Journal of Classification 17:297-314 (2000)
DOI: 10.1007/s003570000023

Clusters of Time Series

Elizabeth Ann Maharaj

Monash University

Abstract: This paper presents a test of hypotheses to compare two stationary time series as well as an accompanying classification procedure that uses this test of hypotheses to cluster stationary time series. Our hypotheses testing procedure, which unlike the existing tests, does not require the time series to be independent, is based on the differences between estimated parameters of the autoregressive models that are fitted to the time series. The classification procedure is based on the *p-value* of the test that is applied to every pair of given time series.

Keywords: Stationary time series; Autoregressive models; Seemingly unrelated regressions; Clustering algorithm.

1. Introduction

The comparison and classification of time series has applications in various fields including economics, business, demography, geology, medicine, and climatology. By using a classification procedure we could, for example, group together those countries that have similar economic indicators such as Gross Domestic Product (GDP) or those countries that have similar birth rates. Then instead of forecasting each of the given time series, forecasting can then be performed on a representative from each group. This strategy is especially useful if one has to forecast a large

---

This research was supported in part by a grant from the Monash University Research Fund and in part by a grant from the Faculty of Business and Economics of Monash University. I am grateful to the three referees and the editor for their constructive comments and suggestions that enhanced this paper.

Author's Address: Ann Maharaj, Department of Econometrics and Business Statistics, Monash University-Caulfield Campus, P.O. Box 197, Caulfield East, Melbourne, Victoria 3145, Australia.; e-mail: ann.maharaj@buseco.monash.edu.au

number of time series. Given the reduced time and cost, this approach would certainly be more practical than forecasting each and every time series.

Existing 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). Other tests in the literature for the comparison of two independent stationary series involve the use of the estimated spectra of the series. Some relevant papers are by Jenkins (1961), Swanepoel and Van Wyk (1986), Coates and Diggle (1986), and Diggle and Fisher (1991). In practice the application of these tests to real time series is limited because comparisons are often made between time series that are influenced by similar factors.

Existing techniques for classification of time series are discriminant analysis (Shumway 1982) and cluster analysis (Bohte, Cepar, and Kosmelu 1980; Piccolo 1990; Shaw and King 1992; Tong and Dabas 1990). Kosmelj and Batagelj (1990) also take time series into account in their proposed general model for cluster analysis. This procedure can be performed on time varying data by including the time factor as a third dimension to the data matrix that contains the units and variables as the other two dimensions.

Discriminant analysis requires the existence of known groupings before further classification can be carried out. While cluster analysis does not require known groupings, definite conclusions cannot be drawn from the results of cluster analysis because clusters are usually identified on a subjective basis, that is, the choice of clustering method and the distance at which clusters are identified are decided upon by the analyst. The procedure proposed by Maharaj (1996) for clustering stationary time series that are assumed to be independent does not require known groupings beforehand as discriminant analysis does, and definite conclusions can be drawn from the results of this analysis as opposed to conventional cluster analysis. However the assumption of independent time series limits the applicability of this method of classification, because in many situations we may want to classify time series that are influenced by similar factors. By introducing a test of hypotheses that can be applied to related as well as independent time series, we extend the clustering procedure of Maharaj (1996) so that it will be applicable to stationary time series that are not necessarily independent.

In Section 2 we present a test of hypotheses to compare two stationary time series that are not necessarily independent. In Section 3 the results of a simulation[1] study to investigate the distributional properties, size, and power

1. All simulation results and the results of the application have been obtained by programming in *Gauss*.

of this test are reported, and we make power comparisons with some of the existing tests. In Section 4, we discuss the clustering procedure and, from the results of a simulation study, discuss the distribution of a measure of discrepancy between the true number of correct clusters and the number of exactly correct clusters that have been obtained. In Section 5 we apply the clustering procedure to a set of economic time series.

2. Hypotheses Testing Procedure

We now consider the comparison of two stationary time series that are not necessarily independent. We will assume that if the series are not stationary, the same order of differencing will be needed to make each one stationary. We will also assume that the stationary times series can be fitted by linear models. Autoregressive infinity (AR(∞)) models, truncated to order k, are fitted to each series, and the test statistic is based on the difference between the AR(k) estimates, which are generalized least squares estimates. It will be assumed that the disturbances of the models are correlated for series that are not independent and uncorrelated for series that are independent.

Let Z_t be a zero mean univariate stochastic process such that $Z_t \in L$, where L is the class of stationary and invertible *ARMA* models. Using the standard notation of Box and Jenkins (1976, p. 74), such a model is defined as

$$\phi(B)Z_t = \theta(B)a_t,$$

where a_t is a univariate white noise process with mean 0 and variance, σ_a^2 and where

