Monograph:

Billings, S.A. and Tao, Q.H. (1990) Model Validity Tests for Nonlinear Signal Processing Applications. Research Report. ACSE Research Report 401 . Department of Automatic Control and Systems Engineering

MODEL VALIDITY TESTS
FOR NONLINEAR SIGNAL PROCESSING APPLICATIONS

S. A. Billings and Q. H. Tao

Department of Control Engineering

University of Sheffield

Mappin Street

Sheffield S1 3JD

U. K.

Research Report No. 401

August 1990

Model Validity Tests for Nonlinear Signal Processing Applications

S. A. BILLINGS and Q. H. TAO

Department of Control Engineering
University of Sheffield
Sheffield S1 3JD, UK

ABSTRACT

Time series model validity tests based on general correlations are presented in this paper. It is shown that the tests $\Phi_{\xi\xi}(\tau)$, $\Phi_{\xi\xi^2}(\tau)$ and $\Phi_{\xi\xi\xi}(\tau_1,\tau_2)$ can only detect a subset of any possible unmodelled terms in the residuals ξ_t, whereas $\Phi_{\xi^{2'}\xi^{2'}}(\tau)$ detects all possible terms. These basic results are then extended to include functions of process or residual terms as entries in the correlation. Simulation studies are included to demonstrate the effectiveness of the tests when applied to estimated models of both simulated and real data sequences.

1. Introduction

Time series analysis and prediction is important in the study of a wide class of signal processing problems where applications range from the analysis of marketing data, brain wave patterns, signals from vibration testing of mechanical structures through to speech processing. In each case a model which is considered as the best possible approximation to the dynamic system is estimated based on observation data. Model validation is usually conducted as a final procedure after the model has been identified and the parameters estimated, with the aim of checking if the completed model adequately fits the data and is unbiased. Model validation is a critical procedure in signal processing and time series modelling. If the model validation methods are not properly designed there is a strong likelihood of accepting an inadequate and a biased model. This means that whilst the fitted model will predict well over the estimation set it is a poor descriptor of the process and will probably yield poor predictions over other data sequences.

One of the important methods of validation is to diagnostically examine the whiteness of the residual sequence, to detect if any terms in the residuals will cause bias in the parameter estimates. If the underlying process is linear so that it can be described by a linear model, there are celebrated results based on the second order covariance. Box and Pierce (1970) obtained the distribution of the residual autocorrelation function for ARMA (Auto-Regressive Moving Average) models and applied a chi-square statistic to check the lack of fit for these models, and Davies, Triggs and

Newbold (1977) and Ljung and Box (1978) later gave a modified version of this test. Box and Jenkens (1976) used model validation based on the residuals to suggest modifications to the fitted model and to iterate towards a final unbiased estimate.

If the process under study is nonlinear, however, the aforementioned diagnostic checking of the residual sequence using the second order covariance is unfortunately no longer sufficient to test the adequacy of the fitted model. In fact, model validation based on the second order covariance can only detect a subset of possible nonlinear terms which could be present in the residual sequence. As time series analysis has been extended to the nonlinear case various nonlinear models (see e.g. Priestly 1985, Chen and Billings 1989b) have been proposed, and there is an increasing need to develop methods for validating fitted nonlinear time series models. Several authors have addressed this problem for the nonlinear case. Granger and Andersen (1978) suggested the autocorrelation function of the squared residuals as a means of detecting bilinear terms in the residuals. Mcleod and Li (1983) introduced the chi-square statistic for the autocorrelation function of the squared residuals for ARMA models. Subba Rao and Gabr (1984) introduced a test based on a subset of the third order moments. Lawrance and Lewis (1985) showed that the second order covariance can not determine the independence of the residual sequence of the NEAR(2) (Nonlinear Auto-Regressive in Exponential variables (2nd order)) model, although this has a second order autoregressive correlation structure like the ARMA model, and suggested a third order moment of the residuals as a validity test. Recently Kumar (1986) also considered a test based on the third order moments and showed that this could distinguish bilinear time series from an independent sequence.

The present study considers model validation for a general class of nonlinear time series or signal processing models and is an attempt to provide clarification and a unification of previous work in the area. By taking the NARMA (Nonlinear Auto-Regressive Moving Average) model (Chen and Billings 1989b) as a basis, a systematic analysis of nonlinear validation is constructed to determine which tests can be expected to detect when the residual sequence is unpredictable from all past linear and nonlinear combinations of values irrespective of what form of model is estimated. The selection of the form of the test is studied in detail and it is shown that the particular tests $\Phi_{\xi\xi}(\tau)$, $\Phi_{\xi\xi^2}(\tau)$ and $\Phi_{\xi\xi\xi}(\tau_1,\tau_2)$ can detect certain types of terms out of all the possible unmodelled terms in the residuals, and $\Phi_{\xi^{2'}\xi^{2'}}(\tau)$ can detect all possible terms. These basic results are then extended to include other functions of process or residual terms as variables in the correlations. Simulation studies are included to demonstrate the effectiveness of the tests when applied to identify and validate some nonlinear time series models including the bilinear model, the nonlinear AR model, as well as an example of real time series.

2. Correlation based validation

Under some mild conditions a discrete time nonlinear stochastic control system can be described by a unified representation called the NARMAX (Nonlinear AutoRegressive Moving Average with eXogenous inputs) model (Leontaritis and Billings 1985, Chen and Billings 1989a). Selecting the input terms in the NARMAX model formulation to be zero gives the following NARMA model (Chen and Billings 1989b) description:

$$x_t = F(x_{t-1}, x_{t-2}, \cdots, x_{t-n_x}; \varepsilon_{t-1}, \varepsilon_{t-2}, \cdots, \varepsilon_{t-n_\varepsilon}) + \varepsilon_t \tag{2.1}$$

where $\{x_t\}$ denotes a measured time series or signal, $\{\varepsilon_t\}$ is an unobservable zero mean i.i.d. sequence and $F(.)$ is some nonlinear function. A typical choice for $F(.)$ in eqn (2.1) is a polynomial expansion but this is only one choice and $F(.)$ can take on many other more complex forms. This model is chosen because it provides a unified representation for a broad class of nonlinear stochastic processes and is about as far as one can go in terms of specifying a general finite dimensional nonlinear relationship (Chen and Billings 1989b). Various existing time series models, such as the bilinear model (Granger and Andersen 1981), state-dependent model (Priestly 1980) and exponential autoregressive model (Ozaki 1985), etc. are particular parameterizations of the NARMA model.

The conditions for $\{x_t\}$ to be stationary and invertible are investigated in Chen and Billings (1989b) and will not be repeated here. But the following assumptions will be made:

Assumption 1

It is assumed throughout the analysis that $\{x_t\}$ satisfies the following conditions:

(i) x_t is strictly stationary and ergodic.

(ii) x_t is invertible.

Remark:

Strictly stationarity is assumed because higher order moments are essential in the nonlinear case and second order stationarity is obviously inadequate. However in practice this requirement may be weakened by assuming stationarity to a certain order according to the particular system under study.

Let model (2.1) be parametrized with a parameter vector θ of dimension n_θ

$$x_t = F(x_{t-1}, x_{t-2}, \cdots, x_{t-n_x}; \xi_{t-1}(\theta), \xi_{t-2}(\theta), \cdots, \xi_{t-n_\varepsilon}(\theta);\theta) + \xi_t(\theta) \tag{2.2}$$

where $\theta \subset D_M$ and D_M is a subset of n_θ-dimensional Euclidean space, and

$\{\xi_t(\theta)\}$, $t = 1,2,...,$ are the model residuals at θ. Identification consists of selecting a model (a parameter vector θ) within D_M which best describes the recorded data.

Assume the following null hypothesis:

H_0: A true parameter vector θ_0 exists which reduces the residuals $\xi_t(\theta)|_{\theta=\theta_0}$ to an i.i.d. sequence ε_t.

Without loss of generality, it is assumed that the parameters θ are estimated using the prediction error method (Ljung 1987) by minimising a loss function,

$$\left.\begin{aligned} \hat{\theta} &= arg\min_{\theta}\sum_{t=1}^{N}\xi_t^2(\theta) \\ \xi_t(\theta) &= x_t - \hat{x}_{t|t-1}(\theta) \end{aligned}\right\} \tag{2.3}$$

to yield the estimates $\hat{\theta}$, where $\hat{x}_{t|t-1}(\theta)$ is the optimal (in the mean square sense) one step ahead prediction of x_t.

Let the vector $y^t(\hat{\theta})$ contain all the process terms and all the residuals at $\hat{\theta}$ up to the time t

$$y^t(\hat{\theta}) = \begin{bmatrix} x^{t-1} \\ \xi^{t-1}(\hat{\theta}) \end{bmatrix} \tag{2.4}$$

where

$$\left.\begin{aligned} x^t &= [x_t \; x_{t-1} \; \cdots \; x_1]^T \\ \xi^t(\hat{\theta}) &= [\xi_t(\hat{\theta}) \; \xi_{t-1}(\hat{\theta}) \; \cdots \; \xi_1(\hat{\theta})]^T \end{aligned}\right\} \tag{2.5}$$

Now introduce a correlation based test

$$r = \frac{1}{\sqrt{N}}\sum_{t=1}^{N} Z_t(\hat{\theta}) \; \xi_t^f(\hat{\theta}) \tag{2.6}$$

where $Z_t(\hat{\theta})$ is a vector of dimension s and depends only on the vector y^t,

$$Z_t(\hat{\theta}) = Z(y^{t-1}(\hat{\theta})) \tag{2.7}$$

and $\xi_t^f(\hat{\theta})$ is some smooth function of $\xi_t(\hat{\theta})$ with the mean deleted

$$\xi_t^f(\hat{\theta}) = f(\xi_t(\hat{\theta})) - E[f(\xi_t(\hat{\theta}))] \tag{2.8}$$

in which typically

$$f(\zeta) = \zeta \tag{2.9}$$

or

$$f(\zeta) = \zeta^2 \tag{2.10}$$

according to the particular case. If $f(\zeta) = \zeta$, r reduces to the test introduced by Bohlin (1978), Leontaritis and Billings(1987) and Söderström and Stoica (1988). Define

$$\varepsilon_t^f = \xi_t^f(\theta)\,|_{\theta=\theta_0} = f(\varepsilon_t) - Ef(\varepsilon_t) \tag{2.11}$$

When $f(\zeta) = \zeta$, Söderström and Stoica (1988) have shown that under H_0 r converges almost surely to the normal distribution with mean zero and covariance matrix

$$\Psi_1 = \lambda_1^2(\Phi_{zz} - \Phi_{z\psi}\Phi_{\psi\psi}^{-1}\Phi_{\psi z}) \tag{2.12}$$

where

$$\lambda_1^2 = E(\varepsilon_t^2) \tag{2.13}$$

$$\Phi_{zz} = E[Z_t(\theta_0)Z_t^T(\theta_0)] \tag{2.14}$$

$$\Phi_{\psi\psi} = E[\psi_t\psi_t^T] \tag{2.15}$$

$$\psi_t = -(\frac{\partial\xi_t(\theta)}{\partial\theta}|_{\theta=\theta_0})^T \tag{2.16}$$

$$\Phi_{z\psi} = E[Z_t\psi_t^T] = \Phi_{\psi z}^T \tag{2.17}$$

It can also be shown that in the general case when $f(.)$ is some smooth function, for the parametrized NARMA model (2.2), under H_0 and assumptions (i), (ii), r converges almost surely to the normal distribution with zero mean and covariance matrix similar to (2.12). When typically $f(\zeta) = \zeta^2$, the covariance matrix is equal to

$$\Psi_2 = \lambda_2^2\Phi_{zz} \tag{2.18}$$

where

$$\lambda_2^2 = E[(\varepsilon_t^f)^2] \tag{2.19}$$

Thus correlation based validity tests can be constructed using the above asymptotic distribution. Define

$$\mu = \begin{cases} r^T\,[\hat\lambda_1^2(\hat\Phi_{zz} - \hat\Phi_{z\psi}\hat\Phi_{\psi\psi}^{-1}\hat\Phi_{\psi z})]^{-1}\,r, & \text{if } f(\zeta) = \zeta \\ r^T\,[\hat\lambda_2^2\hat\Phi_{zz}]^{-1}\,r, & \text{if } f(\zeta) = \zeta^2 \end{cases} \tag{2.20}$$

where

$$\hat\lambda_1^2 = \frac{1}{N}\sum_{t=1}^{N}\varepsilon_t^2 \tag{2.21}$$

$$\hat\lambda_2^2 = \frac{1}{N}\sum_{t=1}^{N}(\varepsilon_t^f)^2 \tag{2.22}$$

$$\hat\Phi_{zz} = \frac{1}{N}\sum_{t=1}^{N}Z_t(\hat\theta)Z_t^T(\hat\theta) \tag{2.23}$$

$$\hat{\Phi}_{\psi\psi} = \frac{1}{N}\sum_{t=1}^{N}\psi_t(\hat{\theta})\psi_t^T(\hat{\theta}) \tag{2.24}$$

$$\hat{\Phi}_{z\psi} = \frac{1}{N}\sum_{t=1}^{N}Z_t(\hat{\theta})\psi_t^T(\hat{\theta}) = \hat{\Phi}_{\psi z}^T \tag{2.25}$$

So that μ is chi-square distributed with s degrees of freedom under H_0, that is

$$\mu \overset{a.s.}{=} \chi^2(m) \tag{2.26}$$

If the value of μ is within the acceptance region for a given level of significance α, that is

$$\mu < \kappa_\alpha(m) \tag{2.27}$$

where $\kappa_\alpha(m)$ is the critical value of the chi-squared distribution, the null hypothesis holds and the fitted model is accepted with a risk of α. Otherwise the model is rejected. Note that the new term $\Phi_{z\psi}\Phi_{\psi\psi}^{-1}\Phi_{\psi z}$ is included in the covariance matrix Ψ_1 in eqn.(2.12) when $f(\zeta) = \zeta$ and the corresponding chi-square test μ is an exact test compared with the approximate test which arises if this term is omitted, in the sense that the former has exactly the first type of risk α (Söderström and Stoica 1988). Notice that the derivatives of the residuals $\psi(t,\hat{\theta})$ and $\hat{\Phi}_{\psi\psi}^{-1}$ are computed as part of the prediction error parameter estimation algorithm, and the inversion of the matrix $\hat{\Phi}_{\psi\psi}$ can be obtained, for example, by the Choleski factorisation method.

Another way of implementing the correlation based test r as defined in eqn.(2.6) which is computationally simpler than the chi-square tests is to directly employ the asymptotically normal distribution property of r. With a significance level of 0.05, every element of the normalised r should fall within a confidence band of $\pm 1.96/N^{1/2}$. The tests using this method which are usually referred to as correlation tests are however severe and occasionally a few correlations can exceed the confidents limits even when the model is valid (Leontaritis and Billings 1987). The chi-square method is therefore preferred.

3. Formulating the tests

3.1 Preliminaries

The problem remaining from the last section is how to choose the vector $Z_t(\hat{\theta})$ and how to decide the form of $f(.)$ in the general correlation based test (2.6). It is evident that the traditional linear method corresponds to a choice of the vector $Z_t(\hat{\theta})$ and $f(\xi_t(\hat{\theta}))$ given by

$$Z_t(\hat{\theta}) = [\xi_{t-1}(\hat{\theta})\ \xi_{t-2}(\hat{\theta})\ \cdots\ \xi_{t-t_d}(\hat{\theta})]^T \tag{3.1}$$

where t_d is the maximum delay in the correlation, and

$$f(\xi_t(\hat{\theta})) = \xi_t(\hat{\theta}) \tag{3.2}$$

respectively. In the nonlinear case correlation tests or the corresponding chi-square tests based on the second order covariance fail to detect all possible missing terms in the residuals because they can not distinguish between some nonlinear terms and an independent sequence, and is therefore not sufficient for nonlinear time series validation. Generally the elements of the vector of $Z_t(\hat{\theta})$ will need to be nonlinear in the elements of the vector y^t, and the function $f(.)$ may take linear or nonlinear forms. Various nonlinear forms can be taken for $Z_t(\hat{\theta})$ and $f(.)$, but tests based on combinations of some particular choices, just like the traditional linear method, may have insignificant power. This occurs because every element of such constructed r tend to be zero when ideally some of them should be nonzero and therefore, the correlations tests or the chi-square tests based on r fail to give the correct information. The problem therefore is to determine a set of tests which are sufficient to detect all possible linear and nonlinear terms of whatever form in the residuals. When the tests are satisfied the residuals should be unpredictable from all linear and nonlinear combinations of past residuals and process terms. There are probably many possible tests, or combinations of tests which would satisfy these objectives and therefore it is important to search for the smallest set of tests which are easy to compute and interpret.

It can be argued that at the model validity stage little will be known about the residual sequence. The residuals $\{\xi_t(\hat{\theta})\}$ (written as ξ_t hereafter) may contain some unmodelled terms. The number and form of these terms, the lags or the type of nonlinearity and the form of probability distributions will all be unknown and consequently the model validity tests that are derived must work for the worst possible combination of properties.

The following general case assumptions will be made throughout the following analysis:

Assumption 2

(i) $\hat{x}_{t|t-1} = 0$, that is, the residual sequence contains any or all possible terms in the NARMA model (2.1)

$$\xi_t = x_t = F(x_{t-1}, x_{t-2}, \cdots, x_{t-n_x}; \varepsilon_{t-1}, \varepsilon_{t-2}, \cdots, \varepsilon_{t-n_\varepsilon}) + \varepsilon_t \tag{3.3}$$

(ii) if the sequence $\{\varepsilon_t\}$ is distributed symmetrically all even moments exist and odd moments are zero.

(iii) if the sequence $\{\varepsilon_t\}$ is distributed asymmetrically all even moments and odd moments exist but it will be assumed that $E\varepsilon_t = 0$.

