ISSN 1440-771X
ISBN 0 7326 1064 8

M O N A S H   U N I V E R S I T Y

AUSTRALIA

Forecasting Time Series From Clusters

Elizabeth A. Maharaj and Brett A. Inder

Working Paper 9/99
June 1999

DEPARTMENT OF ECONOMETRICS AND BUSINESS STATISTICS

Forecasting Time Series from Clusters

Elizabeth A. Maharaj and Brett A. Inder

Monash University

Australia

Abstract

Forecasting large numbers of time series is a costly and time-consuming exercise. Before forecasting a large number of series that are logically connected in some way, we can first cluster them into groups of similar series. In this paper we investigate forecasting the series in each cluster. Similar series are first grouped together using a clustering procedure that is based on a test of hypothesis. The series in each cluster are then pooled together and forecasts are obtained. Simulated results show that this procedure for forecasting similar series performs reasonably well.

Keywords: Autoregressive models, Clustering technique, Mean square forecast error, Pooled series,

1. Introduction

Forecasting large numbers of logically connected time series especially in the short term is a common occurrence in many situations. Some examples are inventory control, stocks and shares, course enrolments at universities. While there is a vast literature on forecasting methodologies, very little research has been done on forecasting similar series.

Shah (1997) discusses model selection in univariate time series forecasting. He mentions that given a number of univariate time series, the forecaster may select one best model for all series or develop a rule that selects the best forecasting model for each series.

This rule is based on the classification technique of discriminant analysis. Discriminant analysis requires known groupings of time series before any further series can be classified. However in many situations, known groupings of the time series under consideration are not available.

Similar time series may be grouped together by some conventional clustering technique (see Everitt 1993). The problem with conventional clustering techniques is its subjective nature. The analyst must decide upon the number of clusters by selecting the distance at which the clusters are to be identified. Maharaj (1996, 1997) proposed a method of clustering stationary time series based on the p-value of a test of hypotheses that there is no difference between the generating processes of every two series under consideration. This method of clustering is far less subjective than the conventional clustering techniques. This method can also be successfully applied to nonstationary time series that can be easily transformed to stationary series.

In this paper we investigate the forecasting of series in each of the clusters which are selected using the above-mentioned clustering technique. Simulated results will show that when the series in a cluster are pooled together and fitted with an AR(k) model, the individual series in this cluster now fitted with the pooled model produce on average more accurate forecasts then when they are fitted with their own models. We will show theoretically that if the pooled and individual models are of the same order k, where $k = 1$ or 2, the mean square forecast error for the pooled model is less than that for the individual model for one-step ahead forecasts.

In Section 2 we briefly discuss the clustering technique mentioned above. In Section 3 we show the theoretical results. The results of the simulation study are given and discussed in Section 4 and in Section 5 we consider an application to a set of economic time series.

2. Clustering of Time Series

The clustering procedure is based on the following test of hypothesis:

H_0: There is no difference between the generating processes of two stationary series

H_A: There is a difference between the generating processes of two stationary series

Truncated AR(¥) models of order *k*, are fitted to each series and the test statistic which is based on the difference between the AR(*k*) estimates is constructed. These estimates are generalised least squares estimates. The order *k* can be selected by criteria such as the Akaike's information criterion (AIC) or the Schwarz's Bayesian information criterion (BIC). A seemingly unrelated regressions model is used to construct the test statistic which follows a chi-square distribution (see Maharaj (1997)).

The clustering procedure as given in Maharaj (1997) has the following steps: First perform the test of hypothesis for every pair of series determining the p-value associated with the test. 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 significance level (for example 0.05 or 0.01). Simulation studies have shown that this clustering procedure performs reasonably well.

3. *Theory*

For $k = 1$ and 2 we will show that when *m* series with equal variance are pooled together and fitted with an AR(*k*) model, this model when fitted to each of the *m* series will produce a smaller mean square forecast error (MSFE) for the one-step ahead forecast than when each of the *m* series is fitted with its own AR(*k*) model. We assume that the *m* series are generated from the same stationary process and hence form a cluster.

Lemma 1

Consider an AR(*k*) model

$$y_t = f_1 y_{t-1} + f_2 y_{t-2} + \dots + f_k y_{t-k} + a_t$$

fitted to the time series $\{y_t, t=1,2,\ldots,T\}$. a_t is a white noise process with mean 0 and variance σ_a^2. Since this is a lag dependent model and since the AR(k) process is stationary

$$plim \frac{1}{\mathrm{T - k}} \sum_{t=k+1}^{T} \boldsymbol{u_t u_t'} = \boldsymbol{Q}$$

where

$$\boldsymbol{u_t'} = [y_{t-1} \quad y_{t-2} \quad . \quad . \quad . \quad y_{t-k}]$$

and

$$\boldsymbol{Q} = \begin{bmatrix} q_{11} & q_{12} & . & . & . & q_{1k} \\ q_{21} & q_{22} & . & . & . & q_{2k} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ q_{k1} & q_{k2} & . & . & . & q_{kk} \end{bmatrix}$$

is a finite positive definite matrix (see Greene (1993)).

Theorem 1: Assume that m series with equal variance are generated from the same stationary process and that each series is fitted with an AR(k) model, where k = 1 or 2. Pool the m series together and assume that this pooled series is also fitted with an AR model of the same order k. Fit each of the m series with the pooled model. Then for one-step ahead forecasts

$$\mathrm{MSFE_P} < \mathrm{MSFE_I}$$

where $\mathrm{MSFE_I}$ and $\mathrm{MSFE_P}$ are the mean square forecast errors when a series is fitted with its own model and with the pooled model respectively.