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p \quad \text{and}$$
$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q,$$

with the usual stationarity and invertibility restrictions on the roots of $\phi(B)$ and $\theta(B)$. Z_t can be expressed as

$$Z_t = \sum_{j=1}^{\infty} \pi_j Z_{t-j} + a_t,$$

where

$$\pi(B) = \phi(B)\theta^{-1}(B) = 1 - \pi_1 B - \pi_2 B^2 - \ldots.$$

Let x_t and y_t, $t = 1, 2, \ldots, T$ be two stationary time series. Then using a definite criterion such as Akaike's Information Criterion (AIC) (cf. Box, Jenkins, and Reinsel 1994, p. 201) or Schwartz's Bayesian Information Criterion (BIC) (cf. Box, Jenkins, and Reinsel 1994, p. 201) for modelling autoregressive structures, truncated AR(∞) models of order k_1 and k_2 can be fitted respectively to x_t and y_t. Define the vector of the AR(k_1) and AR(k_2) parameters of the generating processes X_t and Y_t respectively as

$$\pi_x' = \left[\pi_{1x}\ \pi_{2x} \ldots \pi_{k_1x}\right] \quad \text{and} \quad \pi_y' = \left[\pi_{1y}\ \pi_{2y} \ldots \pi_{k_2y}\right],$$

and the AR(k_1) and AR(k_2) parameter estimates of the series x_t and y_t respectively as $\hat{\pi}_{jx}$, $j = 1, 2, \ldots, k_1$ and $\hat{\pi}_{jy}$, $j = 1, 2, \ldots, k_2$. Let $k = max(k_1, k_2)$. In constructing the test statistic the maximum determined order k will be fitted to both series.

Given the series x_t and y_t, $t = 1, 2, \ldots, T$, the hypotheses to be tested are

H_0: There is no difference between the generating processes of two stationary series; that is, $\pi_x = \pi_y$;

H_1: There is a difference between the generating processes of two stationary series; that is, $\pi_x \neq \pi_y$.

The model to be considered is of the form of "the seemingly unrelated regressions model" as proposed by Zellner (1962). The T-k observations of the models fitted to x_t and y_t can be expressed collectively as

$$\mathbf{x} = \mathbf{W}_x\pi_x + \mathbf{a}_x \quad \text{and} \quad \mathbf{y} = \mathbf{W}_y\pi_y + \mathbf{a}_y, \qquad (2.1)$$

where

$$\mathbf{x}' = \left[x_{k+1} \quad \cdot \quad \cdot \quad \cdot \quad x_{T-1} \quad x_T\right],$$

$$\mathbf{W}_x = \begin{bmatrix} x_k & x_{k-1} & \cdot & \cdot & \cdot & x_1 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ x_{T-2} & x_{T-3} & \cdot & \cdot & \cdot & x_{T-k-1} \\ x_{T-1} & x_{T-2} & \cdot & \cdot & \cdot & x_{T-k} \end{bmatrix},$$

$$\pi_x' = \left[\pi_{1x} \quad \pi_{2x} \quad \cdot \quad \cdot \quad \cdot \quad \pi_{kx}\right], \qquad \text{and}$$

$$\mathbf{a}_x' = \left[a_{k+1x} \quad \cdot \quad \cdot \quad \cdot \quad a_{T-1x} \quad a_{Tx}\right].$$

The quantities $\mathbf{y}'$, $\mathbf{W}_y$, π_y and $\mathbf{a}'_y$ are similarly defined. Furthermore

$$\mathrm{E}[\mathbf{a}_x] = \mathbf{0},\ \mathrm{E}\left[\mathbf{a}_x\mathbf{a}'_x\right] = \sigma_x^2\mathbf{I}_{T\text{-}k},\ \mathrm{E}[\mathbf{a}_y] = \mathbf{0}, \text{ and } \mathrm{E}\left[\mathbf{a}_y\mathbf{a}'_y\right] = \sigma_y^2\mathbf{I}_{T\text{-}k},$$

where $\mathbf{I}_{T\text{-}k}$ is a $(T\text{-}k) \times (T\text{-}k)$ identity matrix. We will assume that the disturbances of the two models are correlated at the same points in time but uncorrelated across observations. That is

$$\mathrm{E}(\mathbf{a}_x\mathbf{a}'_y) = \sigma_{xy}\mathbf{I}_{T\text{-}k}.$$

The dimensions of $\mathbf{x}$, $\mathbf{y}$, $\mathbf{a}_x$, and $\mathbf{a}_y$ are $(T\text{-}k) \times 1$, of π_x and π_y are $k \times 1$, and of $\mathbf{W}_x$ and $\mathbf{W}_y$ are $(T\text{-}k) \times k$.

Then assuming that a total of $2(T\text{-}k)$ observations are used in estimating the parameters of the two equations in (2.1) the combined model may be expressed as

$$\mathbf{Z} = \mathbf{W}\pi + \mathbf{a}, \tag{2.2}$$

Where

$$\mathbf{Z} = \begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix},\ \mathbf{W} = \begin{bmatrix} \mathbf{W}_x & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_y \end{bmatrix},\ \pi = \begin{bmatrix} \pi_x \\ \pi_y \end{bmatrix},\ \mathbf{a} = \begin{bmatrix} \mathbf{a}_x \\ \mathbf{a}_y \end{bmatrix},$$

and

$$E(\mathbf{a}) = \mathbf{0}, \text{ and } E(\mathbf{aa}') = \mathbf{V} = \Sigma \otimes \mathbf{I}_{T\text{-}k},$$

where

$$\Sigma = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix}.$$

Thus, the generalized least squares estimator is

$$\hat{\pi} = \left[\mathbf{W}'\mathbf{V}^{-1}\mathbf{W}\right]^{-1}\mathbf{W}'\mathbf{V}^{-1}\mathbf{Z}.$$