To derive the tests it is necessary to select a form of $F(.)$ in eqn (3.3). In the present study a Volterra series parameterisation is adopted such that (3.3) can be expressed as

$$\xi_t = \sum_{u=1}^{\infty} g_u \varepsilon_{t-u} + \sum_{u=1}^{\infty}\sum_{v=1}^{\infty} g_{uv}\varepsilon_{t-u}\varepsilon_{t-v} + \cdots + \varepsilon_t \tag{3.4}$$

where g_u , g_{uv} , ... , are the Volterra kernels. Notice that this choice does not restrict the application of the results to just the Volterra model. The results will hold for all analytic nonlinear systems whatever form of model is used to parameterise the system. The choice of the Volterra model just simplifies the derivation of the results.

Using the operational notation developed by Brilliant (1958) and George (1959), with obvious extensions to describe the time series case, eqn (3.4) can be represented as

$$\xi_t = \sum_{n=1}^{\infty} G_n[\varepsilon_t] + \varepsilon_t = \sum_{n=1}^{\infty} G_n(\varepsilon_t^n) + \varepsilon_t \tag{3.5}$$

where the square brackets indicate the G operates on ε_t, while the parentheses depict the actual relationship.

Any term in eqn (3.4) $g_{uv} \ldots \varepsilon_{t-u}\varepsilon_{t-v} \cdots$ will be defined as an even term, if the time-lagged monomial in ε_t can be expressed as

$$\eta_{2n}^e = \varepsilon_{t-t_1}^{i_1}\varepsilon_{t-t_2}^{i_2} \cdots \varepsilon_{t-t_m}^{i_m} \tag{3.6}$$

in which the time lags $t_1 \neq t_2 \neq \cdots \neq t_m > 0$, the powers $i_1, i_2, \cdots, i_m$ are all even, and $\sum i = 2n$, n=1, ..., ∞. For example, a typical even term η_{2n}^e might be $K\varepsilon_{t-1}^2\varepsilon_{t-3}^4\varepsilon_{t-7}^2$, where K is the corresponding Volterra kernel.

A term will be referred to as odd, if the time-lagged monomial in ε_t can be expressed concisely as

$$\eta_n^o = \varepsilon_{t-t_1}^{i_1}\varepsilon_{t-t_2}^{i_2} \cdots \varepsilon_{t-t_m}^{i_m}\eta_{2n'}^e \tag{3.7}$$

where time lags $t_1 \neq t_2 \neq \cdots \neq t_m > 0$, the powers $i_1, i_2, \cdots, i_m$, are all odd, $1 \le m \le n$, $\eta_{2n'}^e$ is a possible even component in the odd term, $0 \le 2n' < n$, and $\sum i + 2n' = n$, where n=1, $\cdots$, ∞. For instance, a typical odd term η_n^o might be $K_1\varepsilon_{t-2}\varepsilon_{t-3}^3$ or $K_2\varepsilon_{t-1}^3\varepsilon_{t-3}^3\varepsilon_{t-5}^4$, where K_1,K_2 are Volterra kernels.

It should be noted that the definitions of even and odd are based on the powers of the ε_t monomials, not on whether the Volterra kernels are odd or even.

By using the even and odd definitions, Eqn (3.5) can be further denoted as

$$\xi_t = \xi_t^e + \xi_t^o + \varepsilon_t = \sum_{n=1}^{\infty} G_{2n}^e[\varepsilon_t] + \sum_{n=1}^{\infty} G_n^o[\varepsilon_t] + \varepsilon_t.$$

$$= \sum_{n=1}^{\infty} G_{2n}^{e}(\varepsilon_t^{2n}) + \sum_{n=1}^{\infty} G_n^{o}(\varepsilon_t^{n}) + \varepsilon_t \tag{3.8}$$

where $\xi_t^e = \sum_{n=1}^{\infty} G_{2n}^{e}[\varepsilon_t]$ contains all even terms defined before

$$G_{2n}^{e}[\varepsilon_t] = \sum_{t_1=1}^{\infty}\sum_{t_2=1}^{\infty} \cdots \sum_{t_m=1}^{\infty} g_{t_1^{i_1} t_2^{i_2} \cdots t_m^{i_m}} \varepsilon_{t-t_1}^{i_1} \varepsilon_{t-t_2}^{i_2} \cdots \varepsilon_{t-t_m}^{i_m} \tag{3.9}$$

where $g_{t_1^{i_1} t_2^{i_2} \cdots t_m^{i_m}}$ are the corresponding Volterra kernels. All odd terms are contained in $\xi_t^o = \sum_{n=1}^{\infty} G_n^{o}[\varepsilon_t]$ where

$$G_n^{o}[\varepsilon_t] = \sum_{t_1=1}^{\infty}\sum_{t_2=1}^{\infty} \cdots \sum_{t_m=1}^{\infty} g_{t_1^{i_1} t_2^{i_2} \cdots t_m^{i_m} t_1'^{i_1'} t_2'^{i_2'} \cdots t_m^{i_m}} \varepsilon_{t-t_1}^{i_1} \varepsilon_{t-t_2}^{i_2} \cdots \varepsilon_{t-t_m}^{i_m} \eta_{2n'}^{e} \tag{3.10}$$

The second to forth correlations of the residuals expanded as a Volterra series can now be evaluated with the objectives of finding the smallest subset of tests which detect all possible omitted terms from the model. Notice that the following analysis of correlations of the residuals applies to both correlation tests and chi-square tests which are based on these correlations, because a correlation test and its corresponding chi-square test are based on the same correlation structure. As the chi-square method has statistical advantages compared with correlations, only the chi-square method is employed in this study. All tests in the following context, therefore, should be referred to as the chi-square tests which are constructed using the corresponding correlations.

3.2 Autocorrelation of the residuals $\Phi_{\xi\xi}(\tau)$

Define

$$\Phi_{\xi\xi}(\tau) = E[(\xi_t - E(\xi_t))(\xi_{t-\tau} - E(\xi_t))] \tag{3.11}$$

If H_0 holds, ξ_t will be an independent sequence $\{\varepsilon_t\}$, $\Phi_{\xi\xi}(\tau) = c\delta(\tau)$, where c is a constant. Ideally $\Phi_{\xi\xi}(\tau) \neq 0$ for some $\tau \neq 0$, if any term other than ε_t remains unmodelled in the residual sequence. Substituting the residual sequence ξ_t expressed as odd and even terms from (3.8) into (3.11) gives

$$\begin{aligned}\Phi_{\xi\xi}(\tau) &= E[(\xi_t - E(\xi_t))(\xi_{t-\tau} - E(\xi_t))] \\ &= E[(\xi_t^o - E(\xi_t^o) + \xi_t^e - E(\xi_t^e) + \varepsilon_t)(\xi_{t-\tau}^o - E(\xi_t^o) + \xi_{t-\tau}^e - E(\xi_t^e) + \varepsilon_{t-\tau})] \\ &= \Phi_{\xi^e\xi^e}(\tau) + \Phi_{\xi^o\xi^o}(\tau) + \Phi_{\xi^e\xi^o}(\tau) + \Phi_{\xi^o\xi^e}(\tau) \\ &+ \Phi_{\xi^e\xi^\varepsilon}(\tau) + \Phi_{\xi^\varepsilon\xi^e}(\tau) + \Phi_{\xi^o\xi^\varepsilon}(\tau) + \Phi_{\xi^\varepsilon\xi^o}(\tau) + \Phi_{\xi^\varepsilon\xi^\varepsilon}(\tau)\end{aligned} \tag{3.12}$$

where the polynomial correlation $\Phi_{\xi^e\xi^e}(\tau)$ is defined as

$$\Phi_{\xi^e\xi^e}(\tau) = E[(\xi_t^e - E(\xi_t^e))(\xi_{t-\tau}^e - E(\xi_t^e))]$$

$$= E\{(\sum_{n=1}^{\infty} G_{2n}^e[\varepsilon_t] - E\{\sum_{n=1}^{\infty} G_{2n}^e[\varepsilon_t]\})(\sum_{n=1}^{\infty} G_{2n}^e[\varepsilon_{t-\tau}] - E\{\sum_{n=1}^{\infty} G_{2n}^e[\varepsilon_t]\})\}$$

$$= E\{(G_2^e[\varepsilon_t] - E(G_2^e[\varepsilon_t]))(G_2[\varepsilon_{t-\tau}] - E(G_2[\varepsilon_t]))$$

$$+ (G_4^e[\varepsilon_t] - E(G_4^e[\varepsilon_t]))(G_4^e[\varepsilon_{t-\tau}] - E(G_4^e[\varepsilon_t])) + \cdots \} \qquad (3.13)$$

and $\Phi_{\xi^o\xi^\varepsilon}(\tau)$ is a polynomial crosscorrelation between odd terms and the prediction error ε_t

$$\Phi_{\xi^o\xi^\varepsilon}(\tau) = E[(\xi_t^o - E(\xi_t^o))(\varepsilon_{t-\tau})]$$

$$= E\{(\sum_{n=1}^{\infty} G_n^o[\varepsilon_t] - E\{(\sum_{n=1}^{\infty} G_n^o[\varepsilon_t]\})(\varepsilon_{t-\tau})\}$$

$$= E\{(G_1[\varepsilon_t] - E\{G_1[\varepsilon_t]\})(\varepsilon_{t-\tau}) + (G_3[\varepsilon_t] - E\{(G_3[\varepsilon_t])\}(\varepsilon_{t-\tau})) + \cdots \} \qquad (3.14)$$

The other terms are defined in a similar manner.

The following sections consider each group of polynomial correlations which are induced by either even or odd terms only, or by both even and odd terms together, plus correlations caused by the prediction error term ε_t which are always present in the residuals because all identification procedures aim to reduce the residuals to an independent sequence.

If the probability density function of the residual sequence is asymmetrical such that all even and odd moments except $E(\varepsilon_t)$ exist. Then it can be shown that

(i) $\Phi_{\xi^e\xi^e}(\tau) + \Phi_{\xi^e\xi^\varepsilon}(\tau) + \Phi_{\xi^\varepsilon\xi^e}(\tau) \neq 0$, when $\tau \neq 0$, for any possible single even term. That is, any single even term can be detected by $\Phi_{\xi\xi}(\tau)$.

(ii) $\Phi_{\xi^e\xi^o}(\tau)$, $\Phi_{\xi^o\xi^e}(\tau)$ are crosscorrelations between even and odd terms which are neglected in the analysis, as any single even term can be detected.

(iii) $\Phi_{\xi^o\xi^o}(\tau) + \Phi_{\xi^o\xi^\varepsilon}(\tau) + \Phi_{\xi^\varepsilon\xi^o}(\tau) \neq 0$, where $\tau \neq 0$, for any possible single odd terms, except the type of single odd terms of the form

$$K\varepsilon_{t-t_1}\varepsilon_{t-t_2} \cdots \varepsilon_{t-t_m}\varepsilon_{t-t_{m-1}}^{i_{m-1}} \cdots \varepsilon_{t-t_p}^{i_p}\eta_q^e \qquad (3.15)$$

which will not be detected, where K is the corresponding Volterra kernel, $p>m\geq 2$, $i_{m+1},\ldots,i_p = 0$ or odd(>1), and η_q^e is a possible even component in the odd term, $q \geq 0$.

(iv) $\Phi_{\xi\xi}(\tau)$ can detect a group of two odd terms defined by (3.11) in the residuals

$$\left.\begin{array}{l} K'\varepsilon_{t-t_1'}\varepsilon_{t-t_2'} \cdots \varepsilon_{t-t_m'}\varepsilon_{t-t_{m+1}'}^{i'_{m+1}} \cdots \varepsilon_{t-t_p'}^{i'_p}\eta_{q'}^e \\ K''\varepsilon_{t-t_1''}\varepsilon_{t-t_2''} \cdots \varepsilon_{t-t_m''}\varepsilon_{t-t_{m+1}''}^{i''_{m+1}} \cdots \varepsilon_{t-t_p''}^{i''_p}\eta_{q''}^e \end{array}\right\} \qquad (3.16)$$

where K' and K'' are the corresponding Volterra kernels. $p>m>=2$, and $i'_{m+1}, \cdots i'_p, i''_{m+1}, \cdots i''_p = 0$ or odd(>1). $\eta^e_{q'}, \eta^e_{q''}$ are the possible even components in the odd terms, $q', q'' \geq 0$, and

$$t'_1 - t'_2 = t''_1 - t''_2 = \cdots = t'_{p-1} - t'_p = t''_{m-1} - t''_m \tag{3.17}$$

For example,

$$\xi_t = \beta\varepsilon_{t-1}\varepsilon_{t-2}\varepsilon^2_{t-3} + \cdots + \varepsilon_t \tag{3.18}$$

where the single odd term $\beta\varepsilon_{t-1}\varepsilon_{t-2}\varepsilon^2_{t-3}$ which is of the type described by (3.15) can not be detected by $\Phi_{\xi\xi}(\tau)$. But $\Phi_{\xi\xi}(\tau)$ works in the case of

$$\xi_t = \beta_1\varepsilon_{t-1}\varepsilon_{t-2} + \beta_2\varepsilon_{t-3}\varepsilon_{t-4} \cdots + \varepsilon_t \tag{3.19}$$

where the two terms $\beta_1\varepsilon_{t-1}\varepsilon_{t-2}$, $\beta_2\varepsilon_{t-3}\varepsilon_{t-4}$ are as described by (3.16) and can be detected together by $\Phi_{\xi\xi}(\tau)$. But two such terms can not be found in superdiagonal bilinear type models as discussed in Granger and Andersen (1978), and Kumar (1986). For example,

$$\xi_t = \beta\varepsilon_{t-1}\xi_{t-2} + \varepsilon_t \tag{3.20}$$

which can be represented by the Volterra sequence

$$\xi_t = \varepsilon_t + \beta\varepsilon_{t-1}\varepsilon_{t-2} + \beta^2\varepsilon_{t-1}\varepsilon_{t-2}\varepsilon_{t-3} + \cdots \tag{3.21}$$

where each single term in (3.21) belongs to terms of the type in (3.15) which can not be detected by $\Phi_{\xi\xi}(\tau)$. Therefore as a whole $\Phi_{\xi\xi}(\tau)$ fails to distinguish the superdiagonal bilinear term from an independent sequence. This result can be generalised to conclude that $\Phi_{\xi\xi}(\tau)$ can not detect any single odd term in which there is one or more superdiagonal bilinear components $\varepsilon_{t-t_1}\xi_{t-t_2}$, where $t_1 < t_2$. For example, $K\varepsilon_{t-2}\xi_{t-5}\varepsilon^2_{t-7}$, where there is a superdiagonal bilinear component $\varepsilon_{t-2}\xi_{t-5}$ in the term.

If the probability density function of the residual sequence is symmetrical all moments of odd terms are zero. This puts a more severe condition on the terms which can be distinguished by the test. In a similar way it can be shown that

(i) $\Phi_{\xi^e\xi^e}(\tau) + \Phi_{\xi^e\xi^o}(\tau) + \Phi_{\xi^o\xi^e}(\tau) \neq 0$, when $\tau \neq 0$, for all possible single and groups of even terms except the following cases:

(a) when there is only one isolated even term like

$$K\varepsilon^{i_1}_{t-t_1} \tag{3.22}$$

where K is the Volterra kernel, and i_1 is even.

(b) more than one isolated term as defined by (3.16) but each term with the same time lag

$$\left.\begin{array}{l} K_1\varepsilon_{t-t_1}^{i_1} \\ K_2\varepsilon_{t-t_1}^{i'_1} \\ K_3\varepsilon_{t-t_1}^{i''_1} \\ \cdots \end{array}\right\} \tag{3.23}$$

where K_1, K_2, K_3 are the Volterra kernels, and i_1, i_2, i_3 are all even. In the above two cases, the test $\Phi_{\xi\xi}(\tau)$ does not distinguish unmodelled terms from an independent sequence.

(ii) $\Phi_{\xi^o\xi^o}(\tau) + \Phi_{\xi^o\xi^e}(\tau) + \Phi_{\xi^e\xi^o}(\tau) \neq 0$, when $\tau \neq 0$, only in the following cases:
(a) when there exists a single odd term in the residuals

$$K\varepsilon_{t-t_1}^{i_1}\eta_q^e \tag{3.24}$$

where K is the corresponding Volterra kernel, and i_1 odd. η_q^e is the possible even component in the odd term, $q \geq 0$; (b) when there exists a group of two odd terms in the residuals

$$\left.\begin{array}{l} K'\varepsilon_{t-t'_1}^{i_1}\varepsilon_{t-t'_2}^{i_2}\cdots\varepsilon_{t-t'_p}^{i_p}\eta_{q'}^e \\ K''\varepsilon_{t-t''_1}^{i'_1}\varepsilon_{t-t''_2}^{i'_2}\cdots\varepsilon_{t-t''_p}^{i'_p}\eta_{q''}^e \end{array}\right\} \tag{3.25}$$

where K' and K'' are the corresponding Volterra kernels. $\eta_{q'}^e, \eta_{q''}^e$ are the possible even components in the odd terms, $q', q'' \geq 0$. $i_1, i_2, \cdots, i_p, i'_1, i'_2, \cdots, i'_p$, are all odd, and

$$t'_1 - t'_2 = t''_1 - t''_2 = \cdots = t'_{p-1} - t'_p = t''_{m-1} - t''_m \tag{3.26}$$

(iii) $\Phi_{\xi^e\xi^o}(\tau), \Phi_{\xi^o\xi^e}(\tau) = 0$, when $\tau \neq 0$. So that these components make no contribution to the test.

For example,

$$\xi_t = \beta\varepsilon_{t-3}^3\varepsilon_{t-1}^2 + \cdots + \varepsilon_t \tag{3.27}$$

where the single odd term $\beta\varepsilon_{t-3}^3\varepsilon_{t-1}^2$ can be detected by $\Phi_{\xi\xi}(\tau)$. It is obvious that any single linear term is a special case of (3.24) where the even component in the odd term does not exist. Linear terms are therefore detected by computing the autocorrelations of the residuals. However $\Phi_{\xi\xi}(\tau)$ can not distinguish between an independent sequence and residuals of the form

$$\xi_t = \xi_{t-p}^r\varepsilon_{t-q}^s + \varepsilon_t \tag{3.28}$$

for all r, s odd, and $p>q$. Terms like $\xi_{t-p}^r\varepsilon_{t-q}^s$ can be viewed as an extension to the superdiagonal bilinear type. Following the same argument as in the asymmetrical residuals case, it can be shown that $\Phi_{\xi\xi}(\tau)$ can not detect any single odd term which contains one or more components of the form $\xi_{t-p}^{r}\varepsilon_{t-q}^{s}$, where $p>q$, r, s odd.

