MONASH
UNIVERSITY

AUSTRALIA

A SIGNIFICANCE TEST FOR CLASSIFYING ARMA MODELS

Elizabeth Ann Maharaj

Working Paper No. 18/94

September 1994

DEPARTMENT OF ECONOMETRICS

ISSN 1032-3813

ISBN 0 7326 0759 0

A SIGNIFICANCE TEST FOR CLASSIFYING ARMA MODELS

Elizabeth Ann Maharaj

Working Paper No. 18/94

September 1994

DEPARTMENT OF ECONOMETRICS

MONASH UNIVERSITY, CLAYTON, VICTORIA 3168, AUSTRALIA.

A SIGNIFICANCE TEST FOR CLASSIFYING ARMA MODELS

By *ELIZABETH ANN MAHARAJ*

Department of Econometrics

MONASH UNIVERSITY

ABSTRACT

Given that the Euclidean distance between the parameter estimates of autoregressive expansions of autoregressive moving average models can be used to classify stationary time series into groups, a test is proposed to determine whether or not two stationary time series in a particular group have significantly different generating processes. The results of computer simulations are given.

1. INTRODUCTION

The classification of time series has applications in various fields, some of which are geology, economics, oceanography, psychology and engineering. In particular classification of time series using cluster analysis has been demonstrated by various authors.

Bohte *et al.* (1980) defines a number of distance measures which are based on autocorrelations and/or cross correlations of empirical time series. The time series are classified into groups according to one of the distance measures and occurs over several stages. At each stage a group is separated from the rest of the time series and a specific ARIMA model is then adopted for that group. The model is then fitted to one representative of that group. This method of clustering avoids having to fit ARIMA models to a large number of time series.

Piccolo (1990) proposes fitting ARIMA models to all time series in a given set and then classifying these fitted time series into groups according to a distance measure that is based on the coefficients of the

AR(∞) operator of the fitted ARIMA model. Tong *et al.* (1990) uses various measures of similarity and dissimilarity based on the residuals of the fitted ARIMA and bilinear models and uses them to classify various time series. Shaw *et al.* (1992) determines power spectra of various time series and then obtains Euclidean distances between the power spectra using 256 frequency values. Cluster Analysis is then applied to these Euclidean distances.

In this paper we will consider the distance measure proposed by Piccolo (1990) but base it instead on stationary and invertible ARMA models and develop a related test of significance.

2. DEFINITIONS

2.1 Distance Measure

Let Z_t be a zero mean univariate stochastic process and a_t be a univariate Gaussian white noise process i.e $a_t \sim IN(0,\sigma_a^2)$. Then Z_t is such that $Z_t \in L$, where L is the class stationary and invertible ARMA models. Using the standard notation of Box and Jenkins (1976), such a model is defined as

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

where

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots\dots\dots - \phi_p B^p ,$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots\dots\dots - \theta_q B^q .$$

Z_t can be expressed as

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

where $\Pi(B)$ is the the AR(∞) operator and is defined as

$$\Pi(B) = \phi(B)\theta^{-1}(B) = 1 - \pi_1 B - \pi_2 B^2 - \dots\dots\dots$$

By Piccolo (1990) a measure of structural diversity between $X_t \in L$ and $Y_t \in L$ can be obtained by comparing respective sequences and this assigns a metric on L, which is the distance measure

$$d(X,Y) = \left[\sum_{j=1}^{\infty} (\pi_{jx} - \pi_{jy})^2 \right]^{1/2}$$

where π_{jx} and π_{jy} are the coefficients of the AR(∞) operators of X_t and Y_t respectively.

Given several time series to be classified, instead of ARMA modelling of each time series, automatic modeling of AR structures by means of a definite criterion like Akaike's Information Criterion is reasonable. Hence time series can be clustered by grouping fitted AR models directly from a matrix of distance measures of the form d(X,Y).

Define the vector of AR(k_1) and AR(k_2) parameters for the processes X_t and Y_t respectively as

$$\Pi_x = [\ \pi_{1x} \quad \pi_{2x} \quad \dots\dots\pi_{k_1x}]^T \qquad \text{and}$$

$$\Pi_y = [\ \pi_{1y} \quad \pi_{2y} \quad \dots\dots\pi_{k_2y}]^T , \qquad \text{respectively,}$$

and the vector of AR(k_1) and AR(k_2) parameter estimates of the series x_t and y_t respectively as

$$\hat{\Pi}_x = [\ \hat{\pi}_{1x} \quad \hat{\pi}_{2x} \quad \dots\dots\hat{\pi}_{k_1x}]^T \qquad \text{and}$$

$$\hat{\Pi}_y = [\ \hat{\pi}_{1y} \quad \hat{\pi}_{2y} \quad \dots\dots\hat{\pi}_{k_2y}]^T , \qquad \text{respectively,}$$

Hence the distance measure $d(X,Y)$ becomes

$$d(X,Y) = \left[(\hat{\Pi}_x - \hat{\Pi}_y)^T (\hat{\Pi}_x - \hat{\Pi}_y)\right]^{1/2}$$

$$= \left[\sum_{j=1}^{k} (\hat{\pi}_{jx} - \hat{\pi}_{jy})^2 \right]^{1/2}$$

where $k = \max(k_1, k_2)$ and k_1 and k_2 are the orders of the AR models fitted to the series x_t and y_t respectively.

If $k = k_1$ then $\hat{\pi}_{jy} = 0$ for $j = k_2 + 1, k_2 + 2, \ldots\ldots, k$.

If $k = k_2$ then $\hat{\pi}_{jx} = 0$ for $j = k_1 + 1, k_1 + 2, \ldots\ldots, k$.

2.2 Test of Hypothesis

We now propose a significance test to determine whether or not two finite stationary series in a particular group have significantly different generating processes. If there is no significant difference between the generating processes of all the series in a particular group, then any series in the group can be regarded as a representative of the group. The test is of the hypotheses:

H_0: There is no difference between the generating processes of two stationary series i.e. $\Pi_x = \Pi_y$.

H_1: There is a difference between the generating series of two stationary series i.e. $\Pi_x \neq \Pi_y$.

Berk (1974) has truncated the infinite order AR process to order k and has obtained the AR estimates by the method of least squares. This has been done by assuming that k is chosen as a function of T, such that

$$\frac{k^3}{T} \rightarrow 0 \quad \text{and} \quad \sqrt{T} \sum_{j=k+1}^{\infty} |\pi_{jx}| \rightarrow 0$$

as $T \to \infty$, where T is the length of the stationary series x_t to which the AR(k) model is fitted.