Proof: ***(a)*** ***k* = 1**

Consider an AR(1) model fitted to the time series $\{y_t, t=1,2,\ldots,T\}$

$$y_t = \phi_1 y_{t-1} + a_t \tag{3.1}$$

where a_t is a white noise process with mean 0 and variance s_a^2. The one step ahead forecast for y_t is

$$\hat{y}_t = \hat{f}_1 y_{t-1} . \tag{3.2}$$

Hence the forecast error is

$$e_t = y_t - \hat{y}_t = \left(f_1 - \hat{f}_1\right) y_{t-1} + a_t .$$

The mean square forecast error is

$$\begin{aligned} \text{MSFE}_\text{I} &= E\left(e_t^2\right) = E\left(y_t - \hat{y}_t\right)^2 \\ &= E\left[y_{t-1}^2 \left(\hat{f} - f\right)^2 \right] + s_a^2 \end{aligned} \tag{3.3}$$

Then

$$\left(T-1\right)\left(\text{MSFE}_\text{I} - s_a^2\right) = E\left[y_{t-1}^2 \left(T-1\right)\left(\hat{f} - f\right)^2 \right].$$

Taking the limit as $T \to \infty$

$$\begin{aligned} lim\left(T-1\right)\left(\text{MSFE}_\text{I} - s_a^2\right) &= lim\left[E\left(y_{t-1}^2 \left(T-1\right)\left(\hat{f}_1 - f_1\right)^2 \right)\right] \\ &= E\left[lim\left(y_{t-1}^2 \left(T-1\right)\left(\hat{f}_1 - f_1\right)^2 \right)\right] \\ &= E\left[lim\, y_{t-1}^2 \, lim\left(T-1\right)\left(\hat{f}_1 - f_1\right)^2 \right] \\ &= lim\, E\left(y_{t-1}^2\right) lim\, E\left(\left(T-1\right)\left(\hat{f}_1 - f_1\right)^2 \right). \end{aligned} \tag{3.4}$$

Now since $\{y_t, t=1, 2, . . . T\}$ is a stationary series

$$lim\, E\left(y_{t-1}^2\right) = c_1$$

where c_1 is a constant. Furthermore

$$\begin{aligned} lim\, E\left[\left(\left(T-1\right)\left(\hat{f}_1 - f_1\right)^2 \right)\right] &= plim\, E\left[\left(\left(T-1\right)\left(\hat{f}_1 - f_1\right)^2 \right)\right] \\ &= plim\ \left(T-1\right) s_a^2 \left(\boldsymbol{X'X}\right)^{-1} \end{aligned} \tag{3.5}$$

where

$$X' = [y_1 \quad y_2 \quad . \quad . \quad . \quad y_{T-1}].$$

Hence

$$plim\ (T-1)s_a^2(X'X)^{-1} = \frac{s_a^2}{plim \frac{1}{T-1}\sum_{t=2}^{T} y_{t-1}^2},$$

and by Lemma 1

$$plim \frac{1}{T-1}\sum_{t=2}^{T} y_{t-1}^2 = q_{11} = c_1.$$

Thus Equation (3.4) becomes

$$lim\,(T\text{-}1)\left(\text{MSFE}_\text{I} - s_a^2\right) = \frac{c_1 s_a^2}{c_1} = s_a^2. \tag{3.6}$$

We now pool together m stationary time series $\boldsymbol{y_1}, \boldsymbol{y_2}, \ldots, \boldsymbol{y_m}$, that are generated for the same process, where

$$\boldsymbol{y_i}' = [y_{i1} \quad y_{i2} \quad . \quad . \quad . \quad y_{iT}]$$

for $i = 1, 2, \ldots m$. Fit the pooled series with an AR(1) model. The model for the pooled series is now of the form

$$\boldsymbol{Z} = \boldsymbol{W} f_{ip} + \boldsymbol{b} \tag{3.7}$$

where

$$\begin{aligned} \boldsymbol{Z}' &= [y_{12} \quad y_{13} \quad . \quad . \quad y_{1T} \quad y_{22} \quad y_{23} \quad . \quad . \quad y_{2T} \quad . \quad . \quad y_{m2} \quad y_{m3} \quad . \quad . \quad y_{mT}] \\ &= [z_2 \quad z_3 \quad . \quad . \quad z_T \quad z_{T+1} \quad z_{T+2} \quad . \quad . \quad z_{2T-1} \quad . \quad . \quad z_{(m-1)T+1} \quad z_{(m-1)T+2} \quad . \quad . \quad z_{mT-m+1}] \end{aligned}$$

$$\begin{aligned} \boldsymbol{W}' &= [y_{11} \quad y_{12} \quad . \quad . \quad y_{1T-1} \quad y_{21} \quad y_{22} \quad . \quad . \quad y_{2T-1} \quad . \quad . \quad y_{m1} \quad y_{m2} \quad . \quad . \quad y_{mT-1}] \\ &= [z_1 \quad z_2 \quad . \quad . \quad z_{T-1} \quad z_T \quad z_{T+1} \quad . \quad . \quad z_{2(T-1)} \quad . \quad . \quad z_{(m-1)T} \quad z_{(m-1)T+1} \quad . \quad . \quad z_{m(T-1)}] \end{aligned}$$

$$\begin{aligned} \boldsymbol{b}' &= [a_{12} \quad a_{13} \quad . \quad . \quad a_{1T} \quad a_{22} \quad a_{23} \quad . \quad . \quad a_{2T} \quad . \quad . \quad a_{m2} \quad a_{m3} \quad . \quad . \quad a_{mT}] \\ &= [b_2 \quad b_3 \quad . \quad . \quad b_T \quad b_{T+1} \quad b_{T+2} \quad . \quad . \quad b_{2T-1} \quad . \quad . \quad b_{(m-1)T+1} \quad b_{(m-1)T+2} \quad . \quad . \quad b_{mT-m+1}] \end{aligned}$$

Since the series which are pooled together this manner are expected to be logically connected, we assume that the disturbances are correlated across series. That is

$$E\left[a_{it}a_{jt}\right] = s_{ij}\boldsymbol{I} \qquad i, j = 1, 2, \ldots, m$$

where $\boldsymbol{I}$ is a $(T\text{-}1) \times (T\text{-}1)$ identity matrix. Hence the variance of $\boldsymbol{b}$ is

$$E[\boldsymbol{bb}'] = \boldsymbol{V} = \boldsymbol{Ó} \otimes \boldsymbol{I}$$

where

$$\boldsymbol{Ó} = \begin{bmatrix} s_{11} & s_{12} & . & . & . & s_{1m} \\ s_{12} & s_{22} & . & . & . & s_{2m} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ s_{m1} & s_{m2} & . & . & . & s_{mm} \end{bmatrix}$$

and where it is assumed that

$$s_{11} = s_{22} = \ldots = s_{mm} = s_a^2$$

It is also assume that the observations of the m series are uncorrelated with each other across time.