The above results can be summarized by

Lemma 3.1

Consider the residual sequence defined by eqn (3.4), if ε_t is distributed asymmetrically, $\Phi_{\xi\xi}(\tau)$ detects any possible single even terms, any possible single odd terms except the type of single odd terms defined by (3.15), as well as a group of two terms defined by (3.16). If ε_t is distributed symmetrically, $\Phi_{\xi\xi}(\tau)$ will detect any possible single or group of even terms, except those defined by (3.22) and (3.23), but only single odd terms defined by (3.24) and a group of two terms defined by (3.25).

Therefore $\Phi_{\xi\xi}(\tau)$ works properly for the linear case, but it is only valid for the nonlinear case for even terms and the majority of odd terms if the residual's probabilistic asymmetry holds. The major failure of the test is in the cases of single odd terms with a superdiagonal bilinear component and some even terms and the majority of odd terms when the residual sequence has a symmetrical probability density function.

3.3 Third order correlation function

Define the third order correlation function

$$\Phi_{\xi\xi\xi}(\tau_1,\tau_2) = E[(\xi_t - E(\xi_t))(\xi_{t-\tau_1} - E(\xi_t))(\xi_{t-\tau_2} - E(\xi_t))] \tag{3.29}$$

which reduces to

$$\Phi_{\xi\xi^2}(\tau) = E[(\xi_t - E(\xi_t))(\xi_{t-\tau} - E(\xi_t)^2)] \tag{3.30}$$

when $\tau_1 \equiv \tau_2$. These two cases will be considered separately.

3.3.1 $\Phi_{\xi\xi^2}(\tau)$

By definition

$$\begin{aligned}
\Phi_{\xi\xi^2}(\tau) &= E[(\xi_t - E(\xi_t))(\xi_{t-\tau} - E(\xi_t))^2] \\
&= E[(\xi_t^e - E(\xi_t^e) + \xi_t^o - E(\xi_t^o) + \varepsilon_t)(\xi_{t-\tau}^e - E(\xi_t^e) + \xi_{t-\tau}^o - E(\xi_t^o) + \varepsilon_{t-\tau})^2] \\
&= \Phi_{\xi^e\xi^{e2}}(\tau) + \Phi_{\xi^o\xi^{o2}}(\tau) + \Phi_{\xi^e\xi^{o2}}(\tau) + \Phi_{\xi^o\xi^{e2}}(\tau) \\
&\quad + \Phi_{\xi^e\xi^{(o,e)}}(\tau) + \Phi_{\xi^o\xi^{(o,e)}}(\tau) + \Phi_{\xi^\varepsilon\xi^{e2}}(\tau) + \Phi_{\xi^\varepsilon\xi^{o2}}(\tau) \\
&\quad + \Phi_{\xi^\varepsilon\xi^{(o,e)}}(\tau) + \Phi_{\xi^\varepsilon\xi^{(e,\varepsilon)}}(\tau) + \Phi_{\xi^e\xi^{(e,\varepsilon)}}(\tau) + \Phi_{\xi^o\xi^{(e,\varepsilon)}}(\tau) \\
&\quad + \Phi_{\xi^e\xi^{(o,\varepsilon)}}(\tau) + \Phi_{\xi^o\xi^{(o,\varepsilon)}}(\tau) + \Phi_{\xi^\varepsilon\xi^{(o,\varepsilon)}}(\tau) + \Phi_{\xi^e\xi^{\varepsilon 2}}(\tau) \\
&\quad + \Phi_{\xi^o\xi^{\varepsilon 2}}(\tau) + \Phi_{\xi^\varepsilon\xi^{\varepsilon 2}}(\tau)
\end{aligned} \tag{3.31}$$

where polynomial correlation $\Phi_{\xi^o\xi^{(o,e)}}(\tau)$ is defined as

$$
\begin{aligned}
\Phi_{\xi^o\xi^{(o,e)}}(\tau) &= E[(\xi_t^o - E(\xi_t^o))(\xi_{t-\tau}^o - E(\xi_t^o))(\xi_{t-\tau}^e - E(\xi_t^e))] \\
&= E\{(\sum_{n=1}^{\infty} G_n^o[\varepsilon_t] - \sum_{n=1}^{\infty} E(G_n^o[\varepsilon_t]))(\sum_{n=1}^{\infty} G_n^o[\varepsilon_{t-\tau}] - \sum_{n=1}^{\infty} E(G_n^o[\varepsilon_t])) \\
&\quad \cdot(\sum_{n=1}^{\infty} G_{2n}^e[\varepsilon_{t-\tau}] - \sum_{n=1}^{\infty} E(G_{2n}^e[\varepsilon_t]))\} \\
&= E\{(G_1^o[\varepsilon_t] - E(G_1^o[\varepsilon_t]))(G_1^o[\varepsilon_{t-\tau}] - E(G_1^o[\varepsilon_t]))(G_2^e[\varepsilon_{t-\tau}] - E(G_2^e[\varepsilon_t])) \\
&\quad + (G_1^o[\varepsilon_t] - E(G_1^o[\varepsilon_t]))(G_3^o[\varepsilon_{t-\tau}] - G_3^o[\varepsilon_t])(G_4^e[\varepsilon_t] - E(G_4^e[\varepsilon_t])) \\
&\quad + \cdots\}
\end{aligned} \tag{3.32}
$$

and $\Phi_{\xi^e\xi^{(o,\varepsilon)}}(\tau)$ is defined as

$$
\begin{aligned}
\Phi_{\xi^e\xi^{(o,\varepsilon)}}(\tau) &= E[(\xi_t^e - E(\xi_t^e))(\xi_{t-\tau}^o - E(\xi_t^o))(\varepsilon_{t-\tau})] \\
&= E\{(\sum_{n=1}^{\infty} G_{2n}^e[\varepsilon_t] - \sum_{n=1}^{\infty} E(G_{2n}^e[\varepsilon_t]))(\sum_{n=1}^{\infty} G_n^o[\varepsilon_{t-\tau}] - \sum_{n=1}^{\infty} E(G_n^o[\varepsilon_t]))(\varepsilon_{t-\tau})\} \\
&= E\{(G_2^e[\varepsilon_t] - E(G_2^e[\varepsilon_t]))(G_1^o[\varepsilon_{t-\tau}] - E(G_1^o[\varepsilon_t]))(\varepsilon_{t-\tau})\} \\
&\quad + (G_4^e[\varepsilon_t] - E(G_4^e[\varepsilon_t]))(G_3^o[\varepsilon_{t-\tau}] - E(G_3^o[\varepsilon_t]))(\varepsilon_{t-\tau}) + \cdots\}
\end{aligned} \tag{3.33}
$$

with the other terms defined in an analogous manner. $\Phi_{\xi\xi^2}(\tau) = c\delta(\tau)$, where c is a constant, when ε_t is a zero mean independent sequence. If $\Phi_{\xi\xi^2}(\tau)$ were an ideal test, then $\Phi_{\xi\xi^2}(\tau) \neq 0$, when $\tau > 0$ for the case of any possible unmodelled terms in the residual sequence.

Next consider each group of correlations which exist when either even terms or odd terms, or both even and odd terms remain unmodelled in the residuals.

If ε_t is asymmetrical, it can be proved that

(i) $\Phi_{\xi^o\xi^{e2}}(\tau)$, $\Phi_{\xi^e\xi^{o2}}(\tau)$, $\Phi_{\xi^e\xi^{(o,\varepsilon)}}(\tau)$, $\Phi_{\xi^o\xi^{(e,\varepsilon)}}(\tau)$, $\Phi_{\xi^e\xi^{(o,\varepsilon)}}(\tau)$, $\Phi_{\xi^o\xi^{e2}}(\tau) = 0$, when $\tau \neq 0$. These correlations do not contribute to the test.

(ii) $\Phi_{\xi^e\xi^{e2}}(\tau) + \Phi_{\xi^e\xi^{e2}}(\tau) + \Phi_{\xi^e\xi^{(e,\varepsilon)}}(\tau) \neq 0$, when $\tau \neq 0$. These correlations are induced by any possible even terms in the residuals, so that any single even terms can be detected.

(iii) $\Phi_{\xi^o\xi^{o2}}(\tau) + \Phi_{\xi^o\xi^{(o,\varepsilon)}}(\tau) + \Phi_{\xi^o\xi^{\varepsilon 2}}(\tau) \neq 0$, when $\tau \neq 0$ only on the following conditions:

(a) when there exists a single odd term in the residual

$$K\varepsilon_{t-t_1}\varepsilon_{t-t_2}^{i_2} \cdots \varepsilon_{t-t_m}^{i_m}\eta_q^e \tag{3.34}$$

where K is the corresponding Volterra kernel, and $i_2,\dots,i_m$, $m \geq 2$, $= 0$ or all odd (>1). η_q^e is the possible even component in the odd term, $q \geq 0$. Or equivalently $\Phi_{\xi\xi^2}(\tau)$ fails when the term contains one or more of the superdiagonal bilinear components $\varepsilon_{t-t_1}\xi_{t-t_2}$, $t_1 < t_2$.

(b) when there are two odd terms together in the residuals

$$\left.\begin{array}{l} K\varepsilon_{t-t_1}\varepsilon_{t-t_2} \cdots \varepsilon_{t-t_a}\varepsilon_{t-t_{a+1}}\varepsilon_{t-t_{a+2}}^{i_{a+2}}\varepsilon_{t-t_{a+3}}^{i_{a+3}} \cdots \varepsilon_{t-t_b}^{i_b}\eta_c^e \\ K'\varepsilon_{t-t'_1}\varepsilon_{t-t'_2} \cdots \varepsilon_{t-t'_a}\varepsilon_{t-t'_{a+1}}^{i'_{a+1}}\varepsilon_{t-t'_{a+2}}^{i'_{a+2}} \cdots \varepsilon_{t-t'_d}^{i'_b}\eta_{c'}^e \end{array}\right\} \tag{3.35}$$

where K, K' are the corresponding Volterra kernels. i_{a+2}, i_{a+3}, $\cdots$ i_b, i'_{a+1}, $\cdots$ i'_b $= 0$ or odd(>1), $a \geq 1$, $b \geq 3$, and $\eta_c^e, \eta_{c'}^e$ are possible even components in the two odd terms.

$$\left.\begin{array}{l} t_1 - t_2 = t'_1 - t'_2 = \cdots = t_{a-1} - t_a = t'_{a-1} - t'_a \\ t_{a+1} = t_1 - t'_1 \end{array}\right\} \tag{3.36}$$

(c) when three odd terms coexist in the residuals

$$\left.\begin{array}{l} K\varepsilon_{t-t_1} \cdots \varepsilon_{t-t_a}\varepsilon_{t-t_{a-1}} \cdots \varepsilon_{t-t_b}\varepsilon_{t-t_{b-1}}^{i_{b-1}} \cdots \varepsilon_{t-t_c}^{i_c}\eta_q^e \\ K'\varepsilon_{t-t'_1}\varepsilon_{t-t'_2} \cdots \varepsilon_{t-t'_a}\varepsilon_{t-t'_{b-1}}^{i'_{b-1}}\varepsilon_{t-t'_{b-2}}^{i'_{b-2}} \cdots \varepsilon_{t-t'_c}^{i'_c}\eta_{q'}^e \\ K''\varepsilon_{t-t''_{a-1}}\varepsilon_{t-t''_{a-2}} \cdots \varepsilon_{t-t''_b}\varepsilon_{t-t''_{b+1}}^{i''_{b+1}} \cdots \varepsilon_{t-t''_c}^{i''_c}\eta_{q''}^e \end{array}\right\} \tag{3.37}$$

where K, K', K'' are the three corresponding Volterra kernels, $a \geq 1$, $b \geq 2$ and $i_{b+1}, \dots, i_c$; $i'_{b+1}, \dots, i'_c$; $i''_{b+1}, \dots, i''_c = 0$ or odd (>1). η_q^e, $\eta_{q'}^e$ $\eta_{q''}^e$ are the possible even components in each of the three odd terms, q, $q_,$, $q'' \geq 0$.

$$\left.\begin{array}{l} t_1 - t_2 = t'_1 - t'_2 = \cdots = t_{a-1} - t_a = t'_{a-1} - t'_a \\ = t_{a+1} - t_{a+2} = t''_{a+1} - t''_{a+2} = \cdots = t_{b-1} - t_b = t''_{b-1} - t''_b \end{array}\right\} \tag{3.38}$$

If ε_t is symmetrical, then it can similarly be shown that

(i) $\Phi_{\xi^\varepsilon\xi^{e2}}(\tau)$, $\Phi_{\xi^\varepsilon\xi^{o2}}(\tau)$, $\Phi_{\xi^\varepsilon\xi^{(o,e)}}(\tau)$, $\Phi_{\xi^\varepsilon\xi^{(e,\varepsilon)}}(\tau)$, $\Phi_{\xi\xi}(\tau)$, $\Phi_{\xi^\varepsilon\xi^{\varepsilon 2}}(\tau)$, $\Phi_{\xi^o\xi^{e2}}(\tau)$, $\Phi_{\xi^o\xi^{\varepsilon 2}}(\tau) = 0$, for all $\tau \neq 0$. These correlations do not contribute to the test.

(ii) $\Phi_{\xi^e\xi^{e2}}(\tau) + \Phi_{\xi^e\xi^{e2}}(\tau) + \Phi_{\xi^e\xi^{(e,\varepsilon)}}(\tau) \neq 0$, when $\tau \neq 0$. So that any single even term can be detected.

(iii) $\Phi_{\xi^o\xi^{o2}}(\tau) + \Phi_{\xi^o\xi^{(o,\varepsilon)}}(\tau) \neq 0$, when $\tau \neq 0$ only on the following conditions

(a) when there exist two odd terms together in the residuals

$$\left.\begin{array}{l} K\varepsilon_{t-t_1}^{i_1}\varepsilon_{t-t_2}^{i_2}\cdots\varepsilon_{t-t_a}^{i_a}\varepsilon_{t-t_{a+1}}^{i_{a+1}}\eta_q^e \\ K'\varepsilon_{t-t'_1}^{i'_1}\varepsilon_{t-t'_2}^{i'_2}\cdots\varepsilon_{t-t'_a}^{i'_a}\eta_{q'}^e \end{array}\right\} \quad (3.39)$$

where K, K' are the corresponding Volterra kernels. $i_1, i_2, \cdots i_a, i_{a+1}, i'_1, i'_2, \cdots i'_a$ are all odd, $a \geq 1$, and $\eta_c^e, \eta_{c'}^e$ are possible even components in the three odd terms, $q, q' \geq 0$.

$$\left.\begin{array}{l} t_1 - t_2 = t'_1 - t'_2 = \cdots = t_{a-1} - t_a = t'_{a-1} - t'_a \\ t_{a+1} = t_1 - t'_1 \end{array}\right\} \quad (3.40)$$

(b) when three odd terms as follows in the residuals

$$\left.\begin{array}{l} K\varepsilon_{t-t_1}^{i_1}\cdots\varepsilon_{t-t_a}^{i_a}\varepsilon_{t-t_{a-1}}^{i_{a-1}}\cdots\varepsilon_{t-t_b}^{i_b}\eta_q^e \\ K'\varepsilon_{t-t'_1}^{i'_1}\varepsilon_{t-t'_2}^{i'_2}\cdots\varepsilon_{t-t'_a}^{i'_a}\eta_{q'}^e \\ K''\varepsilon_{t-t''_{a-1}}^{i_{a-1}}\varepsilon_{t-t''_{a+2}}^{i_{a+2}}\cdots\varepsilon_{t-t''_b}^{i_b}\eta_{q''}^e \end{array}\right\} \quad (3.41)$$

where K, K', K'' are the three corresponding Volterra kernels, $a \geq 1, b \geq 2$. $\eta_q^e, \eta_{q'}^e, \eta_{q''}^e$ are the possible even components in each of the three odd terms, $q, q', q'' \geq 0$.

$$\left.\begin{array}{l} t_1 - t_2 = t'_1 - t'_2 = \cdots = t_{a-1} - t_a = t'_{a-1} - t'_a \\ = t_{a+1} - t_{a+2} = t''_{a+1} - t''_{a+2} = \cdots = t_{b-1} - t_b = t''_{b-1} - t''_b \end{array}\right\} \quad (3.42)$$

Lemma 3.2

Consider the residual sequence defined by eqn (3.4), if ε_t is distributed asymmetrically, $\Phi_{\xi\xi^2}(\tau)$ detects any possible single even terms, single odd terms defined by (3.34), . and a group of two terms defined by (3.35) or three odd terms defined by (3.37). If ε_t is distributed symmetrically, $\Phi_{\xi\xi^2}(\tau_1,\tau_2)$ will detect any possible single even terms, none of any single odd terms, and a group of two defined by (3.39) or of three odd terms defined by (3.41) in the residuals.

For instance, no single or group of linear terms belong to the category of terms which can be detected by $\Phi_{\xi\xi^2}(\tau)$. $\Phi_{\xi\xi^2}(\tau)$ does not detect the superdiagonal bilinear types terms defined by eqn (3.20).

