evident in a plot of x_t versus t. A test based upon continuity in phase space has been proposed by Daniel Kaplan. For a detailed discussion of the implementation used in this competition, see Barnett et al. (1996).[28]

Briefly he has used the fact that deterministic solution paths, unlike stochastic processes, have the following property: points that are nearby are also nearby under their image in phase space.[29] Using this fact, he has produced a test statistic, which has a strictly positive lower bound for a stochastic process, but not for a deterministic solution path.[30] By computing the test statistic from an adequately large number of linear processes that plausibly might have produced the data, the approach can be used to test for linearity against the alternative of noisy nonlinear dynamics. The procedure involves producing linear stochastic process surrogates for the data and determining whether the surrogates or a noisy continuous nonlinear dynamical solution path better describe the data. Linearity is rejected, if the value of the test statistic from the surrogates is never small enough relative to the value of the statistic computed from the data.

More formally stated, the procedure is the following. If we define the vector $\boldsymbol{x}_t = (x_t, x_{t-\tau}, x_{t-2\tau}, \cdots, x_{t-(m-1)\tau})$ embedded in m-dimensional 'phase space', then there is a recursive function giving $x_{t+\tau} = f(\boldsymbol{x}_t)$ with the fixed positive integer time delay τ. Here $x_{t+\tau}$ is called the 'image' of the point $\boldsymbol{x}_t$ in phase space. For perfectly deterministic systems with a continuous f, nearby points in m-dimensional phase space will have nearby images. For a stochastic system, on the other hand, nearby points in phase space may have very different images.

In terms of the delta–epsilon proofs of continuity familiar from calculus, distance in phase space plays the role of δ, and distance of the images plays the role of ε. For a given choice of embedding dimension m, one calculates $\delta_{ij} = |\boldsymbol{x}_i - \boldsymbol{x}_j|$ and $\varepsilon_{ij} = |\boldsymbol{x}_{i+\tau} - \boldsymbol{x}_{j+\tau}|$, for all pairs of time subscripts (i, j). The average of the values of ε_{ij} over those (i, j) satisfying $\delta_{ij} < r$ is defined to be $\mathrm{E}(r)$. For a perfectly deterministic system with continuous f, one expects to have $\mathrm{E}(r) \to 0$ as $r \to 0$.

Kaplan's test statistic K is the limit of $\mathrm{E}(r)$ as $r \to 0$. For a system that is not perfectly deterministic, one way of interpreting the nonzero value of K is as a goodness of fit measure from fitting a continuous model of some fixed order to

[28] The implementation of his test described and used in Barnett et al. (1995) differs somewhat from that used by Kaplan in this competition.

[29] That is, if x_t and y_t are close to each other and their lagged values also are close to each other, then x_{t+1} and y_{t+1} also are close to each other.

[30] Producing results on statistical significance requires multiple Monte Carlo trials with the process.

an infinite amount of data (so that overfitting was not an issue). If this measure of fit is smaller for the data than for surrogate data generated from a model that satisfies a stated null hypothesis, then there is evidence that the null hypothesis should be rejected. In order to test for linear dynamics, Kaplan generated 20 linear surrogates, produced to have the same histogram and a similar autocorrelation function as the actual simulated data used in this competition. The time series were embedded in 1, 2, 3, and 4 dimensional spaces.[31]

Two methods exist for computing the minimum value of K consistent with the surrogates. The simplest method is to compute the minimum value of K from the finite number of surrogates, and impute that to the population of surrogates consistent with the procedure. A more appealing approach is to compute the mean and standard error of the values of K from the sample of 20 surrogates and then subtract a multiple (conventionally 2 or 3) of the standard error from the mean to get an estimate of the population minimum. Using a multiple of 2, the conclusions reached below from the Kaplan test are the same for either of the two methods. In the tabulated results, we provide both the minimum value of K from the 20 surrogates and the mean and variance of K from the surrogates. Under the assumption of normality of the distribution of K from the population of surrogates, conclusions could be reached about statistical significance. But we do not provide such an interpretation, since the normality assumption may be a poor approximation, and not enough surrogates were generated to produce a Monte Carlo inference about statistical inference.[32]

9. Results

9.1. Overview

The following is a summary of the successes and failures of each of the tests in the competition, with each test judged relative to the null that it is designed to test. More detailed discussion follows.

---

[31] Kaplan's convention for defining embedding dimension differs from that used by BDS. Add 1.0 to Kaplan's or NEGM's embedding dimension to get the embedding dimension using the BDS convention. In Kaplan's and NEGM's convention, the embedding dimension is the dimension of the space in which δ_{ij} is calculated. The procedure that Kaplan used to produce the surrogates and to approximate his test statistic K with the actual and surrogate data are described in Barnett et al. (1996). Also see Kaplan (1994).

[32] The decision on the initial number of surrogates used was made by Kaplan during the competition. After the fact, it would be possible to run more replications to produce inferences about statistical significance, but one of the rules of the competition was that no additional computations or modifications to conclusions were permitted after the competition was closed and the identities of the generating models revealed. Hence the generation of further surrogates at this point (which in fact was offered by Kaplan) is precluded by the rules.

The Hinich bispectrum test is a test of the null hypothesis of lack of third-order nonlinear dependence. With the small sample, the test was correct in three out of the five cases and failed in two of the cases. With the large sample, the test was correct in three of the five cases, failed in one case, and was ambiguous in one case. The associated Gaussianity test, is a test of a necessary and not sufficient condition for Gaussianity and hence can reject but not accept. Judging the test on its rejections of Gaussianity, the small sample results produced only two rejections, and both were correct rejections. With the small sample, the test produced four rejections, and all four were valid rejections.

The BDS test entered into this competition is a test of the null hypothesis of linearity of the process.[33] With the small sample, the test was correct in two cases out of five and ambiguous in the other three. With the large sample, the test was correct in all five cases.

The NEGM test is a test of the null hypothesis of chaos. The test was correct in all five small sample cases and all five large sample cases.

White's test is a test of the null of linearity in the mean. In the small sample cases, the test was correct in four out of the five cases, and failed in the remaining case. In the large sample cases, White's test again was correct in four out of the five cases, and failed in one case.

