MONASH
E'METRICS
WP
11/97

ISSN 1032-3813
ISBN 0 7326 1034 6

MONASH UNIVERSITY

AUSTRALIA

COMPARISON AND CLASSIFICATION OF STATIONARY MULTIVARIATE TIME SERIES

Ann Maharaj

Working Paper 11/97
November 1997

DEPARTMENT OF ECONOMETRICS
AND BUSINESS STATISTICS

COMPARISON AND CLASSIFICATION OF STATIONARY MULTIVARIATE TIME SERIES

Ann Maharaj

Department of Econometrics and Business Statistics

Monash University, Australia

ABSTRACT

Time series often have patterns that form a basis for comparing them or classifying them into groups. Pattern recognition of time series arises in a number of practical situations. Procedures for the comparison and classification of univariate stationary series already exist in the literature. A famous application is the comparison and classification of earthquake and nuclear explosion waveforms - Shumway (1982). In this paper we present procedures to compare and classify stationary multivariate time series. Simulations studies show that the procedures perform fairly well for reasonably long series.

1. INTRODUCTION

Pattern recognition of time series involves the use of comparison as well as classification techniques. Most tests in the literature for the comparison of independent stationary series involve the use of the estimated spectra of the series. Some relevant papers are Swanepoel and Van Wyk (1986), Coates and Diggle (1986) and Diggle and Fisher (1991). Hypotheses tests designed to compare two stationary independent time series involving the use of fitted parameter estimates were considered by De Souza and

Thomson (1982) and Maharaj (1996). Maharaj and Inder (1997) extended this testing procedure to compare two related stationary time series. We now extend this testing procedure to compare two stationary multivariate time series which are not necessarily independent.

The classification of time series has applications in various fields including Economics, Geology, Psychology, Oceanography and Engineering. Shumway (1982) used a combination of spectral and discriminant analysis to classify both univariate and multivatiate time series that are stationary. Various authors including Bohte et al.(1980), Tong et al. (1990), Shaw et al. (1992) and Piccolo (1990) have used cluster analysis to classify time series. The idea is to investigate similarities of the time series in the identified clusters. Maharaj (1996) proposes a method of clustering that has the property of uniquely identifying groups of stationary univariate time series. None of the existing clustering techniques in the literature have this property. We now extend this procedure to classifying stationary multivariate time series that are not necessarily independent. This procedure is based of the p-values of the test that is proposed for the comparison of two stationary multivariate time series.

Given two stationary multivariate time series, vector autoregressive and moving average (VARMA) models converted to truncated vector autoregressive infinite order (VAR(∞)) models of order k, are fitted to each series. The test statistic is based on the differences between the VAR(k) estimates of the two sets of series under consideration. For related series, it is assumed that the disturbances of the two models are correlated. A test statistic to test for significant differences between the generating processes of these series is based on generalised least squares estimates of the VAR parameters.

Using bivariate series we investigate the distributional properties, size and power of this test statistic, which has an asymptotic chi-square distribution, for finite sample sizes by a Monte Carlo study. We also assess the performance of the p-value clustering procedure when applied to bivariate series.

2. *HYPOTHESIS TESTING PROCEDURE*

Consider a stationary m-dimensional time series

$$x_t' = [\, x_{1t}, x_{2t}, \ldots, x_{mt} \,]$$

that is generated by an infinite order zero mean vector autoregressive process (ie. a VAR(∞) process):

$$x_t = \sum_{i=1}^{\infty} \Pi_i x_{t-i} + a_t$$

where the Π_i are m x m matrices of coefficients and a_t is the m-dimensional white noise process with mean vector 0 and covariance matrix Σ_a . Just as a finite order AR model may be fitted to a univariate series of length T that is assumed to be generated by an AR(∞) process, a finite order VAR model may be fitted to a multivariate series of length T, that is assumed to be generated by a VAR(∞) process. Using the notation of Lutkepohl (1991) define the collection of coefficient matrices of the VAR(k) process as

$$\Pi(k) = [\, \Pi_1, \Pi_2, \ldots, \Pi_k \,] .$$

Fitting a VAR(k) model to a multivariate series x_t, the *i*th estimated coefficient matrix is denoted by $\hat{\Pi}_i$ and the collection of estimated coefficient matrices is