Using the results of Berk (1974), Bhansali (1978) has derived the asymptotic normal distribution of the AR estimates which is

$$\sqrt{T}\ (\hat{\Pi}_x - \Pi_x) \sim N(0,\ \sigma^2_{ax} R^{-1}_x(k))$$

where σ^2_{ax} is the variance of a_{xt}, the white noise process associated with the series x_t and $R_x(k)$ is the upper kxk submatrix of an infinite dimensional covariance matrix

$$R = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1k} & \cdots \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2k} & \cdots \\ \vdots & \vdots & & \vdots & \\ \sigma_{k1} & \sigma_{k2} & \cdots & \sigma_{kk} & \cdots \\ \vdots & \vdots & & \vdots & \cdots \\ \vdots & \vdots & & \vdots & \cdots \end{bmatrix} .$$

For two independent series to which truncated AR(k) models are fitted

$$\hat{\Pi}_x \stackrel{A}{\sim} N(\Pi_x,\ \sigma^2_{ax} R^{-1}_x(k)/T)$$

$$\hat{\Pi}_y \stackrel{A}{\sim} N(\Pi_y,\ \sigma^2_{ay} R^{-1}_y(k)/T)$$

Therefore

$$\hat{\Pi}_x - \hat{\Pi}_y \stackrel{A}{\sim} N(\Pi_x - \Pi_y\ ,\ V)$$

where

$$V = 1/T\ (\sigma^2_{ax} R^{-1}_x(k) + \sigma^2_{ay} R^{-1}_y(k))$$

$R_x^{-1}(k)$ and $R_y^{-1}(k)$ are both positive definite and since

$$\sigma_{ax}^2 > 0 \quad \text{and} \quad \sigma_{ay}^2 > 0$$

$\sigma_{ax}^2 R_x^{-1}(k)$ and $\sigma_{ay}^2 R_y^{-1}(k)$ are both positive definite. Therefore V is positive definite.

Hence $(\hat{\Pi}_x - \hat{\Pi}_y)^T V^{-1} (\hat{\Pi}_x - \hat{\Pi}_y)$ follows a noncentral chi-squared distribution with k degrees of freedom and noncentrality parameter

$$\tau = (\Pi_x - \Pi_y)^T V^{-1} (\Pi_x - \Pi_y).$$

I.e. $(\hat{\Pi}_x - \hat{\Pi}_y)^T V^{-1} (\hat{\Pi}_x - \hat{\Pi}_y) \overset{A}{\sim} \chi^2(k,\tau)$.

Hence the proposed test statistic is

$$D(X,Y) = (\hat{\Pi}_x - \hat{\Pi}_y)^T \hat{V}^{-1} (\hat{\Pi}_x - \hat{\Pi}_y)$$

Now since V is nonsingular and

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

$$D(X,Y) \overset{A}{\sim} \chi^2(k,\tau) .$$

Under H_0 $\tau = 0$, therefore $D(X,Y) \overset{A}{\sim} \chi^2(k)$.

Hence H_0 is rejected at the 100α% level of significance if $D(X,Y) > \chi^2_\alpha(k)$, where $\chi^2_\alpha(k)$ is the $(1-\alpha)$th quantile of the chi-square distribution with k degrees of freedom. If H_0 is rejected, we conclude that the generating processes of the series x_t and y_t are significantly different from each other.

Since D(X,Y) satisfies the properties of non-negativity and symmetry it can also be used as a distance measure by which series may be clustered.

3. ILLUSTRATION

To illustrate the use of this test statistic both in hypothesis testing and as a measure of classification, twenty one series of 200 observations each were simulated from AR(1), MA(1), AR(2), MA(2) and ARMA(1,1) models. Each of these series was then fitted with truncated AR(k) models, with the order k selected using Akaike's Information Criterion. For every pair of series, the Euclidean distance measures d(X,Y) and value of the test statistic D(X,Y) were calculated. Clustering was performed using the d(X,Y) as well as the D(X,Y) values. Hierarchical methods of clustering such as average, simple and complete methods and the Ward method were considered.

The series were labelled as follows: if a series was simulated from an MA(1) model with parameter 0.3, then the label was MA13; if a series was simulated from an ARMA(1,1) model, with parameters -0.2 and 0.5, then the label was ARMA_25.

Power calculations for $\alpha = 0.05$ and $\alpha = 0.01$ were performed for the test when series were generated from the following pairs of models:

AR(1) $\phi = -0.3$ and MA(1) $\theta = 0.3$

AR(1) $\phi = -0.3$ and AR(1) $\theta = 0.3$.

3.1 Results

The estimates of the series after they were fitted with truncated AR(k) models, the Euclidean distance values d(X,Y) and the test statistics D(X,Y) for comparing every pair of series appear in Tables 1,2 and 3 in the Appendix. Agglomeration schedules for each of the four methods of clustering performed on d(X,Y) and D(X,Y) measures appear in Tables 5 to 11. The corresponding dendrograms appear in Figures 1 to 8 and the graphs of the power functions are shown in Figures 9 to 12 in the Appendix.

3.2 Comments.

By examining the dendrograms in the Appendix, it is clear the two distinct clusters form, regardless of which method of clustering was used and regardless of whether d(X,Y) of D(X,Y) was used. These clusters are as a result of the amalgamation of 5 to 7 smaller clusters. We shall examine these smaller clusters to determine whether or not there are significant differences between the generating processes of the series in these clusters.

Consider for example the clusters obtained by the average linkage method using d(X,Y) and D(X,Y). These appear in Figures 1 and 5. In what follows the series shall be referred to by their numbers instead of their generating processes.

Measure	Clusters
d(X,Y)	(1,2,3) (4,5) (6,7,8) (9,10) (11,12,13) (14,15,16,17) (18,19,20,21)
D(X,Y)	(18,19,20,11,12,21) (14,15,16,17) (13) (6,8) (1,2,3,4,9,5,7) (10)

Consider the cluster (18,19,20,11,12,21) which forms using the D(X,Y) measures. From the agglomeration schedule it can be seen that this cluster forms at stage 14. Comparing the D(X,Y) values with the appropriate chi square critical values, it can be seen that there is no significant difference between the generating processes of the following pairs in this cluster: (18,19), (18,20), (18,11), (18,21), (19,20), (21,12) and (11,12). However there is a significant difference between the generating processes of each of the series in the cluster

(18,19,20,21,11,12) and all the other series in the other clusters. At stage 14, when using the d(X,Y) measure, the series 18,19,20,21,11,12 appear in two clusters,i.e. (11,12,13) and (18,19,20,21). As well as there being no significant difference in generating processes between some pairs in each cluster, there is also no significant difference between the generating processes of one or more series of (11,12,13) and one or more series of (18,19,20,21) at this stage. For example, there is no significant difference between the generating process of 18 and 11, but they appear in two different clusters. However these clusters merge at stage 16.

Consider the cluster (1,2,3,4,9,5,7) which forms when using the D(X,Y) measures. From the agglomeration schedule it can be seen that this cluster forms at stage 13. Comparing the D(X,Y) values to the appropriate chi-square critical values, it can be seen that there is no significant difference between the generating processes of the following pairs in this cluster: (1,2), (1,3), (1,4), (2,3), (2,4), (3,4), (5,9) and (5,7). However there is a significant difference between the generating processes of each of the series in the cluster (1,2,3,4,9,5,7) and all the series in the other clusters. However at stage 13, when using the d(X,Y) measures, the series 1,2,3,4,9,5,7 appear in four clusters (1,2,3), (4,5), (6,7,8) and (9,10). A number of pairs of series which have no significant difference in their generating processes appear in different clusters at this stage. However these clusters merge at stage 19.