Kaplan's test is a test of the null hypothesis of linearity of the process.[34] The test was correct in all five cases both with small samples and large samples.

9.2. *Results with the Hinich test*

Tables 1 and 2 provide the results of the Hinich test without prewhitening. The tests are one sided, so that the hypotheses are rejected if the test statistics are 'large', perhaps exceeding 2 or 3 by conventional standards. Recall that the null for the Hinich 'linearity' test actually is lack of third-order nonlinear dependence, and ARCH and GARCH processes with Gaussian innovations do not exhibit third-order nonlinear dependence. Hence in the discussion below and the table, the word 'linearity' should be understood to mean absence of third-order nonlinear dependence. Also recall that the Gaussianity test is a test for a necessary but not sufficient condition for Gaussianity, so that strictly speaking

[33] This conclusion follows from the fact that the prefiltering of the data was with an estimated ARMA process. If the larger class of linear in the mean processes had been filtered out of the data before running the test, the test would have had linearity in the mean as its null.

[34] This conclusion follows from the fact that he used only linear filters among his surrogates. If he had also included linear in the mean processes, such as ARCH and GARCH, his test could have been used to test the null of linearity in the mean. With Kaplan's test the null is defined by the class of models used in producing the surrogates.

Table 1
Hinich bispectral test with sample size = 380

Process	Gaussianity H	Linearity Z	Comments
I (Feig)	1.20	− 2.84	Weakly accept Gaussianity and strongly accept linearity
II (GARCH)	1.89	− 1.85	Weakly accept Gaussianity and strongly accept linearity
III (NLMA)	9.79	0.01	Strongly reject Gaussianity and accept linearity
IV (ARCH)	2.00	− 1.03	Reject Gaussianity and accept linearity
V (ARMA)	− 8.10	− 9.35	Strongly accept linearity and Gaussianity

Note: The linearity test is more formally a test of lack of third-order nonlinear dependence. The Gaussianity test is a test of a necessary but not sufficient condition for Gaussianity, and hence the word 'accept' for this test should be interpreted to mean 'not reject'. The data were not prewhitened. Framesize = 11. The word strongly accept is used when the tail area of the test far exceeds 0.10.

Table 2
Hinich bispectral test with sample size = 2000

Process	Gaussianity H	Linearity Z	Comments
I (Feig)	18.37	− 12.15	Strongly reject Gaussianity and strongly accept linearity
II (GARCH)	3.74	− 0.61	Reject Gaussianity and accept linearity
III (NLMA)	13.64	1.84	Strongly reject Gaussianity and marginally accept linearity
IV (ARCH)	38.05	0.41	Strongly reject Gaussianity and accept linearity
V (ARMA)	− 8.17	− 12.03	Strongly accept linearity and Gaussianity

Note: The linearity test is more formally a test of lack of third-order nonlinear dependence. The Gaussianity test is a test of a necessary but not sufficient condition for Gaussianity, and hence the word 'accept' for this test should be interpreted to mean 'not reject'. The data were not prewhitened. Framesize = 21.

the test can reject but cannot accept Gaussianity. We nevertheless shall use the word accept, since 'not reject' is awkward, but with the qualification that accept really means not reject in the case of the Hinich Gaussianity test.

First consider the small sample results in Table 1. Gaussianity is rejected in case III. The Gaussianity test results are also dramatic in case V. In that case 'acceptance' of Gaussianity is very strong. Regarding the linearity test, again the most dramatic case is case V, in which linearity is very strongly accepted. Since case V is the linear process, the fact that both Gaussianity and linearity are most strongly accepted in that case is a favorable result for the Hinich test.

Lack of third-order nonlinear dependence is accepted in each of the cases, II, III, and IV, but in a less extreme manner than with the linear process, V. That conclusion is correct in cases II and IV, but not in case III. The Gaussianity test results are especially mixed in those three cases, with a very strong (and correct) rejection in case III and a marginal 'acceptance' in cases II and IV.[35]

The results with case I may seem to be surprising, since case I is the purely deterministic and chaotic Feigenbaum map. Despite the deep nonlinearity of that generating mechanism, and despite the fact that no noise was introduced into that data, the Hinich test accepted linearity and weakly accepted Gaussianity, although the acceptances were not as dramatic as with the linear process, case V. However, an explanation does exist. The bispectrum test is known to have low power against certain forms of chaos that produce irregular and widely spaced spikes in the bispectrum. Such singular spikes can be difficult for the Hinich test statistic to detect, although those become evident from visual inspection of the bispectrum. See, e.g., Ashley and Patterson (1989, p. 690). Our case I data were produced from a chaotic map that Ashley and Patterson have found to generate a form of chaos that is difficult for the bispectrum test to detect without direct inspection of the bispectrum plot itself. Since we structured this competition in the form of a controlled competition, we did not permit the use of such informal inspection of plots as a means of generating conclusions. We insisted that the bispectrum test results be based solely upon the use of the scalar Hinich test statistic.

Now consider the large sample results in Table 2. Again the clearest result is the acceptance of linearity and Gaussianity in case V, which indeed was produced from a Gaussian, linear process. In the other cases, the results are similar to those with the small sample, but stronger. In particular, the test continues not to detect the nonlinearity in the chaotic data, but now very strongly rejects Gaussianity. In the nonchaotic nonlinear cases, II–IV, the test correctly concludes that ARCH and GARCH do not exhibit third-order nonlinear dependence, but incorrectly accepts lack of third-order nonlinear dependence in case III, although only marginally. However, with the larger sample the test correctly and strongly rejects Gaussianity with the GARCH data and very strongly rejects Gaussianity with the ARCH and nonlinear moving average data.