$$\hat{\Pi}(k) = [\, \hat{\Pi}_1, \hat{\Pi}_2, \ldots, \hat{\Pi}_k \,] .$$

Suppose that $\{x_t, t = 1, 2, \ldots, T\}$ and $\{y_t, t = 1, 2, \ldots, T\}$ are stationary multivariate series which are assumed to be generated by VAR(∞) processes. Using a definite criterion such as Schwarz's BIC for modelling VAR structures, truncated VAR(∞) models of order k_1 and k_2 can be fitted to $\{x_t\}$ and $\{y_t\}$ respectively. Define the matrices of VAR(k_1) and VAR(k_2) parameters of the generating processes $\{X_t\}$ and $\{Y_t\}$ respectively as Π_{ix}, $i = 1, 2, \ldots, k_1$, and Π_{iy} $i = 1, 2, \ldots, k_2$

Let

$$k = \max(k_1, k_2).$$

Then define

$$\Pi_x = \left[\, \Pi_{1x}, \Pi_{2x}, \ldots, \Pi_{kx} \,\right]$$

$$\Pi_y = \left[\, \Pi_{1y}, \Pi_{2y}, \ldots, \Pi_{ky} \,\right].$$

Given two multivariate stationary series of the same dimension $\{x_t\}$ and $\{y_t\}$, the hypotheses to be tested are:

H_0: There is no significant difference between the generating processes of two stationary m-dimensional times series (ie. Vec Π_x = Vec Π_y).

H_1: There is a significant difference between the generating processes of two stationary m- dimensional time series (ie. Vec $\Pi_x \neq$ Vec Π_y).

Lewis and Reinsel (1985) have obtained results on least squares estimates of a finite order VAR process that has been truncated from a VAR(∞) process. They assume that the truncated order k depends on the length T of the m-dimensional series $\mathbf{x}_t$ such that

$$\frac{k^3}{T} \to 0 \quad \text{and} \quad \sqrt{T}\sum_{i=k+1}^{\infty}\|\Pi_i\| \to 0 \quad \text{as} \quad T \to \infty$$

They have shown consistency and asymptotic normality of these least squares estimates (see Lutkepohl (1993) pages 306-307).

The model to be considered is of the form of the multivariate analogue of the seemingly unrelated regressions model proposed by Zellner (1962). The m(T-k) equations fitted to $\{x_t\}$ and $\{y_t\}$ can be expressed collectively as

$$\begin{aligned} X &= \Pi_x B_x + A_x \\ Y &= \Pi_y B_y + A_y \end{aligned} \qquad (2.1)$$

where

$$\Pi_x = [\, \Pi_{1x}, \Pi_{2x}, \ldots, \Pi_{kx} \,]$$

$$B_x = \begin{bmatrix} x_{1\,k} & x_{1\,k-1} & \cdots & x_{1\,1} \\ x_{2\,k} & x_{2\,k-1} & \cdots & x_{2\,1} \\ \vdots & \vdots & & \vdots \\ x_{m\,k} & x_{m\,k-1} & \cdots & x_{m\,1} \\ \vdots & \vdots & & \vdots \\ x_{1\,T-2} & x_{1\,T-3} & \cdots & x_{1\,T-k+1} \\ x_{2\,T-2} & x_{2\,T-3} & \cdots & x_{2\,T-k+1} \\ \vdots & \vdots & & \vdots \\ x_{m\,T-2} & x_{m\,T-3} & \cdots & x_{m\,T-k+1} \\ x_{1\,T-1} & x_{1\,T-2} & \cdots & x_{1\,T-k} \\ x_{2\,T-1} & x_{2\,T-2} & \cdots & x_{2\,T-k} \\ \vdots & \vdots & & \vdots \\ x_{m\,T-1} & x_{m\,T-2} & \cdots & x_{m\,T-k} \end{bmatrix}$$

$$X = \begin{bmatrix} x_{1\,k+1} & . & . & . & x_{1\,T-1} & x_{1\,T} \\ x_{2\,k+1} & . & . & . & x_{2\,T-1} & x_{2\,T} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ x_{m\,k+1} & . & . & . & x_{m\,T-1} & x_{m\,T} \end{bmatrix}$$

$$A_x = \begin{bmatrix} a_{1\,k+1\,x} & . & . & . & a_{1\,T-1\,x} & a_{1\,Tx} \\ a_{2\,k+1\,x} & . & . & . & a_{2\,T-1\,x} & a_{2\,Tx} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ a_{m\,k+1\,x} & . & . & . & a_{m\,T-1\,x} & a_{m\,Tx} \end{bmatrix}$$