Consider the cluster (14,15,16,17) which forms when using the D(X,Y) measures. From the agglomeration schedule it can be seen that this cluster forms at stage 12. Comparing the D(X.Y) values to the appropriate chi-square critical values, it can be seen that there is no significant difference between the generating process of the pairs (14,15) and (15,16). However there is a significant difference between the generating processes of each of the series in the cluster (14,15,16,17) and all the other series in the other clusters. Using the d(X,Y) measures this cluster forms at stage 10.

Consider the cluster (6,8) which forms when using the D(X,Y) measure. There is a significant difference between the generating processes of these series. There is also a significant difference between the generating processes of each of the series in the cluster (6,8) and all the other series in the other clusters. Using the d(X,Y) measure these two series appear together in the cluster (6,7,8).

Using the D(X,Y) measure, the series 10 and 13 each appear on their own. There is a significant difference between each of these series and every other series in the other clusters. Using the d(X,Y) measure, 10 appears in the cluster (9,10) and 13 appears in the cluster (11,12,13).

It is clear from the above results that in some cases it is possible to identify more homogeneous clusters at an earlier stage using the D(X,Y) rather than the d(X,Y) measure. More or less similar results were obtained when the other methods of clustering are used.

From the graphs of the power functions in the Appendix, it can be seen that, for the pairs of generating processes considered, the test is reasonably powerful.

4. CONCLUSION

This simulation study shows that

i) the test statistic D(X,Y) has the ability to test for significance since any two series whose generating processes are quite different from each other, for example AR(1), $\phi = 0.3$ and AR(1), $\phi = -0.3$, are deemed to have significantly different generating processes whereas many series with like generating processes, for example AR(1), $\phi = -0.3$ and MA(1),$\theta = 0.3$, are deemed to have generating processes that are not significantly different from each other.

ii) when either the d(X,Y) or D(X,Y) measures are used, there are pairs of series in some clusters when H_0 is rejected.

iii) when the D(X,Y) measure is used, it is possible to identify more homogeneous clusters, in some cases at earlier stages, than when the d(X,Y) measure is used.

If given a large number of series, one just wishes to identify groups of similar series, then clustering using the distance measure based on parameter estimates of ARIMA models as suggested by Piccolo (1990) is sufficient. However we believe that if there is a need to use one of the series in a cluster as a representative of that cluster, on which further analysis is to be carried out, then there should not be a significant difference between the generating processes of all pairs of series in that cluster.

ACKNOWLEDGEMENT

I am grateful to Dr. N. Singhforhismany helpful comments and suggestions.

REFERENCES

Berk K.N.(1974) Consistent Autoregressive Spectral Estimates, *The Annals of Statistics* 2 (3) 489 - 502.

Bhansali R.J. (1978) Linear Prediction by Autoregressive Model Fitting in the Time Domain. *The Annals of Statistics* 6 (1) 224 - 231.

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

Box G.E.P. and Jenkins G.M. (1976) *Time Series Analysis: Forecastingand Control*. San Francisco, CA: Holden Day.

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.

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

APPENDIX

TABLE 1

AR ESTIMATES OF THE 21 SIMULATED SERIES

1	AR1_3	-0.3123	0.0000	0.0000	0.0000	0.0000
2	MA13	-0.3693	0.0000	0.0000	0.0000	0.0000
3	AR1_5	-0.4883	0.0000	0.0000	0.0000	0.0000
4	MA15	-0.5162	-0.2531	-0.1810	-0.1830	0.0000
5	ARMA_25	-0.7139	-0.4224	-0.2332	0.0000	0.0000
6	AR1_9	-0.9192	0.0000	0.0000	0.0000	0.0000
7	MA27_4	-0.7546	-0.2194	0.0000	0.0000	0.0000
8	MA28_6	-0.9078	-0.1335	0.3265	0.1816	0.0000
9	MA16	-0.6463	-0.4836	-0.3875	-0.2753	-0.1528
10	MA19	-0.7214	-0.7660	-0.5668	-0.3884	-0.1308
11	MA1_3	0.3942	-0.2848	0.1446	0.0000	0.0000
12	MA1_6	0.5145	-0.3849	0.1974	-0.1613	0.0718
13	AR28_7	0.6950	-0.5824	0.0000	0.0000	0.0000
14	AR17	0.7360	0.0000	0.0000	0.0000	0.0000
15	AR18	0.7936	0.0000	0.0000	0.0000	0.0000
16	AR19	0.9001	0.0000	0.0000	0.0000	0.0000
17	AR29_2	0.9348	-0.2982	0.0000	0.0000	0.0000
18	AR13	0.3392	0.0000	0.0000	0.0000	0.0000
19	AR15	0.4378	0.0000	0.0000	0.0000	0.0000
20	MA1_5	0.4723	-0.1366	0.0000	0.0000	0.0000
21	ARMA63	0.2150	0.1501	0.1769	0.0000	0.0000

TABLE 2

EUCLIDEAN DISTANCES d(X,Y)

	AR1_3	MA13	AR1_5	MA15	ARMA_25	AR1_9	MA27_4
MA13	0.0570						
AR1_5	0.1760	0.1190					
MA15	0.4146	0.3897	0.3621				
ARMA_25	0.6278	0.5929	0.5326	0.3224			
AR1_9	0.6069	0.5499	0.4309	0.5410	0.5244		
MA27_4	0.4937	0.4434	0.3450	0.3524	0.3118	0.2743	
MA28_6	0.7156	0.6689	0.5774	0.7471	0.6836	0.3969	0.4128
MA16	0.7712	0.7483	0.7128	0.3802	0.3623	0.7468	0.5752
MA19	1.1151	1.0954	1.0632	0.7165	0.6304	1.0560	0.8883
MA1_3	0.7753	0.8276	0.9385	0.9845	1.1788	1.3517	1.1597
MA1_6	0.9497	0.9997	1.1063	1.1084	1.3141	1.5079	1.3070
AR28_7	1.1635	1.2132	1.3189	1.2813	1.4370	1.7161	1.4944
AR17	1.0483	1.1053	1.2243	1.3032	1.5281	1.6552	1.5067
AR18	1.1059	1.1629	1.2819	1.3586	1.5828	1.7128	1.5637
AR19	1.2124	1.2694	1.3884	1.4616	1.6846	1.8193	1.6692
AR29_2	1.2823	1.3378	1.4540	1.4743	1.6697	1.8778	1.6912
AR13	0.6515	0.7085	0.8275	0.9284	1.1584	1.2584	1.1156
AR15	0.7501	0.8071	0.9261	1.0200	1.2487	1.3570	1.2124
MA1_5	0.7964	0.8526	0.9703	1.0281	1.2422	1.3982	1.2297
ARMA63	0.5761	0.6287	0.7406	0.9267	1.1657	1.1577	1.0526