It appears that a rejection of linearity with the Hinich test would provide very dramatic support for the conclusion of nonlinearity, but acceptance of the null of linearity with that test provides only weak support for the linearity, since the test, as currently constructed, actually tests the broader null of absence of

[35] See Dalle Molle and Hinich (1991, 1995) and Walden and Williams (1993) regarding the trispectrum test which has high power against those alternatives.

third-order nonlinear dependence. Hence if 'linearity' is accepted with that test, further testing by other means would seem to be in order.[36]

9.3. Results with the BDS test

Results with the BDS test are reported in Tables 3 and 4. The data were prewhitened by Box–Jenkins estimation of an ARIMA model, as a means of removing linear dependence. Hence, with the exception of case V, the BDS test with the large samples appears to be detecting nonlinearity in all of our data series. In addition, the rejection of linearity in case I is extreme. This is a very favorable result for the BDS test, since case V was the only linear case, and case I is the chaotic Feigenbaum map data.

The results are similar with the smaller sample in the two extreme cases of linearity and chaos, but not as successful in the nonchaotic nonlinear cases. In particular the test's results with the small sample are ambiguous in all of the nonchaotic nonlinear cases, since the test statistic is too unstable against variation of the embedding dimension in those cases to produce an unambiguous conclusion. However, the rejection of linearity with the chaotic case I data remains extremely strong, and the acceptance of linearity with the case V ARMA data is fairly clear, although some ambiguity is introduced by the result at $m = 6$.

In both the small sample and large sample cases, the evidence of nonlinearity is stronger with the ARCH data than with the GARCH data. Although this result is somewhat surprising, the Kaplan test produced the same conclusion, as discussed below. Perhaps both tests have somewhat higher power against ARCH than GARCH.

These results are consistent with the prior findings of high power of the BDS test against a vast class of nonlinear alternatives. Evidently the test is triggered by any evidence of nonlinearity in the data. If the null is rejected, other tests should be used to permit the class of relevant alternatives to be narrowed. If the null is accepted, there is little point to continue further, since an acceptance of linearity by the BDS test is a strong result.

Much of the Monte Carlo research that has been published on the BDS test (e.g., Hsieh and Le Baron (1991)) has emphasized the pretesting issue and the potential dependence of the properties of the test on the prior linear filter. The results in Tables 3 and 4 suggest that the past emphasis on those concerns was well directed, since the prior linear filter selected in both the large sample and small sample linear case (case V) was not correct. Some of the test's sensitivity to

[36] In that regard, an important new related test in the time domain has been proposed by Hinich (1996). But as discussed above, we have not included that newer test in this competition.

Table 3
BDS test Z statistics. Residuals from ARIMA fit to simulated data with 380 observations

Process	Fitted ARIMA order (i, j, k)	Epsilon	Embedding dimension	BDS Z Statistic	Decision
I (Feig)	(0,0,0)	0.122	2	82.33	Reject linearity
			3	156.37	(very strongly)
			4	270.50	
			5	507.63	
			6	994.15	
			7	2032.00	
			8	4286.00	
II (GARCH)	(0,0,0)	0.084	2	0.35	Ambiguous
			3	1.68	
			4	2.56	
			5	3.03	
			6	2.91	
			7	− 8.31	
			8	− 4.16	
III (NLMA)	(0,0,0)	0.078	2	3.57	Ambiguous
			3	4.76	(weakly reject linearity)
			4	4.03	
			5	2.85	
			6	2.29	
			7	0.39	
			8	0.59	
IV (ARCH)	(0,0,0)	0.076	2	4.26	Ambiguous
			3	4.49	(weakly reject linearity)
			4	3.93	
			5	3.72	
			6	3.31	
			7	1.62	
			8	− 1.11	
V (ARMA)	(2,0,0)	0.074	2	− 0.99	Accept linearity
			3	− 1.34	
			4	0.24	
			5	1.31	
			6	2.50	
			7	1.50	
			8	− 0.78	

Note: The order of the fitted ARIMA process was acquired by Box–Jenkens methodology. The ARIMA fit detected and filtered out linear structure only in Process V. The resulting estimated coefficient of the AR(1) term was 1.08025, and the estimated coefficient of the AR(2) term was − 0.12002. The BDS Z statistic is asymptotically standard normal under the null of whiteness, and the test is one sided, with rejection if Z is large (perhaps exceeding 2).

Table 4
BDS test Z statistics, Residuals from ARIMA fit to simulated data with 2000 observations

Process	Fitted ARIMA order (*i*, *j*, *k*)	Epsilon	Embedding dimension	BDS *Z* statistic	Decision
I (Feig)	(0,0,0)	0.012	2	262.15	Reject linearity
			3	528.82	(very strongly)
			4	1065.80	
			5	2383.60	
			6	5631.60	
			7	13,904.00	
			8	35,434.00	
II (GARCH)	(0,0,0)	0.060	2	3.45	Reject linearity
			3	4.99	
			4	6.66	
			5	7.65	
			6	8.81	
			7	10.02	
			8	9.47	
III (NLMA)	(0,0,0)	0.053	2	8.55	Reject linearity
			3	11.84	(strongly)
			4	13.76	
			5	14.81	
			6	16.07	
			7	19.07	
			8	23.82	
IV (ARCH)	(0,0,0)	0.032	2	16.65	Reject linearity
			3	16.06	(strongly)
			4	15.73	
			5	15.48	
			6	16.31	
			7	17.52	
			8	17.10	
V (ARMA)	(1,0,0)	0.079	2	1.15	Accept linearity
			3	1.51	
			4	1.10	
			5	0.77	
			6	1.03	
			7	0.14	
			8	0.99	

Note: The order of the fitted ARIMA process was acquired by Box–Jenkens methodology. The ARIMA fit detected and filtered out linear structure only in Process V. The resulting estimated coefficient of the AR(1) term was 0.96963. The BDS Z statistic is asymptotically standard normal under the null of whiteness, and the test is one sided, with rejection if Z is large (perhaps exceeding 2).

nonlinearity could be a result of remaining linear dynamics in the data. However, the BDS test in this competition did successfully accept linearity in the linear cases, despite the fact that the test's prior linear filter in the linear cases was never estimated to be the correct ARMA (2,1) process.

In short, the BDS test and the Hinich test have very different capabilities. While a rejection of linearity is a dramatic result with the Hinich test, which is not easily triggered, the BDS test's null is rejected over a vast range of alternatives.

