Journal of Econometrics 30 (1985) 269–288. North-Holland

IDENTIFICATION OF THE COEFFICIENTS IN A NON-LINEAR TIME SERIES OF THE QUADRATIC TYPE

Melvin J. HINICH

University of Texas, Austin, TX 78712, USA

Douglas M. PATTERSON

Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA

This paper presents a method for identifying the structure of the quadratic model $x(t) = e(t) + \sum_{r=1}^{R} \sum_{s=0}^{S} a(r,s)e(t-s)e(t-r-s)$, where the $e(t)$ are independent, identically distributed zero mean random variables. This finite parameter model is a simplified second-order Volterra expansion of a stable non-linear filter. The standard deviation of $e(t)$ is not estimable from observations of the $x(t)$. It is shown that the $(m, m+n)$th third-order cumulant (called the bicovariance) of $\{x(t)\}$ is equal to $a(m,n)$ times the square of the variance of $e(t)$. The large sample distribution of the sample bicovariance is used to determine which quadratic coefficients are significantly different from zero. The method is illustrated using artificial and real data; daily stock price series for several securities.

1. Introduction

There is no compelling reason why time series found in the physical sciences and various social science applications should conform to linear time series models. Linear models, usually having a finite number of parameters, are employed for ease of statistical analysis, just as the assumption that variables are normally distributed is made for convenience and then is often accepted as a fact when data are interpreted. Indeed, a stationary *non-Gaussian* time series is usually the result of some non-linear operation on a Gaussian input process. The reason why time series analysis has historically relied so heavily on linear models is the dearth of non-linear statistical tools.

Recently, a few statistical methods designed to detect certain types of non-linearity in a time series have appeared in the literature. For the most part, these methods are based on the behavior of the third-order cumulant function of a non-linear time series. For example, Subba Rao and Gabr (1980) present a test for non-linearity using a sample estimate of the bispectrum of the time series. (The bispectrum is defined as the double Fourier transform of the third-order cumulant function.) Hinich (1982) presents a non-parametric test that also uses the sample bispectrum, but which takes advantage of the asymptotic properties of the bispectrum estimator. Maravall (1983) developed

0304-4076/85/$3.30

an easily applied test for detecting non-linearity in the estimated residuals of a time series which uses the autocorrelation function of the squared residuals. Also, Hinich and Patterson (1985) applied the Hinich test to realizations of daily stock returns for a sample of stocks covering the period July 1962 through December 1977. They rejected linearity for each of the fifteen series.

Although time series theorists have made considerable progress in developing the theoretical properties of non-linear models, an efficient statistical method for estimating these models in a parametric form from a finite set of observations remains elusive. A practical estimation method will, no doubt, consist of at least four iterative steps: (1) Detection of non-linearity in the data at hand. As pointed out in the previous paragraph, some progress has been made in this direction, at least for the case where the underlying process has a non-zero third-order cumulant function. (2) Identification, through use of the data, of candidate models to be tentatively considered. This step might also provide rough estimates of model parameters. (3) Estimation of candidate model parameters using appropriate statistical methods. This step may require that the candidate model be inverted; that is, to express the innovations as a function of past values of the non-linear process. In some situations inversion may not be necessary; see Haggan et al. (1984) for a concrete example of this approach. (4) Diagnostic checks. Here lack of fit of the data to the candidate model is discovered. Obviously, the tools employed in steps (1) and (2) can be profitably turned to the task of uncovering model inadequacy and suggesting model improvements.

In this article we present an applied method of model identification for that class of non-linear models featuring a non-zero third-order cumulant function. Specifically, the method presented below is designed to identify the coefficients in a non-linear time series of the quadratic type with a simple linear term. This class of non-linear models can be written as

$$x(t)=\varepsilon(t)+\sum_{m=1}^{R}\sum_{n=0}^{S}a_m(n)\varepsilon(t-n)\varepsilon(t-m-n),$$

where the $\varepsilon(t)$ are independent, identically distributed random deviates with zero mean. These models are characterized as having a 'simple linear term' because the linear component is uncorrelated; it consists of the innovation only, rather than a linear transformation of the innovation. The quadratic non-linear component for this model class restricts all of the non-linear variation to the third-order cumulant function.

The article is organized as follows. Section 2 provides a brief introduction to the general non-linear model. The reader will find a much more thorough presentation in the references. Section 3 presents our identification method, and briefly describes certain estimators and their sampling properties which we have developed for application to non-linear identification. In section 4 we demonstrate the identification method through the use of artificial data. The final section of the article applies the identification method to realizations of daily stock returns for a sample of fifteen randomly selected stocks, with some interesting results.

2. Non-linear stationary time series models

A general non-linear causal time series model can be regarded as the 'output' of a non-linear system whose 'input' is a stationary random process $\{\varepsilon(t)\}$. The output is of the form

$$x(t)=f[\varepsilon(t),\varepsilon(t-1),\ldots], \tag{2.1}$$

where f is a non-linear function that does not depend on t. Non-linear time series are analogous to ordinary non-linear differential equations in that they can exhibit limit cycles, clipping, hysteresis, and so on.

An important insight to the nature of the general non-linear model is provided by expressing (2.1) as a discrete time Volterra series expansion:

$$\begin{aligned} x(t)=&\sum_{i=0}^{\infty}a(i)\varepsilon(t-i)+\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}a(i,j)\varepsilon(t-i)\varepsilon(t-j)\\ &+\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sum_{k=0}^{\infty}a(i,j,k)\varepsilon(t-i)\varepsilon(t-j)\varepsilon(t-k)+\cdots. \end{aligned} \tag{2.2}$$

The $a(i,j,k,\ldots)$ are the coefficients of the series and the $\varepsilon(t)$ input process is usually considered to be unobservable. Wiener (1958), in a classic study of non-linear systems, used the continuous time version of (2.2) to express the non-linear relationship between the input and output of a physical system. When $\{\varepsilon(t)\}$ is a purely random process with zero mean, the first term in (2.2) is a general linear model, and the successive terms are usually referred to as the 'quadratic', 'cubic',... components.