	MA28_6	MA16	MA19	MA1_3	MA1_6	AR28_7
MA16	0.9658					
MA19	1.2549	0.3617				
MA1_3	1.3357	1.2266	1.4663			
MA1_6	1.4918	1.3277	1.5327	0.2418		
AR28_7	1.7059	1.4346	1.5903	0.4472	0.3765	
AR17	1.6910	1.5472	1.7889	0.4678	0.5171	0.5838
AR18	1.7470	1.5989	1.8361	0.5114	0.5442	0.5907
AR19	1.8509	1.6954	1.9249	0.5983	0.6058	0.6175
AR29_2	1.8873	1.6684	1.8577	0.5598	0.5043	0.3719
AR13	1.3086	1.2060	1.4835	0.3241	0.4990	0.6825
AR15	1.4029	1.2878	1.5555	0.3224	0.4735	0.6367
MA1_5	1.4298	1.2732	1.5200	0.2213	0.3655	0.4983
ARMA63	1.1817	1.2494	1.5611	0.4715	0.6384	0.8934

	AR17	AR18	AR19	AR29_2	AR13	AR15	MA1_5
AR18	0.0576						
AR19	0.1641	0.1065					
AR29_2	0.3584	0.3299	0.3002				
AR13	0.3968	0.4544	0.5609	0.6661			
AR15	0.2982	0.3558	0.4623	0.5796	0.0986		
MA1_5	0.2970	0.3491	0.4491	0.4899	0.1907	0.1409	
ARMA63	0.5703	0.6234	0.7233	0.8662	0.2632	0.3217	0.4239

TABLE 3

TEST STATISTICS D(X,Y)

	AR1_3	MA13	AR1_5	MA15	ARMA_25	AR1_9	MA27_4
MA13	0.3660						
AR1_5	3.7441	1.7325					
MA15	9.8624	7.6324	8.4464				
ARMA_25	22.5608	18.5458	15.7344	10.2739			
AR1_9	69.2090	58.9670	40.2047	55.8865	44.1818		
MA27_4	19.9282	15.2225	10.2311	11.3855	7.9497	32.9655	
MA28_6	46.3183	40.0623	26.7249	42.2334	33.9738	32.0773	10.1108
MA16	25.3122	21.6465	20.8010	6.2253	7.0230	61.1761	17.5964
MA19	51.0537	48.1274	49.3625	20.9524	17.0741	97.6873	38.4936
MA1_3	57.1314	66.5841	88.9541	89.7197	127.0540	208.3570	139.5760
MA1_6	75.3375	86.5357	110.5490	108.6710	151.9040	237.0410	166.2750
AR28_7	160.3040	169.5880	189.5000	222.2790	308.8770	312.6110	293.9150
AR17	160.6220	183.7710	244.2020	173.3190	237.0200	885.9760	237.0780
AR18	191.2540	218.0880	288.7500	328.4380	250.4200	192.0710	258.3760
AR19	266.1230	302.3450	266.7180	229.1270	302.5280	1865.7500	305.8700
AR29_2	163.2890	179.0700	215.3530	218.1680	292.2330	394.0450	305.2210
AR13	47.2576	57.1213	82.7206	79.4615	119.5010	303.6550	122.2910
AR15	65.3561	77.4092	108.4920	98.5649	143.3220	378.9330	146.3640
MA1_5	62.6606	69.0674	95.5856	100.8510	146.0130	206.9780	154.2570
ARMA63	34.0681	38.5621	55.2863	75.3253	112.4480	147.2510	97.3383

	MA28_6	MA16	MA19	MA1_3	MA1_6	AR28_7	AR17
MA16	381.6780						
MA19	84.5307	7.6288					
MA1_3	196.4700	112.1680	135.4170				
MA1_6	288.6310	132.9630	161.2190	3.8684			
AR28_7	325.7460	287.8290	385.5470	25.4920	18.7242		
AR17	283.3830	218.2160	276.1630	35.6465	34.6535	47.3619	
AR18	291.8240	230.6600	284.8820	40.7754	43.4134	44.6588	0.7966
AR19	343.6950	281.2050	349.7240	65.0075	64.7981	40.8713	8.1757
AR29_2	362.4210	263.9170	316.0740	42.6361	121.4440	93.1458	9.3639
AR13	168.9920	107.9300	143.2090	7.8528	13.2457	72.2692	23.3190
AR15	191.8890	129.3150	166.2840	9.7610	14.0132	64.8537	17.0698
MA1_5	208.2340	138.4860	164.3330	5.0507	6.9134	85.2442	59.6413
ARMA63	126.6640	102.1350	139.3650	21.2158	32.5862	79.1096	30.6013

	AR18	AR19	AR29_2	AR13	AR15	MA1_5
AR19	3.9843					
AR29_2	10.9834	22.1573				
AR13	32.7485	57.8994	37.0860			
AR15	20.0783	42.2288	25.9216	0.1412		
MA1_5	27.7051	49.1927	23.9095	2.6482	1.9401	
ARMA63	35.5314	55.1550	60.4560	7.6475	9.5138	17.3889

TABLE 4

AGGLOMERATION SCHEDULE USING AVERAGE LINKAGE FOR d(X,Y) MEASURES

	Clusters Combined			Stage Cluster 1st Appears		Next
Stage	Cluster 1	Cluster 2	Coefficient	Cluster 1	Cluster 2	Stage
1	1	2	.057000	0	0	5
2	14	15	.057600	0	0	4
3	18	19	.098600	0	0	6
4	14	16	.135300	2	0	10
5	1	3	.147500	1	0	15
6	18	20	.165806	3	0	11
7	11	12	.241770	0	0	14
8	6	7	.274280	0	0	13
9	4	5	.322431	0	0	15
10	14	17	.329515	4	0	17
11	18	21	.336238	6	0	16
12	9	10	.361682	0	0	19
13	6	8	.404868	8	0	18
14	11	13	.411815	7	0	16
15	1	4	.486616	5	9	18
16	11	18	.502209	14	11	17
17	11	14	.521962	16	10	20
18	1	6	.532803	15	13	19
19	1	9	.817640	18	12	20
20	1	11	1.308777	19	17	0