9.4. *Results with the NEGM test*

With the NEGM Lyapunov exponent test, the GCV estimates of the parameter triple, (L, d, q), are displayed in Table 5. The dominant Lyapunov exponent estimates are provided in Table 6. According to this test, only case I appears chaotic. The same conclusion was reached with both the large and the small sample. This result is very favorable for the NEGM test, since case I is the only case of chaotic data. Since the test is a test of chaos rather than of general nonlinearity, comparisons among the results with cases II–IV are not meaningful, aside from the fact that the test correctly recognized the fact that the nonlinearity in that data was not chaotic. Figs. 1 and 2 indicate the sensitivity of the Lyapunov exponent estimate to variations in the parameters for case I. See Section 6.3 for details of the construction and interpretation of those plots.

The NEGM sensitivity plots for the small sample chaotic case, case I, are displayed in Fig. 1. According to Table 5, the GCV estimate for (L, d) with the small sample Feigenbaum data is (1, 1). Observing the box corresponding to $(L, d) = (1, 1)$ in Fig. 1, we see that the entire range of the box is above zero.

Table 5
Dominant Liapunov exponent estimation: Selection of delay, number of lags, and number of hidden units

Process	(L, d, q) Triple that minimizes GCV	
	380 Observations	2000 Observations
I (Feig)	(1,1,2)	(1,2,4)
II (GARCH)	(4,3,1)	(4,4,2)
III (NLMA)	(1,2,3)	(1,3,8)
IV (ARCH)	(1,6,2)	(1,6,3)
V (ARMA)	(1,1,1)	(1,3,1)

Note: Each entry in the table is the GCV selection (minimizing the generalized cross validation criterion) of the triple, (L, d, q), where L is the time delay parameter, d is the number of lags in the autoregression, and q is the number of units in the hidden layer of the neural net. The data were not prewhitened.

Table 6
Dominant Liapunov Exponent Point Estimate

Process	Dominant Liapunov Exponent		Conclusion	
	380 observations	2000 observations	380 observations	2000 observations
I (Feig)	0.0168	0.0130	Accept chaos	Accept chaos
II (GARCH)	− 1.3379	− 0.394	Reject chaos	Reject chaos
III (NLMA)	− 0.3716	− 0.298	Reject chaos	Reject chaos
IV (ARCH)	− 0.9634	− 0.517	Reject chaos	Reject chaos
V (ARMA)	− 0.0539	− 0.038	Reject chaos	Reject chaos

Note: Data was not prewhitened. The Liapunov exponent was computed from the fitted time series using a neural net nonparametric fit. Logarithms are natural logarithms.

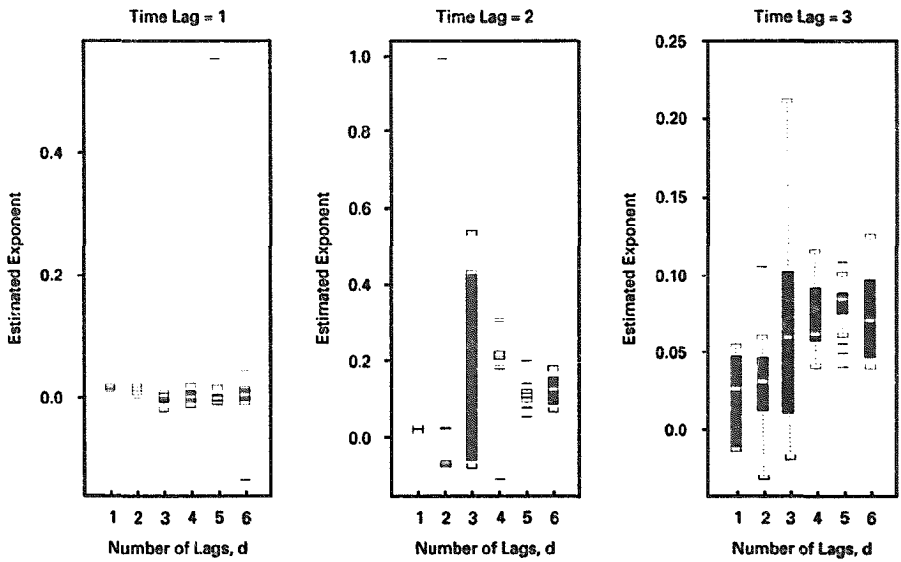

Fig. 1. NEGM sensitivity plot with Feigenbaum small sample: Indicates sensitivity of the λ estimate to the initial condition for θ and to variations In (L, d).

Hence the inference of chaos is robust to variations in the parameter vector θ within the 20 cases selected by the nested iteration. Furthermore, observe that the inference of positive Lyapunov exponent is robust to an increase in either the time lag, L, or the number of lags, d, but not to a simultaneous increase in both. If d and L are simultaneously increased by 1, so that $(L, d) = (2,3)$, the sign of the

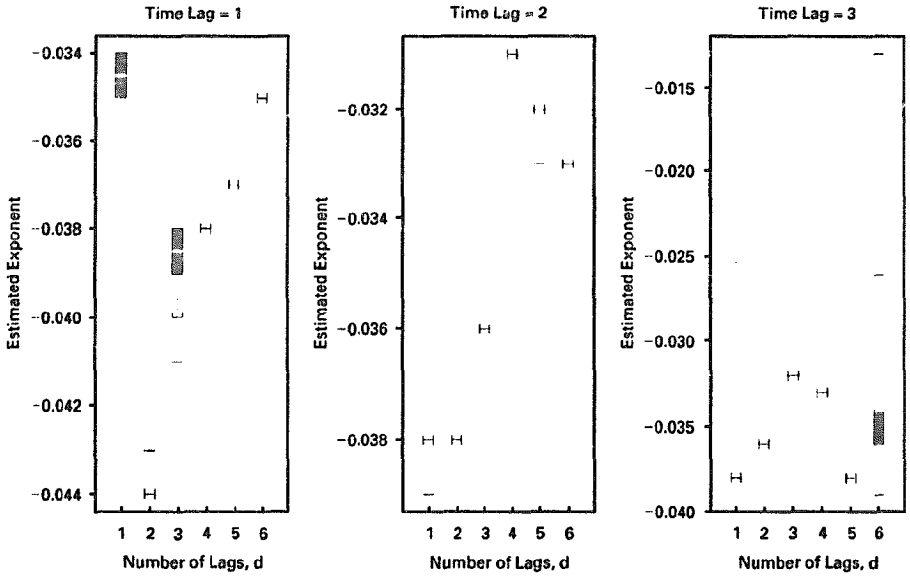

Fig. 2. NEGM sensitivity plot with ARIMA large sample: Indicates sensitivity of the λ estimate to the initial condition for θ and to variations in (L, d).