The dimensions of Π_x, X, B_x and A_x are m x mk, m x (t-k), mk x (t-k) and m x (t-k) respectively. Π_y, Y, B_y and A_y are similarly defined. We make the following assumptions for the error term:

$$E[A_x] = 0 \qquad E\left[\ A_x A_x'\ \right] = \Sigma_x \otimes I_{T-k}$$

$$E[A_x] = 0 \qquad E\left[\ A_y A_y'\ \right] = \Sigma_y \otimes I_{T-k}$$

$$E\left[\ A_x A_y'\ \right] = \Sigma_{xy} \otimes I_{T-k}$$

where I_{T-k} is a (T-k) x (T-k) identity matrix and

$$\Sigma_x = \begin{bmatrix} \sigma_{x1}^2 & \sigma_{x1\,x2} & . & . & . & \sigma_{x1\,xm} \\ \sigma_{x2\,x1} & \sigma_{x2}^2 & . & . & . & \sigma_{x2xm} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ \sigma_{xm\,x1} & \sigma_{xm\,x2} & . & . & . & \sigma_{xm}^2 \end{bmatrix}$$

$$\Sigma_y = \begin{bmatrix} \sigma_{y1}^2 & \sigma_{y1\,y2} & \cdot & \cdot & \cdot & \sigma_{y1\,ym} \\ \sigma_{y2\,y1} & \sigma_{y2}^2 & \cdot & \cdot & \cdot & \sigma_{y2\,ym} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \sigma_{ym\,y1} & \sigma_{ym\,y2} & \cdot & \cdot & \cdot & \sigma_{ym}^2 \end{bmatrix}$$

$$\Sigma_{xy} = \begin{bmatrix} \sigma_{x1\,y1} & \sigma_{x1\,y2} & \cdot & \cdot & \cdot & \sigma_{x1\,ym} \\ \sigma_{x2\,y1} & \sigma_{x2\,y2} & \cdot & \cdot & \cdot & \sigma_{x2\,ym} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \sigma_{xm\,y1} & \sigma_{xm\,y2} & \cdot & \cdot & \cdot & \sigma_{xm\,ym} \end{bmatrix}.$$

The combined model may be expressed as

$$Z = \Pi B + A \qquad (2.2)$$

where

$$Z = \begin{bmatrix} X \\ Y \end{bmatrix}, \quad B = \begin{bmatrix} B_x & 0 \\ 0 & B_y \end{bmatrix}$$

$$\Pi = \begin{bmatrix} \Pi_x \\ \Pi_y \end{bmatrix}, \quad A = \begin{bmatrix} A_x \\ A_y \end{bmatrix}$$

and

$$E[AA'] = V = \begin{bmatrix} \Sigma_x \otimes I_{T-k} & \Sigma_{xy} \otimes I_{T-k} \\ \Sigma_{xy} \otimes I_{T-k} & \Sigma_y \otimes I_{T-k} \end{bmatrix}.$$

The dimensions of V are 2m(T-k) x 2m(T-k). The models in (2.2) can also be expressed as

$$\mathrm{Vec}(X) = (B_x' \otimes I_m)\mathrm{Vec}(\Pi_x) + \mathrm{Vec}(A_x)$$

$$\mathrm{Vec}(Y) = (B_y' \otimes I_m)\mathrm{Vec}(\Pi_y) + \mathrm{Vec}(A_y) \qquad (2.3)$$

where I_m is a m x m identity matrix. Hence the combined model in (2.2) can also be expressed as

$$Z_v = B_v \Pi_v + A_v \tag{2.4}$$

where

$$Z_v = \begin{bmatrix} Vec(X) \\ Vec(Y) \end{bmatrix}, \quad B_v = \begin{bmatrix} B_x' \otimes I_m & 0 \\ 0 & B_y' \otimes I_m \end{bmatrix}$$

$$\Pi_v = \begin{bmatrix} Vec(\Pi_x) \\ Vec(\Pi_y) \end{bmatrix}, \quad A_v = \begin{bmatrix} Vec(A_x) \\ Vec(A_y) \end{bmatrix}.$$