3.3.2 $\Phi_{\xi\xi\xi}(\tau_1,\tau_2)$

By definition

$$\Phi_{\xi\xi\xi}(\tau_1,\tau_2) = E[(\xi_t - E(\xi_t))(\xi_{t-\tau_1} - E(\xi_t))(\xi_{t-\tau_2} - E(\xi_t))]$$

$$= E[(\xi_t^e - E(\xi_t^e) + \xi_t^o - E(\xi_t^o) + \varepsilon_t)(\xi_{t-\tau_1}^e - E(\xi_t^e) + \xi_{t-\tau_1}^o - E(\xi_t^o) + \varepsilon_{t-\tau_1})$$
$$\cdot(\xi_{t-\tau_2}^e - E(\xi_t^e) + \xi_{t-\tau_2}^o - E(\xi_t^o) + \varepsilon_{t-\tau_2})]$$

$$= \Phi_{\xi^e\xi^e\xi^e}(\tau_1,\tau_2) + \Phi_{\xi^o\xi^o\xi^o}(\tau_1,\tau_2) + \Phi_{\xi^e\xi^o\xi^o}(\tau_1,\tau_2) + \Phi_{\xi^o\xi^o\xi^e}(\tau_1,\tau_2)$$
$$+ \Phi_{\xi^o\xi^e\xi^o}(\tau_1,\tau_2) + \Phi_{\xi^o\xi^e\xi^e}(\tau_1,\tau_2) + \Phi_{\xi^e\xi^o\xi^e}(\tau_1,\tau_2) + \Phi_{\xi^e\xi^e\xi^o}(\tau_1,\tau_2)$$
$$+ \Phi_{\xi^e\xi^e\xi^\varepsilon}(\tau_1,\tau_2) + \Phi_{\xi^e\xi^\varepsilon\xi^e}(\tau_1,\tau_2) + \Phi_{\xi^e\xi^\varepsilon\xi^\varepsilon}(\tau_1,\tau_2) + \Phi_{\xi^\varepsilon\xi^e\xi^\varepsilon}(\tau_1,\tau_2)$$
$$+ \Phi_{\xi^\varepsilon\xi^e\xi^e}(\tau_1,\tau_2) + \Phi_{\xi^\varepsilon\xi^\varepsilon\xi^e}(\tau_1,\tau_2) + \Phi_{\xi^o\xi^o\xi^\varepsilon}(\tau_1,\tau_2) + \Phi_{\xi^o\xi^\varepsilon\xi^o}(\tau_1,\tau_2)$$
$$+ \Phi_{\xi^o\xi^\varepsilon\xi^\varepsilon}(\tau_1,\tau_2) + \Phi_{\xi^\varepsilon\xi^o\xi^\varepsilon}(\tau_1,\tau_2) + \Phi_{\xi^\varepsilon\xi^o\xi^o}(\tau_1,\tau_2) + \Phi_{\xi^\varepsilon\xi^\varepsilon\xi^o}(\tau_1,\tau_2)$$
$$+ \Phi_{\xi^e\xi^o\xi^\varepsilon}(\tau_1,\tau_2) + \Phi_{\xi^e\xi^\varepsilon\xi^o}(\tau_1,\tau_2) + \Phi_{\xi^o\xi^e\xi^\varepsilon}(\tau_1,\tau_2) + \Phi_{\xi^o\xi^\varepsilon\xi^e}(\tau_1,\tau_2)$$
$$+ \Phi_{\xi^\varepsilon\xi^\varepsilon\xi^\varepsilon}(\tau_1,\tau_2) \tag{3.43}$$

where the polynomial correlation $\Phi_{\xi^e\xi^o\xi^e}(\tau_1,\tau_2)$ is defined as

$$\Phi_{\xi^e\xi^o\xi^e}(\tau_1,\tau_2) = E[(\xi_t^e - E(\xi_t^e))(\xi_{t-\tau_1}^o - E(\xi_t^o))(\xi_{t-\tau_2}^e - E(\xi_t^e))]$$

$$= E\{(\sum_{n=1}^{\infty} G_{2n}^e[\varepsilon_t] - \sum_{n=1}^{\infty} E(G_{2n}^e[\varepsilon_t]))(\sum_{n=1}^{\infty} G_n^o[\varepsilon_{t-\tau_1}] - \sum_{n=1}^{\infty} E(G_n^o[\varepsilon_t]))$$
$$\cdot(\sum_{n=1}^{\infty} G_{2n}^e[\varepsilon_{t-\tau_2}] - \sum_{n=1}^{\infty} E(G_{2n}^e[\varepsilon_t]))\}$$

$$= E\{(G_2^e[\varepsilon_t] - E(G_2^e[\varepsilon_t]))(G_1^o[\varepsilon_{t-\tau_1}] - E(G_1^o[\varepsilon_t]))(G_2^e[\varepsilon_{t-\tau_2}] - E(G_2^e[\varepsilon_t]))$$

$$+ (G [\xi_t] - E(G [\xi_t]))(G [\xi_{t}]_{\tau_1} - E(G [\xi_t]))(G [\xi_{t}]_{\tau_2} - E(G [\xi_t]))$$

$$+ \cdots \} \tag{3.44}$$

and $\Phi_{\xi^e\xi^\varepsilon\xi^e}(\tau_1,\tau_2)$ is defined as

$$\Phi_{\xi^e\xi^\varepsilon\xi^e}(\tau_1,\tau_2) = E[(\xi_t^e - E(\xi_t^e))(\varepsilon_{t-\tau_1})(\xi_{t-\tau_2}^e - E(\xi_t^e))]$$

$$= E\{(\sum_{n=1}^{\infty} G_{2n}^e[\varepsilon_t] - \sum_{n=1}^{\infty} E(G_{2n}^e[\varepsilon_t]))(\varepsilon_{t-\tau_1})(\sum_{n=1}^{\infty} G_{2n}^e[\varepsilon_{t-\tau_2}] - \sum_{n=1}^{\infty} E(G_{2n}^e[\varepsilon_t]))\}$$

$$= E\{(G_2^e[\varepsilon_t] - E(G_2^e[\varepsilon_t]))(\varepsilon_{t-\tau_1})(G_2^e[\varepsilon_{t-\tau_2}] - E(G_2^e[\varepsilon_t]))$$

$$+ (G_4^e[\varepsilon_t] - E(G_4^e[\varepsilon_t]))(\varepsilon_{t-\tau_1})(G_4^e[\varepsilon_{t-\tau_2}] - E(G_4^e[\varepsilon_t])) + \cdots\} \quad (3.45)$$

Others are defined in an analogous manner. If the residual sequence is independent, $\Phi_{\xi\xi\xi}(\tau_1,\tau_2)$, $= c\delta(\tau)$, where c is a constant, but ideally if any unmodelled terms exist in the residuals, then it is required that $\Phi_{\xi\xi\xi}(\tau_1,\tau_2)$, $\neq 0$, when τ_1, $\tau_2 \neq 0$.

If ε_t is asymmetrical, it can be proved that

(i) $\Phi_{\xi^e\xi^e\xi^e}(\tau_1,\tau_2)$, $\Phi_{\xi^e\xi^e\xi^e}(\tau_1,\tau_2)$, $\Phi_{\xi^e\xi^e\xi^e}(\tau_1,\tau_2)$, $\Phi_{\xi^e\xi^o\xi^o}(\tau_1,\tau_2)$, $\Phi_{\xi^e\xi^e\xi^e}(\tau_1,\tau_2) = 0$, for any $\tau_1,\tau_2 \neq 0$. These correlations do not contribute to the test.

(ii) $\Phi_{\xi^e\xi^e\xi^e}(\tau_1,\tau_2) + \Phi_{\xi^e\xi^e\xi^e}(\tau_1,\tau_2) \neq 0$, when $\tau_1,\tau_2 \neq 0$, for any possible even terms. So that any single even term can be detected.

(iii) $\Phi_{\xi^o\xi^o\xi^o}(\tau_1,\tau_2) + \Phi_{\xi^o\xi^o\xi^e}(\tau_1,\tau_2) + \Phi_{\xi^o\xi^e\xi^o}(\tau_1,\tau_2) + \Phi_{\xi^o\xi^e\xi^e}(\tau_1,\tau_2) + \Phi_{\xi^e\xi^o\xi^e}(\tau_1,\tau_2)$ $+ \Phi_{\xi^e\xi^e\xi^o}(\tau_1,\tau_2) \neq 0$, when τ_1, $\tau_2 \neq 0$ for all possible odd terms except the case when there exists a single odd term in the residuals

$$K\varepsilon_{t-t_1}\varepsilon_{t-t_2}\cdots\varepsilon_{t-t_m}\varepsilon_{t-t_{m-1}}^{i_{m-1}}\cdots\varepsilon_{t-t_p}^{i_p}\eta_q^e \quad (3.46)$$

where K is the corresponding Volterra kernel, $m \geq 3$, $i_{m+1},\ldots,i_p = 0$ or odd (>1) and η_q^e is the possible even component in the odd term, $q \geq 0$. In this case, the test does not distinguish between single odd terms and an independent sequence.

(iv) $\Phi_{\xi\xi\xi}(\tau_1,\tau_2)$ can detect a group of two odd terms of the type of (3.46) together in the residuals

$$\left.\begin{array}{l} K\varepsilon_{t-t_1}\varepsilon_{t-t_2}\cdots\varepsilon_{t-t_a}\varepsilon_{t-t_{a-1}}^{i_{a-1}}\varepsilon_{t-t_{a-2}}^{i_{a+2}}\varepsilon_{t-t_{a-3}}^{i_{a-3}}\cdots\varepsilon_{t-t_b}^{i_b}\eta_c^e \\ K'\varepsilon_{t-t'_1}\varepsilon_{t-t'_2}\cdots\varepsilon_{t-t'_a}\varepsilon_{t-t_{a-1}}^{i'_{a-1}}\varepsilon_{t-t_{a+2}}^{i'_{a-2}}\cdots\varepsilon_{t-t'_b}^{i'_b}\eta_{c'}^e \end{array}\right\} \quad (3.47)$$

where K, K' are the corresponding Volterra kernels. $a \geq 3$, i_{a+1} is odd, $i_{a+2}, i_{a+3}, \cdots i_b, i'_{a+1}, \cdots i'_b = 0$ or odd(>1), and $\eta_c^e, \eta_{c'}^e$ are possible even components in the three odd terms.

$$t_1 - t_2 = t'_1 - t'_2 = \cdots = t_{a-1} - t_a = t'_{a-1} - t'_a \quad (3.48)$$

(v) $\Phi_{\xi\xi\xi}(\tau_1,\tau_2)$ detects a group of tree odd terms of the types defined by (3.46)

$$\left.\begin{array}{l} K\varepsilon_{t-t_1}\cdots\varepsilon_{t-t_a}\varepsilon_{t-t_{a+1}}\cdots\varepsilon_{t-t_b}\varepsilon_{t-t_{b-1}}^{i_{b-1}}\cdots\varepsilon_{t-t_c}^{i_c}\eta_q^e \\ K'\varepsilon_{t-t'_1}\varepsilon_{t-t'_2}\cdots\varepsilon_{t-t'_a}\varepsilon_{t-t'_{b-1}}^{i'_{b-1}}\varepsilon_{t-t'_{b-2}}^{i'_{b-2}}\cdots\varepsilon_{t-t'_c}^{i'_c}\eta_{q'}^e \\ K''\varepsilon_{t-t''_{a-1}}\varepsilon_{t-t''_{a-2}}\cdots\varepsilon_{t-t''_b}\varepsilon_{t-t''_{b+1}}^{i''_{b-1}}\cdots\varepsilon_{t-t''_c}^{i''_c}\eta_{q''}^e \end{array}\right\} \quad (3.49)$$

where K, K', K'' are the three corresponding Volterra kernels, $a \geq 3$, $b \geq 5$ and $i_{b+1}, \ldots, i_c;\ i'_{b+1}, \ldots, i'_c;\ i''_{b+1}, \ldots, i''_c = 0$ or odd (>1). η^e_q, $\eta^e_{q'}$ $\eta^e_{q''}$ are the possible even components in each of the three odd terms, $q, q_{,}, q'' \geq 0$.

$$\left.\begin{aligned} t_1 - t_2 = t'_1 - t'_2 = \cdots = t_{a-1} - t_a = t'_{a-1} - t'_a \\ t_{a+1} - t_{a+2} = t''_{a+1} - t''_{a+2} = \cdots = t_{b-1} - t_b = t''_{b-1} - t''_b \end{aligned}\right\} \tag{3.50}$$

If ε_t is symmetrical, similarly it can be shown that

(i) $\Phi_{\xi^e\xi^e\xi^e}(\tau_1,\tau_2)$, $\Phi_{\xi^e\xi^e\xi^e}(\tau_1,\tau_2)$, $\Phi_{\xi^\varepsilon\xi^e\xi^e}(\tau_1,\tau_2)$, $\Phi_{\xi^\varepsilon\xi^o\xi^o}(\tau_1,\tau_2)$, $\Phi_{\xi^\varepsilon\xi^\varepsilon\xi^\varepsilon}(\tau_1,\tau_2) = 0$, for all $\tau_1, \tau_2 \neq 0$. These correlations do not therefore contribute to the test.

(ii) $\Phi_{\xi^\varepsilon\xi^e\xi^e}(\tau_1,\tau_2) + \Phi_{\xi^e\xi^\varepsilon\xi^\varepsilon}(\tau_1,\tau_2) \neq 0$, where $\tau_1, \tau_2 \neq 0$, for any possible even terms. So that any single even term can be detected.

(iii) $\Phi_{\xi^o\xi^o\xi^o}(\tau_1,\tau_2) + \Phi_{\xi^o\xi^o\xi^\varepsilon}(\tau_1,\tau_2) + \Phi_{\xi^o\xi^\varepsilon\xi^o}(\tau_1,\tau_2) + \Phi_{\xi^o\xi^\varepsilon\xi^\varepsilon}(\tau_1,\tau_2) + \Phi_{\xi^\varepsilon\xi^o\xi^\varepsilon}(\tau_1,\tau_2)$ $+ \Phi_{\xi^\varepsilon\xi^\varepsilon\xi^o}(\tau_1,\tau_2) \neq 0$, for some τ_1, τ_2 only on the following conditions:

(a) when there exists a single odd term in the residuals

$$K\varepsilon^{i_1}_{t-t_1}\varepsilon^{i_2}_{t-t_2}\eta^e_q \tag{3.51}$$

where K is the corresponding Volterra kernel, i_1, i_2 odd and η^e_q is a possible even component in the odd term, $q \geq 0$.

(b) when there exists a group of two odd terms

$$\left.\begin{aligned} K'\varepsilon^{i'_1}_{t-t'_1}\varepsilon^{i'_2}_{t-t'_2}\cdots\varepsilon^{i'_a}_{t-t'_a}\varepsilon^{i'_{a+1}}_{t-t'_{a+1}}\eta^e_{q'} \\ K''\varepsilon^{i''_1}_{t-t''_1}\varepsilon^{i''_2}_{t-t''_2}\cdots\varepsilon^{i''_a}_{t-t''_a}\eta^e_{q''} \end{aligned}\right\} \tag{3.52}$$

where K', K'' are the corresponding Volterra kernels. $a \geq 2$, $i'_{a+1}, i'_2, \cdots, i'_a, i'_{a+1}, i''_1, i''_2, \cdots, i''_a$ are all odd, and $\eta^e_{q'}, \eta^e_{q''}$ are possible even components in the three odd terms.

$$t'_1 - t'_2 = t''_1 - t''_2 = \cdots = t'_{a-1} - t'_a = t''_{a-1} - t''_a \tag{3.53}$$

(c) when there exists a group of three odd terms

$$\left.\begin{aligned} K\varepsilon^{i_1}_{t-t_1}\cdots\varepsilon^{i_a}_{t-t_a}\varepsilon^{i_{a+1}}_{t-t_{a+1}}\cdots\varepsilon^{i_b}_{t-t_b}\eta^e_q \\ K'\varepsilon^{i'_1}_{t-t'_1}\varepsilon^{i'_2}_{t-t'_2}\cdots\varepsilon^{i'_a}_{t-t'_a}\eta^e_{q'} \\ K''\varepsilon^{i''_{a+1}}_{t-t''_{a+1}}\varepsilon^{i''_{a+2}}_{t-t''_{a+2}}\cdots\varepsilon_{t-t''_b}\eta^e_{q''} \end{aligned}\right\} \tag{3.54}$$

where K, K', K'' are the three corresponding Volterra kernels, $a \geq 3$, $b \geq 5$ and $i_1, \cdots, i_a;\ i_{a+1}, \cdots, i_b;\ i'_1, \cdots, i'_a;\ i''_{a+1}, \ldots, i''_b$ are all odd. η^e_q, $\eta^e_{q'}$ $\eta^e_{q''}$ are the possible even components in each of the three odd terms, $q,\ q',\ q'' \geq 0$.

$$\left.\begin{array}{l} t_1 - t_2 = t_1' - t_2' = \cdots = t_{a-1} - t_a = t_{a-1}' - t_a' \\ t_{a+1} - t_{a+2} = t_{a+1}'' - t_{a+2}'' = \cdots = t_{b-1} - t_b = t_{b-1}'' - t_b'' \end{array}\right\} \tag{3.55}$$

These three categories of single or group of terms can be detected by $\Phi_{\xi\xi\xi}(\tau_1,\tau_2)$.

Lemma 3.3

Consider the residual sequence defined by eqn (3.4), if ε_t is distributed asymmetrically, $\Phi_{\xi\xi\xi}(\tau_1,\tau_2)$ detects any possible even term, all possible odd terms except the type of single term defined by (3.46), a group of two odd terms defined by (3.47) or three odd terms defined by (3.48). If ε_t is distributed symmetrically, $\Phi_{\xi\xi\xi}(\tau_1,\tau_2)$ detects any possible even terms, one type of odd terms defined by (3.51), and a group of two odd terms defined by (3.52), or of three odd terms defined by (3.54) in the residuals.

For instance, no linear terms can be detected by $\Phi_{\xi\xi\xi}(\tau_1,\tau_2)$. But for the superdiagonal bilinear type terms as defined by (3.20) and (3.21), it is seen that there exists a group of two odd terms $\varepsilon_{t-1}\varepsilon_{t-2}$, $\varepsilon_{t-1}\varepsilon_{t-2}\varepsilon_{t-3}$ which are of the type defined by (3.47) when ε_t is asymmetrical, or by (3.52) when ε_t is symmetrical, and can be detected by $\Phi_{\xi\xi\xi}(\tau_1,\tau_2)$.