Now assuming that the white noise process a_t is normally distributed, then by results in Anderson (1971, p. 189) and Amemiya (1985, p. 187), $\hat{\pi}$ is asymptotically normally distributed with mean π and covariance matrix

$$\lim_{T\to\infty} Var\left(\sqrt{T}\hat{\pi}\right) = plim\left(\frac{\mathbf{W}'\mathbf{V}^{-1}\mathbf{W}}{T}\right)^{-1}.$$

Now H_0: $\pi_x = \pi_y$ may be expressed as H_0: $\mathbf{R}\pi = \mathbf{0}$, where $\mathbf{R} = [\mathbf{I}_k\ \ {-\mathbf{I}_k}]$, and $\mathbf{I}_k$ is a $k \times k$ identity matrix. Hence, $\mathbf{R}\hat{\pi}$ is asymptotically normally distributed with mean $\mathbf{R}\pi$ and covariance matrix

$$\lim_{T\to\infty} Var\left(\sqrt{T}\mathbf{R}\hat{\pi}\right) = plim\,\mathbf{R}\left(\frac{\mathbf{W}'\mathbf{V}^{-1}\mathbf{W}}{T}\right)^{-1}\mathbf{R}'.$$

It can be shown (cf. Maharaj 1997) that under H_0,

$$D = (\mathbf{R}\hat{\pi})' \left[\mathbf{R}\left(\mathbf{W}'\hat{\mathbf{V}}^{-1}\mathbf{W}\right)^{-1}\mathbf{R}'\right]^{-1}(\mathbf{R}\hat{\pi})$$

is one possible test statistic, and it is asymptotically distributed as chi-square with k degrees of freedom. $\mathbf{V}$ is estimated by $\hat{\mathbf{V}} = \hat{\Sigma} \otimes \mathbf{I}$ where according to Zellner (1962), least squares residuals may be used to estimate the elements of Σ consistently.

It has been shown in Maharaj (1997) that this testing procedure can be extended to multiple comparisons of stationary time series. However, multiple testing is not required for the clustering procedure that follows in Section 4.

3. Simulation Study

3.1. Assessment of the Test

To check on the suitability of the test statistic D, which is the cornerstone of the clustering procedure that can be applied to related series, we investigate the finite sample behavior of D. Series of lengths 50 and 200 were simulated from a number of ARMA processes. Distributional properties of the test based on D were checked by obtaining estimates of the mean, variance, and skewness of the test statistic, and the size of the test. This approach was implemented by applying the test to pairs of series simulated from the AR(1) processes: $\phi = 0, 0.1, 0.5, 0.9$, the MA(1) processes: $\theta = 0.1, 0.5, 0.9$, the AR(2) process: $\phi_1 = 0.6$ and $\phi_2 = 0.2$, the MA(2) process: $\theta_1 = 0.8$ and $\theta_2 = -0.6$, and the ARMA(1,1) process: $\phi = 0.8$ and $\theta = 0.2$. It was assumed that the correlation between disturbances of each pair of processes from which the series were generated, were in turn 0, 0.5 and 0.9. Estimates of size were obtained for the 5% and 1% significance levels. Estimates of power for the 5% and 1% significance levels were obtained by applying the test to series generated from the following processes: AR(1) $\phi = 0$ versus AR(1), $\phi > 0$, AR(1) $\phi = 0$ versus AR(2) $\phi_1 = 0$, $\phi_2 > 0$, and AR(1) $\phi = 0.5$ versus AR(1) $\phi \neq 0.5$. This procedure was repeated while assuming that the correlation between disturbances of each pair of processes from which the series were generated was in turn 0, 0.5 and 0.9.

The order (up to 10) of the truncated autoregressive model to be fitted to each series was determined by Schwartz's BIC (cf. Box, Jenkins and Reinsel 1994, p. 201). However in estimating the model in (2.2), the maximum determined order k was fitted to both the series in each pair. The test statistic D was then obtained. This process was repeated 2000 times. As

Table 1

Overall Estimates of Size for $T = 200$

Process	Level of Significance	Correlation 0	0.5	0.9
AR(1) ϕ=0	5%	0.073*	0.073*	0.051
	1%	0.018*	0.019*	0.011
ϕ=0.1	5%	0.074*	0.077*	0.053
	1%	0.018	0.020*	0.100
ϕ=0.5	5%	0.074*	0.074*	0.046
	1%	0.022*	0.020*	0.013
ϕ=0.9	5%	0.083*	0.068*	0.070*
	1%	0.023*	0.013	0.020*
MA(1)θ=0.1	5%	0.084*	0.060	0.061*
	1%	0.027*	0.017	0.013*
θ=0.5	5%	0.080*	0.069*	0.058
	1%	0.015	0.019*	0.013
θ=0.9	5%	0.088*	0.099*	0.086*
	1%	0.022*	0.025*	0.026*
AR(2)	5%	0.070*	0.068*	0.063*
ϕ_1=0.6 ϕ_2=0.2	1%	0.019	0.021*	0.010
MA(2)	5%	0.088*	0.077*	0.075*
θ_1= 0.8 θ_2=0.6	1%	0.022*	0.022*	0.015*
ARMA(1,1)	5%	0.085*	0.072*	0.072*
ϕ=0.8 θ=0.2	1%	0.024*	0.016	0.018*

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

well as obtaining size and power estimates for the various degrees of freedom, overall estimates of power and size were also obtained by aggregating the respective estimates over the various degrees of freedom.

