Biometrika (1985), **72**, 1, *pp.* 39–44
Printed in Great Britain

A Tukey nonadditivity-type test for time series nonlinearity

By DANIEL MacRAE KEENAN
Division of Applied Mathematics, Brown University, Providence, Rhode Island 02912, U.S.A.

Summary

The most general form of a nonlinear strictly stationary process is that referred to as a Volterra expansion; this is to a linear process what a polynomial is to a linear function. Because of this similarity, an analogue of Tukey's one degree of freedom for non-additivity test is constructed as a test for linearity versus a second-order Volterra expansion.

Some key words: Nonlinear time series; Tukey's nonadditivity test; Volterra expansion.

1. Introduction

The past few years have seen a revived interest in the study of nonlinear stationary time series. The most general form of a nonlinear stationary process is that referred to as a Volterra expansion. This is to a linear process what a higher order polynomial is to a linear function. The so-called bilinear model is a special case which is quite broad and is defined by a small number of parameters. Volterra expansions involve more than second-order theory and require higher-order cumulant spectra.

Various tests for linearity have been proposed although none is totally satisfactory. Because of the similarity of Volterra expansions to polynomials, an analogue of Tukey's (1949) one degree of freedom for nonadditivity test seems quite reasonable as a diagnostic for linearity versus a second-order Volterra expansion. Such a test would be time domain based and computationally less complex than the frequency domain based alternatives, directly generalizable to higher order than second and possibly suggestive of a power transformation toward linearity.

2. Background

The most general form of nonlinear, stationary time series models is that referred to as Volterra expansions (Wiener, 1958, lecture 10; Brillinger, 1970). These models are

$$Y_t = \mu + \sum_{u=-\infty}^{\infty} a_u \varepsilon_{t-u} + \sum_{u,v=-\infty}^{\infty} a_{uv} \varepsilon_{t-u} \varepsilon_{t-v} + \sum_{u,v,w=-\infty}^{\infty} a_{uvw} \varepsilon_{t-u} \varepsilon_{t-v} \varepsilon_{t-w} + \ldots, \qquad (2{\cdot}1)$$

where $\{\varepsilon_t, -\infty < t < \infty\}$ is a strictly stationary process, for our purposes assumed to be a mean zero, independent and identically distributed sequence. The terms of the expansion are usually referred to as the linear, quadratic, cubic, etc., respectively. The major drawback of such models is the multitude of parameters. One alternative is to assume $a_u, a_{uv}, \ldots$ are functions of a small number of parameters, a generalization of what is ordinarily done in studying arbitrary linear processes. The kth-order ($k \geqslant 2$) cumulant spectrum is defined as

$$f(\lambda_1, \ldots, \lambda_{k-1}) = (2\pi)^{-k+1} \sum_{v_1, \ldots, v_{k-1} = -\infty}^{\infty} c(v_1, \ldots, v_{k-1}) \exp\left(-i \sum_{j=1}^{k-1} \lambda_j v_j\right),$$

where $c(v_1, \ldots, v_{k-1})$ is the kth joint cumulant of $\{Y_0, Y_{v_1}, \ldots, Y_{v_{k-1}}\}$.

Various tests for linearity have been proposed. They have primarily been spectral approaches based upon the fact that if the process is linear then the third-order cumulant spectrum has the representation

$$f(\alpha, \beta) = \mu_3 f(\alpha) f(\beta) \bar{f}(\alpha+\beta), \tag{2·2}$$

where $\bar{f}$ denotes the complex conjugate of f and μ_3 is the third moment of the noise process. Subba Rao & Gabr (1980) have proposed a statistic based on this in which estimates of the second- and third-order spectra are obtained, the ratios are taken, and because of (2·2) the constancy of the statistic is used as a measure of linearity. The actual test statistic is based upon the complex-valued analogue of Hotelling's T^2 test of the mean vector lying on the equiangular line. Subba Rao & Gabr use the asymptotic complex-normality of the bispectrum in a certain triangular region, Δ, of $[0, 2\pi]^2$ (Van Ness, 1966). They, however, use as their estimated covariance matrix the usual sample second moment matrix of multivariate analysis, treating the bispectral estimates as the data, rather than using the known asymptotic covariance matrix of the bispectral estimates. Hinich (1982) improved on their test by using the known asymptotic covariance matrix of the bispectral estimates and also proposed a test based on the interquartile range of square modulus of the sample bispectrum over Δ. Using the autocorrelation function from the squared fitted residuals, McLeod & Li (1983) considered a portmanteau-type goodness-of-fit statistic for diagnosing autoregressive-moving average models.