As opposed to ordinary least squares estimation of f_1 in (3.1), f_{1P} is now estimated by generalised least squares. That is

$$\hat{f}_{1P} = \left[\boldsymbol{W}'\boldsymbol{V}^{-1}\boldsymbol{W}\right]^{-1}\left[\boldsymbol{W}'\boldsymbol{V}^{-1}\boldsymbol{Z}\right].$$

Premultiplying the model in (3.7) by $\boldsymbol{V}^{-1/2}$ gives

$$\boldsymbol{V}^{-1/2}\boldsymbol{Z} = \boldsymbol{V}^{-1/2}\boldsymbol{W}\, f_{1P} + \boldsymbol{V}^{-1/2}\boldsymbol{b}$$

or

$$\boldsymbol{Z}^* = \boldsymbol{W}^*\, f_{1P} + \boldsymbol{b}^*$$

The variance of $\boldsymbol{b}^*$ is

$$E\left[\boldsymbol{b}^*\boldsymbol{b}^{*\prime}\right] = s_a^2\boldsymbol{I} \ .$$

Therefore

$$\hat{f}_{1P} = \left[\boldsymbol{W}^{*\prime}\boldsymbol{W}^*\right]^{-1}\left[\boldsymbol{W}^*\boldsymbol{Z}^*\right].$$

Hence fitting the pooled AR(1) model to each of the *m* series gives

$$y_t = f_{1P}\,y_{t-1} + a_t \ .$$

The one step-ahead forecast for y_t is

$$\hat{y}_t = \hat{f}_{1P}\,y_{t-1} \, .$$

Following through steps similar to those in equations (3.3) to (3.5)

$$\begin{aligned} lim\,(m(T-1))\left(\text{MSFE}_\text{P} - s_a^2\right) &= c_1\ plim\,E\left[\left(m(T\text{-}1)\left(\hat{f}_{1\text{P}} - f_{1\text{P}}\right)^2\right)\right] \\ &= c_1\ plim\ (m(T-1))s_a^2\left(\boldsymbol{W}^{*\prime}\boldsymbol{W}^*\right)^{-1} \\ &= \frac{c_1 s_\text{a}^2}{plim\dfrac{1}{m(T\text{-}1)}\displaystyle\sum_{t=2}^{m(T-1)+1} z_{t-1}^{*2}} \, . \end{aligned}$$

As a consequence of Lemma 1

$$plim\frac{1}{m(T\text{-}1)}\sum_{t=2}^{m(T-1)+1} z_{t-1}^{*2} = q_{11} = c_1 \, .$$

Thus

$$lim\,(T-1)\left(\text{MSFE}_\text{P} - s_a^2\right) = \frac{c_1 s_a^2}{m\,c_1} = \frac{s_a^2}{m} \ . \tag{3.8}$$

Taking the ratio of (3.6) to (3.8)

$$lim\frac{\left(\text{MSFE}_\text{I} - s_a^2\right)}{\left(\text{MSFE}_\text{P} - s_a^2\right)} = m \ . \tag{3.9}$$

Thus

$$\text{MSFE}_\text{P} < \text{MSFE}_\text{I}$$

for large samples, with the mean square forecast error ratio proportional to the number of series that are pooled together.

b) ***k=2***

Consider an AR(2) model fitted to the time series $\{y_t, t=1,2,\ldots,T\}$.

$$y_t = f_1 y_{t-1} + f_2 y_{t-2} + a_t$$

where a_t is a white noise process. The one step ahead forecast for y_t is

$$\hat{y}_t = \hat{f}_1 y_{t-1} + \hat{f}_2 y_{t-2}.$$

Hence the forecast error is

$$e_t = y_t - \hat{y}_t = \left(f_1 - \hat{f}_1\right) y_{t-1} + \left(f_2 - \hat{f}_2\right) y_{t-2} + a_t.$$

The mean square forecast error

$$\begin{aligned}\text{MSFE} &= E\left(e_t^2\right) = E\left(y_t - \hat{y}_t\right)^2 \\ &= E\left[y_{t-1}^2\left(\hat{f}_1 - f_1\right)^2\right] + E\left[y_{t-2}^2\left(\hat{f}_2 - f_2\right)^2\right] \\ &\quad + 2E\left[y_{t-1}y_{t-2}\left(\hat{f}_1 - f_1\right)\left(\hat{f}_2 - f_2\right)\right] + s_a^2.\end{aligned} \tag{3.9}$$

Then taking the limit as $T \to \infty$ and as a consequence of Lemma 1

$$\begin{aligned}&lim\,(T-2)\left(\text{MSFE}_\text{I} - s_a^2\right) \\ &= d_1\ plim\,E\left[(T-2)\left(\hat{f}_1 - f_1\right)^2\right] + d_2\ plim\,E\left[(T-2)\left(\hat{f}_2 - f_2\right)^2\right] + 2d_{12}\ plim\,E\left[(T-2)\left(\hat{f}_1 - f_1\right)\left(\hat{f}_2 - f_2\right)\right] \\ &= d_1\ s_a^2\ plim\,(T-2)S_{11}^{-1} + d_2\ s_a^2\ plim\,(T-2)S_{22}^{-1} + 2d_{12}\ s_a^2\ plim(T-2)S_{12}^{-1}\end{aligned} \tag{3.10}$$

where $d_1 = q_{11}$, $d_2 = q_{22}$, $d_{12} = q_{12}$, S_{ij}^{-1} , i,j =1, 2 is the ij^{th} element of the matrix $\left(\boldsymbol{X'X}\right)^{-1}$ and

$$\boldsymbol{X'} = \begin{bmatrix} y_2 & y_3 & . & . & . & y_{T-2} \\ y_1 & y_2 & . & . & . & y_{T-1}\end{bmatrix}.$$

Since both d_1 and d_2 represent the asymptotic variance of the time series, $d_1 = d_2$.