The third order correlations $\Phi_{\xi\xi\xi}(\tau)$, including $\Phi_{\xi\xi^2}(\tau)$, therefore provide very similar information as the autocorrelation function $\Phi_{\xi\xi}(\tau)$ for even terms. $\Phi_{\xi\xi\xi}(\tau)$ detects a wider class of odd terms including bilinear types of terms, but with more computation than $\Phi_{\xi\xi^2}(\tau)$. Consequently the third order correlations cannot be used to provide a general validity test for nonlinear systems.

3.4 Fourth order correlations

Only the special subset of fourth order correlations (Granger and Andersen 1981) needs to be considered. Define

$$\begin{aligned}
\Phi_{\xi^{2'}\xi^{2'}}(\tau) &= E[(\xi_t^2 - E(\xi_t^2))(\xi_{t-\tau}^2 - E(\xi_t^2))] \\
&= E[(\xi_t^e + \xi_t^o + \varepsilon_t)^2 - E(\xi_t^e + \xi_t^o + \varepsilon_t)^2) \\
&\qquad \cdot(\xi_{t-\tau}^e + \xi_{t-\tau}^o + \varepsilon_{t-\tau})^2 - E(\xi_t^e + \xi_t^o + \varepsilon_t)^2)] \\
&= \Phi_{\xi^{e2'}\xi^{e2'}}(\tau) + \Phi_{\xi^{o2'}\xi^{o2'}}(\tau) + \Phi_{\xi^{o2'}\xi^{e2'}}(\tau) + \Phi_{\xi^{e2'}\xi^{o2'}}(\tau) + \Phi_{\xi^{(oe)}\xi^{(oe)}}(\tau) \\
&\quad + \Phi_{\xi^{(oe)}\xi^{o2'}}(\tau) + \Phi_{\xi^{(oe)}\xi^{e2'}}(\tau) + \Phi_{\xi^{o2'}\xi^{(oe)}}(\tau) + \Phi_{\xi^{e2'}\xi^{(oe)}}(\tau) \\
&\quad + \Phi_{\xi^{\varepsilon 2'}\xi^{\varepsilon 2'}}(\tau) + \Phi_{\xi^{e2'}\xi^{\varepsilon 2'}}(\tau) + \Phi_{\xi^{\varepsilon 2'}\xi^{e2'}}(\tau) + \Phi_{\xi^{o2'}\xi^{\varepsilon 2'}}(\tau) + \Phi_{\xi^{\varepsilon 2'}\xi^{o2'}}(\tau)
\end{aligned}$$

$$+ \Phi_{\xi^{e2'}\xi^{(oe)}}(\tau) + \Phi_{\xi^{(oe)}\xi^{e2'}}(\tau) + \Phi_{\xi^{e2'}\xi^{(o\varepsilon)}}(\tau) + \Phi_{\xi^{(o\varepsilon)}\xi^{e2'}}(\tau) + \Phi_{\xi^{o2'}\xi^{(o\varepsilon)}}(\tau)$$

$$+ \Phi_{\xi^{(o\varepsilon)}\xi^{o2'}}(\tau) + \Phi_{\xi^{e2'}\xi^{(oe)}}(\tau) + \Phi_{\xi^{(oe)}\xi^{e2'}}(\tau) + \Phi_{\xi^{e2'}\xi^{(e\varepsilon)}}(\tau) + \Phi_{\xi^{(e\varepsilon)}\xi^{e2'}}(\tau)$$

$$+ \Phi_{\xi^{o2'}\xi^{(e\varepsilon)}}(\tau) + \Phi_{\xi^{(e\varepsilon)}\xi^{o2'}}(\tau) + \Phi_{\xi^{e2'}\xi^{(e\varepsilon)}}(\tau) + \Phi_{\xi^{(e\varepsilon)}\xi^{e2'}}(\tau) \qquad (3.56)$$

where polynomial correlation $\Phi_{\xi^{o2'}\xi^{oe}}(\tau)$ is given by

$$\Phi_{\xi^{o2'}\xi^{oe}}(\tau) = E[(\xi_t^o)^2 - E((\xi_t^o)^2))(\xi_{t-\tau}^o \xi_{t-\tau}^e) - E(\xi_t^o \xi_t^e)]$$

$$= E\{[(\sum_{n=1}^{\infty} G_n^o[\varepsilon_t])^2 - E\{(\sum_{n=1}^{\infty} G_n^o[\varepsilon_t])^2\}]$$

$$\cdot [(\sum_{n=1}^{\infty} G_n^o[\varepsilon_{t-\tau}])(\sum_{n=1}^{\infty} G_{2n}^e[\varepsilon_{t-\tau}]) - E\{(\sum_{n=1}^{\infty} G_n^o[\varepsilon_t])(\sum_{n=1}^{\infty} G_{2n}^e[\varepsilon_t])\}]\}$$

$$= E\{[(G_1^o[\varepsilon_t])^2 - E((G_1^o[\varepsilon_t])^2)][G_1^o[\varepsilon_{t-\tau}G_2^e[\varepsilon_{t-\tau}] - E(G_1^o[\varepsilon_t]G_2^e[\varepsilon_t])]$$

$$+ [(G_1^o[\varepsilon_t])^2 - E((G_1^o[\varepsilon_t])^2)][G_2^o[\varepsilon_{t-\tau}]G_4^e[\varepsilon_{t-\tau}] - E(G_2^o[\varepsilon_t]G_4^e[\varepsilon_t])]$$

$$+ \cdots \} \qquad (3.57)$$

and the others are defined in a similar manner. It is easy to see that if ξ_t is independent, $\Phi_{\xi^{2'}\xi^{2'}}(\tau) = c\delta(\tau)$, where c is a constant. If any unmodelled terms remain in the residual sequence, therefore this should be indicated by $\Phi_{\xi^{2'}\xi^{2'}}(\tau) \neq 0$, when $\tau \neq 0$.

By analysing eqn(3.56) it is sufficient to consider only the polynomial square autocorrelations of even and odd terms and their square crosscorrelations with the prediction error ε_t, $\Phi_{\xi^{e2'}\xi^{e2'}}(\tau)$, $\Phi_{\xi^{o2'}\xi^{o2'}}(\tau)$, $\Phi_{\xi^{\varepsilon2'}\xi^{\varepsilon2'}}(\tau)$, $\Phi_{\xi^{e2'}\xi^{\varepsilon2'}}(\tau)$, $\Phi_{\xi^{\varepsilon2'}\xi^{e2'}}(\tau)$, $\Phi_{\xi^{\varepsilon2'}\xi^{o2'}}(\tau)$, $\Phi_{\xi^{o2'}\xi^{\varepsilon2'}}(\tau)$. Any other polynomial correlations will be neglected in the analysis because whether they are zero or non-zero does not affect the results of the analysis. When ε_t is distributed both asymmetrically and symmetrically, it follows that

(i) $\Phi_{\xi^{\varepsilon2'}\xi^{e2'}}(\tau) + \Phi_{\xi^{e2'}\xi^{\varepsilon2'}}(\tau) + \Phi_{\xi^{\varepsilon2'}\xi^{\varepsilon2'}}(\tau) + \Phi_{\xi^{e2'}\xi^{e2'}}(\tau) \neq 0$, when $\tau \neq 0$, for any possible even terms. So that any kind of single even terms will be detected.

(ii) $\Phi_{\xi^{o2'}\xi^{o2'}}(\tau) + \Phi_{\xi^{o2'}\xi^{\varepsilon2'}}(\tau) + \Phi_{\xi^{\varepsilon2'}\xi^{o2'}}(\tau) + \Phi_{\xi^{\varepsilon2'}\xi^{\varepsilon2'}}(\tau) \neq 0$, when $\tau \neq 0$. That is, any single odd term will be detected by $\Phi_{\xi^{2'}\xi^{2'}}(\tau)$.

Lemma 3.4

Consider the residual sequence defined by eqn (3.4) $\Phi_{\xi^{2'}\xi^{2'}}(\tau)$ detects any possible terms that could be present in the residuals.

3.5 Summary

The results obtained above under the assumptions defined in Section 2 are briefly summarised in Fig.3.1, in terms of the unmodelled single term which can not be detected by the corresponding test. The relationships between these tests and the $Z_t(\hat{\theta})$ and $f(.)$ can be easily established. For example, $\Phi_{\xi^{2'}\xi^{2'}}(\tau)$ corresponds to the choice of $Z_t(\hat{\theta})$ equal to

$$Z_t(\hat{\theta}) = [\xi^{2'}_{t-1}, \xi^{2'}_{t-2}, ..., \xi^{2'}_{t-d}]^T \tag{3.58}$$

where $'$ represents mean deleted, and $f(.)$ is defined by eqn.(2.10). It is seen that in the worst possible case when all odd moments are zero, any single even term can be detected by $\Phi_{\xi\xi^2}(\tau)$ and by $\Phi_{\xi^{2'}\xi^{2'}}(\tau)$, which All the correlation functions except $\Phi_{\xi^{2'}\xi^{2'}}(\tau)$ in Fig.3.1 can only detect some odd terms out of all possible kinds of single odd terms in the residuals. There are of course many other possible choices for $Z_t(\hat{\theta})$, a general and convenient choice is the monomials of the elements of the vector $y^t(\hat{\theta})$ for $Z_t(\hat{\theta})$ (Leontaritis and Billings 1987)

$$Z_t(\hat{\theta}) = [m_t, m_{t-2}, ..., m_{t-t_d}]^T \tag{3.59}$$

where t_d as defined before is the maximum delay, m_t is a monomial of elements of the vector y^t, such as $\xi_{t-1}\xi_{t-3}\varepsilon^2_{t-3}$ and $f(.)$ is defined by (2.9) or (2.10). These general monomial correlations tests involve measurements of moments or cumulants to a certain order and can be viewed as an extension of the tests based on the second to the fourth order correlations investigated above. Therefore some of the properties obtained for the tests based on the second to the fourth order correlations can be apllied to the the tests based on general monomial correlations. For example, it can be concluded that like $\Phi_{\xi^{2'}\xi^{2'}}(\tau)$, $\Phi_{\xi^{2'}\xi^{4'}}(\tau)$ will detect any possible terms. In a similar way assume that the p.d.f of the residuals is symmetric, then to detect a single odd term $\varepsilon^{i_1}_{t-t_1}\varepsilon^{i_2}_{t-t_2} \cdots \varepsilon^{i_m}_{t-t_m}\eta^{e}_{2n'}$ defined by eqn.(3.7), a sufficient choice of m_t for $Z_t(\hat{\theta})$ would be

$$m_t = \xi_{t-t_1}\xi_{t-t_2} \cdots \xi_{t-t_m} \tag{3.60}$$

with $f(.)$ defined by eqn.(2.9). Tests thus formed are generalisations to $\Phi_{\xi\xi\xi}(\tau_1,\tau_2)$, where τ_1,τ_2 take some particular values. The monomial form of the elements of $y^t(\hat{\theta})$ defined by eqn.(3.59) in the correlation tests can be further extended to include other functions, such as

$$m_t = \begin{cases} abs(\xi_t) \\ sin(\xi_t) \\ \cos(\xi_t) \\ \exp(\xi_t) \\ \exp(-\xi_t^2) \\ \mathrm{atan}(\xi_t) \\ \sinh(\xi_t) \\ \tanh(\xi_t) \end{cases} \tag{3.61}$$

It is seen that tests based on correlations of these functions are combinations of the monomial ones, and their properties are therefore decided by those of the corresponding main monomial correlations. For example,

$$\Phi_{\xi^{2'}\sin^{2'}(\xi)}(\tau) = \Phi_{\xi^{2'}\xi^{2'}}(\tau) + 1/9\ \Phi_{\xi^{2'}\xi^{6'}}(\tau) + \cdots \tag{3.62}$$

with properties similar to $\Phi_{\xi^{2'}\xi^{2'}}(\tau)$, so that this choice will be a useful alternative to $\Phi_{\xi^{2'}\xi^{2'}}(\tau)$.

If there are clues regarding the kind of terms which could be unmodelled in the residuals, from knowledge of the system under study perhaps, or certain types of terms are of special interest, a sufficient test can be designed according to the above results. For example, $\Phi_{\xi\xi\xi}(\tau_1,\tau_2)$, where τ_1,τ_2 take some particular values, can be employed to detect bilinear terms. However if nothing is known about the residuals, a necessary and sufficient choice is $\Phi_{\xi^{2'}\xi^{2'}}(\tau)$ which is the correlation of lowest possible order that can detect any possible single term in the residuals. It should be noted however that $\Phi_{\xi^{2'}\xi^{2'}}(\tau)$ and related tests are aggregated tests in the sense that they are based on the correlations of the squared residuals which is an aggregated measurement and on a few occasions the tests can present so small a value that this can not be distinguished from zero for some missing terms whose coefficients are very small. In these situations the other tests derived in this study should be used to complement $\Phi_{\xi^{2'}\xi^{2'}}(\tau)$. For example, for residuals generated by

$$\xi_t = 0.1\varepsilon_{t-1}\varepsilon_{t-2} + \varepsilon_t \tag{3.63}$$

the test $\Phi_{\xi\xi\xi}(\tau_1,\tau_2)$ has larger power than $\Phi_{\xi^{2'}\xi^{2'}}(\tau)$. As a result, it is advisable to use a combination of several forms of test consisting of $\Phi_{\xi^{2'}\xi^{2'}}(\tau)$, $\Phi_{\xi\xi^2}(\tau)$, and tests defined by eqn.(3.60) which include the test $\Phi_{\xi\xi}(\tau)$. $\Phi_{\xi^{2'}\xi^{2'}}(\tau)$ is employed for general checking, and if necessary $\Phi_{\xi\xi^2}(\tau)$, the test of lowest possible order that can detect all possible single even terms, is used to detect omitted even terms with small coefficients and a number of tests defined by eqn.(3.60) are used for small odd terms. For example, $\Phi_{\xi\xi}(\tau)$ should be used to detect linear terms and $\Phi_{\xi\xi\xi\xi}(\tau_1,\tau_2,\tau_3)$ for odd terms like

$\beta\varepsilon_{t-\tau_1}\varepsilon_{t-\tau_2}\varepsilon_{t-\tau_3}$, where the coefficient β may be small. Generally $\Phi_{\xi^{2'}\xi^{2'}}(\tau)$ alone is sufficient, but other specific tests can be added according to the particular requirements.

Fig. 3.1 Correlation based model validity test: A summary

Tests		Cases when the tests fail	
		Odd terms	**Even terms**
$\Phi_{\xi\xi}(\tau)$	P.d.f. is Asymmetrical	$\varepsilon_{t-t_1}\xi_{t-t_2}\cdots$, $t_1 < t_2$	None
	P.d.f. is Symmetrical	$\varepsilon^{p}_{t-t_1}\xi^{q}_{t-t_2}\ \cdots$, $t_1 < t_2$, p, q *odd*	$\varepsilon^{i_1}_{t-t_1}$, i_1 *even*
$\Phi_{\xi\xi^2}(\tau)$	P.d.f. is Asymmetrical	$\varepsilon_{t-t_1}\xi_{t-t_2}\cdots$, $t_1 < t_2$	None
	P.d.f. is Symmetrical	Any	None
$\Phi_{\xi\xi\xi}(\tau_1,\tau_2)$	P.d.f. is Asymmetrical	$\varepsilon_{t-t_1}\varepsilon_{t-t_2}\varepsilon_{t-t_3}\cdots$,	None
	P.d.f. is Symmetrical	$\varepsilon^{i_1}_{t-t_1}\varepsilon^{i_2}_{t-t_2}\varepsilon^{i_3}_{t-t_3}\cdots$; or $\varepsilon^{i_1}_{t-t_1}\cdots$, i_1, i_2, i_3 *odd*	None
$\Phi_{\xi^{2'}\xi^{2'}}(\tau)$	P.d.f. is Asymmetrical	None	None
	P.d.f. is Symmetrical	None	None

4. Simulation studies

The parameters of all models used in the simulation studies have been estimated using a prediction error method based on Newton's method with line search. The initial values of the parameters were obtained by using an orthogonal least square estimator. The prediction error method combined with an orthogonal least squares estimator is both accurate and fast for nonlinear time series modelling and will be reported in forthcoming publications.

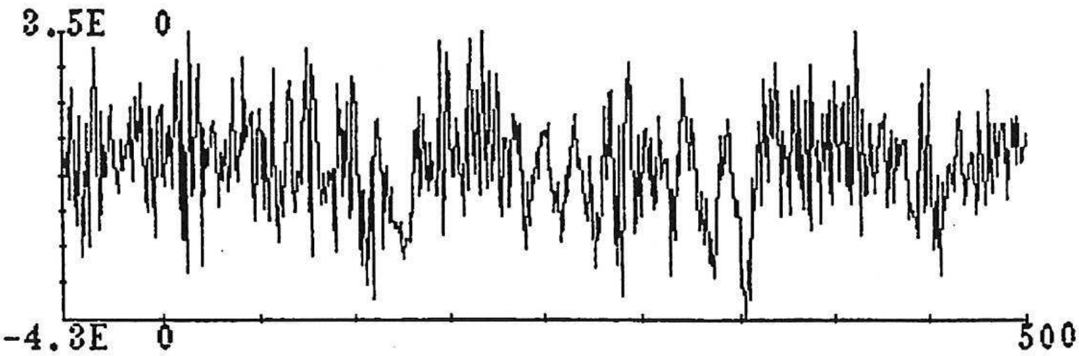
Example 1

A data sequence of 500 points was generated from the following bilinear time series model

$$x_t = 0.4x_{t-1} - 0.25x_{t-2} + 0.3x_{t-3} - 0.5x_{t-3}\varepsilon_{t-2} + \varepsilon_t \tag{4.1}$$

where ε_t is an i.i.d. Gaussian sequence with zero mean and unit variance. This set of data is illustrated in Fig. 4.1.