3. A TEST OF LINEARITY

A test of linearity is equivalent to a test of no multiplicative terms in (2·1). There is a striking resemblance of this to the framework of Tukey's one degree of freedom test for nonadditivity. Tukey's test in a regression setting is based upon the Fourier expansion of the residual. If the mean function is linear then the Fourier coefficients with respect to an orthogonal polynomial expansion are zero from the second term on out; Tukey uses a particular second-degree orthogonal polynomial representation, one based upon the square of the linear fit, with its linear fit removed.

An analogous framework can be established for diagnosing time series nonlinearity; this test is also based upon a Fourier expansion of the residual. The regression setting with a quadratic mean function is replaced by a standard Hilbert space formulation of prediction for a quadratic Volterra process, $\{Y_s\}$. If the process is nonlinear and does not contain a variable in a nonlinear term without also having that variable in a linear term, the nonlinearity will be distributionally reflected in the diagnostics. The difficulty with this formulation is that necessary projections into subspaces involving the entire past cannot be calculated from a partial realization; if they could be calculated and the process were linear and Gaussian, then by the same arguments as Scheffé (1959, Exercise 4·19) and Seber (1977, pp. 269–70), the same exact distributional results as in Tukey's nonadditivity test are available in the present context. Since we only observe a partial realization, $(Y_1, \ldots, Y_n)$, for large n and a large, fixed M, the various projections can be well approximated by regressions on the previous MY's. The steps of the procedure are as follows.

(i) Regress Y_s on $\{1, Y_{s-1}, \ldots, Y_{s-M}\}$ and calculate the fitted values $\{\hat{Y}_s\}$ and the residuals, $\{\hat{e}_s\}$, for $s = M+1, \ldots, n$, and the residual sum of squares, $\langle \hat{e}, \hat{e} \rangle = \Sigma \hat{e}_s^2$.

(ii) Regress $\hat{Y}_s^2$ on $\{1, Y_{s-1}, \ldots, Y_{s-M}\}$ and calculate the residuals $\{\hat{\xi}_s\}$, for

$$s = M+1, \ldots, n.$$

(iii) Regress $\hat{e} = (\hat{e}_{M+1}, \ldots, \hat{e}_n)$ on $\hat{\xi} = (\hat{\xi}_{M+1}, \ldots, \hat{\xi}_n)$ and obtain $\hat{\eta}$ and $\hat{F}$ via

$$\hat{\eta} = \hat{\eta}_0 \left(\sum_{t=M+1}^{n} \hat{\xi}_t^2 \right)^{\frac{1}{2}}, \tag{3·1}$$

where $\hat{\eta}_0$ is the regression coefficient, and

$$\hat{F}_{1,n-2M-2} = \frac{\hat{\eta}^2(n-2M-2)}{\langle \hat{e}, \hat{e} \rangle - \hat{\eta}^2}, \tag{3·2}$$

where the degrees of freedom associated with $\langle \hat{e}, \hat{e} \rangle$ are $(n-M)-M-1$. Since estimation is often done via approximate least squares in both the autoregressive and moving average cases, the test is easily implementable into standard estimation procedures.