Now

$$S_{11}^{-1} = \frac{\sum_{t=3}^{T} y_{t-2}^2}{\sum_{t=3}^{T} y_{t-1}^2 \sum_{t=3}^{T} y_{t-2}^2 - \left(\sum_{t=3}^{T} y_{t-1} y_{t-2}\right)^2}$$

$$S_{22}^{-1} = \frac{\sum_{t=3}^{T} y_{t-1}^2}{\sum_{t=3}^{T} y_{t-1}^2 \sum_{t=3}^{T} y_{t-2}^2 - \left(\sum_{t=3}^{T} y_{t-1} y_{t-2}\right)^2}$$

$$S_{12}^{-1} = \frac{\sum_{t=3}^{T} y_{t-1} y_{t-2}}{\left(\sum_{t=3}^{T} y_{t-1} y_{t-2}\right)^2 - \sum_{t=3}^{T} y_{t-1}^2 \sum_{t=3}^{T} y_{t-2}^2} .$$

Hence

$$plim(T-2)S_{11}^{-1} = \frac{plim \frac{1}{T-2} \sum_{t=3}^{T} y_{t-2}^2}{plim \frac{1}{T-2} \sum_{t=3}^{T} y_{t-1}^2 \; plim \frac{1}{T-2} \sum_{t=3}^{T} y_{t-2}^2 - plim \frac{1}{T-2} \sum_{t=3}^{T} y_{t-1} y_{t-2} \; plim \frac{1}{T-2} \sum_{t=3}^{T} y_{t-1} y_{t-2}}$$

$$= \frac{d_1}{\left(d_1^2 - d_{12}^2\right)} .$$

Similarly

$$plim\, S_{22}^{-1} = \frac{d_1}{\left(d_1^2 - d_{12}^2\right)}$$

$$plim\, S_{12}^{-1} = \frac{-d_{12}}{\left(d_1^2 - d_{12}^2\right)} .$$

Hence Equation (3.10) becomes

$$lim(T-2)\left(\text{MSFE} - s_a^2\right) = s_a^2 \left[\frac{2d_1^2 - 2d_{12}^2}{d_1^2 - d_{12}^2}\right] \qquad (3.11)$$

$$= 2s_a^2$$

So now if *m* stationary time series $\boldsymbol{y_1, y_2, \ldots, y_m}$ are generated for the same process and are pooled together and fitted with an *AR(2)* model, it can be shown in a similar manner to that in case $k=1$, that when this pooled model is fitted to each of the *m* series

$$\lim\,(\mathrm{T}-2)\left(\mathrm{MSFE_P}-\sigma_a^2\right)=\frac{2\sigma_a^2}{m}\;.$$

Hence

$$\lim\frac{\left(\mathrm{MSFE_I}-\sigma_a^2\right)}{\left(\mathrm{MSFE_P}-\sigma_a^2\right)}=m\;.$$

So again

$$\mathrm{MSFE_P} < \mathrm{MSFE_I}.$$

for large samples, with the mean square forecast error ratio proportional to the number of series that are pooled together.

The results of Theorem 1 can also be extended to series fitted with AR(k) models giving

$$\lim(T-k)\left(\mathrm{MSFE_I}-\sigma_a^2\right)=k\sigma_a^2$$

$$\lim(T-k)\left(\mathrm{MSFE_P}-\sigma_a^2\right)=\frac{k\sigma_a^2}{m}$$

It is clear that as the number of series in a cluster increases the greater the effect of the pooled model on the mean square forecast error. It should however it noted that for all values of k, as T increases, the difference between $\mathrm{MSFE_P}$ and $\mathrm{MSFE_I}$ decreases.

4. Simulation Study

4.1. Design

The simulation study was carried out in two stages :

(a) Two groups of two series each:Two series were generated from each of an AR(1) process with ϕ=0.5 and a MA(1) process with θ = 0.9. Each of the four series were fitted with AR(k) models. The two series generated from the AR(1) model were pooled together and the pooled series was fitted with an AR(*k*) model. Likewise the two series generated from the MA(1) model were pooled together and the pooled series was fitted with an AR(*k*) model.

(b) **Two groups of eight series each**: Eight series were generated from each of an AR(1) process with f=0.5 and a MA(1) process with q = 0.9. Each of the sixteen series were fitted with AR(*k*) models. The eight series generated from the AR(1) model were pooled together and the pooled series was fitted with an AR(*k*) model. Likewise the eight series generated from the MA(1) model were pooled together and the pooled series was fitted with an AR(*k*) model.

For stage (a) series lengths of T=50 and 200 were taken when it was assumed that the correlation between the disturbances of each pair of processes from which the series were generated was in turn 0 and 0.5. For stage (b) series lengths of T=50 and 200 were taken when it was assumed that the correlation between the disturbances of each pair of processes from which the series were generated was 0. When it was assumed that the correlation between the disturbances of each pair of processes from which the series were generated was 0.5, series lengths of only T [1]=50 was taken. One, two and five-step-ahead forecasts were obtained for each series fitted with their own autoregressive models as well as for each series fitted with the relevant pooled autoregressive model. Each time the models were fitted to T-h (h=1,2,5) observations for the individual series and to $m(T-h)$ observations for pooled series, (m=2,8). The mean square forecast error for both the pooled and individual models were determined for the h out of sample values. The Akaike's (AIC) and the Schwarz's Bayesian (BIC) criteria were used to select the order of the AR model. One thousand simulations were carried out for each series length, selection criterion and h (1,2 5) step ahead forecast, for each of stages (a) and (b). Average mean square forecast error were determined for the series in each group, when the series were fitted with their own AR models as well as with the relevant pooled AR model. 95% prediction intervals were also obtained for the forecasts and the average number of times this interval contained the true value was observed for each group.