For $T = 50$, size was considerably overestimated, and estimates of the mean, variance, and skewness did not correspond closely to the respective theoretical values. Because the test statistic has an asymptotic distribution, it seems that series of length 50 are not long enough for the test statistic to display this asymptotic behavior. Hence, no further analysis was performed on series of this length.

For $T = 200$, the results for which there were at least 100 test statistics corresponding to particular degrees of freedom were recorded. The size estimates for series generated from the autoregressive models were fairly close to the prespecified significance levels, when the correct order k was

Table 2

Overall Power Estimates for T=200 (AR(1) ϕ=0 versus AR(1) ϕ>0)

Process	Level of Significance	Correlation 0	0.5	0.9
AR(1) ϕ				
0	5%	0.073	0.073	0.051
	1%	0.018	0.014	0.011
0.1	5%	0.203	0.264	0.854
	1%	0.074	0.118	0.681
0.2	5%	0.549	0.721	1.000
	1%	0.312	0.495	1.000
0.3	5%	0.859	0.972	
	1%	0.672	0.903	
0.4	5%	0.985	1.000	
	1%	0.937	0.993	
0.5	5%	0.998		
	1%	0.994		

fitted. However, size was often overestimated for other values of k. For the series generated from the MA and ARMA models, the size estimates were fairly close to the pre-specified significance levels for some values of k but were overestimated for other values of k. Because of this problem, the overall size was slightly overestimated. These overall size estimates are shown in Table 1 from where it can be seen that in most cases size improves (that is, gets closer to the nominal 5% and 1% levels) as the correlation between disturbances of processes from which the series were generated gets larger.

For those values of k for which reasonably good size estimates were obtained, the estimates of the means, variances, and skewness of the test statistic were very often observed to be fairly close to their respective theoretical values.

Power estimates based on at least 100 test statistics corresponding to particular degrees of freedom and overall power estimates based on all 2000 test statistics were obtained, and from these results it was observed that the test appeared to have reasonably good power for series of length T = 200. Some overall power estimates are given in Table 2, from where it can also be seen that the power of the test improves as the correlation between disturbances of processes from which the series were generated gets larger.

3.2 Power Comparisons

Because other tests in the literature are applicable to independent series only, we compared the power of the test D (which we shall refer to as the autoregressive fitted models test (AR)) for independent series to some of the other tests, namely, those proposed by Jenkins (1961), Diggle and Fisher (1991), and Swanpoel and Van Wyk (1986). There is no evidence in the literature that Jenkins's (1961) test was previously simulated, whereas Diggle and Fisher (1991), and Swanepoel and Van Wyk (1986) simulated their tests and obtained estimates of size and power. All these tests compared two independent stationary time series by comparing their estimated spectra.

When the series were generated from the following processes: MA(1) $\theta = 0$ versus MA(1) $\theta > 0$, and MA(1) $\theta = 0.5$ versus MA(1) $\theta \neq 0.5$, we found that when power comparisons were made, the autoregressive fitted models test had better power than the other tests at both the 5% and 1% levels of significance. Power curves for the 5% level of significance are shown in Figures 1-2.

Comparisons which were also made when the series were generated from the following processes: AR(1) $\phi = 0$ versus AR(1), $\phi > 0$; AR(1) $\phi = 0$ versus AR(2) $\phi_1 = 0$, $\phi_2 > 0$; and AR(1) $\phi = 0.5$ versus AR(1) $\phi \neq 0.5$ produced similar results. While power comparisons using series generated from autoregressive processes favor the AR test over the others because of the autoregressive nature of the test statistic, it is clear the test still performs well in a fairer situation, that is, when the series were generated from the moving average processes.

The reasonably good performance of the test for the simulated scenarios would appear to justify its use in the clustering algorithm that follows in the next section.