Consider the following example of the quadratic model:

$$x(t)=\varepsilon(t)+b\varepsilon(t-1)\varepsilon(t-2). \tag{2.3}$$

The parameter b can take on any non-zero value. Now, $\mathrm{E}[X(t)]=\mu_x=0$ since $\mathrm{E}[\varepsilon(t)]=0$, and the $\varepsilon(t)$ are independent variates. This model is also a white-noise series because its autocovariance function $c_x(m)=\mathrm{E}[x(t+m)\times x(t)]-\mu_x^2=0$ for all $m\neq 0$. Although the $x(t)$ variates are uncorrelated, there is a relationship between $x(t-2)$, $x(t-1)$, and $x(t)$ for all t, because

$E[x(t-2)x(t-1)x(t)] = b\sigma_\varepsilon^4$. This term is one element of the set of third-order cumulants of $\{x(t)\}$, which are defined for each (r, s) as

$$c_{xxx}(r,s) = E[x(t+r)x(t+s)x(t)], \tag{2.4}$$

when $E[x(t)] = 0$. From stationarity the following symmetries hold for all (r, s):

$$c_{xxx}(r,s) = c_{xxx}(s,r) = c_{xxx}(r-s,-s) = c_{xxx}(s-r,-r).$$

Note that for a zero-mean Gaussian time series, $c_{xxx}(r,s) \equiv 0$.

The quadratic time series shown in (2.3) above has six non-zero third-order cumulants, namely $c_{xxx}(1,2) = c_{xxx}(-1,-2) = c_{xxx}(1,-1)$, plus their three permutation symmetries. The general quadratic time series with a simple linear term

$$x(t) = \varepsilon(t) + \sum_{m=1}^{\infty} \sum_{n=0}^{\infty} a_m(n)\varepsilon(t-n)\varepsilon(t-m-n), \tag{2.5}$$

with parameters $a_1(0), a_1(1), \ldots, a_m(n), \ldots$, has many non-zero cumulants and is white-noise if $a_m(n)a_m(n+r) = 0$. These processes are clearly non-Gaussian. From the Volterra series expansion of the general non-linear model in expression (2.2), we see how the general quadratic model can be regarded as an approximation to the general non-linear model when the linear component is simple. However, there are non-linear models where $c_{xxx}(r,s) = 0$ for all (r, s), and it is easy to construct linear models which just have a non-zero contemporaneous third-order cumulant. Hence, it is necessary to be able to distinguish between a linear non-Gaussian process (if the linear process has a non-zero third-order cumulant it is non-Gaussian) and a non-linear process. The test developed by Hinich (1982) accomplishes this task.

Because third-order cumulants are difficult to interpret, and their moment estimates even more difficult to comprehend, the Hinich test employs the statistical properties of the double Fourier transform of the third-order cumulant function (called the *bispectrum*).

The bispectrum, denoted by $B_x(f_1, f_2)$, is defined as follows: For frequencies f_1 and f_2,

$$B_x(f_1,f_2) = \sum_{r=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} c_{xxx}(r,s)\exp[-i2\pi(f_1 r + f_2 s)], \tag{2.6}$$

assuming that $|c_{xxx}(r,s)|$ is summable. The symmetries of $c_{xxx}(r,s)$ translate into symmetries of $B_x(f_1, f_2)$ that produce a principal domain for the bispectrum, which is the triangular set $\Omega = \{0 < f_1 < \frac{1}{2}, 2f_1 + f_2 < 1\}$. The third-order cumulants are expressed in terms of the bispectrum by the inverse Fourier transform,

$$c_{xxx}(r,s) = \int\int_{\Omega} B_x(f_1, f_2)\exp[i2\pi(f_1 r + f_2 s)]\,df_1\,df_2. \tag{2.7}$$

Eq. (2.7) implies that a consistent estimator of the bispectrum yields consistent estimators of the $c_{xxx}(r,s)$. See Brillinger and Rosenblatt (1967), Brillinger (1975), Subba Rao (1983), and Rosenblatt (1983) for a rigorous treatment of the bispectrum.

Suppose that a process $\{x(t)\}$ is generated by a linear model with $\{\varepsilon(t)\}$ a non-Gaussian purely random process. If $\sum_{i=0}^{\infty}|a(i)|$ is finite,

$$B_x(f_1, f_2) = \mu_3 A(f_1)A(f_2)A^*(f_1+f_2), \tag{2.8}$$

where $\mu_3 = \mathrm{E}[\varepsilon^3(t)]$, the asterisk denotes complex conjugate, and $A(f) = \sum_{k=-\infty}^{\infty} b(k)\exp(-i2\pi fk)$. Since the spectrum of $x(t)$ is $S_x(f) = \sigma_\varepsilon^2 |A(f)|^2$, it follows from (2.8) that

$$(|B_x(f_1,f_2)|^2)/(S_x(f_1)S_x(f_2)S_x(f_1+f_2)) = \mu_3^2/\sigma_\varepsilon^6, \tag{2.9}$$

for all (f_1, f_2) in the set Ω. If $\varepsilon(t)$ is Gaussian, $\mu_3 = 0$ and then $B(f_1, f_2) \equiv 0$. The square root of the left side of (2.9) is called the skewness of $\{x(t)\}$. The fact that the skewness of a linear time series is independent of frequency is used by Hinich to test the null hypothesis of linearity using the estimated bispectrum [the estimated bispectrum is used in a similar manner by Subba Rao and Gabr (1980) to test the linearity null hypothesis].

The upshot of this discussion is that if we apply the linearity test to an empirical time series, and if there is sufficient evidence to reject linearity, then we are justified in concluding that $c_{xxx}(r,s) \neq 0$ for some $r \neq 0$ and $s \neq 0$. Furthermore, we can now regard some unspecified form of the general quadratic model, (2.5), as an approximation to the underlying non-linear process.

We now turn to the problem of determining which of the $a_m(n)$ parameters in the general quadratic model are non-zero.

3. Identification of the quadratic parameters

Suppose a time series is adequately approximated by a quadratic time series model with a simple linear term:

$$x(t) = \varepsilon(t) + \sum_{m=1}^{\infty}\sum_{n=0}^{\infty} a_m(n)\varepsilon(t-n)\varepsilon(t-m-n), \tag{3.1}$$

where $\{\varepsilon(t)\}$ is an independent, identically distributed sequence of random variables with mean zero and variance denoted σ_ε^2. The $\{\varepsilon(t)\}$ process is unobservable. We want to determine those parameters $a_m(n)$ in (3.1) which are non-zero. Or, in other words, we want to identify a representational form of (3.1) which is worthy of serious consideration. Our technique for accomplishing this goal will be explained by stating and discussing four key theorems. The proofs of the theorems can be found in Hinich and Patterson (1984).