---

[1] For series length T=200, the simulations were unmanageable because of restrictions on the work space requirements of the computer program.

4.2 Discussion

For both uncorrelated and correlated series, it can be seen from Tables 1, 2, 3 and 4, for stage (a), and from Tables 5, 6 and 7 for stage (b) that the group average mean square forecast error is always smaller when the series were fitted with the pooled model than when the series were fitted with their own models. In most cases the percentage decrease in average mean square forecast error is greatest for the one-step ahead forecasts. The percentage decrease in average mean square forecast error in all cases is greater when the series in the group were generated from the MA(1) process with q = 0.9 than from the AR(1) process with f = 0.5.

< INSERT TABLES 1, 2, 3, 4, 5, 6, 7 HERE>

For both stages (a) and (b), the percentage decrease in the group average mean square forecast error between the individual and pooled models is greater for series length 50 than for series length 200. Comparing the results for the uncorrelated and correlated series in Table 1 with 5, 2 with 6 and 3 with 7, it can be seen that the percentage decrease in the group average mean square forecast error between the individual and pooled models are fairly similar.

Comparing the results of Tables 1, 2 and 3 with that of Tables 5, 6 and 7 respectively, it is clear that in almost all cases the percentage decrease in the group average mean square forecast error is greater for groups with eight series each, than for groups with two series each. These observations are consistent with the theoretical results in Section 3 even though the autoregressive models that were fitted to the individual series and to the pooled series within a group were not necessarily of the same order.

It can also be seen that in almost all cases lower group average mean square forecast errors were obtained for the pooled model than for the corresponding individual model when the AIC was used to select the order of the autoregressive model fitted to the series.

From Tables 8, 9, 10 and 11 for stage (a) and from Tables 12, 13 and 14 for stage (b), it can be seen that in almost all cases, the average number of times the 95% prediction interval in each group contains the true value of the forecast is almost always greater when the pooled model was fitted to the series. However, it is mostly for the one-step-ahead forecasts, for the pooled model cases, that the true value of the forecast lies in the prediction interval close to 95% of the time.

< INSERT TABLES 8,9,10,11,12,13,14 HERE>

5. Application

Consider the time series of the number of dwelling units financed by all lenders (banks and other institutions) in the states and territories of Australia from January 1978 to March 1998. Clearly these series are related since they are all influenced by the same economic factors. The natural logarithm transformation of each series is shown in Figure 1 from where it is clear the all these series[2] are non-stationary. Some series appear to have fairly similar patterns. However 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 of clustering procedure of Maharaj (1997) was then applied to these differenced series.

<INSERT FIGURE 1 HERE>

When the level of significance was set at 5% the algorithm produced the following clusters : (NSW, NT), (OLD, SA, WA), (VIC, ACT) and (TAS). Graphs of the undifferenced series for each of the three clusters are given in Figures 2 - 4, from where it is quite clear that the patterns of the series from each cluster are similar. However it can be seen from these figures that the series from a particular cluster are not necessarily on the same level. The reason for this is that while the test of hypothesis on which the clustering algorithm is based differentiates between stochastic nature of series it does not differentiate between their corresponding deterministic nature. Hence it is possible for series at different levels but similar

[2] ACT:Australian Capital Territory, NSW:New South Wales, NT:Northern Territory, QLD:Queensland,

patterns to cluster together. However this in no way affects any further analysis of the series in

<INSERT FIGURES 2- 4 HERE>

each cluster. For example, forecasts of the original series in a particular cluster will be obtained by reversing the operation of differencing and any other transformation on the stationary series.

The standardised first difference of the logarithm of each the series excluding the March 1998 value, in each of the clusters (NSW, NT), (QLD, SA, WA), (VIC, ACT), were pooled together and fitted with the relevant AR*(k)* model. The residual correlations between the series in each cluster is given in Table 15, from where it is clear that the series are related.

<INSERT TABLE 15 HERE>

The forecasts from the pooled and individual models, percentage decrease in the mean square forecast error between and individual and pooled models of each of the series as well as the percentage decrease in group average mean square forecast error between the pooled and individual models for the one-step-ahead forecasts, (that is for March 1998) are given in Table 16. These results were obtained when the BIC was used to select *k*. Similar results were also obtained when the AIC was used to select *k*. With the exception of series TAS (which forms its own cluster), NT and ACT, all the other series have smaller mean square forecast errors for the pooled model than for the individual model. On average the mean square forecast error of each group is smaller for the pooled than for the individual models.

<INSERT TABLE 16 HERE>

Forecasts were then obtained for the original series in each cluster. The results which are given in Table 17 show that fitting the pooled model to the standardised first difference of the logarithm of the series, produced on average more accurate one-step ahead forecasts than when the series are fitted with their own models.

<INSERT TABLE 17 HERE>

Two and five-step ahead forecasts were also obtained, but the results showed that in these cases there was no advantage in pooling the series in the clusters.

SA:South Australian, VIC:Victoria, WA:Western Australia, *Source: Australian Bureau of Statistics*

6. *Concluding Remarks*

From both the theoretical and simulated results, it is clear that fitting a pooled model to all series in a particular cluster produces more accurate forecasts than when the series are fitted with their own models. The outcome of the application shows that the results are generally consistent with the theory.

So while there isn't overwhelming evidence to support the use of a pooled model over that of an individual model for forecasting, it would appear that it is worthwhile giving some consideration to forecasting each series in a cluster from the pooled model.

Acknowledgments

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.

References

Green, W.H. (1993), *Econometric Analysis*, Second Edition, Macmillan Publishing Company, New York, 1993.

Everitt, B.S. (1993), *Cluster Analysis*, Halsted Press, New York.

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, Monash University.

Shah, C (1997), Model Selection in Univariate Time Series Forecasting using Discriminant Analysis, *International Journal of Forecasting*, 13, 489-500.