4. Clustering Procedure

The clustering procedure that we now consider has the following steps: First perform the test of hypotheses for every pair of series,

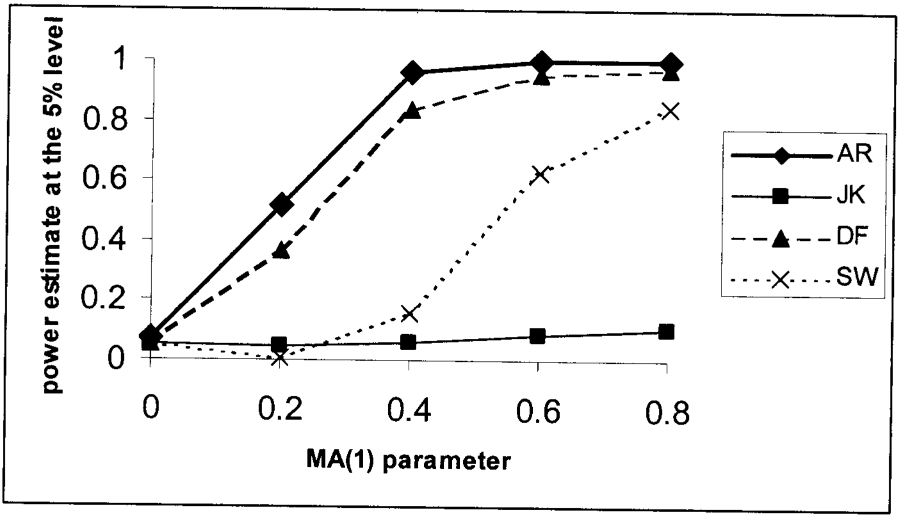

Figure 1. Power Curve for MA(1) θ=0 versus MA(1) θ>0

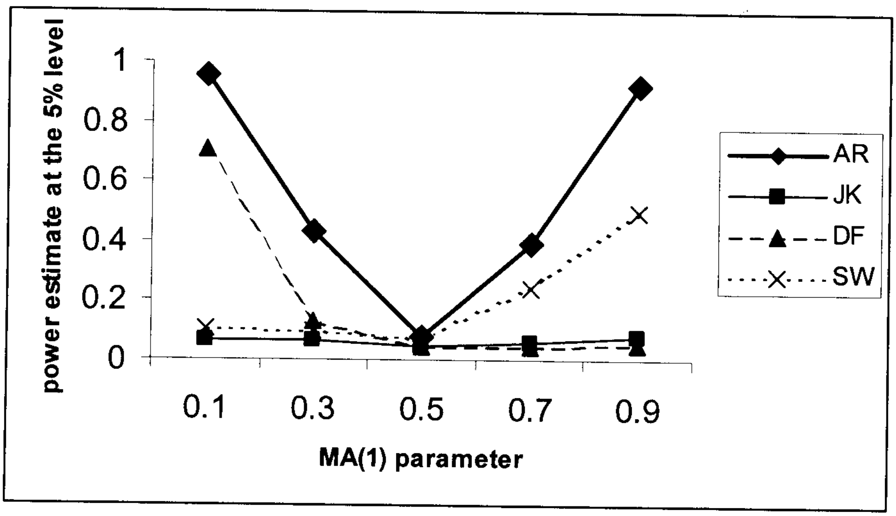

Figure 2. Power Curve for MA(1) θ=0.5 versus MA(1) θ≠0.5

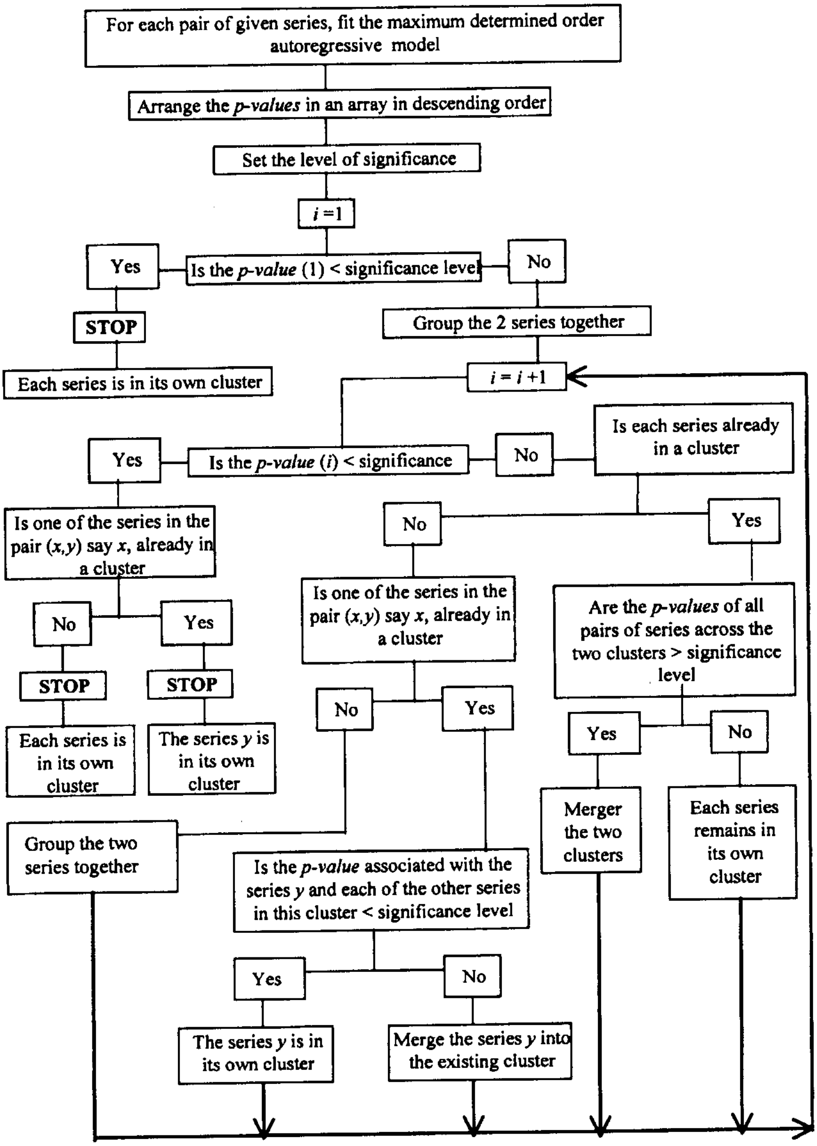

Figure 3. Clustering Algorithm

Table 3

Distribution of MD_1 : Measure of Discrepancy between the True and Obtained Clustering