Fig. 4.1 Simulated data of Example 1



The following parameter estimates were obtained:

terms	estimates	stds
x_{t-1}	0.39742e+00	0.21892e-01
x_{t-2}	-0.28580e+00	0.23297e-01
x_{t-3}	0.30859e+00	0.29220e-01
$x_{t-3}\varepsilon_{t-2}$	-0.49137e+00	0.23160e-01

where stds stands for the standard deviations of the estimates, and the variance of residuals is 1.0323. It is seen that all the above estimates are unbiased, for the true values of the parameters are all within one or two standard deviations of the estimates. The adequacy of the fitted model is confirmed by various forms of validity tests shown in Figs.4.2.1 and 4.2.2, including the aggregated test $\Phi_{\xi^{2'}\xi^{2'}}(\tau)$, similar tests such as $\Phi_{\cos^{2'}(\xi)\cos^{2'}(\xi)}(\tau)$, as well as other forms like $\Phi_{\xi\xi\xi}(\tau_1,\tau_2)$ which are designed to efficiently detect some very small even terms.

If the superbilinear term $x_{t-3}\varepsilon_{t-2}$ is deliberately omitted from the estimated model structure the parameter estimates become

terms	estimates	stds
x_{t-1}	0.41477e+00	0.42336e-01
x_{t-2}	-0.26507e+00	0.44619e-01
x_{t-3}	0.31423e+00	0.42333e-01

where the variance of residuals is 1.5793. This error in the model structure should be reflected in the correlations of the residuals. As expected the results illustrated in Fig.4.3.1.a-d indicate that the bilinear term $x_{t-3}\varepsilon_{t-2}$ which is missing from the fitted model will not be detected by $\Phi_{\xi\xi}(\tau)$, $\Phi_{\xi\xi^2}(\tau)$ and other forms of tests which are extensions or generalisation of these, such as $\Phi_{\xi sin(\xi)}(\tau)$ and $\Phi_{\xi \sin^2(\xi)}(\tau)$. This missing term can be detected by $\Phi_{\xi^2\xi^2}(\tau)$, $\Phi_{\xi\xi\xi}(\tau_1,\tau_2)$, and other similar forms of tests shown in Fig. 4.3.1.e,f and 4.3.2a-f.

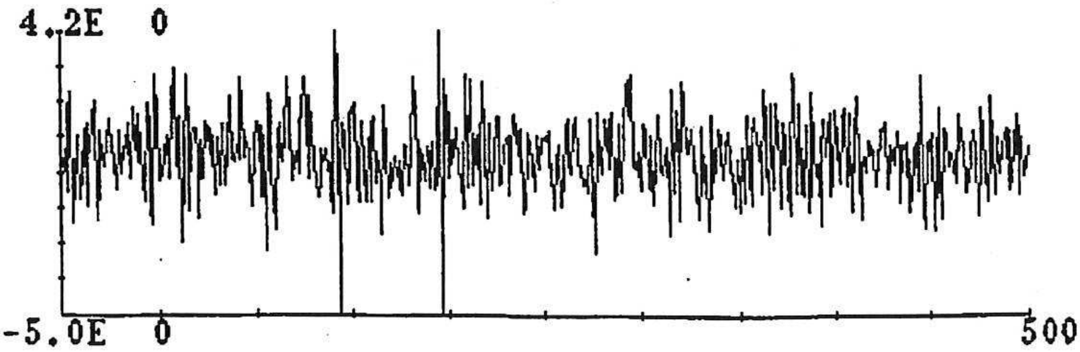
Example 2

A data sequence of 500 points illustrated in Fig. 4.4 was simulated using the non-linear MA model

$$x_t = 0.2\varepsilon_{t-1} - 0.36\varepsilon_{t-2} + 0.15\varepsilon_{t-1}\varepsilon_{t-2} + 0.25\varepsilon_{t-1}^2 + 0.18\varepsilon_{t-2}^2\varepsilon_{t-3} + \varepsilon_t \tag{4.2}$$

where ε_t is an i.i.d. Gaussian N(0,1).

Fig. 4.4 Simulated data of Example 2



The following parameter estimates were obtained:

terms	estimates	stds
ε_{t-1}	0.19306e+00	0.30563e-01
ε_{t-2}	-0.39001e+00	0.32384e-01
ε_{t-1}^2	0.23603e+00	0.15924e-01
$\varepsilon_{t-1}\varepsilon_{t-2}$	0.84242e-01	0.34385e-01
$\varepsilon_{t-2}^2\varepsilon_{t-3}$	0.13752e+00	0.19285e-01

where the variance of residuals is 1.0451. The independence of the residuals associated with this model is shown by the validity tests of various forms in Figs.4.5.1 and 4.5.2.

When the even term ε_{t-1}^2 is excluded from the fitted model structure, the following parameter estimates were obtained:

terms	estimates	stds
ε_{t-1}	0.90149e-01	0.37004e-01
ε_{t-2}	-0.35204e+00	0.45926e-01
$\varepsilon_{t-1}\varepsilon_{t-2}$	-0.47061e-01	0.23771e-01
$\varepsilon_{t-2}^2\varepsilon_{t-3}$	0.62660e-01	0.11845e-01

where the variance of residuals is 1.3061. As pointed out in the theoretical analysis, this inadequate model will not be detected by $\Phi_{\xi\xi}(\tau)$, $\Phi_{\xi\xi^3}(\tau)$, and $\Phi_{\xi sin(\xi)}(\tau)$ shown in Fig. 4.6.1.a,b,c; but is clearly indicated by the specific tests $\Phi_{\xi\xi}(\tau)$, $\Phi_{\xi\xi\xi}(\tau_1,\tau_2)$, and similar types of tests, such as $\Phi_{\xi cos(\xi)}(\tau)$ shown in Fig. 4.6.1.d,e,f. Alternatively the omitted term will be indicated by $\Phi_{\xi^{2'}\xi^{2'}}(\tau)$, $\Phi_{\xi^{2'}\xi^{4'}}(\tau)$, and other similar types of aggregated tests such as $\Phi_{\xi^{2'}abs'(\xi)}(\tau)$ and $\Phi_{\xi^{2'}\xi^{3'}}(\tau)$ in Fig 4.6.2. It is noted that $\Phi_{\xi^{2'}sin^{2'}(\xi)}(\tau)$ and $\Phi_{\xi^{2'}tanh'(\xi)}(\tau)$ illustrated in Fig 4.6.2 fail to detect this model deficiency, probably because these correlations are very small, though they are non zero theoretically.

Now consider the case when the odd term $\varepsilon_{t-2}^2\varepsilon_{t-3}$ is omitted from the model structure. The following parameter estimates were obtained:

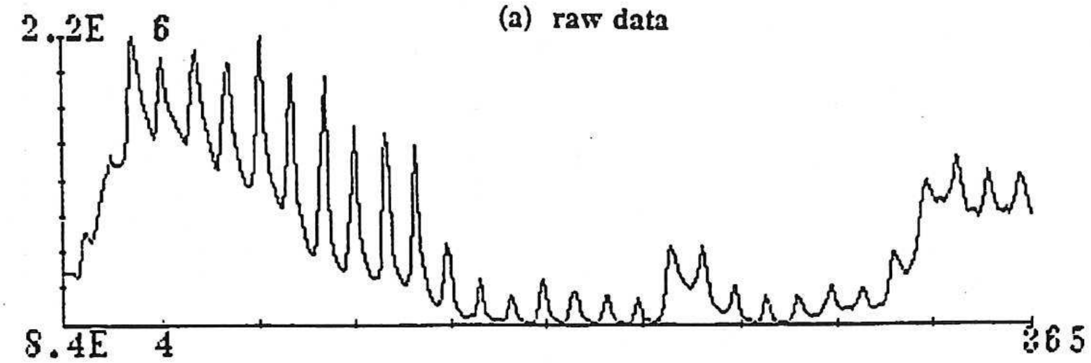
terms	estimates	stds
ε_{t-1}	0.20795e+00	0.39687e-01
ε_{t-2}	-0.44589e+00	0.35819e-01
$\varepsilon_{t-1}\varepsilon_{t-2}$	0.46160e-01	0.38555e-01
ε_{t-1}^2	-0.18971e+00	0.16552e-01

where the variance of residuals is 1.1382. The missing odd term will not be detected by $\Phi_{\xi\xi\xi}(\tau_1,\tau_2)$ and similar kinds of tests such as $\Phi_{\xi abs(\xi)}(\tau)$ illustrated in Fig.4.7.1.a-c. The omitted term will be distinguished from an independent sequence by $\Phi_{\xi\xi}(\tau)$, $\Phi_{\xi sin(\xi)}(\tau)$ and similar tests such as $\Phi_{\xi atan(\xi)}(\tau)$ as shown in Fig. 4.7.1.d-f; or aggregated tests $\Phi_{\xi^{2'}\xi^{2'}}(\tau)$, $\Phi_{\xi^{2'}\xi^{4'}}(\tau)$, and other related tests such as $\Phi_{\xi^{2'}cos^{2'}(\xi)}(\tau)$ as shown in Fig. 4.7.2.a-e. It is also noted that some aggregated tests exhibit small power probably because of the small correlation values such as $\Phi_{\xi^{2'}cos^{2'}(\xi)}(\tau)$ as shown in Fig. 4.7.2.f.

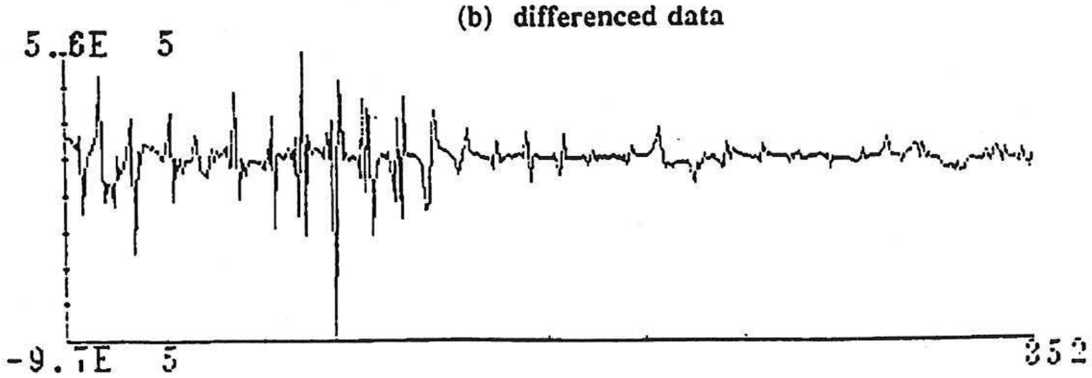
Example 3

Consider a sequence of 389 monthly unemployment figures for West Germany for the period January 1948 to May 1980 illustrated in Fig.4.8.a. Subba Rao and Gabr (1984) built models for the first 365 data and then forecast the remaining 24 values. They fitted a full linear model, a best subset linear model, and a best subset bilinear model to the differenced data, with difference operator $(1-B)(1-B^{12})$, where B is defined as the backward shift operator. The differenced data is plotted in Fig. 4.8.b.

Fig. 4.8 Unemployment figures in West Germany



(a) raw data



(b) differenced data

If the best subset linear and bilinear model structures of Subba Rao and Gabr

(1984) are employed on the first 365 data samples, the estimates were obtained in Table 4.1 using a prediction error routine:

Table 4.1 Subset linear model

terms	estimates	stds
x_{t-1}	-0.97011e-01	0.49505e-01
x_{t-2}	-0.13650e+00	0.45779e-01
x_{t-9}	0.76754e-01	0.42572e-01
x_{t-11}	0.30321e+00	0.43307e-01
x_{t-12}	-0.37413e+00	0.46728e-01

where the variance of residuals is 0.81831e+10. These estimates were identical with Subba Rao and Gabr's. Applying a combination of validity tests of both aggregated and specific forms yields the results shown in Figs.4.9.1 and 4.9.2, and clearly indicate that the above best subset linear model is an inadequate fit to the differenced data.

Estimating a model with the same structure as Subba Rao and Gabr's subset bilinear model provided the results in Table 4.2.

Table 4.2 Subset bilinear model

terms	estimates	stds
const.	-0.45936e+04	0.35902e+04
x_{t-1}	0.87436e-01	0.55230e-01
x_{t-2}	-0.12609e+00	0.14883e-01
x_{t-9}	0.42629e-01	0.16984e-01
x_{t-11}	0.25571e+00	0.28816e-01
x_{t-12}	-0.50658e+00	0.32749e-01
$x_{t-1}\varepsilon_{t-10}$	0.13150e-04	0.50590e-06
$x_{t-2}\varepsilon_{t-5}$	-0.12787e-05	0.17777e-06
$x_{t-5}\varepsilon_{t-4}$	-0.37868e-06	0.10722e-06
$x_{t-11}\varepsilon_{t-7}$	0.19025e-05	0.43059e-06
$x_{t-12}\varepsilon_{t-4}$	0.15129e-05	0.20153e-06
$x_{t-12}\varepsilon_{t-2}$	-0.22638e-05	0.32680e-06
$x_{t-4}\varepsilon_{t-10}$	-0.95078e-06	0.18960e-06
$x_{t-10}\varepsilon_{t-8}$	-0.19498e-05	0.27571e-06
$x_{t-1}\varepsilon_{t-9}$	0.27147e-05	0.72953e-06

where the variance of the residuals of 0.36665e+10 was identical with Subba Rao and Gabr's estimate. However Figs.4.10.1 and 4.10.2 clearly illustrate that this model is also not adequate as shown by a combination of validity tests of various forms. It may therefore be possible to fit a better model of the West German unemployment data and this will be addressed in subsequent publications.

5. Conclusions

Model validity tests for nonlinear time series or signal processing applications based on general correlation functions have been studied. The widely-used test for linear systems based on the second order correlations of the residuals cannot distinguish between all nonlinear terms and an independent sequence. Assuming the worst possible combinations of conditions in which all odd moments are zero and with all possible unmodelled terms in the residual, it has been shown that the traditional second order correlation function $\Phi_{\xi\xi}(\tau)$ and the higher order correlations $\Phi_{\xi\xi\xi}(\tau_1,\tau_2)$, $\Phi_{\xi\xi^2}(\tau)$ can only detect a subset of the missing terms, whereas the autocorrelation function of the squared residuals $\Phi_{\xi^{2'}\xi^{2'}}(\tau)$, and tests based on aggregated measurements of the residuals can detect all possible missing terms. These results apply to validity tests based on correlations using general monomials of the vector y^t as entries and other periodic functions which are combinations of the monomial case. To try to prove that every possible term is detected is algebraically very tedious however it is suggested that a combination of tests be used consisting of an aggregated test $\Phi_{\xi^{2'}\xi^{2'}}(\tau)$ and some specific tests including $\Phi_{\xi\xi^2}(\tau)$ to detect small even terms, and the general correlation test defined by eqn.(3.59) to detect some possibly small odd terms. The results have been illustrated using simulated and real data sequences and the use of a combination of correlation tests in nonlinear time series validation was found to be both efficient and practical.

References

[1] Billings, S.A., and W.S.F. Voon (1986). Correlation based model validity tests for non-linear models. *Int. J. Control,* Vol.44, No.1, pp235-244.

[2] Bohlin, T.(1978). Maximum-power validation of models without higher-order fitting. *Automatica,* Vol.14, pp137-146.

[3] Box, G.E.P., and G.M. Jenkins (1976). *Time Series Analysis: Forecasting and Control.* Holden-Day Inc., San Francisco.

[4] Box, G.E.P., and D.A. Pierce (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *J. Amer. Statist. Ass.* Vol.65, No.332, pp1509-1526.

[5] Brilliant, M.B. (1958). Theory Continuous nonlinear systems. *Technical report,* MIT Research Lab. of Electromics, No.355.

[6] Chen, S., and S.A. Billings (1989a). Representations of non-linear systems: the NARMAX model. *Int. J. Control,* Vol.50, No.6, pp2151-2171.

[7] Chen, S., and S.A. Billings (1989b). Modelling and analysis of non-linear time series. *Int. J. Control,* Vol.50, No.6, pp2151-2171.

[8] Davies, N., C.M. Triggs, and P. Newbold (1977). Significance of the Box-Pierce Portmanteau statistics in finite samples.. *Biometrika,* Vol.64, pp517-522.

[9] George D.A. (1959). Continuous nonlinear systems. *Technical report,* MIT Research Lab. of Electromics, No.355.

[10] Granger, C.W.J., and A.P. Andersen (1978). *An Introduction to Bilinear Time Series Models.* Vandenhoeck and Ruprecht: Göttingen.

[11] Kumar, K. (1986). On the identification of some bilinear time series models. *J. Time Series Analysis,* Vol.7, No.2, pp117-122.

[12] Lawrance, A.J., and P.A.W. Lewis (1985). Modelling autoregressive time series in exponential variables. *J. R. Statist. Soc. B,* Vol.47, No.2, pp165-202.

[13] Leontaritis, I.J., and S.A Billings (1987). Model selection and validation methods for non-linear systems. *Int. J. Control,* Vol.45, No.1, pp311-341.

[14] Ljung, G.M., and G.E.P. Box (1978). on a measure of lack of fit in time series models. *Biometrika,* Vol.65, No.2, pp297-303.

[15] Ljung, L. (1987). *System Identification: Theory for the User.* Prentice-Hall, Englewood Cliffs.

[16] Mcleod, A.I., and W.K. Li (1983). Diagnostic checking ARMA time series models using squared-residual autocorrelations. *J. Time Series Analysis,* Vol.4, No.4, pp269-273.