Table 1: Average mean square forecast error: 2 groups of 2 series, T=50, Corr=0

Generating Process	Model Selection Criterion	Fitted Model Type	h-step		
			1	**2**	**5**
AR(1) ϕ =0.5	BIC	Individual Pooled	0.8723 0.8415	0.9739 0.9153	1.0011 0.9810
		% decrease in MSFE	**4**	**6**	**2**
	AIC	Individual Pooled	0.9877 0.8890	1.1052 0.9793	1.1220 1.0244
		% decrease in MSFE	**10**	**11**	**9**
MA(1) θ =0.9	BIC	Individual Pooled	0.6680 0.6035	0.9291 0.8603	1.0657 0.9906
		% decrease in MSFE	**10**	**7**	**7**
	AIC	Individual Pooled	0.6618 0.5794	0.9889 0.8670	1.1378 1.0074
		% decrease in MSFE	**12**	**12**	**11**

Table 8: Average number of times the 95% prediction interval of the forecast contains the true value per group: 2 groups of 2 series, T=50, Corr=0

Generating Process	Model Selection Criterion	Fitted Model Type	h-step							
			1	**2**		**5**				
AR(1) ϕ =0.5	BIC	Individual	922	967	944	970	949	955	952	951
		Pooled	932	977	958	976	955	960	952	952
	AIC	Individual	932	977	958	976	955	960	952	952
		Pooled	933	973	957	981	959	965	962	963
MA(1) θ =0.9	BIC	Individual	921	986	958	993	971	960	969	973
		Pooled	946	996	969	997	979	971	974	981
	AIC	Individual	915	987	958	995	974	966	968	974
		Pooled	954	997	975	996	984	978	981	983

Table 2: Average Mean Square Forecast Error: 2 Groups of 2 Series, T=200, Corr=0

Generating Process	Model Selection Criterion	Fitted Model Type	h-step		
			1	**2**	**5**
AR(1) ϕ =0.5	BIC	Individual	0.7687	0.8541	0.9201
		Pooled	0.7679	0.8523	0.9169
		% decrease in MSFE	**0.1**	**0.2**	**0.3**
	AIC	Individual	0.7968	0.8759	0.9486
		Pooled	0.7851	0.8698	0.9383
		% decrease in MSFE	**2**	**0.7**	**1**
MA(1) θ =0.9	BIC	Individual	0.5752	0.8328	0.9509
		Pooled	0.5590	0.8231	0.9355
		% decrease in MSFE	**3**	**1**	**2**
	AIC	Individual	0.5621	0.8338	0.9505
		Pooled	0.5501	0.8123	0.9336
		% decrease in MSFE	**2**	**3**	**2**

Table 9: Average number of times the 95% prediction interval of the forecast contains the true value per group: 2 groups of 2 series, T=200, Corr=0

Generating Process	Model Selection Criterion	Fitted Model Type	h-step							
			1	**2**		**5**				
AR(1) ϕ =0.5	BIC	Individual	943	985	963	983	972	974	970	969
		Pooled	944	986	964	985	971	974	968	969
	AIC	Individual	939	983	963	986	974	972	971	974
		Pooled	944	985	966	987	977	974	975	975
MA(1) θ =0.9	BIC	Individual	949	998	983	998	981	983	984	989
		Pooled	951	999	987	999	984	985	987	989
	AIC	Individual	944	999	984	999	982	987	988	991
		Pooled	952	999	989	999	986	987	988	990

Table 3: Average mean square forecast error: 2 groups of 2 series, T=50, Corr=0.5

Generating Process	**Model Selection Criterion**	**Fitted Model Type**	**h-step**		
			1	**2**	**5**
AR(1) ϕ =0.5	BIC	Individual Pooled	0.8407 0.8054	0.9827 0.9292	0.9962 0.9537
		% decrease in MSFE	**4**	**5**	**4**
	AIC	Individual Pooled	0.9714 0.8656	1.1017 1.0032	1.1151 1.0115
		% decrease in MSFE	**11**	**9**	**9**
MA(1) θ =0.9	BIC	Individual Pooled	0.6496 0.5979	0.9489 0.8777	1.0389 0.9842
		% decrease in MSFE	**8**	**8**	**5**
	AIC	Individual Pooled	0.6682 0.5668	1.0063 0.8880	1.1271 1.0085
		% decrease in MSFE	**15**	**12**	**11**

Table 10: Average number of times the 95% prediction interval of the forecast contains the true value per group: 2 groups of 2 series, T=50, Corr=0.5

Generating Process	**Model Selection Criterion**	**Fitted Model Type**	**h-step**		
			1	**2**	**5**
AR(1) ϕ =0.5	BIC	Individual	933	969 940	975 956 954 953 950
		Pooled	944	979 950	980 961 966 957 951
	AIC	Individual	904	956 923	986 952 956 953 943
		Pooled	942	975 947	980 963 970 968 959
MA(1) θ =0.9	BIC	Individual	918	987 951	993 975 972 974 974
		Pooled	949	995 974	994 982 979 984 984
	AIC	Individual	915	990 953	993 974 975 972 975
		Pooled	956	996 976	996 984 987 986 985

Table 4: Average Mean Square Forecast Error: 2 Groups of 2 Series, T=200, Corr=0.5

Generating Process	Model Selection Criterion	Fitted Model Type	h-step		
			1	**2**	**5**
AR(1) ϕ =0.5	BIC	Individual Pooled	0.8036 0.8034	0.8443 0.8414	0.9764 0.9741
		% decrease in MSFE	**0**	**0**	**0**
	AIC	Individual Pooled	0.8294 0.8202	0.8682 0.8550	1.0009 0.9897
		% decrease in MSFE	**1**	**2**	**1**
MA(1) θ =0.9	BIC	Individual Pooled	0.6219 0.5963	0.6219 0.5693	0.9540 0.9407
		% decrease in MSFE	**4**	**4**	**1**
	AIC	Individual Pooled	0.5953 0.5765	0.8429 0.8166	0.9627 0.9435
		% decrease in MSFE	**3**	**3**	**2**