Theorem 1. The expected value of $x(t)$, defined in (3.1), is zero, and the third-order cumulant (bicovariance) is

$$\mathrm{E}[x(t)x(t+m)x(t+m+n)] = \sigma_\varepsilon^4 a_m(n), \tag{3.2}$$

for all $m \geq 1$ and $n \geq 1$. For $n = 0$, $\mathrm{E}[x(t)x^2(t+m)] = 2\sigma_\varepsilon^4 a_m(0)$.

Because $\varepsilon(t)$ is unobservable, we can not estimate σ_ε. As a consequence, we can only determine the coefficients up to the scale factor σ_ε^4, as shown in (3.2).

Suppose that we have a sample $x(1),\ldots,x(N)$ from the process. For each $m, n \geq 1$, the sample bicovariance $(N-m-n)^{-1}\sum_{t=1}^{N-m-n} x(t)x(t+m)x(t+m+n)$ is an unbiased estimator of $a_m(n)$ from Theorem 1. This unbiased estimator, as with the standard sample autocovariances, has a variance that increases with the size of the lags. As $m+n$ increases the number of products decreases in the above sum. The dependence of the variance of the sample bicovariance on $m+n$ is not its only problem. It is inefficient to compute all the sums of lagged products for a large number of m and n values. A more computationally convenient estimator uses the discrete Fourier transform of the data.

Using the convention that the first observation is taken at time zero, the discrete Fourier transform of the data is defined to be

$$X(\omega_j) = \sum_{t=0}^{N-1} x(t)\exp(-\mathrm{i}\omega_j t), \qquad j = 0,1,\ldots,N-1, \tag{3.3}$$

where $\omega_j = 2\pi j/N$ is the angular frequency with period N/j. Any one of the versions of the Fast Fourier transform is an efficient way to compute $X(\omega_j)$ for $j = 0,1,\ldots,N-1$. Since $x(t)$ is real, $X(\omega_{N-j}) = X^*(\omega_j)$ where again the asterisk denotes complex conjugate. The following estimator of $a_m(n)$ is biased, but it has the same variance for all m and $n \geq 1$ and it is computationally more efficient than the sample bicovariance approach for large N. Remember that from Theorem 1, $a_m(n)$ can be estimated only up to a scalar multiple since σ_ε is an unknown parameter.

Theorem 2. Define the estimator

$$\hat{a}_m(n) = N^{-3} \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} X(\omega_j) X(\omega_{k-j}) X^*(\omega_k) \exp\left[\mathrm{i}(\omega_j n + \omega_k m\right]. \tag{3.4}$$

The inverse discrete Fourier transform of $X(\omega_j)$ *as defined in (3.4) gives* $x(t) = N^{-1}\sum_{j=0}^{N-1} X(\omega_j)\exp(\mathrm{i}\omega_j t)$, *where* $x(t+N) = x(t)$ *for each* $t = 0, \ldots, N-1$. *Then,* $\hat{a}_m(n) = N^{-1}\sum_{t=0}^{N-1} x(t)x(t+m)x(t+m+n)$.

The right-hand side of the expression for $\hat{a}_m(n)$ in (3.4) is the discrete double Fourier transform of the complex array $\{N^{-1}X(\omega_j)X(\omega_{k-j})X^*(\omega_k)\}$ for $j, k = 0, \ldots, N-1$. It can be computed by N Fast Fourier transforms of vectors of length N. The estimator is asymptotically normal.

The result in Theorem 2 can be used to efficiently compute the array of sample bicovariances by means of a simple trick to break the circularity in the $x(t)$'s. Suppose that N zeros are added to the end of the data array. The Fourier transform of this extended array of length $2N$ is $X(2\pi j/2N)$ for $j = 0, \ldots, 2N-1$. A double Fourier transform of the $(2N)\times(2N)$ array $\{N^{-1}X(\pi j/N)X(\pi(k-j)/N)X^*(\pi k/N)\}$ will yield the bicovariances since the $x(t)$'s are now periodic with period $2N$ with N consecutive zeros.

It can also be shown that in large samples the estimator $a_m(n)$ defined in Theorem 2 is correlated along both the m and n dimensions, and that the correlation falls off more slowly along the n dimension. In particular, the correlation along the n dimension rolls off at least as fast as the order of the difference in the n's.

The following results give the expected value and variance of $\hat{a}_m(n)$:

Theorem 3. For all $1 \leq m < N$,

$$\sigma_\varepsilon^{-4}\mathrm{E}[\hat{a}_m(0)] = 2(1 - m/N)a_m(0), \tag{3.5}$$

and for $n \geq 1$,

$$\sigma_\varepsilon^{-4}\mathrm{E}[\hat{a}_m(n)] = [1-(m+n)/N]a_m(n) + (n/N)a_{N-m-n}(m) + (m/N)a_n(N-m-n). \tag{3.6}$$

Corollary 3.1. If $m + n \leq N/2$, *then for* $n \leq 1$,

$$\sigma_\varepsilon^{-4}\mathrm{E}[\hat{a}_m(n)] = [1-(m+n)/N]a_m(n). \tag{3.7}$$

Theorem 3 and its corollary show that the estimator of Theorem 2 is downward biased, with the bias following a ramp function. The variance of the estimator is given by Theorem 4.

Theorem 4. Assume that $\sum_{r=1}^{\infty}\sum_{s=0}^{\infty} a_r^2(s)$ *is finite, and let* $c_4 = \mathrm{E}[x^4(t)] - 3\sigma_x^4$ *denote the fourth cumulant of* $x(t)$. *Then,*

$$\operatorname{var}[\hat{a}_m(n)] = \sigma_x^6/N + \mathrm{O}(N^{-2}), \tag{3.8}$$

$$\operatorname{var}[\hat{a}_m(0)] = (c_4 + 3\sigma_x^4)\sigma_x^2/N + \mathrm{O}(N^{-2}). \tag{3.9}$$

We see from Theorem 4 that in testing the significance of an estimate, we must distinguish between the $n = 0$ cases and the $n \geq 1$ cases.