MD_1	Correlation			
	0		0.5	
	Level of Significance			
	5%	1%	5%	1%
	Frequency			
0	1	4	35	57
1	8	6	49	20
2	147	407	381	683
3	385	446	381	213
4	368	133	140	25
5	91	40	14	2

determining the *p-value* associated with the test *D*. Use these *p-values* in an algorithm[2] (see the flow chart in Figure 3) that incorporates the principles of agglomerative algorithmic approach to hierarchical clustering (cf. Sneath and Sokal 1973, Ch. 10) but will only group together those series whose associated *p-values* are greater than some user prespecified significance level (for example, 5% or 1%). If one of the series, say *x*, is already in a cluster, then the other series, say *y*, will merge into this cluster if the *p-value* associated with *y* and every other series in this cluster is greater than the prespecified significance level. If each series from a pair under consideration is in a different cluster, then the two clusters will merge if the *p-values* of all pairs of series across the two clusters are greater the prespecified significance level.

4.1 Simulation Study to Assess the Clustering Procedure

To evaluate the performance of the clustering algorithm, four series of length 200 were simulated from each of the following processes: AR(1): $\phi = 0.5$; MA(1): $\theta = 0.7$; AR(2): $\phi_1 = 0.6$, $\phi_2 = 0.2$; MA(2): $\theta_1 = 0.8$, $\theta_2 = -0.6$; ARMA(1,1): $\phi = 0.8$, $\theta = 0.2$, and the clustering algorithm was applied. This process was repeated 1000 times.

2. The software for this algorithm is available from the author on request.

Table 4

Distribution of MD_2 : Measure of Discrepancy between the True and Obtained Clustering

MD_2	Correlation			
	0		0.5	
	Level of Significance			
	5%	1%	5%	1%
	Frequency			
0	228	666	248	663
1	410	274	398	286
2	265	58	279	47
3	90	2	69	4
4	7	0	6	0

We assess the performance of this clustering procedure for series whose generating processes are assumed to be uncorrelated (correlation between the disturbances of every two generating processes = 0) as well as for series whose generating processes are assumed to be correlated (correlation between the disturbances of every two generating processes = 0.5). The level of significance was in turn set at 5% and 1%.

For each of the 1000 simulations, the number of exactly correct clusters produced at each simulation (clusters containing the four series from the same generating process) was observed. Now, because four series were generated from each of five different processes, the true number of correct clusters is five (that is, one would expect the four series generated from the same process to cluster together). A measure of discrepancy defined as: MD_1 = [5 - number of exactly correct clusters] was determined for each of the 1000 simulations. The distribution of MD_1 for each scenario is given in Table 3. It can be seen that for the uncorrelated scenario, at both the 5% and 1% levels of significance, and for the correlated scenario at the 5% level of significance, the clustering algorithm performs poorly because the discrepancy is too high. However for the correlated scenario, the performance of the clustering algorithm is more reasonable at the 1% level of significance. For both the uncorrelated and correlated scenarios, the algorithm performs better at the 1% than at the 5% level of significance as would be expected, while at each significance level it performs better for related than for independent series.

The simulated results were examined further to try to establish reasons for this algorithm's poor performance. One possible reason is the number of times some series that were generated from the AR(1), AR(2), and ARMA(1,1) processes came together to form a mixed cluster. Statistics on mixed clusters revealed that far too much mixing occurred. Clearly the test of hypothesis cannot always distinguish between these three processes. This conclusion was supported by further simulated results, showing that the test does not have very high power for series generated from the pairs of processes (AR(1); AR(2)), (AR(1); AR(2)), and (AR(2), ARMA(1,1)).

To determine if the algorithm will perform any better for series simulated from processes that are forced to be quite different from each other, the algorithm was applied to a new set of sixteen series, each of length 200. Four series were generated from each of the following processes: AR(1): $\phi = 0$; AR(1): $\phi = 0.5$; MA(1): $\theta = 0.9$; and ARMA(1,1) $\phi = -0.6$ and $\theta = 0.3$. The true number of correct clusters is now four. The distribution of the measure of discrepancy, MD_2 = (4 - number of exactly correct clusters) is given in Table 4.

Clearly, the performance of the algorithm is now much better than it was for the previous set-up. For all four scenarios, the discrepancy is relatively lower than before, with MD_2 having a higher cumulative frequency for 0 and 1 than for 2, 3, and 4. No mixed clusters were produced. For both the uncorrelated and correlated scenarios, the algorithm performs much very better at the 1% than at the 5% level of significance. However there is very little difference in the distribution of MD_2 for related and independent series.