A possible 4th step is to calculate a potential power transformation, $(Y_n + c)^p$, towards linearity

$$p = 1 - (\bar{Y} + c)\,\hat{\eta}_0,$$

where c is a constant. The motivation for the power transformation in this setting is roughly the same as that in the regression setting (Snedecor & Cochran, 1967, p. 332): if $Z_n = (Y_n + c)^p$ is an autoregressive process of order q

$$\sum_{i=1}^{q} a_i (Z_{n-i} - \mu_z) + \delta_n,$$

then inverting, expanding in a Taylor series to the second degree and using first-order approximations, we have

$$e_s \simeq \mu_y^{-1}(1-p) \sum_{i,j=1}^{q} a_i a_j (Y_{s-i} - \mu_Y)(Y_{s-j} - \mu_Y),$$

where $\mu_Y^{-1}(1-p)$ is the regression coefficient of e_s on ξ_s for $s = M+1, \ldots, n$, which is η, by definition. The degree of appropriateness for the above power transformation argument, in general, rests on the relative deviations from the mean, $\mu_z^{-1}(Z_i - \mu_z)$, being small. This is quite reasonable in regression but not necessarily so in the time series setting.

The following is partial justification for the above procedure, steps (i), (ii) and (iii).

LEMMA 3·1. *If Y_t is a stationary invertible autoregressive process of order M with representation*

$$(Y_t - \mu) = \sum_{j=1}^{M} a_j (Y_{t-j} - \mu) + e_t,$$

where the e_t are independent and identically distributed with mean zero, variance σ_e^2, and finite fourth cumulant, then $\hat{\eta}$ given by expression (3·1), *is asymptotically normal* $(0, \sigma_e^2)$.

Proof. Consider the partial realization, $Y_1, \ldots, Y_n$, and define X_t, θ, A_n and V_n by

$$X_t = (1, Y_{t-1}, \ldots, Y_{t-M}), \quad \theta' = (\theta_0, a_1, \ldots, a_M), \quad \theta_0 = \left(\sum_{j=1}^{M} a_j - 1\right)\mu,$$
$$A_n = (n-M)^{-1} \sum_{t=M+1}^{n} X_t' X_t, \quad V_n = (n-M)^{-1} \sum_{t=M+1}^{n} X_t' Y_t, \tag{3·3}$$

where $'$ denotes a matrix transpose. The least squares estimate of θ, the fitted values, and the residuals are

$$\hat{\theta} = A_n^{-1} V_n, \quad \hat{Y}_t = X_t \hat{\theta}, \quad \hat{e}_t = Y_t - \hat{Y}_t = X_t(\theta - \hat{\theta}) + e_t \quad (t = M+1, \ldots, n). \tag{3·4}$$

Define $\hat{\xi}_t$ and q_n by

$$q_n = (n-M)^{-1} \sum_{t=M+1}^{n} X_t' \hat{Y}_t^2 = (n-M)^{-1} \sum_{t=M+1}^{n} X' \hat{\theta}' X_t' X_t \hat{\theta}, \quad \hat{\xi}_t = \hat{Y}_t^2 - X_t A_n^{-1} q_n. \tag{3·5}$$

Fuller (1976, Theorem 8·2·1) shows that $(\hat{\theta} - \theta)$ is $o_p(1)$ and $A_n^{-1} V_n - A^{-1} V$ is $o_p(1)$, where A and V are the nonrandom probability limits of A_n and V_n, respectively. Let q be the vector whose first component is

$$\theta_0^2 + 2\theta_0 \mu \sum_{j=1}^{M} a_j + \sum_{i,j=1}^{M} a_i a_j \gamma(i-j)$$

and whose kth component, for $2 \leqslant k \leqslant M+1$, is

$$\theta_0^2 + 2\theta_0 \sum_{j=1}^{M} a_j \gamma(j-k) + \sum_{i,j=1}^{M} a_i a_j m(i-k, j-k),$$

where $\gamma(\,.\,)$ and $m(\,.\,,.\,)$ are the covariances and third moment lagged function for the Y_n process. Because $(\hat{\theta} - \theta)$ is $o_p(1)$, $(q_n - q)$ is also $o_p(1)$. The numerator of $\hat{\eta}$ is

$$n^{-\frac{1}{2}} \sum_{t=M+1}^{n} \hat{e}_t \hat{\xi}_t \tag{3·6}$$

and substituting expressions (3·4) and (3·5) into (3·6), it follows that it is sufficient to show that

$$n^{-\frac{1}{2}} \sum_{t=M+1}^{n} e_t(\theta' X_t' X_t \theta - X_t A^{-1} q) = n^{\frac{1}{2}} \sum_{t=M+1}^{n} Z_t, \tag{3·7}$$