It can be shown that the signal to signal plus noise ratio ($SSPN$) of the quadratic process (3.1) is given by the sum of squares of the coefficients over the variance of the process cubed. That is,

$$SSPN = \frac{\sigma_\varepsilon^8 \sum_{m=1}^{\infty} \sum_{n=0}^{\infty} a_m^2(n)}{\sigma_\varepsilon^6 \left(1 + \sigma_\varepsilon^2 \sum_{m=1}^{\infty} \sum_{n=0}^{\infty} a_m^2(n)\right)^3} = \frac{\sigma_\varepsilon^2 \sum_{m=1}^{\infty} \sum_{n=0}^{\infty} a_m^2(n)}{\left(1 + \sigma_\varepsilon^2 \sum_{m=1}^{\infty} \sum_{n=0}^{\infty} a_m^2(n)\right)^3}. \tag{3.10}$$

The ratio has a maximum of (4/27), or -8.29 dB. The $SSPN$ ratio measures the relative contribution of the quadratic terms to the variance of the process. The greater the ratio, the greater the quadratic non-linearity in a model.

In order to implement our identification method, we wrote a computer algorithm which can handle up to 1280 observations. The program breaks a time series into blocks of 256 observations, finds the coefficients for each block of 256, with 256 zeros added to the end of the array, and averages them together. This method was chosen because we wanted to limit the program size to less than 1 megabyte of core (storage requirements go up as $8N^2$). The $n = 0$ estimates are divided by 2 to adjust for part of the bias shown in Theorem 3. The program does not adjust for the ramp bias in calculating the coefficients.

With a sample size of 1280 this technique introduces additional correlation into the estimates beyond the large sample correlation mentioned above. This occurs because we are estimating $256 \times 256 = 65536$ coefficients from a total sample of 1280 observations. Obviously, the estimates are not independent. The correlation will not be particularly severe when the underlying quadratic model is sparse in non-zero coefficients. Coefficient sparseness is a reasonable assumption when considering real data, as will be shown in section 5.

Finally, the program compares the estimated coefficients to their standard error, for $n = 0, \ldots, 128$ and $m = 1, \ldots, 128$, and prints out those coefficients which exceed a user specified level of significance. Standard errors for the $n=0$ and $n \geq 1$ cases are calculated using the expressions in Theorem 4 with sample moments replacing population moments.

4. Artificial data experiments

In this section the non-linear identification technique is illustrated through the use of artificial data. Although the technique was tested using a variety of models, we will restrict our attention here to the following model:

$$x(t)=\varepsilon(t)+31\varepsilon(t-9)\varepsilon(t-13)+31\varepsilon(t)\varepsilon(t-95), \tag{4.1}$$

with $\sigma_\varepsilon^2=0.0003$ and $\sigma_x^2=0.0005$. The $\varepsilon(t)$ variates were generated by the GGNML pseudo-random number routine in the *IMSL* library.

A realization of (4.1) with 1280 observations is plotted in fig. 1. For this realization, $\hat{\mu}_x=-0.0001$, $\hat{\sigma}_x^2=0.00044$ and $\hat{\sigma}_\varepsilon^2=0.00030$. Note from fig. 1

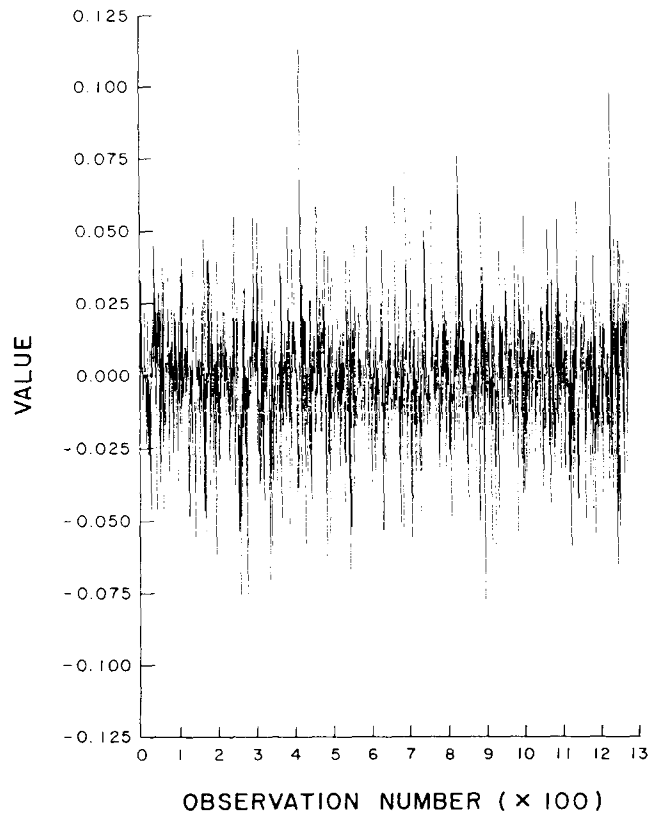

Fig. 1. Artificial data $m=4.95$.

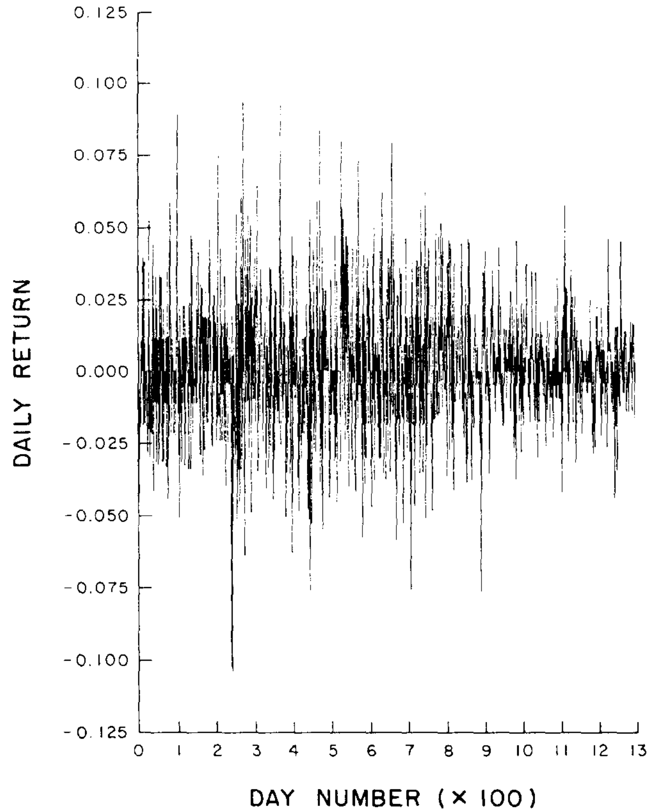

Fig. 2. American Standard 12/5/72–12/30/77.