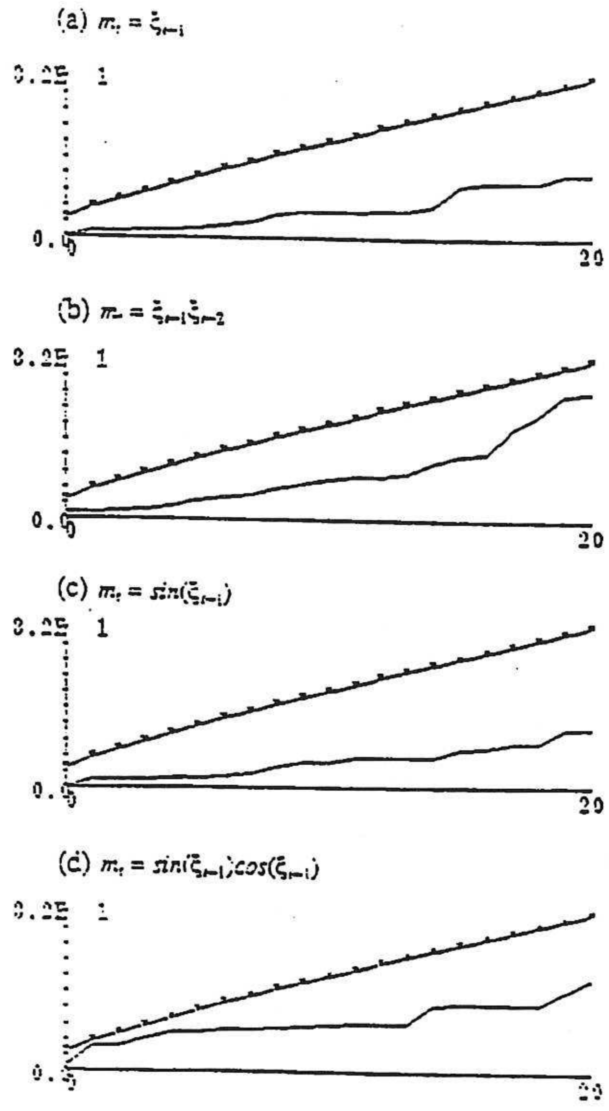
[17] Ozaki, T. (1985). Non-linear time series models and dynamical systems. In: Hannan, E.J., P.R. Krishnaiah and M.M. Rao (Eds), *Handbook of Statistics 5, Time Series in the Time Domain,* pp25-83, North-Holland, Amstredam.

[18] Priestley, M.B. (1980). State-dependent models: a general approach to non-linear time series analysis. *J. Time Series Analysis,* Vol.1, No.1, pp47-71.

[19] Söderström, T., and P. Stoica (1988). On covariance function tests used in system identification. *Report 7308,* Department of Automatic Control, Lund Institute of Technology, Lund, Sweden.

[20] Subba Rao, T., and M.M. Gabr (1984). *An Introduction to Bispectral Analysis and Bilinear Time Series Models*. Lecture Notes in Statistics, Springer-Verlag, New York.

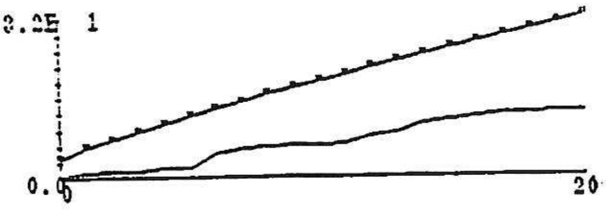
Fig. 4.2.1 Chi-squared tests for Example 1 with $f(\xi_t) = \xi_t$



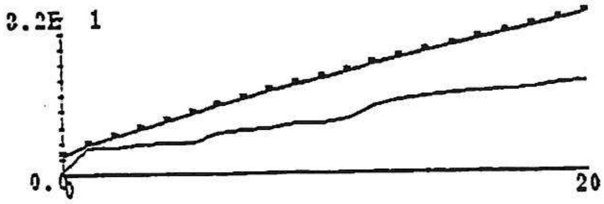
--*-*-* = 95% confidence limit

Fig. 4.2.2 Chi-squared tests for Example 1 with $f(\xi_t) = \xi_t^2$

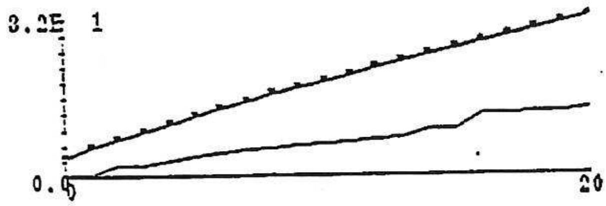
(a) $m_t = \xi_{t-1}\xi_{t-1}$



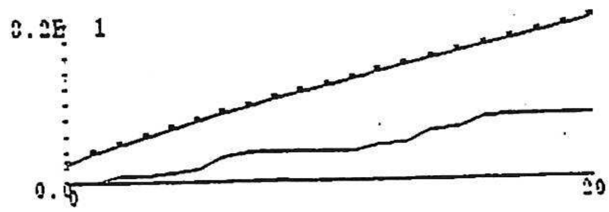
(b) $m_t = \xi_{t-1}\xi_{t-1}\xi_{t-1}\xi_{t-1}$



(c) $m_t = cos(\xi_{t-1})cos(\xi_{t-1})$



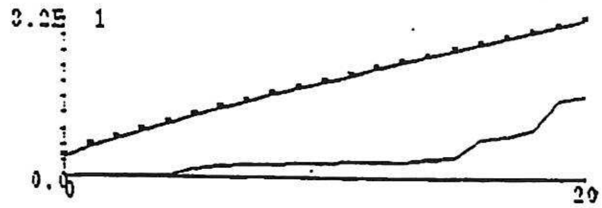
(d) $m_t = atan(\xi_{t-1})atan(\xi_{t-1})$



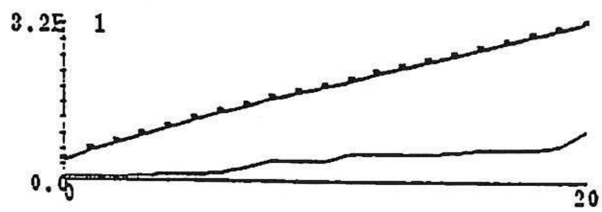
*_*_*_*_* = 95% confidence limit

Fig. 4.3.1 Chi-squared tests for Example 1 with $f(\xi_t) = \xi_t$

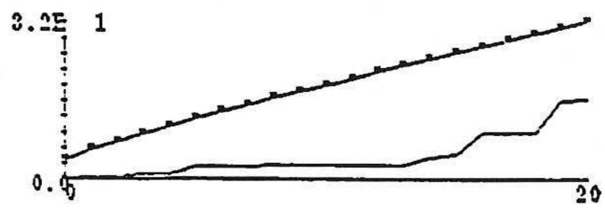
(a) $m_t = \xi_{t-i}$



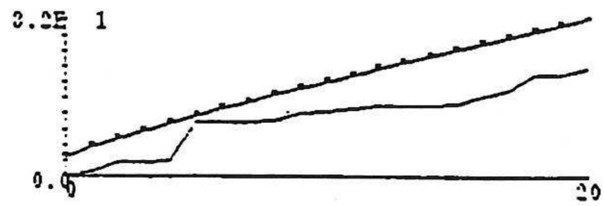
(b) $m_t = \xi_{t-1}\xi_{t-i}$



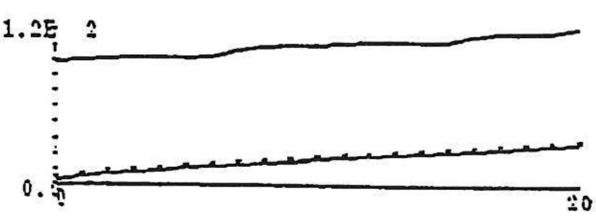
(c) $m_t = sin(\xi_{t-i})$



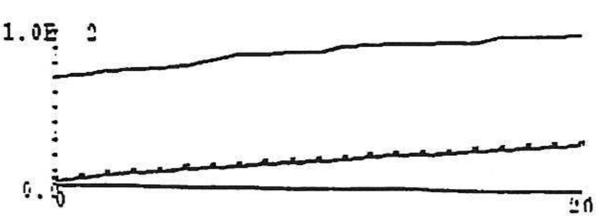
(d) $m_t = sin(\xi_{t-1})sin(\xi_{t-1})$



(e) $m_t = \xi_{t-3}\xi_{t-2}$



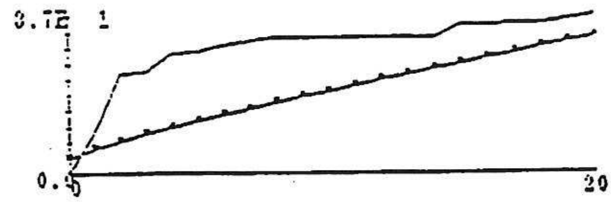
(f) $m_t = sin(\xi_{t-3})sin(\xi_{t-2})$



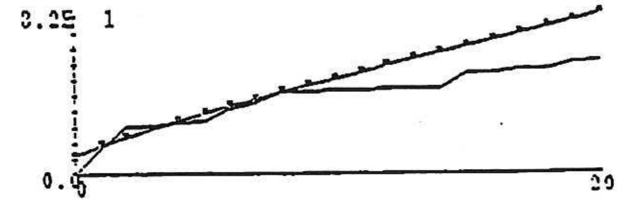
--*-*-* = 95% confidence limit

Fig. 4.3.2 Chi-squared tests for Example 1 with $f(\xi_t) = \xi_t^2$

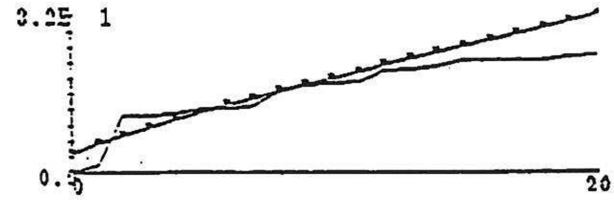
(a) $m_t = \xi_{t-1}\xi_{t-1}$



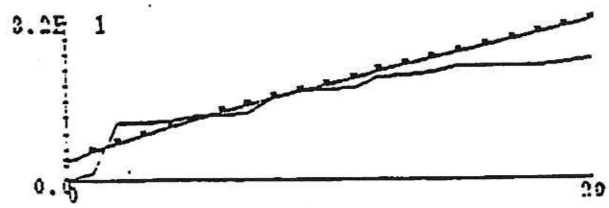
(b) $m_t = \xi_{t-1}\xi_{t-1}\xi_{t-1}\xi_{t-1}$



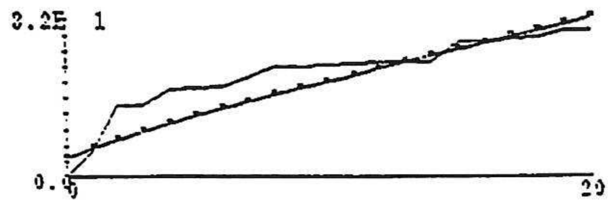
(c) $m_t = sin(\xi_{t-1})sin(\xi_{t-1})$



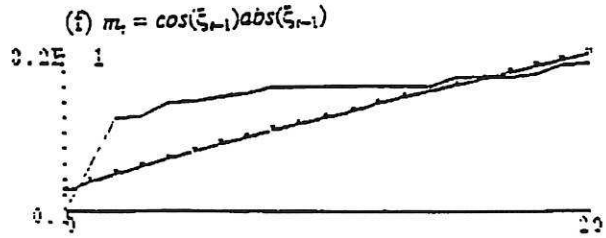
(d) $m_t = cos(\xi_{t-1})cos(\xi_{t-1})$



(e) $m_t = \xi_{t-1}x_{t-1}\xi_{t-1}x_{t-1}$



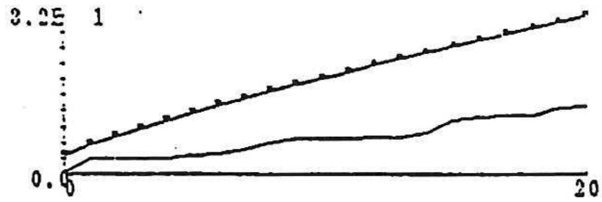
(f) $m_t = cos(\xi_{t-1})abs(\xi_{t-1})$



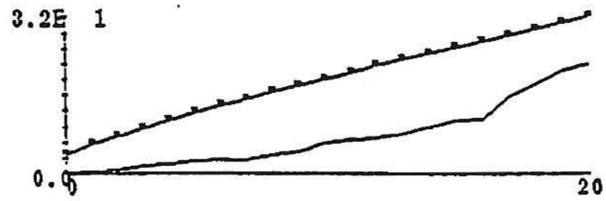
-*-*-*-*-* = 95% confidence limit

Fig. 4.5.1 Chi-squared tests for Example 2 with $f(\xi_t) = \xi_t$

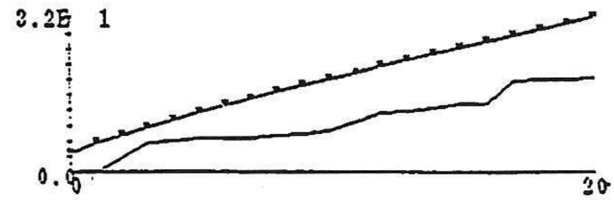
(a) $m_t = \xi_{t-1}$



(b) $m_t = \xi_{t-1}\xi_{t-2}$



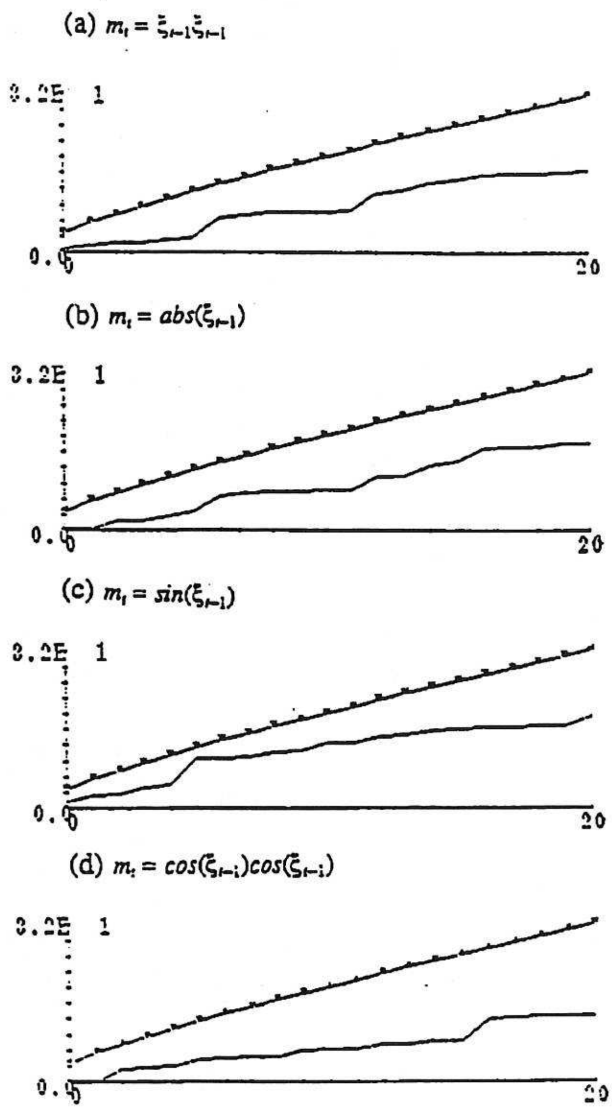
(c) $m_t = \xi_{t-1} \sin(\xi_{t-1}) x_{t-5}$