The dimensions of Z_v, B_v, Π_v and A_v are 2m(T-k) x 1, 2m(t-k) x $2m^2k$, $2m^2k$ x 1 and 2m(T-k) x 1 respectively. Thus using (2.4) the generalised least squares estimator of Π_v is

$$\hat{\Pi}_v = \left[B_v' V^{-1} B_v \right]^{-1} B_v V^{-1} Z_v \tag{2.5}$$

Now assuming that A_v is normally distributed, then

$$\sqrt{T}\left(\hat{\Pi}_v - \Pi_v\right) \to N\left(0, V^*\right),$$

(multivariate analogue of results in Anderson (1971))

where

$$V^* = \lim_{T\to\infty} Var\left(\sqrt{T}\,\hat{\Pi}_v\right) = plim\left(\frac{B_v' V^{-1} B_v}{T}\right)^{-1}.$$

Now the null hypothesis H_0: Vec (Π_x) = Vec (Π_y) may be expressed as

$$R\Pi_v = 0$$

where

$$R = [I \quad -I]$$

and where I is a m^2k x m^2k identity matrix. It is easy to see that under H_0

$$\sqrt{T}\left(R\hat{\Pi}_v - R\Pi_v\right) \rightarrow N\left(0, plim\left(\frac{RB_v'V^{-1}B_vR'}{T}\right)^{-1}\right). \qquad (2.6)$$

Define

$$F = \left(R\left(B_v'V^{-1}B_v\right)^{-1}R'\right)^{-1/2}\left(R\hat{\Pi}_v - R\Pi_v\right).$$

Under H_0

$$F = \left(R\left(B_v'V^{-1}B_v\right)^{-1}R'\right)^{-1/2}R\hat{\Pi}_v.$$

So from (2.6)

$$F \overset{A}{\sim} N\left(0, I_{m^2k}\right)$$

and

$$F'F = \left(R\hat{\Pi}\right)'\left[R\left(B_v'V^{-1}B_v\right)^{-1}R'\right]^{-1}\left(R\hat{\Pi}\right) \overset{A}{\sim} \chi^2\left(m^2k\right).$$

Since Σ_x , Σ_y and Σ_{xy} are unknown, feasible generalised least squares estimators of Π_v have to be obtained. Using the multivariate equivalent to the univariate results of Zellner (1962), ordinary least squares estimates may be used to estimate consistently the elements of Σ_x , Σ_y and Σ_{xy} with

$$\hat{\Sigma}_x = \frac{\hat{A}_x\hat{A}_x'}{m(T-k)}$$

$$\hat{\Sigma}_y = \frac{\hat{A}_y\hat{A}_y'}{m(T-k)}$$

$$\hat{\Sigma}_{xy} = \frac{\hat{A}_x\hat{A}_y'}{m(T-k)}$$

Hence the feasible least squares estimator of Π_v is

$$\hat{\Pi}_v = \left[B_v'\hat{V}^{-1}B_v\right]^{-1} B_v\hat{V}^{-1}Z_v$$

where

$$\hat{V} = \begin{bmatrix} \hat{\Sigma}_x \otimes I_{T-k} & \hat{\Sigma}_{xy} \otimes I_{T-k} \\ \hat{\Sigma}_{xy} \otimes I_{T-k} & \hat{\Sigma}_y \otimes I_{T-k} \end{bmatrix}$$

Since $\hat{V}$ is nonsingular,

$$plim\, \hat{V} = V \ ,$$

so under H_0

$$D = FF' = \left(R\hat{\Pi}_v\right)' \left[R\left(B_v'\hat{V}^{-1}B_v\right)^{-1} R'\right]^{-1} \left(R\hat{\Pi}_v\right) \overset{A}{\sim} \chi^2\left(m^2k\right).$$

2.1 Simulation Study : Assessment of the Test

To investigate the finite sample behaviour of the test statistic D, bivariate series of lengths 50 and 200 are simulated from the following VARMA processes:

AR(1) $\quad \Phi = \begin{bmatrix} 0.5 & 0.1 \\ 0.7 & 0.5 \end{bmatrix}$

MA(1) $\quad \Theta = \begin{bmatrix} 0.3 & 0.7 \\ 0.1 & 0.3 \end{bmatrix}$

AR(2) $\quad \Phi_1 = \begin{bmatrix} 0.5 & 0.1 \\ 0.7 & 0.5 \end{bmatrix} \quad \Phi_2 = \begin{bmatrix} 0 & 0 \\ 0 & -0.3 \end{bmatrix}$

MA(2) $\quad \Theta_1 = \begin{bmatrix} 0.3 & 0.7 \\ 0.1 & 0.3 \end{bmatrix} \quad \Theta_2 = \begin{bmatrix} -0.2 & 0.1 \\ 0.3 & 0.1 \end{bmatrix}$

ARMA(1,1) $\quad \Phi = \begin{bmatrix} 0.5 & 0.1 \\ 0.7 & 0.5 \end{bmatrix} \quad \Theta = \begin{bmatrix} -0.2 & 0.1 \\ 0.3 & 0.1 \end{bmatrix}$

Distributional properties of the test based on D are checked by obtaining estimates of the mean, variance, skewness of the test statistic and the size of the test under H_0. This is done by applying the test to pairs of bivariate series simulated from the same process. It is assumed that the correlation between the disturbances are in turn

(i) $\quad \Sigma_x = \Sigma_y = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad \Sigma_{xy} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$

and

(ii) $\quad \Sigma_x = \Sigma_y = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \text{ and } \quad \Sigma_{xy} = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}.$

Estimates of size are obtained for 5% and 1% significance levels. Estimates of power for 5% and 1% significance levels are obtained by applying the test to every pair of bivariate series .

The order (up to 10) of the truncated VAR model to be fitted to each bivariate series in each pair is determined by Schwarz's BIC. In estimating the model in (3.4), the maximum order is fitted to both the bivariate series in each pair. The test statistic is then obtained. This is repeated 1000 times. As well as obtaining size and power estimates for the various degrees of freedom, overall estimates of power and size are also obtained by aggregating the size estimates over the various degrees of freedom.

For series of length 50, size is considerably overestimated and estimates of the mean, variance and skewness do not correspond closely to the respective theoretical values. Hence no further analysis was performed on series of this length 50. The overall size estimates for T = 50 are shown in Table 2.1.

Table 2.1 Overall Estimates of and Size for T=50

Generating Process	Correlation (i)		Correlation (ii)	
	5%	1%	5%	1%
AR(1)	0.120*	0.055*	0.106*	0.032*
MA(1)	0.139*	0.057*	0.140*	0.054*
AR(2)	0.142*	0.059*	0.134*	0.054*
MA(2)	0.094*	0.089*	0.184*	0.087*
ARMA(1,1)	0.238*	0.120*	0.204*	0.109*

* size differs from nominal size by a significant amount (5% level)

For series of length 200, where a reasonable number of statistics occur for a given degrees of freedom, the estimates of the means, variances and measures of skewness are in many cases fairly close to the corresponding theoretical parameters. The results for which there are at least 100 test statistics are shown in Tables 2.2 and 2.3.

Table 2.2-3.3 Estimates of Mean, Variance, Skewness and Size for T=200 (Bivariate Series)

Table 2.2 Correlation (i)

Process	Order k	df=m^2k	No. Test Stats.	Mean	Var	Skew	Size 5%	Size 1%
AR(1)	1	4	991	4.212	8.434	0.228	0.063	0.000
MA(1)	1	4	140	3.957	7.370	0.215	0.036	0.014
	2	8	774	8.645	20.370	0.190	0.067*	0.022*
AR(2)	2	8	973	8.469	18.363	0.193	0.071*	0.013
MA(2)	2	8	602	8.047	16.746	0.188	0.055	0.013
	3	12	379	13.865	34.810	0.119	0.121*	0.026*
ARMA (1,1)	2	8	237	8.464	21.438	0.212	0.097*	0.021
	3	12	655	12.785	28.104	0.146	0.075*	0.016

* size differs from nominal size by a significant amount (5% level)

Table 2.3 Correlation (ii)