that the model appears to be producing a large number of outliers vis-à-vis what we would expect from a Gaussian process; the sample kurtosis is 4.5, much larger than the reference value of 3 for a Gaussian process. In section 5 we will apply our identification technique to series of daily stock returns of length 1280, so it is worth mentioning that those series, and daily stock returns in general, exhibit leptokurtic behavior. For example, fig. 2 is a plot of the 1280 observations of American Standard, a series used in section 5. The sample kurtosis for this series is 4.74 with a sample variance of 0.0005, the same as the artificial data. For reference purposes, the reader's attention is directed to fig. 3, where we have plotted a purely random Gaussian process (it is linear) with a variance 0.0005. In comparing fig. 3 with figs. 1 and 2 we can easily see a difference between the behavior of the linear Gaussian process and the non-linear processes.

The coefficient estimates for the artificial data are $\hat{a}_4(9) = 0.236\text{E-}5$ and $\hat{a}_{95}(0) = 0.204\text{E-}5$. Because we know $\hat{\sigma}_\varepsilon^2 = 0.3\text{E-}3$, we can use Theorem 3 to

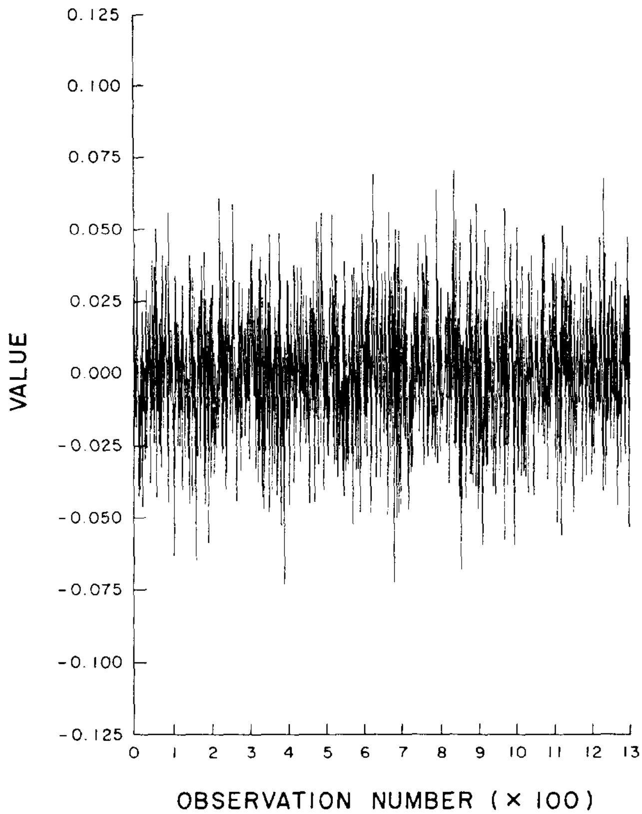

Fig. 3. Purely random Gaussian process.

calculate the implied $a_m(n)$'s: $\hat{a}_4(9) = 28$ and $\hat{a}_{95}(0) = 36$. Note that we use $N = 256$ in these calculations rather than the sample size of 1280, because the program uses blocks of size 256 to calculate the estimators. We can calculate a 95% confidence interval for the implied coefficients by adjusting the standard errors of the estimators in Theorem 4 by σ_ε^8. This gives 95% confidence intervals for the implied estimators of (22, 33) and (24, 48) for $a_4(9)$ and $a_{95}(0)$ respectively, which both cover the true coefficients.

The sample *SSPN* ratio for this particular realization of the process is 0.27, considerably greater than the true value of 0.15. What has happened is that sampling error has caused the program to identify 47 significant coefficients which exceed 2 standard errors in addition to the two 'correct' coefficients. In turn, this has inflated the numerator of the *SSPN* ratio, giving a value of 0.27. Using the result that the estimators are distributed normally, and assuming uncorrelatedness, we would expect to identify about 72 significant coefficients by chance at the 2 standard error level. As a consequence, we believe that the sample *SSPN* ratio is best used for making relative comparisons between sample series.

In our analysis of non-linear time series we have found the following method quite helpful in detecting the important differences, m. Compute for each $m = 1, 2, \ldots, 100$, the sum of squares of the 'significant' coefficients. Let $N(m)$ denote the number of significant $a_m(n)$ for difference m. Then,

$$L(m) = \sum_{j=1}^{N(m)} \hat{a}_m^2(n(j)), \tag{4.2}$$

where $n(j)$ is the lag n for the jth significant coefficient. We have found that a display of the $L(m)$ values, which we call the *Lagstrum*, provides a simple graphical method for detecting the differences of the quadratic model.

Fig. 4 is a plot of the Lagstrum for the artificial data generated by the model (4.1). This Lagstrum is typical of those we observed in other realizations of the

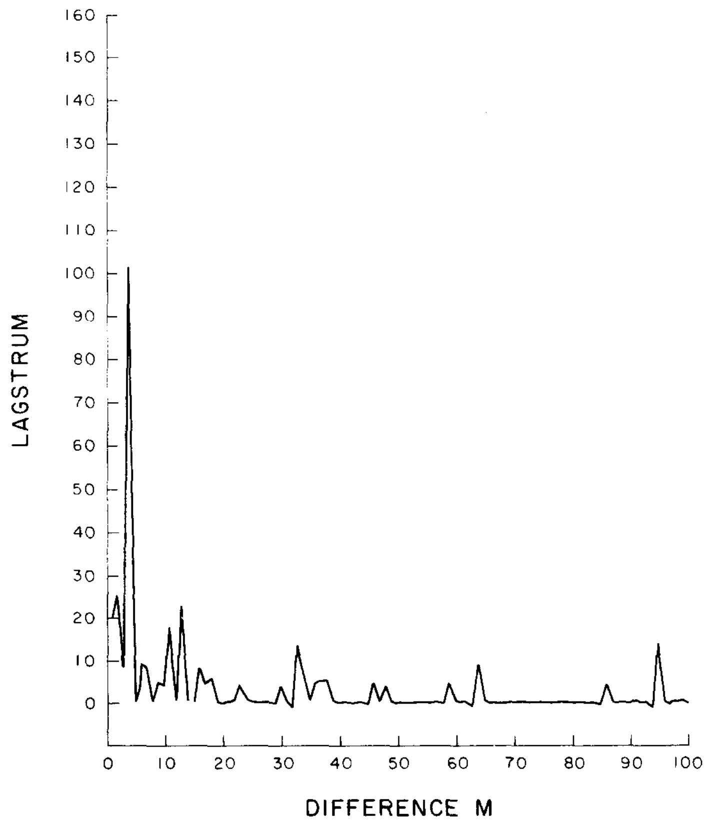

Fig. 4. Artificial data Lagstrum.