dominant Lyapunov exponent becomes heavily dependent upon the parameter vector θ. While the use of the neural net method has some instability (and thereby robustness) problems, the stability of that approach is superior to that of the other methods that have been considered in this context.[37] We do not supply the analogous plot for the large sample Feigenbaum data, since the large and small sample plots are similar in that case.

For comparison, the NEGM sensitivity plots are supplied in Fig. 2 for the large sample ARIMA data. According to Table 5, the GCV estimate for (L, d) with the small sample case V data is (1,3). Observing the box corresponding to $(L, d) = (1,3)$ in Fig. 2, we see that the entire range of the box is below zero. Furthermore, observe that the full range of every box for each setting of (L, d) in that plot is negative. Clearly there is not evidence of chaos in the large sample ARIMA data. The small sample ARIMA data produced a similar plot.

9.5. *Results with White's test*

The results with White's test, displayed in Table 7, provide clear evidence against the hypothesis of linearity in the mean of the growth rate data in case I,

[37] This fact is well established in Nychka et al. (1992).

Table 7
White's test

Series	Value of test statistic		Decision at 5% level	
	$T = 380$	$T = 2000$	$T = 380$	$T = 2000$
I (Feig)	299	1998.7	Reject linearity (strongly)	Reject linearity (strongly)
II (GARCH)	4.95	1.27	Accept linearity	Accept linearity
III (NLMA)	5.29	8.20	Accept linearity	Reject linearity
IV (ARCH)	0.39	0.55	Accept linearity	Accept linearity
V (ARMA)	0.079	7.430	Accept linearity (strongly)	Reject linearity

Note: The test for linearity is not against general nonlinearity of the process but against nonlinearity in the mean.

which was the chaotic Feigenbaum data. The strength of that conclusion is evident from the fact that the critical value of the test at the 0.05 level is 5.99, with rejection for values of the test statistic exceeding that critical value. In that case, White's test strongly rejected linearity in the mean with both the small sample size and the large sample. The test correctly accepted linearity in the mean with both the small sample and the large sample of the ARCH and GARCH processes. Although ARCH and GARCH are nonlinear processes, they are linear in the mean.

In the case of the nonlinear moving average data, White's test was able to reject linearity with the large sample, but not with the small sample. The converse happened with the linear ARMA data. In that case, White's test correctly accepted linearity with the small sample, but then rejected linearity with the large sample. The rejection of linearity in the large sample ARMA case is a puzzling failure of the test.

The direction in which the test looks for nonlinearities is chosen at random. To obtain a feel for the variability inherent in the test itself, White's test was replicated 50 times on each time series. The results are summarized in Table 8. The table focuses on the location as well as the spread of the realizations of the test statistic with each data series. The outcome of the replication experiment implies substantial robustness to the randomly selected direction. In particular, the strong rejection of linearity in the Feigenbaum case continues to hold over the entire range of the test statistic in both the small sample and large sample case. Similarly the acceptance of linearity in the small sample ARIMA case holds over the entire range of the test statistic within the 50 replications.

The test statistic for series V in Table 7 with the large sample is slightly outside the range reported for that generating model with the large sample in Table 8. While odd, this result nevertheless does not represent a contradiction, since the

Table 8
Selected descriptive statistics for the results of 50 replications of White's test

Series	T	Min.	Max.	Mean	Std. dev.
I (Feig)	380	298.8	299	299	0.023
	2000	1998.3	1999	1999	0.145
II (GARCH)	380	4.85	5.04	4.94	0.04
	2000	1.91	1.92	1.41	0.13
III (NLMA)	380	4.26	5.34	4.77	0.24
	2000	6.34	8.29	7.44	0.40
IV (ARCH)	380	0.34	2.32	0.78	0.34
	2000	0.26	1.41	0.36	0.17
V (ARMA)	380	0.004	4.135	0.268	0.299
	2000	0.900	7.381	4.219	2.002

Note: Min, max, and mean refer to the minimum, maximum, and mean of the test statistic over the 50 replications, while std. dev. is the standard deviation of the test statistic over the 50 replications.

Table 9
Kaplan test statistics, results from simulated data with 380 observations

Process	Mean K on surrogates	Min K on surrogates	Std Dev. of K on surrogates	Embedding dimension	K on simulated data	Conclusion
I (Feig)	0.121	0.097	0.0133	1	0.00358	Reject linearity
	0.072	0.044	0.0191	2	0.00365	(strongly)
	0.057	0.026	0.0136	3	0.00356	
	0.049	0.036	0.0098	4	0.00318	
II (GARCH)	3.48	3.404	0.0363	1	3.33	Reject linearity
	3.46	3.376	0.0484	2	3.26	
	3.48	3.367	0.0540	3	3.04	
	3.49	3.316	0.0728	4	2.91	
III (NLMA)	1.412	1.384	0.0124	1	1.303	Reject linearity
	1.413	1.364	0.0200	2	1.133	
	1.421	1.377	0.0222	3	1.141	
	1.426	1.365	0.0325	4	1.134	
IV (ARCH)	1.516	1.492	0.0139	1	1.281	Reject linearity
	1.510	1.453	0.0222	2	1.165	
	1.518	1.462	0.0335	3	1.160	
	1.519	1.433	0.0443	4	1.155	
V (ARMA)	3.632	3.565	0.0525	1	3.713	Accept linearity
	3.633	3.494	0.0782	2	3.739	
	3.597	3.411	0.1211	3	3.481	
	3.531	3.098	0.1981	4	3.482	

Note: K is the Kaplan test statistic. Twenty surrogates were used, and hence the mean, minimum, and standard deviations are over the 20 surrogates. Embedding dimension, m, as defined by Kaplan, is $m = 1$, when embedding dimension is defined as in the BDS or NEGM tests. Hence add 1.0 to each embedding dimension in the table to acquire consistency with the definitions used by BDS and NEGM. Time delay was determined as in Kaplan (1994).

test results reported in Table 7 are not included among the 50 replications used in producing Table 8.