FIGURE 1

DENDROGRAM USING AVERAGE LINKAGE FOR d(X,Y) MEASURES

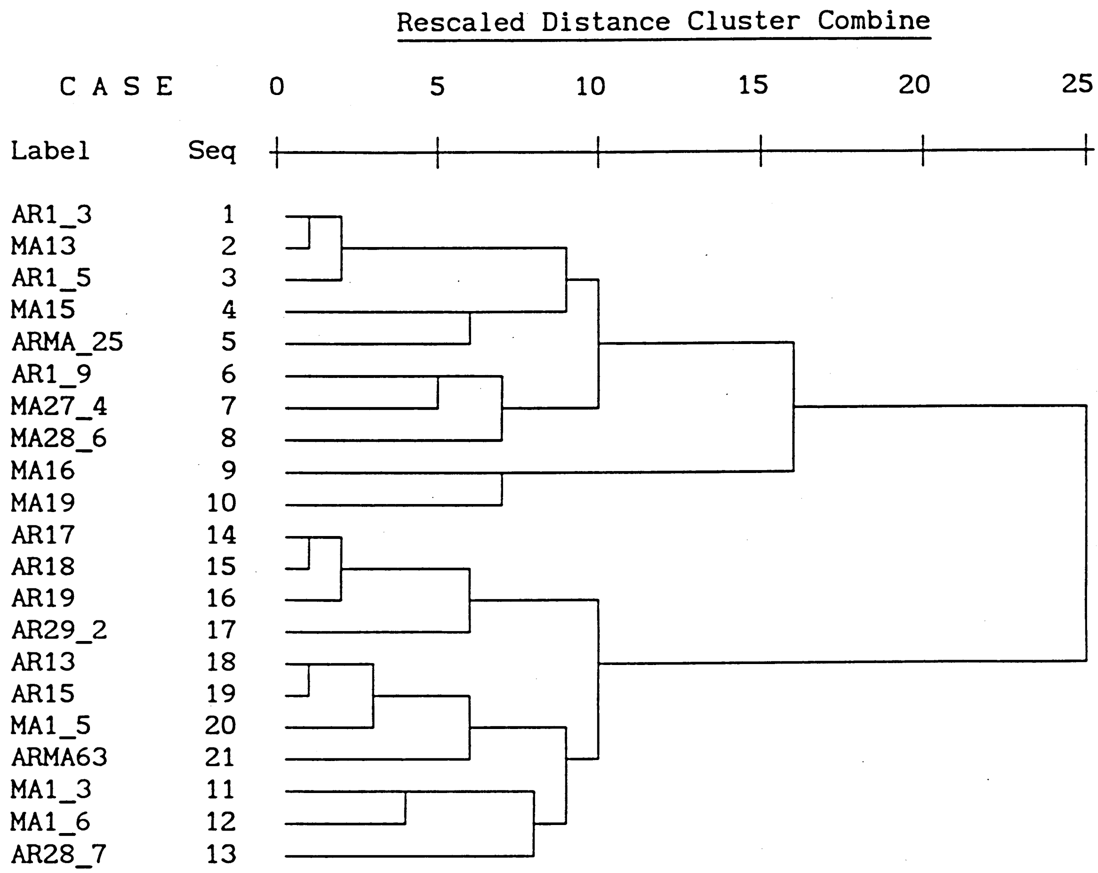

TABLE 5

AGGLOMERATION SCHEDULE USING SINGLE LINKAGE FOR d(X,Y) MEASURES

	Clusters Combined			Stage Cluster 1st Appears		Next
Stage	Cluster 1	Cluster 2	Coefficient	Cluster 1	Cluster 2	Stage
1	1	2	.057000	0	0	5
2	14	15	.057600	0	0	4
3	18	19	.098600	0	0	6
4	14	16	.106500	2	0	11
5	1	3	.119000	1	0	15
6	18	20	.140889	3	0	7
7	11	18	.221296	0	6	8
8	11	12	.241770	7	0	9
9	11	21	.263153	8	0	11
10	6	7	.274280	0	0	13
11	11	14	.296980	9	4	12
12	11	17	.300212	11	0	18
13	5	6	.311846	0	10	14
14	4	5	.322431	0	13	15
15	1	4	.345039	5	14	17
16	9	10	.361682	0	0	17
17	1	9	.362300	15	16	19
18	11	13	.371852	12	0	20
19	1	8	.396904	17	0	20
20	1	11	.576081	19	18	0

FIGURE 2

DENDROGRAM USING SINGLE LINKAGE FOR d(X,Y) MEASURES

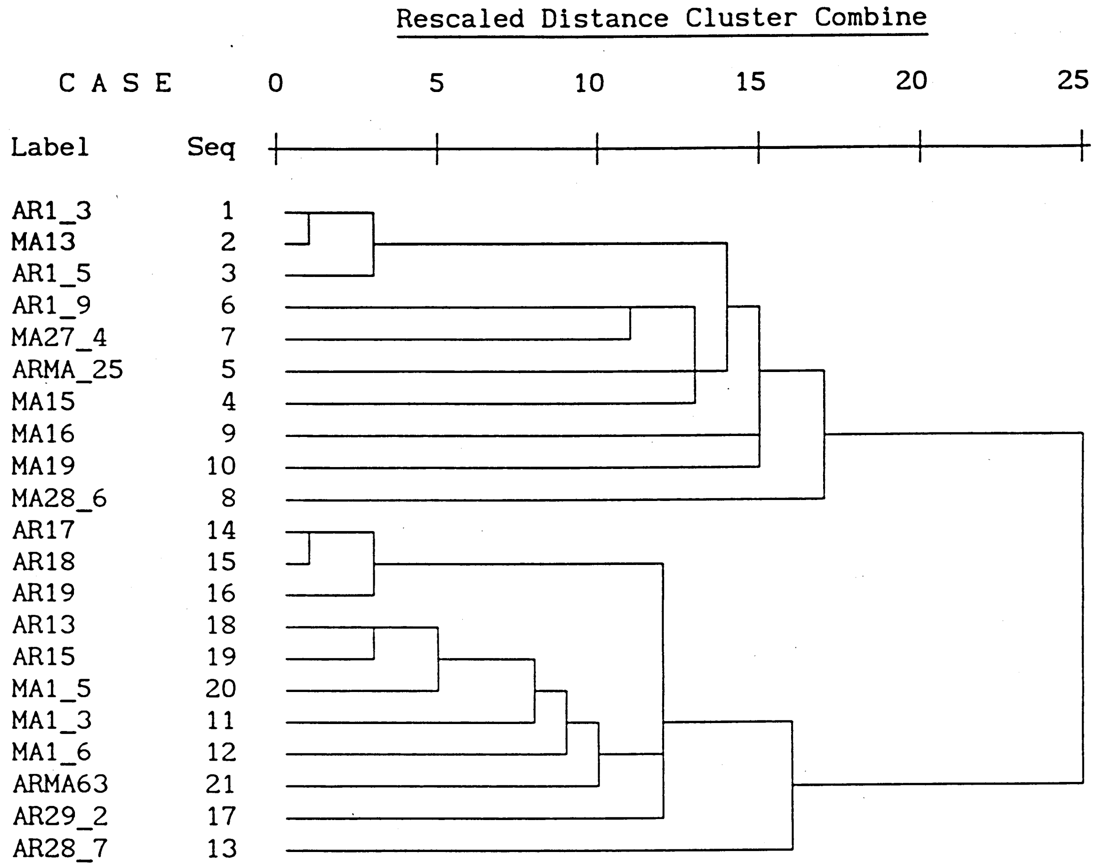

TABLE 6

AGGLOMERATION SCHEDULE USING COMPLETE LINKAGE FOR d(X,Y) MEASURES

	Clusters Combined			Stage Cluster 1st Appears		Next
Stage	Cluster 1	Cluster 2	Coefficient	Cluster 1	Cluster 2	Stage
1	1	2	.057000	0	0	5
2	14	15	.057600	0	0	4
3	18	19	.098600	0	0	6
4	14	16	.164100	2	0	10
5	1	3	.176000	1	0	16
6	18	20	.190723	3	0	13
7	11	12	.241770	0	0	14
8	6	7	.274280	0	0	12
9	4	5	.322431	0	0	16
10	14	17	.358392	4	0	15
11	9	10	.361682	0	0	19
12	6	8	.412832	8	0	17
13	18	21	.423903	6	0	18
14	11	13	.447164	7	0	15
15	11	14	.617459	14	10	18
16	1	4	.627763	5	9	17
17	1	6	.747090	16	12	19
18	11	18	.893448	15	13	20
19	1	9	1.254908	17	11	20
20	1	11	1.924901	19	18	0