process. Note that the peaks at $m = 4$ and $m = 95$ are easily discerned. In calculating this Lagstrum we used a significance level of 2 standard errors. Although the coefficients in (4.1) are equal, the Lagstrum at $m = 95$ is smaller than the one at $m = 4$ because of the ramp bias in the estimator for $a_m(n)$.

5. Results with daily stock returns

Daily stock returns are a convenient source of real data which can be used to further illustrate our non-linear identification method. As mentioned in the Introduction, Hinich and Patterson (1985) tested fifteen randomly selected stock series for non-linearity with a test based on the sample bispectrum. Linearity was rejected for all fifteen series. The sample size for each series was 3881, covering the period July 1962 through December 1977. In this section we

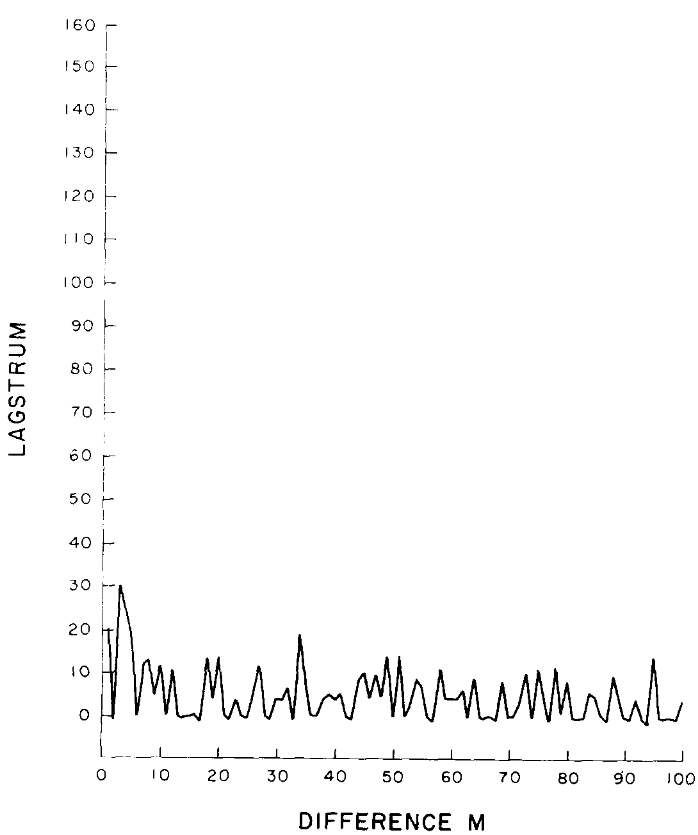

Fig. 5. American Standard.

report the results of non-linear identification for these same stocks, but with a sample size of 1280. The period spanned is December 5, 1972 through December 30, 1977, that is, approximately the last five years of the fifteen and one-half years covered in the earlier study. Because linearity has been rejected for this sample of stocks, our identification procedure becomes appropriate because, as argued above, the non-linear model generating returns can be approximated by a quadratic model.

Before we describe the results, some background information concerning daily stock returns is in order. In the opinion of most financial economists, daily stock returns closely follow a *martingale* or 'fair game'. This opinion is based on a large body of empirical work which suggests that the low levels of autocorrelation found in daily stock returns are not economically significant. To be more specific, daily stock returns tend to have low levels of marginally

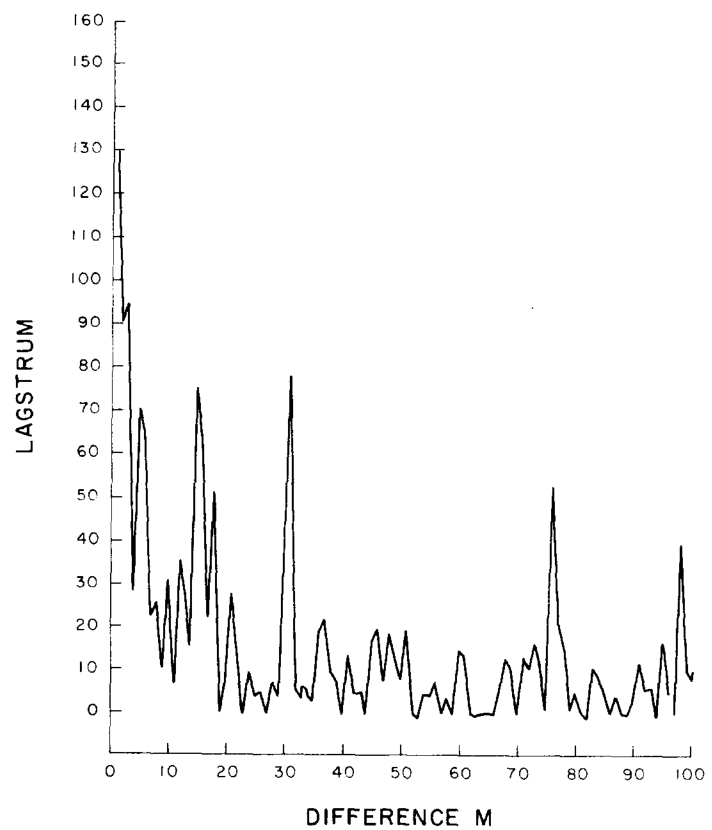

Fig. 6. Carolina Power & Light.

significant autocorrelation for lags varying between one and about ten days. As a consequence, the *martingale* model is not literally true, but writers in this area are quick to point out that the degree of linear dependence is not strong enough to form the basis of a profitable trading rule when transaction costs are included. Hence, the opinion that the stock market can be considered a fair game. Needless to say, we were quite surprised to find high levels of non-linearity in daily stock returns as most forms of non-linearity are inconsistent with the *martingale* hypothesis. It can be shown that the general quadratic model will generate a martingale when the linear term is simple and only $a_m(0)$ coefficients are allowed [see Hinich and Patterson (1984)]. However, as we will see shortly, the vast majority of identified coefficients for this sample of stocks have $n > 0$.