9.6. Results with Kaplan's test

The null hypothesis for Kaplan's test is linearity of the process. The results with Kaplan's test are displayed in Tables 9 and 10. The test was successful in all cases, including all generating models and all sample sizes. Based upon the very low tail area of the test in the case of the Feigenbaum map, Kaplan's test appears to have very strong power against chaos and hence can be expected not to accept linearity when the data is chaotic. However, the test in its current form can either accept of reject linearity, but cannot accept chaos, which is not the test's null hypothesis. In that sense the model is similar to the BDS test, although the success rate of Kaplan's test in this competition exceeded that of the BDS test.

Table 10
Kaplan test statistics, results from simulated data with 2000 observations

Process	Mean K on surrogates	Min K on surrogates	Std Dev. of K on surrogates	Embedding dimension	K on simulated data	Conclusion
I (Feig)	0.163	0.086	0.0200	1	4×10^{-6}	Reject linearity
	0.125	0.110	0.0119	2	3×10^{-6}	(very strongly)
	0.096	0.043	0.0166	3	4×10^{-6}	
	0.064	0.019	0.0221	4	2×10^{-6}	
II (GARCH)	4.003	3.863	0.0738	1	3.905	Reject linearity
	3.983	3.690	0.1300	2	3.661	(marginally)
	4.006	3.624	0.1457	3	3.424	
	4.047	3.701	0.1748	4	3.280	
III (NLMA)	1.470	1.405	0.0412	1	1.394	Reject linearity
	1.473	1.358	0.0559	2	1.240	
	1.457	1.354	0.0639	3	1.135	
	1.458	1.263	0.0869	4	1.162	
IV (ARCH)	1.695	1.608	0.0393	1	1.337	Reject linearity
	1.678	1.581	0.0534	2	1.230	
	1.681	1.543	0.0779	3	1.173	
	1.703	1.483	0.0892	4	1.161	
V (ARMA)	4.382	3.708	0.3148	1	4.089	Accept linearity
	4.542	3.889	0.3972	2	3.790	
	4.436	3.611	0.5381	3	4.355	
	4.181	2.623	0.7026	4	4.885	

Note: K is the Kaplan test statistic. Twenty surrogates were used, and hence the mean, minimum, and standard deviations are over the 20 surrogates. Embedding dimension, m, as defined by Kaplan, is $m-1$, when embedding dimension is defined as in the BDS or NEGM tests. Hence add 1.0 to each embedding dimension in the table to acquire consistency with the definitions used by BDS and NEGM. Time delay was determined as in Kaplan (1994).

Observe the somewhat stronger rejection of linearity in the ARCH case than in the GARCH case. Perhaps the Kaplan test may have somewhat higher power against ARCH than against GARCH. The same result was acquired with the BDS test.

10. Conclusions

We find some consistency in our inferences across methods of inference, although there are some clear differences among the power functions of the tests. It is possible that greater robustness across inference methods might be attained at much greater sample size, although the results with the 2000 observation sample size probably capture much of the characteristics of the tests with large samples.[38] None of these tests, which are among the best of the available tests for nonlinearity and chaos in noisy data, has the ability to isolate the origins of the nonlinearity or chaos to be within the structure of the economy. These tests, which do not condition upon any particular economic structure, could detect deterministic nonlinear or chaotic weather conditions that are transmitted to economic variables through a linear economic structure, as emphasized recently by Day (1992).

Two considerations are important in interpreting the differences in the results among some of these tests. One consideration is the differences in the power functions over alternatives, for fixed null. The other consideration is the differences in null hypotheses of each test. The latter consideration produces a degree of noncomparability of the tests and the possibility that some of the tests could be used jointly. For example, the bispectrum test has no power against those forms of nonlinearity that produce flat bispectrum and non-flat higher order polyspectra. Hence the 'linearity' hypothesis usually viewed as the null of the test actually is correctly interpreted as lack of third-order nonlinear dependence. In fact the bispectrum test also has low power against those forms of chaos that produce irregular and widely spaced spikes in the bispectrum. Such singular and widely spaced spikes can be difficult for the Hinich test statistic to detect, although the spikes become evident from visual inspection of the bispectrum.[39]

---

[38] Relative to the literature on empirical economics, our large sample is very large. Nevertheless, much larger samples are common in the physical sciences, and in some of our results there is evidence that the large sample may not be large enough. For example, White's test in one case did better with the small sample than with the large sample. It is possible that small sample properties are still being seen with the 2000 observation data, and an even larger sample would produce better results.

[39] See, e.g., Ashley and Patterson (1989, p. 690). The problem in those cases is associated with the fact that the test is based upon only the 80th quantile of an empirical distribution function. Using more quantiles, or a Kolmogonov–Smirnov statistic using all of the quantiles, could raise power.

Some of the 'competing' tests could be viewed as complementary, rather than competing. Using all of them jointly can produce deeper insight into the nature of the nonlinearity that may exist in the data.[40] In particular, the BDS and Kaplan tests are omnibus tests that test linearity against all possible alternatives to exact linearity. Those tests seem to be very sensitive to departures from linearity, and the values of the test statistic for the BDS test were dramatically convincing in the extreme cases of linearity and chaos. The Kaplan test's characteristics appear to be similar to those of the BDS test, although the Kaplan test is newer, and its properties have not yet been as extensively investigated as those of the BDS test. However, it is noteworthy to observe that in our experiments the Kaplan test, unlike the BDS test, got the right answer in every case, with both the large and small samples. Hence it would seem that the BDS or Kaplan test, or perhaps both tests, could be the first test run to rule out the narrowest null of exact linearity.