Process	Order k	df $=m^2k$	No. Test Stats.	Mean	Var	Skew	Size 5%	Size 1%
AR(1)	1	4	996	4.198	8.869	0.230	0.063	0.016
MA(1)	1	4	127	4.093	7.672	0.161	0.047	0.008
	2	8	759	8.371	19.148	0.231	0.082*	0.012
	3	12	111	13.781	26.161	0.074	0.090	0.009
AR(2)	2	8	992	8.231	17.014	0.179	0.057	0.015
MA(2)	2	8	211	8.198	18.555	0.213	0.071	0.005
	3	12	670	12.648	27.788	0.105	0.067	0.015
	4	16	117	18.496	43.481	0.020	0.103*	0.034*
ARMA (1,1)	2	8	155	8.714	20.410	0.263	0.097*	0.026
	3	12	740	12.614	29.839	0.147	0.073*	0.020*
	4	16	105	18.059	45.440	0.174	0.133*	0.038*

***size differs from nominal size by a significant amount (5% level)**

In many cases the size estimates are reasonably close to the predetermined significance levels. The overestimation of size in other cases in general caused the overall size to be slightly overestimated. These overall size estimates, as well as overall power estimates, are shown in Tables 2.4 to 2.7.

Table 2.4-2.7 Overall Estimates of Size and Power for T =200

Table 2.4 Correlation (i)

5% level	AR(1)	MA(1)	AR(2)	MA(2)	ARMA(1,1)
AR(1)	0.068*	1.000	0.920	1.000	1.000
MA(1)		0.067*	1.000	0.737	1.000
AR(2)			0.074*	1.000	1.000
MA(2)				0.085*	1.000
ARMA(1,1)					0.088*

*** size differs from nominal size by a significant amount (5% level)**

Table 2.5 Correlation (i)

1% level	AR(1)	MA(1)	AR(2)	MA(2)	ARMA(1,1)
AR(1)	0.015	1.000	0.846	1.000	1.000
MA(1)		0.022*	1.000	0.547	1.000
AR(2)			0.015	1.000	1.000
MA(2)				0.018*	1.000
ARMA(1,1)					0.021*

* size differs from nominal size by a significant amount (5% level)

Table 2.6 Correlation (ii)

5% level	AR(1)	MA(1)	AR(2)	MA(2)	ARMA(1,1)
AR(1)	0.063	1.000	0.998	1.000	1.000
MA(1)		0.078*	1.000	0.990	1.000
AR(2)			0.057	1.000	1.000
MA(2)				0.073*	1.000
ARMA(1,1)					0.083*

* size differs from nominal size by a significant amount (5% level)

Table 2.7 Correlation (ii)

1% level	AR(1)	MA(1)	AR(2)	MA(2)	ARMA(1,1)
AR(1)	0.016	1.000	0.994	1.000	1.000
MA(1)		0.011	1.000	0.965	1.000
AR(2)			0.015	1.000	1.000
MA(2)				0.015	1.000
ARMA(1,1)					0.023*

* size differs from nominal size by a significant amount (5% level)

Overall power estimates reveal that the test is quite powerful when the two bivariate series are generated from vastly different processes. For those degrees of freedom for which reasonably good size estimates were obtained, the estimates of the means, variances and skewness of the test statistic are very often fairly close to the theoretical means, variances and measures of skewness respectively.

3. CLUSTERING PROCEDURE

The method of clustering as proposed by Maharaj (1996) will now to be extended here for use with stationary multivariate time series has the following steps: First perform the test of hypothesis from Section 2.1 for every pair of series determining the p-value associated with the test statistic D. Use these p-values in an algorithm that incorporates the principles of hierarchical clustering but will only group together those series whose associated p-values are greater than some predetermined number (eg. 0.05 or 0.01). Note that the p-value of the test is a measure of similarity and satisfies properties of a semi-metric.

We assess the performance of this clustering procedure for bivariate series in the following section.

3.1 Simulation Study : Assessment of Clustering Procedure

The p-value clustering procedure was applied to 10 bivariate series, 2 of which are each simulated from the VARMA processes listed in Section 2.1. The series are labelled as follows:

AR(1): 1, 2; MA(1): 3, 4; AR(2): 5, 6;

MA(2): 7, 8 ; ARMA(1,1):9, 10.