FIGURE 3

DENDROGRAM USING COMPLETE LINKAGE FOR d(X,Y) MEASURES

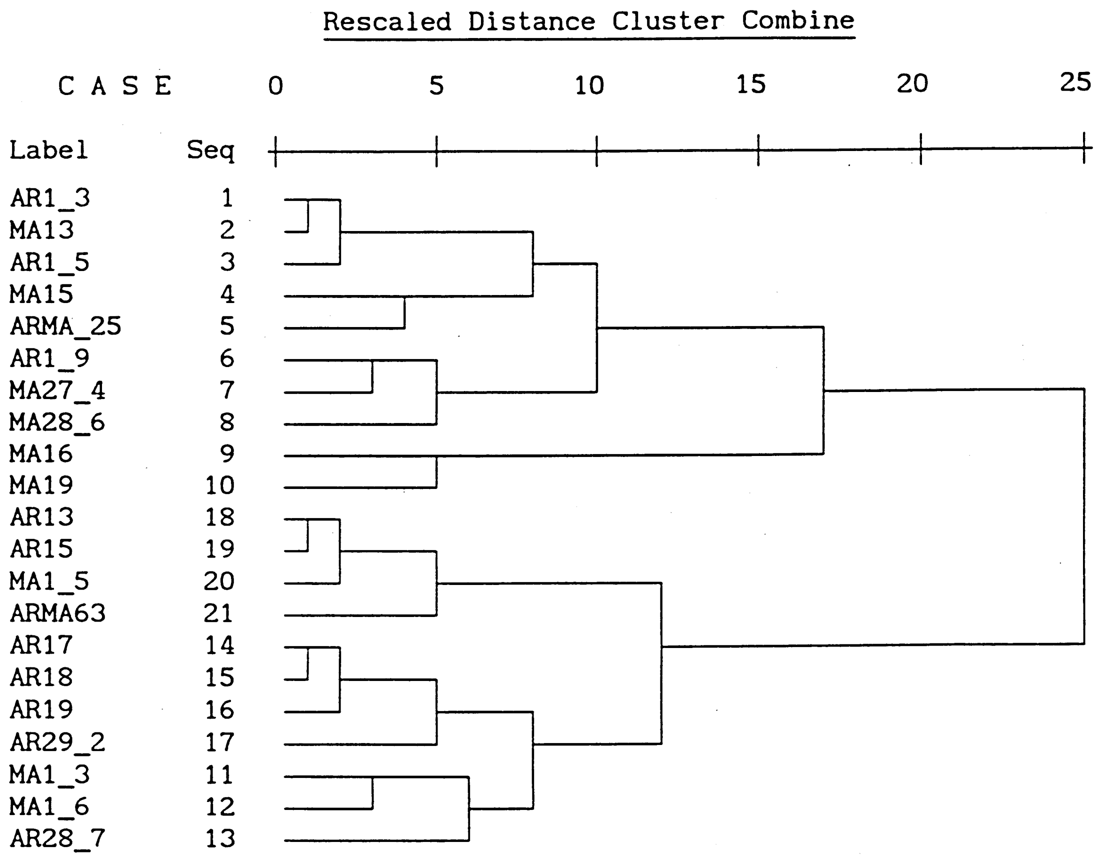

TABLE 7

AGGLOMERATION SCHEDULE USING THE WARD METHOD FOR d(X,Y) MEASURES

	Clusters Combined			Stage Cluster 1st Appears		Next
Stage	Cluster 1	Cluster 2	Coefficient	Cluster 1	Cluster 2	Stage
1	1	2	.001625	0	0	5
2	14	15	.003283	0	0	4
3	18	19	.008144	0	0	6
4	14	16	.020348	2	0	16
5	1	3	.034853	1	0	17
6	18	20	.051974	3	0	12
7	11	12	.081200	0	0	14
8	6	7	.118815	0	0	13
9	4	5	.170796	0	0	15
10	9	10	.236203	0	0	15
11	13	17	.305340	0	0	14
12	18	21	.387946	6	0	18
13	6	8	.484728	8	0	17
14	11	13	.662879	7	11	16
15	4	9	.900812	9	10	19
16	11	14	1.226860	14	4	18
17	1	6	1.603029	5	13	19
18	11	18	2.034028	16	12	20
19	1	4	2.993703	17	15	20
20	1	11	10.967413	19	18	0