Thus, from the results of these simulation studies, it can be seen that the clustering algorithm performs poorly when some groups of series are generated from fairly similar processes whereas it performs reasonably well on groups of series that are generated from clearly distinguishable processes.

5. Application

To see how our clustering algorithm performs on empirical data, we consider time series of the number of dwelling units financed by all lenders (banks and other institutions) in the states and territories of Australia[3] from January 1978 to March 1998. Clearly, these series are related because they are all influenced by the same economic factors. The natural logarithmic transformation was taken for each series, and they are shown in Figure 4

3. ACT:Australian Capital Territory, NSW:New South Wales, NT:Northern Territory, QLD:Queensland, SA:South Australia, TAS: Tasmania, VIC:Victoria, WA:Western Australia (Source: Australian Bureau of Statistics - web address: gopher://gopher.abs.gov.au:70/11/PUBS/finan/56090)

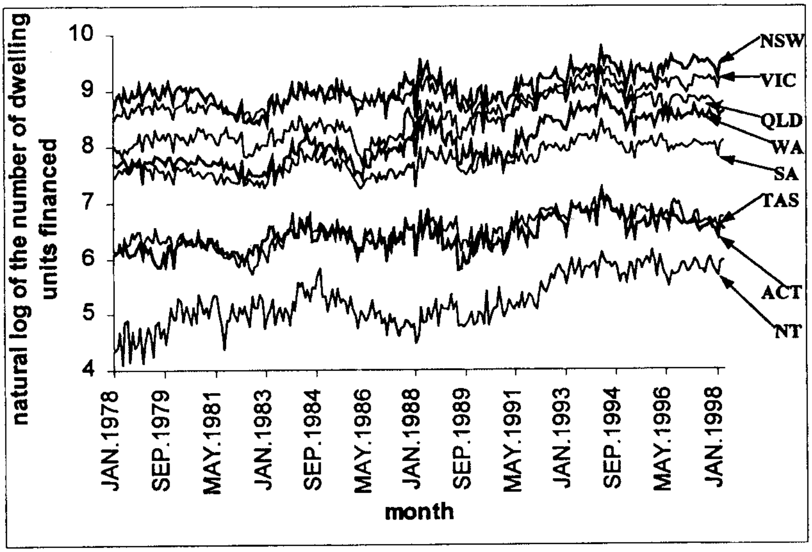

Figure 4. Number of dwelling units financed from January 1978 to March 1998 for all states and territories in Australia

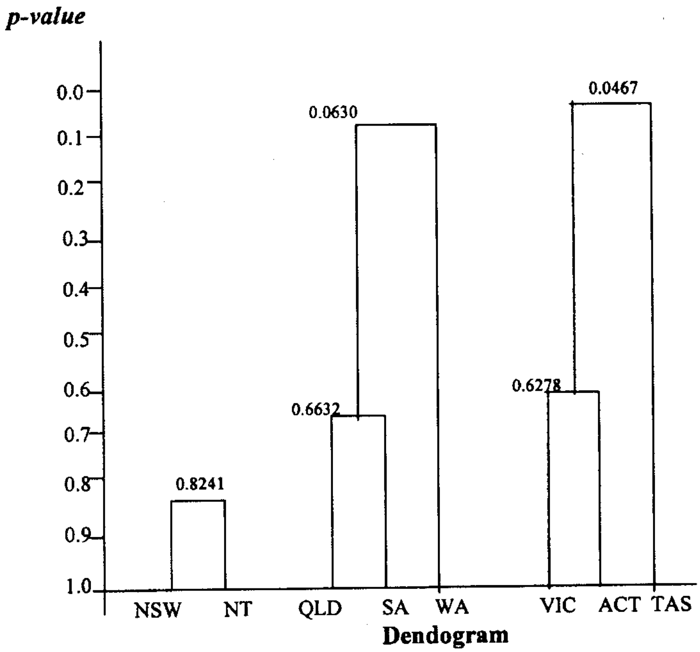

Figure 5. Clusters of stationary series: Number of dwelling units financed from January 1978 to March 1998 for all states and territories in Australia

Table 5

Associated *p-value* for each pair of stationary series: Number of dwelling units financed from January 1978 to March 1998 for all states and territories in Australia

Pair		*p-value*	Pair		*p-value*
NSW	NT	0.8241	NSW	TAS	0.0421
QLD	SA	0.6632	VIC	WA	0.0397
VIC	ACT	0.6278	WA	TAS	0.0342
TAS	ACT	0.6116	VIC	NT	0.0228
QLD	NT	0.4302	NSW	VIC	0.0095
NT	ACT	0.4241	WA	ACT	0.0070
TAS	NT	0.3293	QLD	TAS	0.0064
NSW	QLD	0.2427	SA	NT	0.0054
QLD	ACT	0.1768	SA	ACT	0.0053
NSW	ACT	0.0781	WA	NT	0.0035
SA	WA	0.0720	NSW	SA	0.0026
VIC	QLD	0.0631	NSW	WA	0.0004
QLD	WA	0.0630	SA	TAS	0.0000
VIC	TAS	0.0467	VIC	SA	0.0000

from which it can be seen that the series are not stationary. Some series appear to have fairly similar patterns but one cannot clearly distinguish how similar or how different these patterns are. First differencing of each of these series appeared to render them stationary. The algorithm was then applied to these differenced series.