Table 11: Average number of times the 95% prediction interval of the forecast contains the true value per group: 2 groups of 2 series, T=200, Corr=0.5

Generating Process	Model Selection Criterion	Fitted Model Type	h-step							
			1	**2**		**5**				
AR(1) ϕ =0.5	BIC	Individual	945	977	972	984	972	958	954	959
		Pooled	945	980	974	983	974	960	957	962
	AIC	Individual	936	977	961	986	975	965	959	966
		Pooled	942	981	971	987	977	967	964	969
MA(1) θ =0.9	BIC	Individual	940	999	980	999	988	985	987	987
		Pooled	945	1000	985	999	990	989	988	989
	AIC	Individual	942	1000	1000	999	988	985	987	990
		Pooled	949	981	986	999	993	989	991	992

Table 5: Average mean square forecast error: 2 groups of 8 series, T=50, Corr =0

Generating Process	**Model Selection Criterion**	**Fitted Model Type**	**h-step**		
			1	**2**	**5**
AR(1) ϕ =0.5	BIC	Individual	0.9144	0.9561	1.0110
		Pooled	0.8722	0.9130	0.9808
		% decrease in MSFE	**5**	**5**	**3**
	AIC	Individual	1.0389	1.0792	1.1448
		Pooled	0.9016	0.9302	1.0004
		% decrease in MSFE	**13**	**14**	**13**
MA(1) θ =0.9	BIC	Individual	0.6744	0.9721	1.0505
		Pooled	0.5617	0.8500	0.9515
		% decrease in MSFE	**17**	**13**	**9**
	AIC	Individual	0.6764	1.0318	1.1355
		Pooled	0.5498	0.8493	0.9536
		% decrease in MSFE	**19**	**18**	**16**

Table 12: Average number of times the 95% prediction interval of the forecast contains the true value per group: 2 groups of 8 series, T=50, Corr=0

Generating Process	**Model Selection Criterion**	**Fitted Model Type**	**h-step**							
			1	**2**		**5**				
AR(1) ϕ =0.5	BIC	Individual	915	964	948	970	953	953	952	952
		Pooled	931	974	961	978	961	961	960	956
	AIC	Individual	894	949	932	968	950	951	952	951
		Pooled	946	981	970	986	976	976	972	970
MA(1) θ =0.9	BIC	Individual	922	988	955	992	971	972	972	975
		Pooled	963	997	980	997	986	985	984	984
	AIC	Individual	918	988	956	993	975	973	972	974
		Pooled	970	998	985	998	989	989	987	990

Table 6: Average Mean Square Forecast Error: 2 Groups of 8 Series, T=200, Corr=0

Generating Process	Model Selection Criterion	Fitted Model Type	h-step		
			1	**2**	**5**
AR(1) ϕ =0.5	BIC	Individual	0.7543	0.8653	0.9456
		Pooled	0.7507	0.8623	0.9451
		% decrease in MSFE	**0.5**	**0.3**	**0.05**
	AIC	Individual	0.7804	0.8911	0.9729
		Pooled	0.7614	0.8729	0.9548
		% decrease in MSFE	**2**	**2**	**2**
MA(1) θ =0.9	BIC	Individual	0.5766	0.8017	0.9564
		Pooled	0.5423	0.1195	0.9316
		% decrease in MSFE	**6**	**4**	**3**
	AIC	Individual	0.5577	0.8103	0.9608
		Pooled	0.5388	0.7776	0.9606
		% decrease in MSFE	**3**	**4**	**3**

Table 13: Average number of times the 95% prediction interval of the forecast contains the true value per group: 2 groups of 8 series, T=200, Corr=0

Generating Process	Model Selection Criterion	Fitted Model Type	h-step							
			1	**2**		**5**				
AR(1) ϕ =0.5	BIC	Individual	947	981	966	982	967	970	966	967
		Pooled	948	980	967	982	970	970	967	970
	AIC	Individual	942	978	963	982	970	972	970	971
		Pooled	947	981	969	988	978	977	975	978
MA(1) θ =0.9	BIC	Individual	950	998	987	998	985	985	989	988
		Pooled	961	999	992	999	988	989	991	991
	AIC	Individual	948	998	987	998	985	986	989	989
		Pooled	962	999	992	999	989	989	991	992

Table 7: Average mean square forecast error: 2 groups of 8 series, T=50, Corr =0.5

Generating Process	Model Selection Criterion	Fitted Model Type	h-step		
			1	**2**	**5**
AR(1) ϕ =0.5	BIC	Individual	0.9144	0.9561	1.0110
		Pooled	0.8594	0.9056	0.9746
		% decrease in MSFE	**6**	**5**	**4**
	AIC	Individual	1.0389	1.0792	1.1448
		Pooled	0.8812	0.9187	0.9904
		% decrease in MSFE	**15**	**15**	**13**
MA(1) θ =0.9	BIC	Individual	0.6744	0.9720	1.0505
		Pooled	0.5425	0.8400	0.9441
		% decrease in MSFE	**20**	**14**	**10**
	AIC	Individual	0.6764	1.0318	1.1358
		Pooled	0.5171	0.8362	0.9448
		% decrease in MSFE	**24**	**19**	**17**

Table 14: Average number of times the 95% prediction interval of the forecast contains the true value per group: 2 groups of 8 series, T=50, Corr=0.5

Generating Process	Model Selection Criterion	Fitted Model Type	h-step		
			1	**2**	**5**
AR(1) ϕ =0.5	BIC	Individual	915	964 947	969 952 952 952 951
		Pooled	934	976 965	979 964 962 960 957
	AIC	Individual	893	984 931	967 950 950 952 951
		Pooled	948	983 975	988 979 976 974 974
MA(1) θ =0.9	BIC	Individual	922	987 954	991 971 971 971 974
		Pooled	964	998 986	998 990 988 989 990
	AIC	Individual	918	988 954	993 974 972 972 973
		Pooled	976	999 991	999 993 991 992 993