FIGURE 4

DENDROGRAM USING THE WARD METHOD FOR d(X,Y) MEASURES

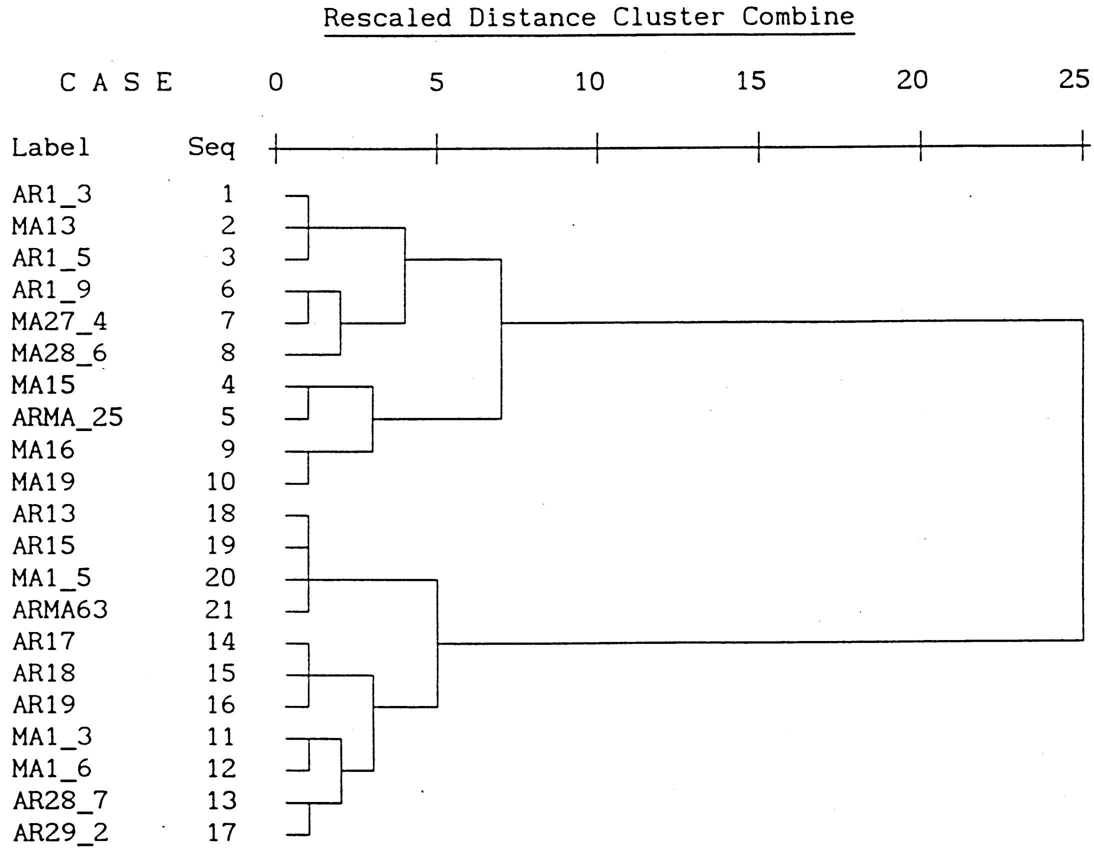

TABLE 8

AGGLOMERATION SCHEDULE USING AVERAGE LINKAGE FOR D(X,Y) MEASURES

	Clusters Combined			Stage Cluster 1st Appears		Next
Stage	Cluster 1	Cluster 2	Coefficient	Cluster 1	Cluster 2	Stage
1	18	19	.140000	0	0	4
2	1	2	.370000	0	0	5
3	14	15	.800000	0	0	7
4	18	20	2.295000	1	0	10
5	1	3	2.735000	2	0	13
6	11	12	3.870000	0	0	10
7	14	16	6.080000	3	0	12
8	4	9	6.230000	0	0	11
9	5	7	7.950000	0	0	11
10	11	18	9.471666	6	4	14
11	4	5	11.570001	8	9	13
12	14	17	14.166667	7	0	17
13	1	4	16.326666	5	11	16
14	11	21	17.672001	10	0	17
15	6	8	32.080002	0	0	19
16	1	10	33.240002	13	0	19
17	11	14	43.580830	14	12	18
18	11	13	57.171997	17	0	20
19	1	6	70.369370	16	15	20
20	1	11	209.701462	19	18	0

FIGURE 5

DENDROGRAM USING AVERAGE LINKAGE FOR D(X,Y) MEASURES

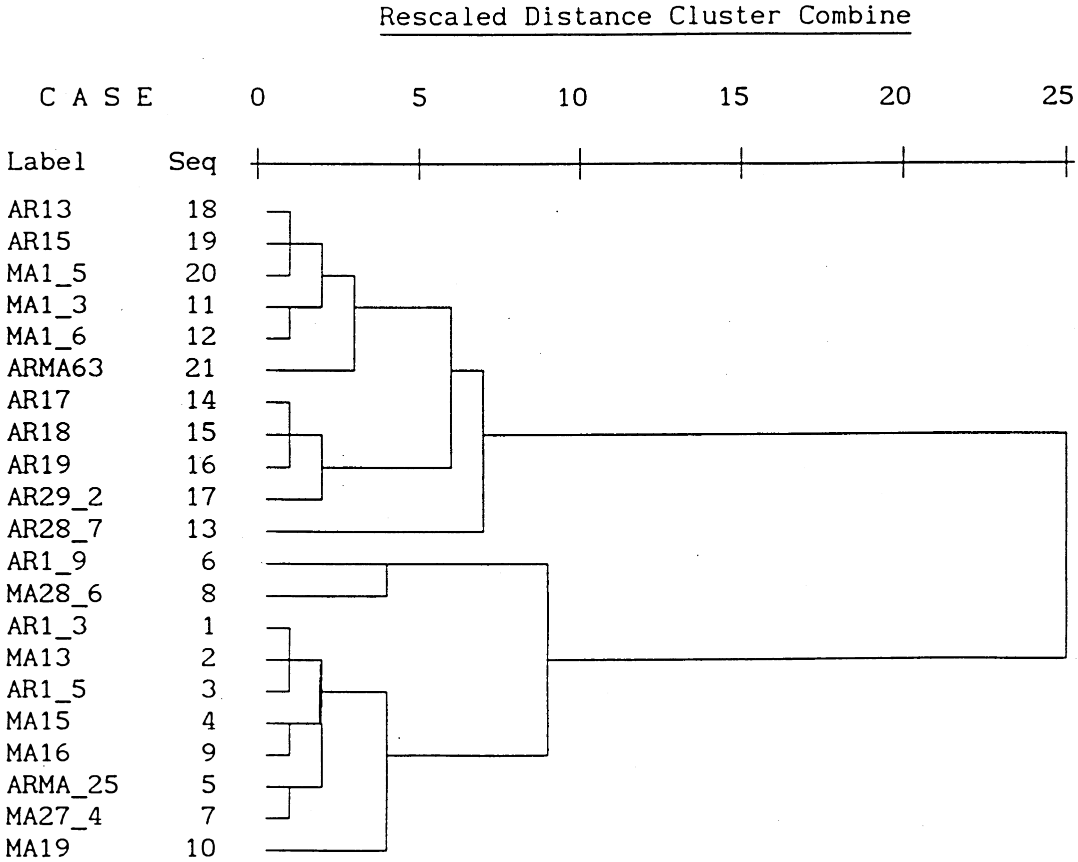

TABLE 9

AGGLOMERATION SCHEDULE USING SINGLE LINKAGE FOR D(X,Y) MEASURES

	Clusters Combined			Stage Cluster 1st Appears		Next
Stage	Cluster 1	Cluster 2	Coefficient	Cluster 1	Cluster 2	Stage
1	18	19	.140000	0	0	5
2	1	2	.370000	0	0	4
3	14	15	.800000	0	0	7
4	1	3	1.730000	2	0	12
5	18	20	1.940000	1	0	8
6	11	12	3.870000	0	0	8
7	14	16	3.980000	3	0	15
8	11	18	5.050000	6	5	13
9	4	9	6.230000	0	0	10
10	4	5	7.020000	9	0	11
11	4	10	7.630000	10	0	12
12	1	4	7.630000	4	11	14
13	11	21	7.650000	8	0	17
14	1	7	7.950000	12	0	16
15	14	17	9.360000	7	0	17
16	1	8	10.110000	14	0	19
17	11	14	17.070000	13	15	18