(d) $m_t = \exp(\xi_{t-1})$

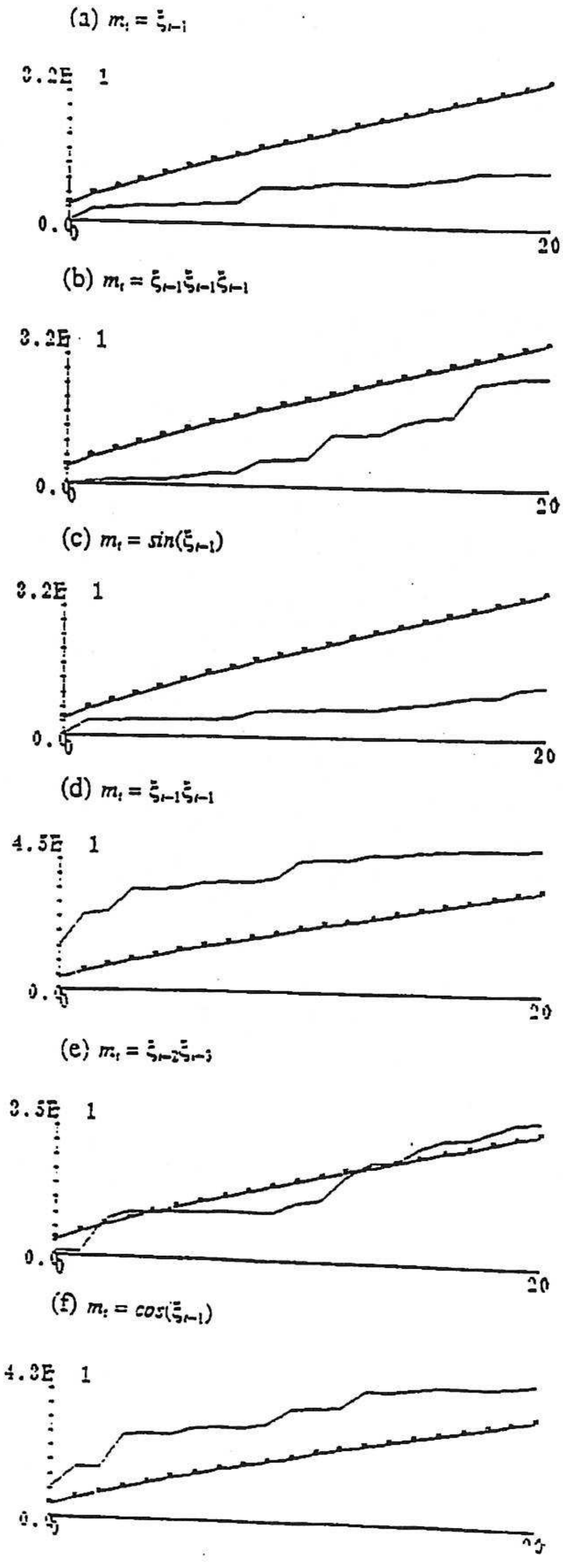
*_*_*_*_* = 95% confidence limit

Fig. 4.5.2 Chi-squared tests for Example 2 with $f(\xi_t) = \xi_t^2$



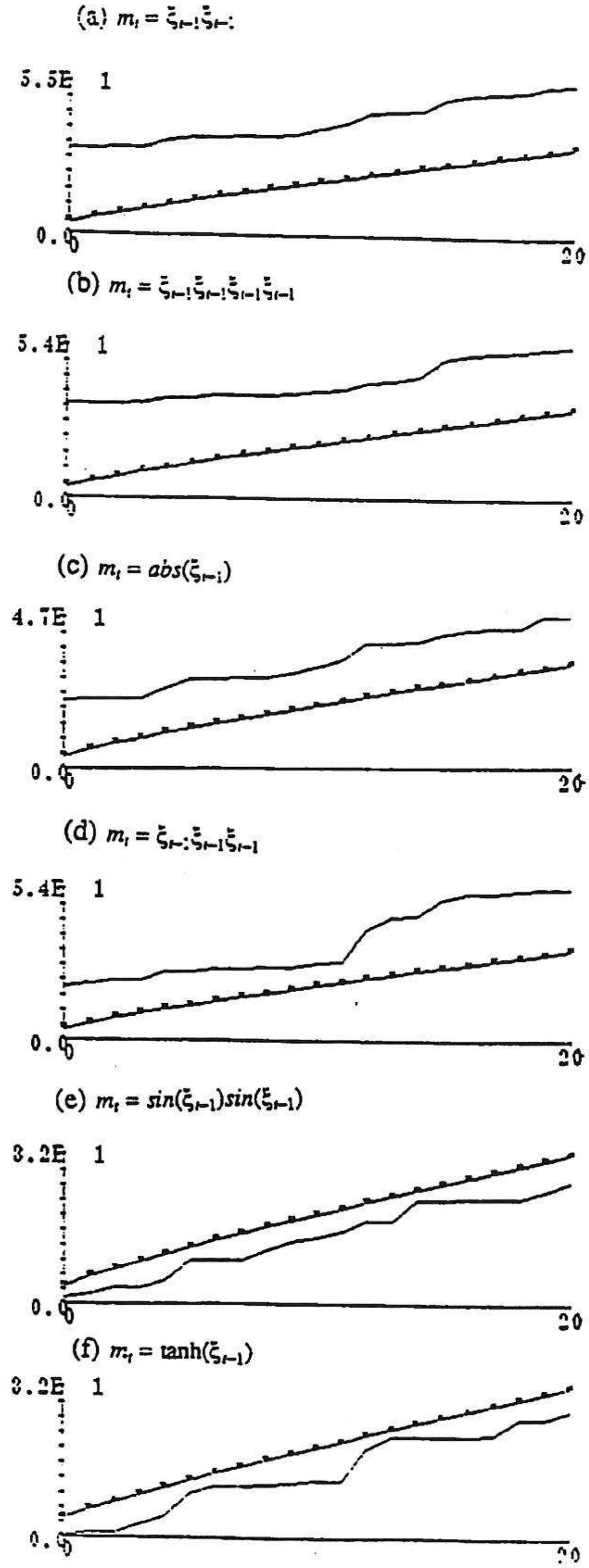
*_*_*_*_* = 95% confidence limit

Fig. 4.6.1 Chi-squared tests for Example 2 with $f(\xi_t) = \xi_t$



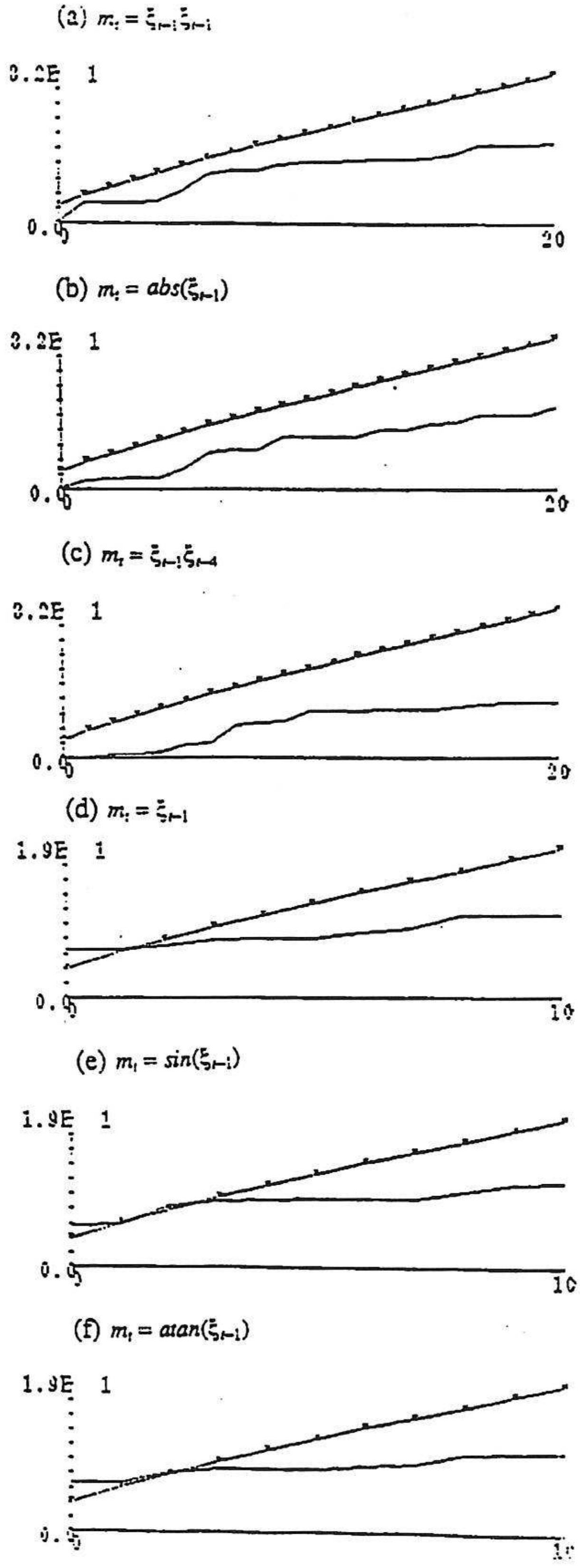
--*-*-* = 95% confidence limit

Fig. 4.6.2 Chi-squared tests for Example 2 with $f(\xi_t) = \xi_t^2$



--*-*-* = 95% confidence limit

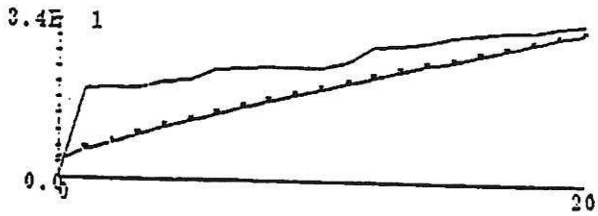
Fig. 4.7.1 Chi-squared tests for Example 2 with $f(\xi_t) = \xi_t$



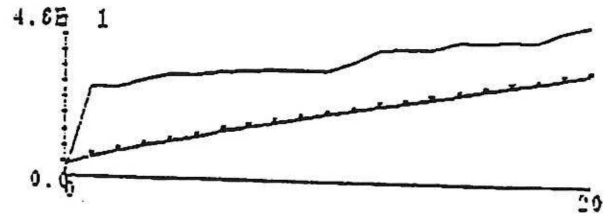
--*-*-* = 95% confidence limit

Fig. 4.7.2 Chi-squared tests for Example 2 with $f(\xi_t) = \xi_t^2$

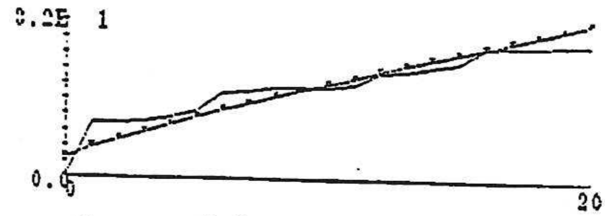
(a) $m_t = \xi_{t-1}\xi_{t-1}$



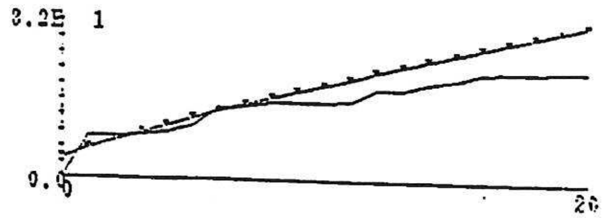
(b) $m_t = \xi_{t-1}\xi_{t-1}\xi_{t-1}\xi_{t-1}$



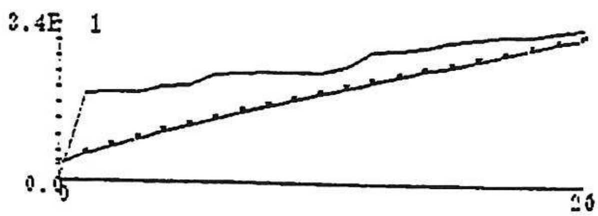
(c) $m_t = abs(\xi_{t-1})$.



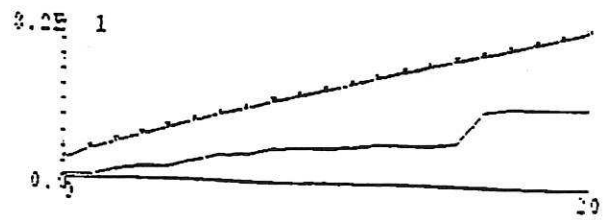
(d) $m_t = cos(\xi_{t-1})$



(e) $m_t = abs(\xi_{t-1})abs(\xi_{t-1})$



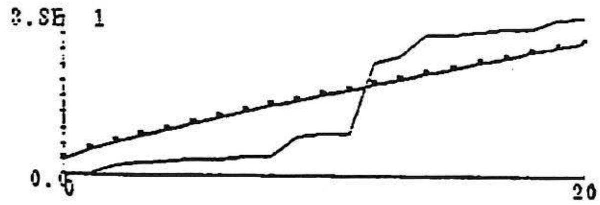
(f) $m_t = \cos(\xi_{t-1})\cos(\xi_{t-1})$



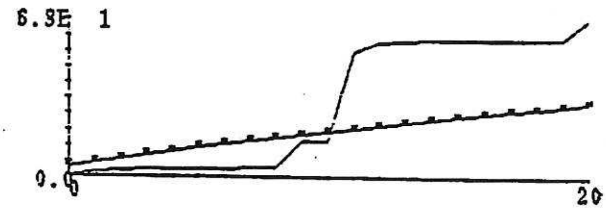
--*-*-* = 95% confidence limit

Fig. 4.9.1 Chi-squared tests for Example 3 with $f(\xi_t) = \xi_t$

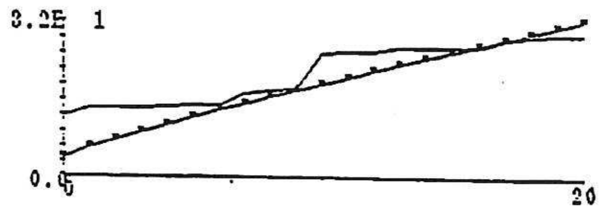
(a) $m_t = \xi_{t-1}$



(b) $m_t = \xi_{t-1}\xi_{t-1}$



(c) $m_t = \xi_{t-1}\xi_{t-3}$

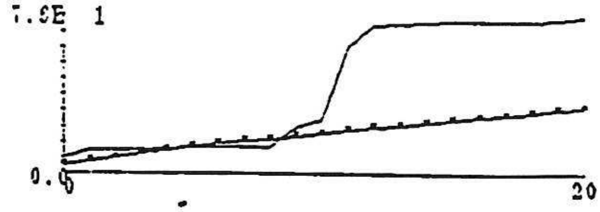


(d) $m_t = sin(\xi_{t-1})$

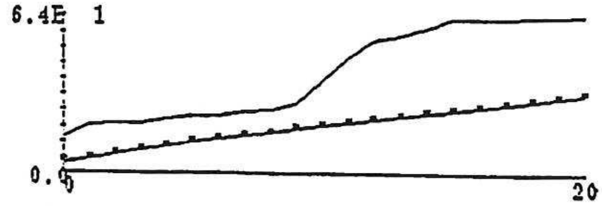
--*-*-* = 95% confidence limit

Fig. 4.9.2 Chi-squared tests for Example 3 with $f(\xi_t) = \xi_t^2$

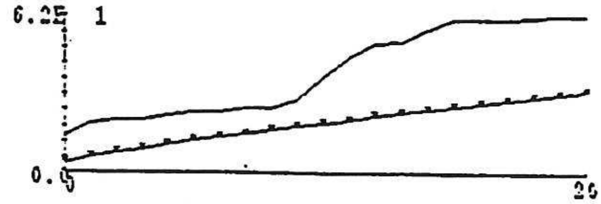
(a) $m_t = \xi_{t-1}\xi_{t-1}$



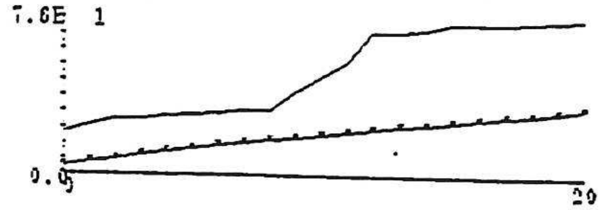
(b) $m_t = sin(\xi_{t-1})sin(\xi_{t-1})$



(c) $m_t = cos(\xi_{t-1})cos(\xi_{t-1})$

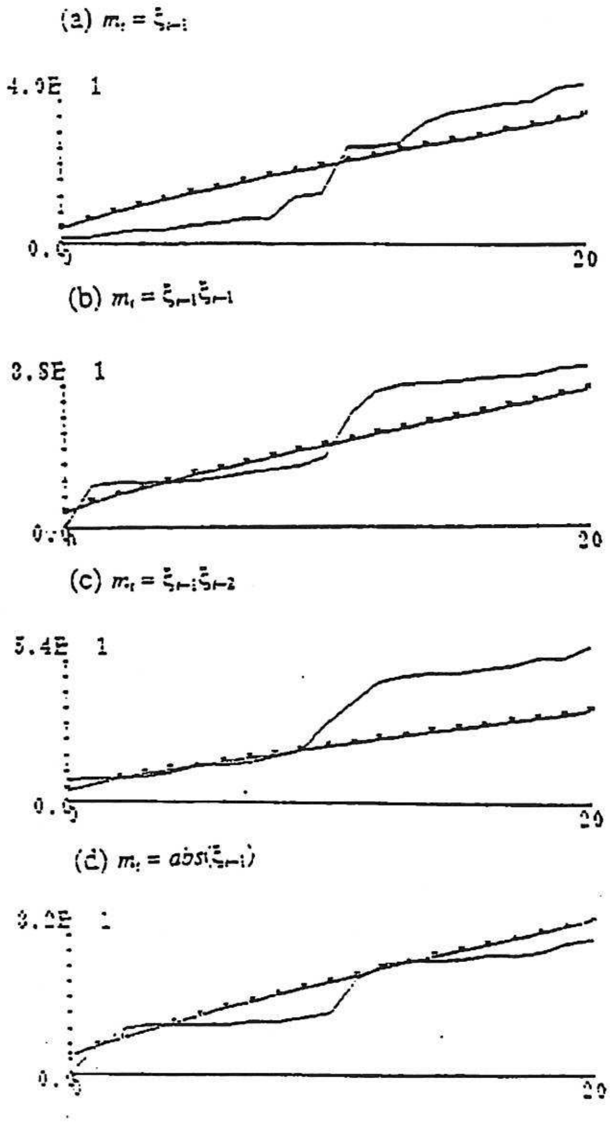


(d) $m_t = atan(\xi_{t-1})atan(\xi_{t-1})$



--*-*-* = 95% confidence limit

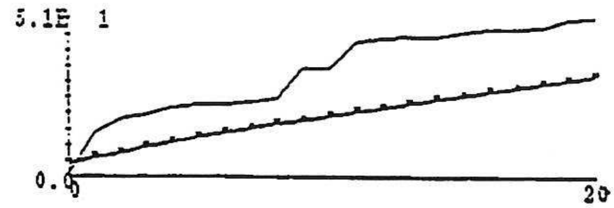
Fig. 4.10.1 Chi-squared tests for Example 3 with $f(\xi_t) = \xi_t$



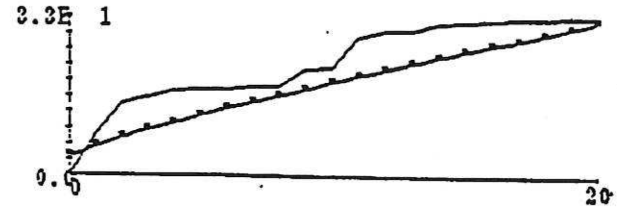
*_*_*_*_* = 95% confidence limit

Fig. 4.10.2 Chi-squared tests for Example 3 with $f(\xi_t) = \xi_t^3$

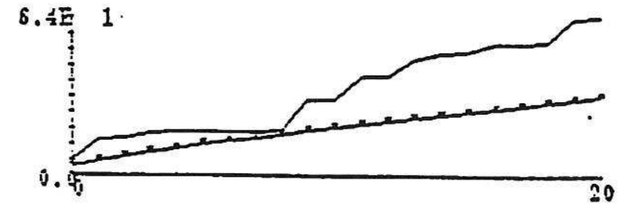
(a) $m_t = \xi_{t-1}\xi_{t-1}$



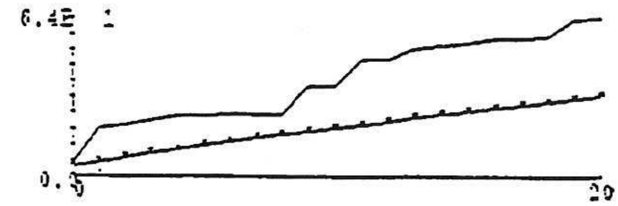
(b) $m_t = \xi_{t-1}\xi_{t-1}\xi_{t-1}\xi_{t-1}$



(c) $m_t = cos(\xi_{t-1})cos(\xi_{t-1})$



(d) $m_t = tanh(\xi_{t-1})tanh(\xi_{t-1})$



--*-*-* = 95% confidence limit