Table 15: Residual correlations between series in each cluster

Cluster	Correlation
1 NSW, NT	0.4854
2 QLD, WA	0.7209
QLD, SA	0.7333
WA, SA	0.6964
3 VIC, ACT	0.5844

Table 16: Forecast statistics for March 1998 of the standardised first difference of the logarithm of the series

	Actual value of the standardised first difference of the log of original series	Forecast from individual model	Forecast from pooled model	% Decrease in MSFE between individual and pooled models
NSW	-0.4205	0.3923	0.1699	47
VIC	-0.1038	0.2789	0.1977	38
QLD	-0.1845	0.1721	-0.0149	77
SA	-0.4554	0.4485	-0.0315	78
WA	-0.1429	0.2838	-0.0804	98
TAS	0.0372	0.2483	0.2483	0
NT	-0.0967	0.1782	0.1825	- 3
ACT	0.5477	0.4251	0.4190	-10

Group	% Decrease in the group average MSFE between the individual and pooled models
(NSW, NT)	42
(QLD,SA,WA)	81
(VIC,ACT)	33

Table 17 Foreecast statistics for March 1998 of the original series

	Actual value of original series	Forecast from individual model	Forecast from pooled model	% Decrease in MSFE between individual and pooled models
NSW	14048	12322	12772	45
VIC	10147	9584	9701	37
QLD	6568	6214	6397	77
SA	3321	2925	3129	77
WA	5491	5154	5440	98
TAS	393	819	819	0
NT	682	366	366	0
ACT	848	694	695	1

Group	% Decrease in the group average MSFE between the individual and pooled models
(NSW, NT)	44
(QLD,WA,SA)	83
(VIC, ACT)	35

Figure 1 Number of dwelling units financed from January 1978 to March 1998 for all states and territories

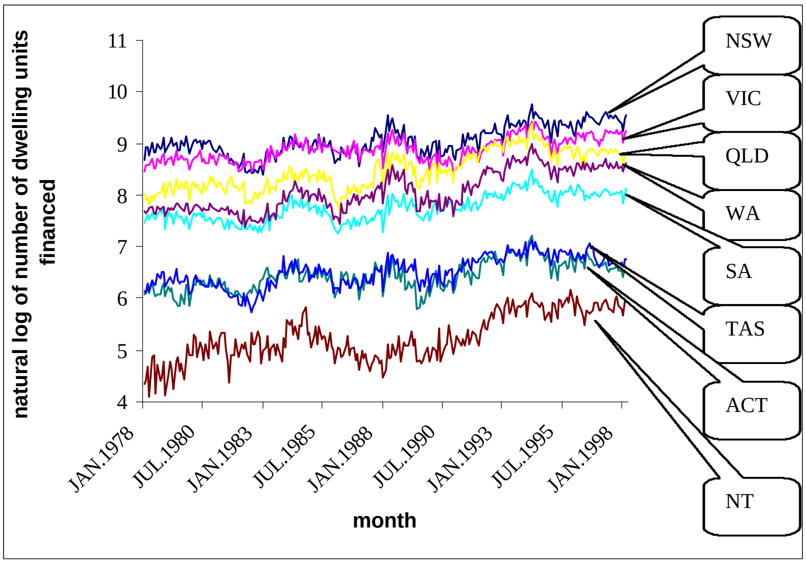

Figure 2 Number of dwelling units financed from January1978 to March 1998 for New South Wales and the Northern Territory

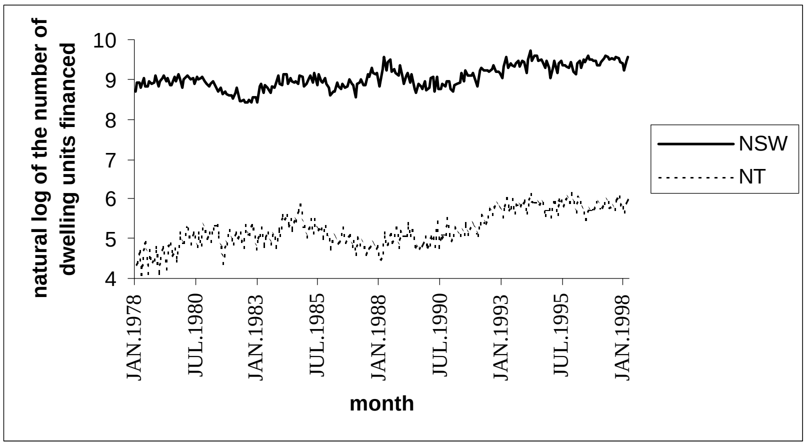

Figure 3 Number of dwelling units financed from January 1978 to March 1998 for Queensland, South Australia and Western Australia

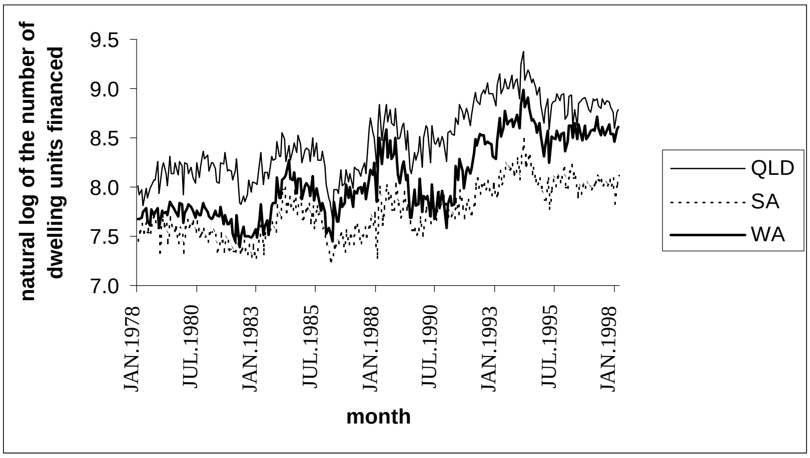

Figure 4 Number of dwelling units financed from January 1978 to March 1998 for Victoria, Australian Capital Territory and Tasmania