say, converges in distribution, since then expressions (3·6) and (3·7) will differ by $o_p(1)$. Because e_t is independent of $\{e_s, s \leqslant t-1\}$ and $\{Y_s, s \leqslant t-1\}$, the $\{Z_t\}$ forms a stationary and ergodic martingale difference process with constant variance $\operatorname{var}(Z_t) = \sigma_e^2 \delta$, where $\delta = E\{(\theta' X_t' X_t \theta - X_t A^{-1} q)^2\}$ and consequently using a martingale central limit theorem (Billingsley, 1961), expression (3·7) is asymptotically normal $(0, \sigma_e^2 \delta)$. The denominator of $\hat{\eta}$ is $(n^{-1} \Sigma \hat{\xi}_t^2)^{\frac{1}{2}}$, where the sum is over $t = M+1, \ldots, n$, and the term within parentheses will converge in probability if

$$n^{-1} \sum_{t=M+1}^{n} (\theta' X_t' X_t \theta - X_t A^{-1} q)^2 \tag{3·8}$$

converges in probability; the terms of the summand in (3·8) are from a stationary process satisfying the conditions for ergodicity, and consequently by Corollary 6·1·1·1 of Fuller (1976), expression (3·8) converges in probability to δ. □

COROLLARY 3·1. *Under the conditions of Lemma* 3·1,

$$\hat{F}_{1,n-2M-2} = \hat{\eta}^2\{(\langle \hat{e}, \hat{e}\rangle - \hat{\eta}^2)/(n-2M-2)\}^{-1} = \chi_1^2 + o_p(1). \tag{3·9}$$

Proof. Fuller (1976, Theorem 8·2·2) shows that

$$\langle \hat{e}, \hat{e}\rangle/(n-2M-1) = \langle e, e\rangle/(n-M) + o_p(1), \tag{3·10}$$

where $e = (e_{M+1}, \ldots, e_n)$. By the weak law of large numbers, the right-hand side of (3·10) differs from σ_e^2 by $o_p(1)$. From Lemma 3·1 we have that $\hat{\eta}^2 = \sigma_e^2 \chi_1^2 + o_p(1)$ from which (3·9) follows. □

4. APPLICATIONS

To illustrate the procedure, the distribution of (3·2) was simulated for various linear and nonlinear models. The models were:

Model 1, $Y_t = e_t - 0·4e_{t-1} + 0·3e_{t-2}$;
Model 2, $Y_t = e_t - 0·4e_{t-1} + 0·3e_{t-2} + 0·5e_t e_{t-2}$;
Model 3, $Y_t = e_t - 0·3e_{t-1} + 0·2e_{t-2} + 0·4e_{t-1} - 0·25e_{t-1}^2$;
Model 4, $Y_t = 0·4Y_{t-1} - 0·3Y_{t-2} + e_t$;
Model 5, $Y_t = 0·4Y_{t-1} - 0·3Y_{t-2} + 0·5Y_{t-1}e_{t-1} + e_t$;
Model 6, $Y_t = 0·4Y_{t-1} - 0·3Y_{t-2} + 0·5Y_{t-1}e_{t-1} + 0·8e_{t-1} + e_t$.

Models 1 and 4 are linear models, second-order moving average and second-order autoregressive, respectively. The others are all nonlinear models; Models 2 and 5 are designed to be slightly different from 1 and 4, respectively, with the addition of one nonlinear term. Model 5 has the additional property of containing a nonlinear term without a lower order term, e_{t-1}; Model 6 differs from Model 5 by including the lower order term. Model 3 includes two nonlinear terms $e_{t-1}e_{t-2}$ and e_{t-1}^2. The parameter values of the above models are fairly representative, in the sense of not being close to the boundary.