Table 1 lists the names of the fifteen common stocks as well as the sample variance, sample third cumulant, and sample signal to the signal plus noise

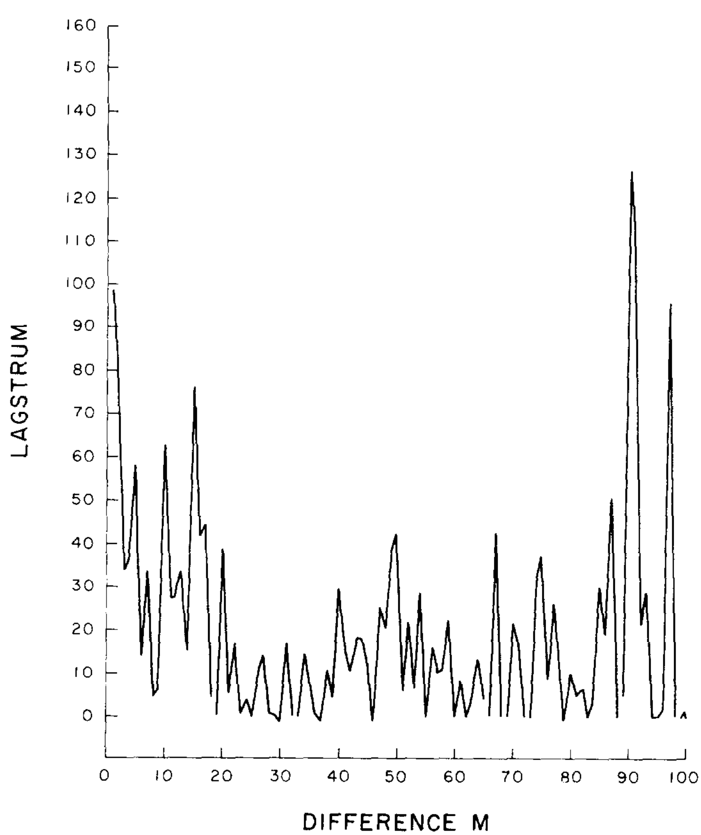

Fig. 7. General Public Utilities.

Table 1

Descriptive statistics.

Firm	Return variance	Third cumulant	Signal to signal plus noise
American Standard	5.0E-4	1.5E-6	0.486
Armstrong Rubber	4.4E-4	−4.9E-6	0.797
Brunswick Corp.	7.8E-4	7.2E-6	0.769
Carolina Power & Light	2.6E-4	3.7E-6	2.253
Chessie Systems, Inc.	5.7E-4	2.1E-6	0.580
Ex Cell 0	5.3E-4	7.5E-6	0.671
General Public Utilities	4.1E-4	5.6E-6	3.058
Holly Sugar	5.8E-4	3.8E-6	1.709
Kaiser Ind.	1.0E-3	1.6E-5	1.393
Liberty Fabrics	2.7E-3	1.4E-4	0.801
Pennzoil	5.2E-4	9.4E-7	1.566
Thiokol Chemical	6.5E-4	2.7E-6	0.794
Union Electric	1.6E-4	6.1E-7	2.473
United Park City Mines	3.0E-3	1.3E-4	0.448
Seagrave Corp	1.1E-3	2.2E-5	0.583

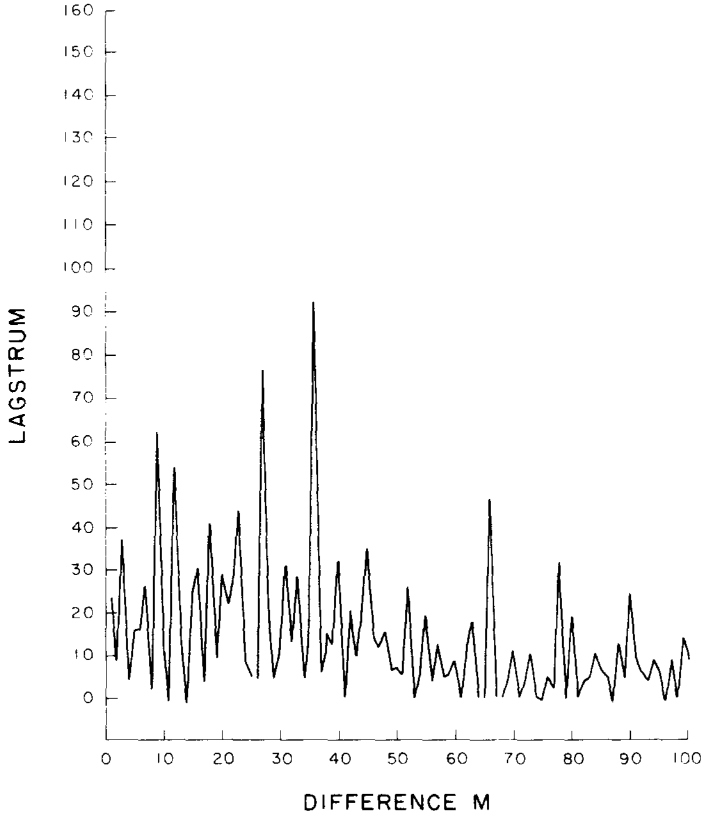

Fig. 8. Pennzoil.

ratio for each time series. Using the *SSPN* ratio for relative comparisons of quadratic non-linearity, we see that the series exhibiting the greatest degree of non-linearity is General Public Utilities.

The Lagstrums for five representative stock series are plotted in fig. 5 through 9. The Lagstrums were computed using those coefficients which were significant at 2 standard errors. In examining the Lagstrums, the first thing we noticed is that the stock Lagstrums tend to be much more cluttered than the artificial data Lagstrum. We were surprised to see that many of the Lagstrums have large peaks in the area of 75 to 100 days.