When the level of significance was set at 5%, the algorithm produced the following clusters (QLD, SA, WA), (NSW, NT), (VIC, ACT), and (TAS). However when the significance level was set at 1%, the following clusters (QLD, SA, WA), (NSW, NT), (VIC, ACT, TAS) were produced. Inspection of the *p-values* of the test of each of the pairs of series (VIC, TAS), (VIC, ACT), and (ACT, TAS) revealed that for the pair (VIC, TAS), the *p-value* was 0.0467. Hence at the 1% level of significance, TAS merged with VIC and ACT. Table 5 shows the *p-values* associated with every pair of series, and Figure 5 which is a variation of the conventional dendogram shows how the clusters form.

These results show that it is possible for series on different levels but having similar patterns to cluster together. The reason is that while the test of hypotheses differentiates between the stochastic nature of series it does not differentiate between their corresponding deterministic natures. This result in no way affects any further analysis of the series in each cluster. For example, if we are analyzing a large number of series and we obtain forecasts of a representative from each cluster, then these forecasts will apply to all series in a particular cluster and in turn to all the corresponding original series. This advantage will, of course, be achieved by reversing the operation of differencing and any other transformation of the stationary series that would have been carried out to render the original series stationary in the first place.

6. Concluding Remarks

The test of hypotheses as well as the clustering technique can be applied to independent as well as to related time series of reasonable length. The simulation results show that, for series of reasonable length, the distributional approximations of the test statistic to the chi-square distribution are fairly adequate and that the clustering algorithm performs reasonably well under some circumstances only. The results of the application of the algorithm to empirical data appear to be quite reasonable.

Conventional hierarchical clustering methods have no rule for deciding on the number of clusters; that is, the decision is usually subjective. By contrast, our clustering algorithm has a rule for deciding on the number of clusters to choose, even though the choice of the significance level remains somewhat subjective.

This application of the algorithm is restricted to reasonably long stationary series that can be modeled as linear processes. Because in practice, we may encounter short or stationary series that cannot be fitted by linear models or nonstationary series that cannot be easily tranformed to stationary series, further research is being undertaken to extend our clustering algorithm so that it can be more generally applied.

References

AMEMIYA, T. (1985), *Advanced Econometrics*, Cambridge, MA: Harvard University Press.

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

BOHTE, Z. D., CEPAR, D., AND KOSMELU K. (1980), "Clustering of Time Series", *Proceedings in Computational Statistics 1980,* 587-593.

BOX, G. E. P., and JENKINS, G. M. (1976), *Time Series Analysis: Forecasting and Control*, San Francisco, CA: Holden Day.

BOX, G. E. P., JENKINS, G. M., and REINSEL, G.C. (1994), *Time Series Analysis: Forecasting and Control*, 3rd edition, Englewood Cliffs, NJ: Prentice Hall.

COATES, D. S., AND DIGGLE, P. J. (1986), "Test for Comparing Two Estimated Spectral Densities", *Journal of Time Series Analysis*, 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, 30*, 2, 304-315.

DIGGLE, P. J., AND FISHER, N. I. (1991), "Nonparametric Comparison of Cumulative Periodograms", *Applied Statistics*, 40, 423-434.

KOSMELJ, K., and BATAGELJ, V. (1990), "Cross-Sectional Approach for Clustering Time-Varying Data", *Journal of Classification*, 7, 99-109.

JENKINS, G. M., (1961), "General Considerations in the Analysis of Spectra" *Technometrics*, 3, 133-166.

MAHARAJ, E. A. (1996), "A Significance Test for Classifying ARMA Models", *Journal of Statistical Computation and Simulation*, 54, 305-331.

MAHARAJ, E. A. (1997), *Pattern Recognition Techniques for Time Series*, unpublished Ph.D. thesis, Melbourne, Australia: Department of Econometrics, Monash University.

PICCOLO, D. (1990), "A Distance Measure for Classifying ARIMA Models", *Journal of Time Series Analysis,* 11, 2, 153 - 164.

SHAW, C. T., AND KING G. P. (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, Amsterdam: North-Holland.

SNEATH, P. H., and SOKAL, R. R. (1973), *Numerical Taxonomy*, San Francisco: Freeman.

SWANEPOEL, J. W. H., and VAN WYK, J. W. J. (1986), "The Comparison of Two Spectral Density Functions using the Bootstrap", *Journal of Statistical Computation and Simulation*, 24, 271-282.

TONG, H., and DABAS P. (1990), "Clusters of Time Series Models: An Example", *Journal of Applied Statistics* 17, 187-198.

ZELLNER, A. (1962), "Estimators for Seemingly Unrelated Regressions Equations and Test of Aggregation Bias", *Journal of the American Statistical Association*, 57, 500-509.