An important practical matter is the question of how to appropriately choose the order of the autoregressive approximation, M. A rough rule seems to be to take M in the range of 4 to 8, although even bigger values seem to do as well, if not better. To analyse the effect of sample size, N, and choice of M on the procedure, 350 replications were performed for the four combinations of $N = 204$, $M = 4$ and $M = 8$, and $N = 70$, $M = 4$ and $M = 8$. Table 1 shows the simulated pth quantiles of $\hat{F}_{1,N-2M-2}$ for the four

Table 1. *pth quantiles of* $\hat{F}_{1,N-2M-2}$ *for* $N = 70$ *and* 204, $M = 4$ *and* 8, 350 *replications*

		Probability, p									
		0·5	0·75	0·9	0·95	0·99	0·5	0·75	0·9	0·95	0·99
	M			$N = 70$					$N = 204$		
Model 3 (NL)	4	3·51	7·88	13·16	18·29	25·65	12·07	26·78	35·61	43·49	55·82
	8	1·73	4·76	9·06	12·12	23·89	12·35	19·02	28·10	32·88	46·18
Model 2 (NL)	4	0·73	2·24	4·39	6·29	9·75	0·75	2·55	5·66	7·54	11·48
	8	1·37	3·55	6·45	8·38	14·83	1·26	3·54	6·45	9·77	14·38
Model 1 (L)	4	0·43	1·18	2·55	3·51	5·57	0·49	1·27	2·55	3·82	8·18
	8	0·42	1·23	2·62	3·83	5·66	0·47	1·43	2·66	3·93	6·44
χ_1^2		0·46	1·32	2·71	3·84	6·63	0·46	1·32	2·71	3·84	6·63
Model 4 (L)	4	0·50	1·41	2·86	3·84	7·37	0·40	1·33	2·53	3·41	6·02
	8	0·46	1·41	2·97	3·70	7·07	0·43	1·38	2·66	3·88	6·10
Model 5 (NL)	4	1·09	3·57	6·32	8·37	13·49	1·62	5·09	8·58	13·15	26·49
	8	0·82	2·82	5·79	8·10	14·83	1·68	3·66	7·74	11·70	18·84
Model 6 (NL)	4	4·15	8·81	16·50	22·86	41·67	11·59	21·72	28·44	34·68	62·25
	8	4·27	8·90	18·17	24·60	43·22	13·13	20·89	29·58	36·60	63·80

NL, nonlinear; L, linear.

combinations with the centre row of the table being the pth quantiles of $F_{1,N-2M-2}$ which for large N is approximately χ_1^2. The ability for the procedure to discriminate between linear and nonlinear models appears to be quite good.

ACKNOWLEDGEMENTS

This work was partially supported by a National Science Foundation Mathematical Sciences Postdoctoral Fellowship at Brown University.

REFERENCES

BILLINGSLEY, P. (1961). The Lindeberg–Lévy theorem for martingales. *Proc. Am. Math. Soc.* **12**, 788–92.

BRILLINGER, D. R. (1970). The identification of polynomial systems by means of higher order spectra. *J. Sound. Vib.* **12**, 301–13.

FULLER, W. A. (1976). *Introduction to Statistical Times Series*. New York: Wiley.

HINICH, M. J. (1982). Testing for Gaussianity and linearity of a stationary times series. *J. Time Series Anal.* **3**, 169–76.

MCLEOD, A. I. & LI, W. K. (1983). Diagnostic checking ARMA time series models using squared-residual autocorrelations. *J. Time Series Anal.* **4**, 269–73.

SCHEFFÉ, H. (1959). *Analysis of Variance*. New York: Wiley.

SEBER, G. A. F. (1977). *Linear Regression Analysis*. New York: Wiley.

SNEDECOR, G. W. & COCHRAN, W. G. (1967). *Statistical Methods*, 6th ed. Ames, Iowa: Iowa State University Press.

SUBBA RAO, T. & GABR, M. M. (1980). A test for linearity of stationary time series. *J. Time Series Anal.* **1**, 145–58.

TUKEY, J. W. (1949). One degree of freedom for non-additivity. *Biometrics* **5**, 232–42.

VAN NESS, J. W. (1966). Asymptotic normality of bispectral estimates. *Ann. Math. Statist.* **37**, 1257–75.

WIENER, N. (1958). *Non-linear Problems in Random Theory*. Cambridge, Mass: M.I.T. Press.

[*Received May* 1984. *Revised August* 1984]