An example of the significant differences m and lags n is contained in table 2 for the General Public Utilities stock. In this table we display the number of significant coefficients within five- and six-day ranges for m and n. At the bottom of the table we show the total number of significant coefficients

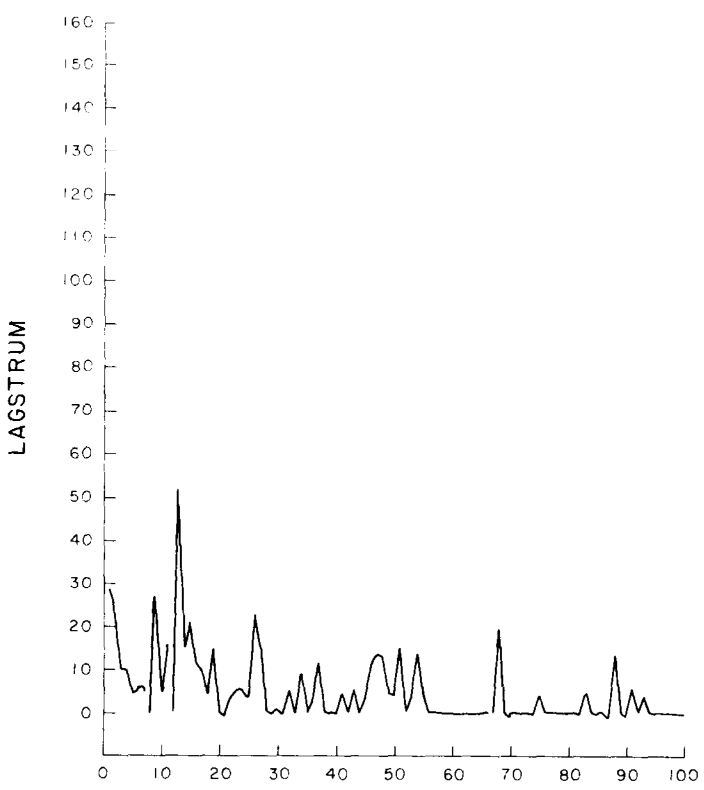

Fig. 9. United Park City Mines.

Table 2

General Public Utilities for differences m, 329 significant coefficients.

Lag n	1–5	6–10	11–15	16–20	21–25	26–30	31–35	36–40	41–45	46–50	51–55	56–60	61–65
0– 5	7	6	4	2		2		3	1	3		1	
6– 10	9	2	2	2		1	2		3	2	3	1	
11– 15	6		3	5	1		1	1	2	2			
16– 20	2	3	7	3	2	1		2	2	1		1	1
21– 25		3	2	1	1					2		1	2
26– 30	4		1	3					1	1			
31– 35	4		3	3			2				1	1	1
36– 40	3	1	1	1			1			1	2	1	1
41– 45	2	2	2							2	1	3	2
46– 50	4	1						1	2		2	3	
51– 55										4	1		
56– 60									1	1			
61– 65													
66– 70													
71– 75													
76– 80													
81– 85													
86– 90			1										
91– 95			1										
96–100													
101–105													
106–110													
111–115													
116–120													
121–126													

Table 2 (continued)

Lag n	66–70	71–75	76–80	81–85	86–90	91–95	96–100	101–105	106–110	111–115	116–120	121–126
0– 5		2			4	2	4	5	1	2	2	
6– 10		1		2	3	8	3	3	3		1	
11– 15		2	4		5	5	1	3	5			
16– 20	2	3	1	1	5	1	1	7	1	1		
21– 25	2	1	1	3	4	2	2	1	1			
26– 30	1	1	1		1	5	1	1	1			
31– 35	1		1	2	2			1	1			
36– 40	4					1		1				
41– 45	1				1	1						
46– 50						1						
51– 55						1						
56– 60												
61– 65												
66– 70												
71– 75												
76– 80												
81– 85												
86– 90												
91– 95												
96–100												
101–105												
106–110												
111–115												
116–120												
121–126												

identified at 2 standard errors. General Public Utilities has the greatest number of significant coefficients, 329, and American Standard and United Park City Mines the lowest, at 98. These results suggest that the non-linear models generating daily stock returns are rather complex. They also suggest that the process generating these returns is not a martingale.

6. Conclusions

When some of the coefficients in the quadratic term of the Volterra series expansion of the general non-linear model are non-zero, the bispectrum of the model will also be non-zero. This particular form of non-linearity can be detected through a statistical test based on the estimated bispectrum. We have shown that the non-linear model can then be approximated by a quadratic model with a simple linear term, and we have shown how the coefficients in the quadratic approximation can be identified. Given that the identified quadratic model can be inverted, it then becomes possible to estimate a parametric version of the model. At the estimation stage, any linear components will be swept into the estimated residuals, and at this point the methods for linear identification and estimation become applicable.

In conclusion, the methods presented here offer the data analyst a useful technique for identifying the significant coefficients in a non-linear quadratic model. The daily stock return series suggest the possibility of many interesting applications.

References

Brillinger, D.R., 1975, Time series: Data analysis and theory (Holt, Rinehart and Winston, New York).

Haggan, V., S.M. Heravi and M.B. Priestley, 1984, A study of the application of state-dependent models in non-linear time series analysis, Journal of Time Series Analysis 5, 69–102.

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

Hinich, M.J. and D.M. Patterson, 1984, A bispectrum based test of the martingale model, Unpublished working paper (Virginia Polytechnic Institute and State University, Blacksburg, VA).

Hinich, M.J. and D.M. Patterson, 1985, Evidence of nonlinearity in daily stock returns, Journal of Business and Economic Statistics 3, 69–77.

Maravall, A., 1983, An application of nonlinear time series forecasting, Journal of Business and Economic Statistics 1, 66–74.

Priestley, M.B., 1981, Spectral analysis and time series, Vol. 2 (Academic Press, New York).

Rosenblatt, M., 1983, Cumulants and cumulant spectra, in: D.R. Brillinger and P.R. Krishnaiah, eds., Handbook of statistics, Vol. 3 (North-Holland, Amsterdam) 369–382.

Subba Rao, T., 1983, The bispectral analysis of nonlinear time series with reference to bilinear time series models, in: D.R. Brillinger and P.R. Krishnaiah, eds., Handbook of statistics, Vol. 3 (North-Holland, Amsterdam) 293–319.

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

Wiener, N., 1958, Non-linear problems in random theory (M.I.T. Press, Cambridge, MA).