FIGURE 6

DENDROGRAM USING SINGLE LINKAGE FOR D(X,Y) MEASURES

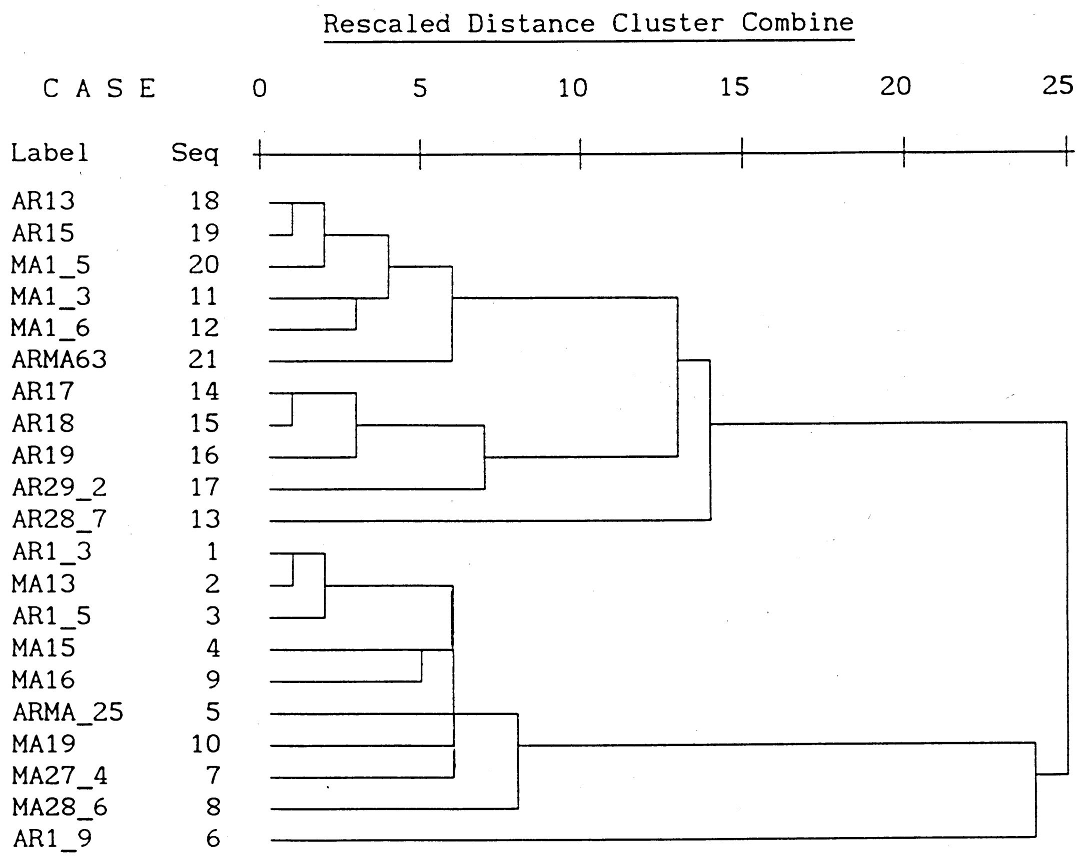

TABLE 10

AGGLOMERATION SCHEDULE USING COMPLETE LINKAGE FOR D(X,Y) MEASURES

	Clusters Combined			Stage Cluster 1st Appears		Next
Stage	Cluster 1	Cluster 2	Coefficient	Cluster 1	Cluster 2	Stage
1	18	19	.140000	0	0	4
2	1	2	.370000	0	0	5
3	14	15	.800000	0	0	9
4	18	20	2.650000	1	0	10
5	1	3	3.740000	2	0	13
6	11	12	3.870000	0	0	10
7	4	9	6.230000	0	0	11
8	5	7	7.950000	0	0	11
9	14	16	8.180000	3	0	12
10	11	18	14.010000	6	4	15
11	4	5	17.600000	7	8	13
12	14	17	22.160000	9	0	18
13	1	4	25.309999	5	11	16
14	6	8	32.080002	0	0	19
15	11	21	32.590000	10	0	17
16	1	10	51.049999	13	0	19
17	11	13	85.239998	15	0	18
18	11	14	121.440002	17	12	20
19	1	6	381.679993	16	14	20
20	1	11	1865.750000	19	18	0

FIGURE 7

DENDROGRAM USING COMPLETE LINKAGE FOR D(X,Y) MEASURES

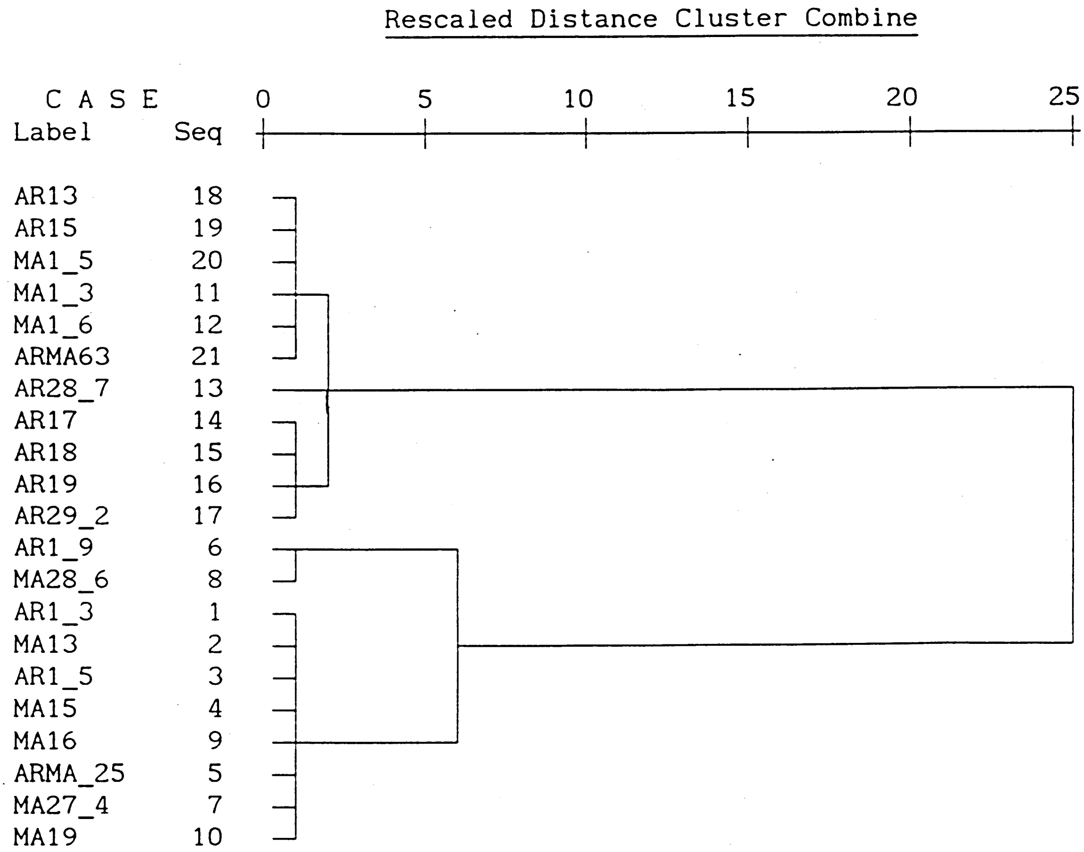

TABLE 11

AGGLOMERATION SCHEDULE USING THE WARD METHOD FOR D(X,Y) MEASURES

	Clusters Combined			Stage Cluster 1st Appears		Next
Stage	Cluster 1	Cluster 2	Coefficient	Cluster 1	Cluster 2	Stage
1	18	19	.070000	0	0	4
2	1	2	.255000	0	0	5
3	14	15	.655000	0	0	8
4	18	20	2.161667	1	0	11
5	1	3	3.923333	2	0	16
6	11	12	5.858333	0	0	14
7	4	9	8.973333	0	0	10
8	14	16	12.893333	3	0	12
9	5	7	16.868334	0	0	10
10	4	5	24.893333	7	9	13
11	18	21	33.136665	4	0	14
12	14	17	42.681664	8	0	17
13	4	10	56.486664	10	0	16
14	11	18	70.363327	6	11	18
15	6	8	86.403328	0	0	19
16	1	4	117.399162	5	13	19
17	13	14	159.834167	0	12	18

FIGURE 8

DENDROGRAM USING THE WARD METHOD FOR D(X,Y) MEASURES