The correlation between the disturbances are the same as those given in scenarios (i) and (ii) in Section 2.1. The minimum p-value is in turn set at 0.05 and 0.01 and clusters are simulated 100 times for each of these minimum p-value settings. For each simulation, statistics are then obtained on the number of clusters produced each of the 100 times, the number of exactly correct clusters produced at each simulation (clusters containing the 4 series from the same generating process) , the number of clusters of mixed series which are produced at each simulation (clusters containing series from different generating processes) and the number of occasions that series from the same generating processes fail to come together when no mixing occurs. The results are shown in Tables 3.1 to 3.4.

From Table 3.1 for correlation (i), that is

$$\Sigma_x = \Sigma_y = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad \Sigma_{xy} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix},$$

with the minimum p-value set at 0.05 we observe the follow: The four cluster solution occurs most frequently (55% of the time) , followed by the five cluster solution which occurs 36% of the time. At least one exactly correct cluster is produced 100% of the time, with exactly three correct clusters occurring most frequently (74% of the time) and exactly five correct clusters occurring only 1% of the time. One cluster of mixed series occurs 91% of the time. In 80% of simulations there were no cases of processes failing to come together and one failure occurs 18% of the time.

From Table 3.2 it can be seen that for correlation (i) with the minimum p-value set at 0.01, the four cluster solution occurs more frequently than in the previous case (87% of the time compared to 55% of the time) . This was followed by the five cluster solution which occurs 12% of the time as compared to 36% of the time in the previous case. At least one exactly correct cluster is produced 100% of the time, with exactly

Table 3.1 -3.2 Cluster Statistics for T=200, Correlation (i) (Bivariate Series)

Table 3.1 p-value=0.05

Number of clusters	% freq.	Number of exactly correct clusters	% freq.
4	55	0	0
5	36	1	6
6	8	2	18
7	1	3	74
		4	1
		5	1
	100		100
Number of clusters of mixed series	% freq	Number of occasions ,series from the same generating process failed to come together in the absence of mixing	% freq
0	2	0	80
1	91	1	18
2	7	2	2
	100.00		100.00

Table 3.2 p-value=0.01

Number of clusters	% freq.	Number of exactly correct clusters	% freq.
4	87	0	0
5	12	1	5
6	1	2	5
		3	90
	100.00		100.00
Number of clusters of mixed series	% freq	Number of occasions ,series from the same generating process failed to come together in the absence of mixing	% freq
0	0	0	96
1	94	1	4
2	6		
	100		100

Table 3.3-3.4. Cluster Statistics for T=200, Correlation (ii)
(Bivariate Series)

Table 3.3 **p-value=0.05**

Number of clusters	% freq.	Number of exactly correct clusters	% freq.
4	2	0	0
5	77	1	0
6	18	2	3
7	2	3	26
8	1	4	15
		5	56
	100		100
Number of clusters of mixed series	% freq	Number of occasions ,series from the same generating process failed to come together in the absence of mixing	% freq
0	76	0	79
1	24	1	20
		2	0
		3	1
	100		100

Table 3.4 **p-value=0.01**

Number of clusters	% freq.	Number of exactly correct clusters	% FREQ.
4	14	0	0
5	79	1	0
6	7	2	4
		3	31
		4	3
		5	62
	100.00		100.00
Number of clusters of mixed series	% freq	Number of occasions ,series from the same generating process failed to come together in the absence of mixing	% FREQ
0	65	0	93
1	35	1	7
	100.00		100.00

three correct clusters occurring 90% of the time compared with 74% of the time in the previous case. Neither four nor five correct clusters occur at all. One cluster of mixed series occurs 94% of the time as compared to 91% of the time in the previous case. In 96% of simulations there are no cases of processes failing to come together compared with 80% in the previous case and one failure occurs 4% of the time compared with 18% in the previous case.

From Table 3.3 it can be seen that for correlation (ii), that is

$$\Sigma_x = \Sigma_y = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \quad \text{and} \quad \Sigma_{xy} = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix},$$

and when the minimum p-value is set at 0.05, the five cluster solution occurs most frequently - 77% of the time, followed by the six cluster solution which occurs 18% of the time. There are at least two exactly correct clusters, with exactly five correct clusters occurring most frequently (56% of the time). 76% of the time there were no and one cluster of mixed series occurs 24% of the time. In 79% of simulations there were no cases of processes failing to come together and one failure occurs 20% of the time.