If linearity is rejected with the BDS and Kaplan test, it becomes reasonable to use more focused tests to try to distinguish among the possible forms of nonlinearity. For example, the bispectral test could be used to distinguish between third-order nonlinear dependence and other forms of nonlinearity, if linearity already has been rejected by the BDS or Kaplan test. White's test has very high power against chaos and can be used to distinguish among those nonlinear processes that are nonlinear in the mean (such as the NLMA) and those that are not (such as ARCH and GARCH). Hence before proceeding to the NEGM test, which is focused specifically on chaos as the null, White's test could be run. If linearity is rejected with White's test, the computationally difficult and very focused NEGM test becomes well worth running.

If used jointly in this manner, problems of pretesting arise, including questions regarding statistical significance of tests run conditionally upon the results of prior tests. Nevertheless, we believe that few alternatives currently exist to sequential learning from data in that manner, since many specific forms of nonlinear structure are worth investigating, including chaos. Simply rejecting linearity is not likely to exhaust the useful information in the data about nonlinear structure.

Finally it should be observed that we have by no means exhausted all possible interesting cases in our competition. For example, the competition would have benefited from the inclusion of (1) a higher dimensional case to permit investigation of the properties of the order determination algorithm used in some of the tests, (2) an even larger sample to permit determination of whether or not the

[40] We are indebted to William Brock for suggesting this idea to us in a private correspondence with William Barnett.

2000 observation case was large enough to support the use of asymptotic inference, and (3) the inclusion of a noisy chaotic case. But the computational burdens upon the participants in this competition were already pressing the limits that could reasonably be expected of those courageous enough to subject their tests to this professionally risky competition.

References

Ashley, R.A., Patterson, D.M., 1989. Linear versus nonlinear macroeconomies: a statistical test. International Economic Review 30, 685–704.

Ashley, R.A., Patterson, D.M., Hinich, M., 1986. A diagnostic test for nonlinear serial dependence in time series fitting errors. Journal of Time Series Analysis 73, 165–178.

Barnett, W.A., Chen, P., 1986. Economic theory as a generator of measurable attractors. Mondes en Developpement 14(453), 13–28; reprinted In: Prigogine, I. and Sanglier, M. (Eds.), Laws of Nature and Human Conduct: Specificities and Unifying Themes, G.O.R.D.E.S., Brussels, pp. 209–224.

Barnett, W.A., Chen, P., 1988a. The aggregation-theoretic monetary aggregates are chaotic and have strange attractors: an econometric application of mathematical chaos. In: Barnett, W., Berndt, E., White, H. (Eds.), Dynamic Econometric Modeling, Proceedings 3rd International Symposium in Economic Theory and Econometrics. Cambridge University Press, Cambridge, pp. 199–246.

Barnett, W.A., Chen, P., 1988b. Deterministic chaos and fractal attractors as tools for nonparametric dynamical econometric inference. Mathematical Computer Modeling 10, 275–296.

Barnett, W., Hinich, M.J., 1992. Empirical chaotic dynamics in economics. Annals of Operations Research 37, 1–15.

Barnett, W., Hinich, M.J., 1993. Has chaos been discovered with economic data. In: Chen, P., Day, R. (Eds.), Evolutionary Dynamics and Nonlinear Economics. Oxford University Press, Oxford, pp. 254–263.

Barnett, W.A., Gallant, A.R., Hinich, M.J., Jungeilges, J., Kaplan, D., Jensen, M.J., 1995. Robustness of nonlinearity and chaos tests to measurement error, inference method, and sample size. Journal of Economic Behavior and Organization 27, 301–320.

Barnett, W.A., Gallant, A.R., Hinich, M.J., Jensen, M.J., Jungeilges, J., 1996a. Comparisons of the available tests for nonlinearity and chaos. In: Barnett, W.A., Gandolfo, G., Hillinger, C. (Eds.), Dynamic Disequilibrium Modeling: Theory and Applications. Cambridge University Press, Cambridge, pp. 313–346.

Barnett, W.A., Gallant, A.R., Hinich, M.J., Jungeilges, J., Kaplan, D., Jensen, M.J., 1996b. An experimental design to compare tests of nonlinearity and chaos. In: Barnett, W., Kirman, A., Salmon, M. (Eds.), Nonlinear Dynamics in Economics. Cambridge University Press, Cambridge, pp. 163–191.

Bickel, P.J., Bühlmann, P., 1996. What is a Linear Process?, Proceedings of the National Academy of Science, USA, Statistics Section, 93, pp. 12,128–12,131.

Bierens, H. 1990. A consistent conditional moment test of functional form, Econometrica 58, 1443–1458.

Brillinger, D.R., 1965. An introduction to polyspectrum, Annals of Mathematical Statistics 36, 1351–1374.

Brock, W. A., Dechert, W.D., LeBaron, B., Scheinkman, J., 1996. A test for independence based on the correlation dimension. Econometric Reviews 15, 197–235.

Brock, W.A., Hsieh, D.A., LeBaron, B., 1991. Nonlinear Dynamics, Chaos, and Instability: Statistical Theory and Economic Evidence. MIT Press, Cambridge, MA.

Brock, W.A., Lakonishok, J., LeBaron, B., 1992. Simple technical trading rules and the stochastic properties of stock returns. Journal of Finance, December, 1731–1764.

Broomhead, D.S., Huke, J.P., Muldoon, M.R., 1992. Linear filters and non-linear systems. Journal of the Royal Statistical Society B 54, 373–382.

Casdagli, M., Eubank, S., Farmer, J.D., Gibson, J., 1991. State space reconstruction in the presence of noise. Physica D 51, 52–98.

Dalle Molle, J.W., Hinich, M.J., 1995. Trispectral analysis of stationary random time series. Journal of the Acoustical Society of America 97, 2963–2978.

Dalle Molle, J.W., Hinich, M.J., 1991. Cumulant spectra-based tests for the detection of coherent signal in noise. In: Lacoume, J.L., Lagunas, M.A., Nikias, C.L. (Eds.), Proceedings of the International Signal Processing Workshop on Higher Order Statistics, pp. 151–154.