From Table 3.4 it can be seen that when for correlation (ii) with the minimum p-value set at 0.01, the five cluster solution occurs most frequently (79% of the time), as compared to 77% in the previous case. This is followed by the four cluster solution which occurs 14% of the time. In the previous case the four cluster solution occurs only 2% of the time. At least one exactly correct cluster is produced 100% of the time, with exactly 5 correct clusters occurring 62% of the time compared with 56% of the time in the previous case. For 65% of the cases, no clusters of mixed series. In 93% of simulations there are no cases of processes failing to come together compared to 79% in the previous case and one failure occurs 7% of the time compared to 20% of the time in the previous case.

For both correlation scenarios (i) and (ii) the performance of the clustering procedure clearly improves when the minimum p-value is set at 0.01 since the number of cases where series from the same generating process fail to come together in the absence of mixing is reduced. Furthermore there is no increase in the number of mixed clusters occurring.

For the correlation (ii) scenario, the clustering procedure performs very much better than the correlation (i) scenario for both minimum settings of the p-value. The five cluster solution (which is the correct one) occurs 77% (p-value 0.05) and 79% (p-value 0.01) of the time as compared to 36% (p-value 0.05) and 12% (p-value 0.01) of the time.

4. CONCLUDING REMARKS

From the simulation study it is clear that the test procedure works quite well for reasonably large sample sizes. The distributional approximations to the chi-square distribution are fairly adequate and size estimates reveal that the test statistic gives an approximately valid test. The test appears to have fairly good power in that it clearly distinguishes between series that were generated from vastly different generating processes.

This procedure can also be extended to test for significant differences between the generating processes of more than two stationary multivariate processes. However one will require reasonably large sample sizes to obtain reliable estimates of size and power.

From the results of clustering it is clear that the p-value algorithm performs reasonably well. While all these simulation results are valid for bivariate stationary series, it is expected that they would be valid for higher dimensional stationary series as well, although larger sample sizes may be needed to get reasonable power.

REFERENCES

Anderson, T. (1971), *The Statistical Analysis of Time Series*, New York, Wiley.

Bohte, Z. D., Cepar and Kosmelu K. (1980), Clustering of Time Series, *Compstat,* 80, 587-593.

Coates, D.S. and Diggle, P.J.(1986), Test for Comparing Two Estimated Spectral Densities, *J. Time Series Anal.*, 7, 7-20.

De Souza, P. and Thomson, P.J. (1982), LPC Distance Measures and Statistical Tests with Particular Reference to the Likelihood Ratio, *IEEE Transactions on Acoustics, Speech and Signal Processing, ASSP-30*, 2, 304-315.

Diggle, P.J. and Fisher, N.I.(1991), Nonparametric Comparison of Cumulative Periodograms, *Appl.Statist.*, 40, 423-434.

Lewis, R. and G. C. Reinsel (1985), Prediction of Multivarite Time Series by Autoregressive Model Fitting, *Journal of Multivariate Analysis*, 16, 393-411.

Lutkepohl, H. (1991), *Introduction to Multiple Time Series Analysis*, Springer-Verlag, Berlin.

Maharaj, E.A. (1996), A Significant Test for Classifying ARMA Models, *J. Statist. Comput. Simul.*, 54, 305-331.

Maharaj, E.A. and B.A. Inder (1997), A Test to Compare Related Stationary Time Series ,*Conference Proceedings, Econometrics Society Australian Meeting*, July 1997.

Piccolo, D. (1990), A distance measure for classifying ARIMA models, *Journal of Time Series* 11, 2, 153 - 164.

Shaw, C. T. and G. P. King (1992), Using cluster analysis to classify time series, *Physica D, Non Linear Phenomena*, 58, 288-298.

Shumway, R.H. (1982), Discriminant Analysis for Time Series, *Handbook of Statistics*, Volume 2. (Editors: P.R.Krishnaiah and L.N. Kanal), 1-46, North Holland Publishing Company, Amsterdam.

Swanepoel, J.W.H. and J.W.J. Van Wyk (1986), The Comparison of Two Spectral Density Functions using the Bootstrap, *J. Statist. Comput. Simul.*, 24, 271-282.

Tong, H. and P. Dabas (1990), Cluster of Time Series Models: An example, *Journal of Applied Statistics* 17, 2, 187-198.

Zellner, A., (1962), Estimators for Seemingly Unrelated Regressions Equations and Test of Aggregation Bias, *JASA*, 57, 500-509.