DeCoster, G.P., Mitchell, D.W., 1991a. Nonlinear monetary dynamics. Journal of Business and Economic Statistics 9, 455–462.

DeCoster, G.P., Mitchell, D.W., 1991b. The efficacy of the correlation dimension technique in detecting determinism in small samples. Journal of Statistical Computer Simulation 39, 221–229.

DeCoster, G.P., Mitchell, D.W., 1992. Dynamic implications of chaotic monetary policy. Journal of Macroeconomics 14, 267–287.

DeCoster, G.P., Mitchell, D.W., 1994. Reply. Journal of Business and Economic Statistics 12, 136–137.

Eckmann, J.-P., Ruelle, D., 1985. Ergodic theory of chaos and strange attractors, Review of Modern Physics 57, 617–656.

Ellner, S., Gallant, A.R., McCaffrey, D., Nychka, D., 1991. Convergence rates and data requirements for Jacobian-based estimates of Liapunov exponents from data. Physics Letters A 153, 357–363.

Gallant, A.R., White, H., 1988. There exists a neural network that does not make avoidable mistakes. In: Proceedings 2nd IEEE International Conference on Neural Networks, 24–27 July SOS Printing, San Diego, 1.657–1.664.

Gallant, A.R., White, H., 1992. On learning the derivatives of an unknown mapping with multilayer feedforward networks. Neural Networks 5, 129–138.

Gencay, R., Dechert, W.D., 1992. An algorithm for the n Lyapunov exponents of an n-dimensional unknown dynamical system. Physica D 59, 142–157.

Hinich, M.J., 1982. Testing for Gaussianity and linearity of a stationary time series. Journal of Time Series Analysis 3(3), 169–176.

Hinich, M.J., 1996. Testing for dependence in the input to a linear time series model. Nonparametric Statistics 6, 205–221.

Hinich, M.J., Patterson, D., 1985. Identification of the coefficients in a non-linear time series of the quadratic type. Journal of Econometrics 30, 269–288. Reprinted In: Barnett, W., Gallant, R., (Eds.), New Approaches to Modelling, Specification Selection, and Econometric Inference, Proceedings 1st International Symposium in Economic Theory and Econometrics. Cambridge University Press, Cambridge, 1989.

Hinich, M.J., Patterson, D., 1989. Evidence of nonlinearity in the trade-by-trade stock market return generating process. In: Barnett, W., Geweke, J., Shell, K. (Eds.), Economic Complexity: Chaos, Sunspots, Bubbles, and Nonlinearity, Proceedings 4th International Symposium in Economic Theory and Econometrics. Cambridge University Press, Cambridge. pp. 383–409.

Hsieh, D., Le Baron, B., 1991. Finite sample properties of the BDS statistic. In: Brock, H., LeBaron, (Eds.), Nonlinear Dynamics, Chaos, and Instability: Statistical Theory and Economic Evidence. MIT Press, Cambridge, MA.

Jungeilges, J.A., 1996. Operational charactristics of White's test for neglected nonlinearities. In: Barnett, W., Kirman, A., Salmon, M. (Eds.), Nonlinear Dynamics in Economics. Cambridge University Press, Cambridge, pp. 219–266.

Kaplan, D.T., 1994. Exceptional events as evidence for determinism, Physica D 73, 38–48.

Kinderman, A.J., Ramage, J.G., 1976. Computer generation of normal random numbers. Journal of the American Statistical Association 71, 893–896.

Kuan, C., White, H., 1991. Artificial neural networks: an econometric perspective, Working paper, Department of Economics and Institute for Neural Computation, University of California at San Diego, San Diego, CA.

Lee, T.-H., White, H., Granger, C., 1993. Testing for neglected nonlinearities in time series models. Journal of Econometrics 56, 269–290.

McCaffrey, D. F., Ellner, S., Gallant, A.R., Nychka, D.W., 1992. Estimating the Lyapunov exponent of a chaotic system with nonparametric regression. Journal of the American Statistical Association 87, 682–695.

Nychka, D., Ellner, S., Gallant, R., McCaffrey, D., 1992. Finding chaos in noisy systems, Journal of the Royal Statistical Society B 54, 399–426.

Ramsey, J.B., Rothman, P., 1994. Comment on 'nonlinear monetary dynamics' by DeCoster and Mitchell, Journal of Business and Economic Statistics 12, 135–136.

Ramsey, J.B., Sayers, C.L., Rothman, P., 1990. The statistical properties of dimension calculations using small data sets: some economic applications. International Economic Review 31, 991–1020.

Scheinkman, J., LeBaron, B., 1989. Nonlinear dynamics and GNP data. In: Barnett, W., Geweke, J., Shell, K. (Eds.), Economic Complexity: Chaos, Sunspots, Bubbles, and Nonlinearity, Proceedings 4th International Symposium in Economic Theory and Econometrics. Cambridge University Press, Cambridge, pp. 213–227.

Schwarz, G., 1978. Estimating the dimension of a model, Annals of Statistics 6, 461–464.

Serletis, A., 1995. Random walks, breaking trend functions, and the chaotic structure of the velocity of money, Journal of Business and Economic Statistics 4, 453–458.

Smith, R.L., 1992. Estimating dimension in noisy chaotic time series. Journal of the Royal Statistical Society B 54, 329–351.

Subba Rao, T., Gabr, M., 1980. A test for linearity of stationary time series, Journal of Time Series Analysis 1, 145–158.

Walden, A.T., Williams, M.L., 1993. Deconvolution, bandwidth, and the trispectrum. Journal of the American Statistical Association 88, 1323–1329.

White, H., 1989a. Some asymptotic results for learning in single hidden-layer feedforward network models. Journal of the American Statistical Association 84, 1003–1013.

White, H., 1989b. An additional hidden unit test for neglected nonlinearity in multilayer feedforward networks. In: Proceedings of the International Joint Conference on Neural Networks, vol. 2. IEEE Press, New York, pp. 451–455.