WORKING PAPER NO. TSRA-87-001E
Time Series Research & Analysis
Methodology

Cahier de travail TSRA 87-001E
Division de Récherche et Analyse
de séries chronologiques
Direction de la Méthodologie

FORECAST PERFORMANCE OF ARIMA MODELS AS FUNCTION OF THE NOISE CONTENT AND TREND-CYCLE PATTERN OF TIME SERIES

by

Estela Bee Dagum, Guy Huot, Marietta Morry and Kim Chiu
Statistics Canada

This is a preliminary version. Do not quote without authors' permission. Comments are welcome.

Résumé

Dans cet article, nous analysons la performance prévionnelle des quatre modèles ARMMI qui seront incorporés à la nouvelle version du programme de désaisonnalisation X-11-ARMMI. Ces modèles sont le $(0,1,1)(0,1,1)_{12}$, le $(0,1,2)(0,1,1)_{12}$, le $(0,2,2)(0,1,1)_{12}$ et le $(2,1,0)(0,1,1)_{12}$, qui sont ajustés à un échantillon de 120 séries macroéconomiques. La performance prévisionnelle des modèles est calculée en function de la structure globale des séries ainsi que de la tendance-cycle et du bruit présents dans chaque série. La performance prévsionnelle est exprimée en terme d'erreur pourcentuelle absolue moyenne calculée pour quatre horizons de prévision.

FORECAST PERFORMANCE OF ARIMA MODELS AS FUNCTION OF THE NOISE CONTENT AND TREND-CYCLE PATTERN OF TIME SERIES

Estela Bee Dagum, Guy Huot, Marietta Morry and Kim Chiu
Statistics Canada

1. INTRODUCTION

It has been proven theoretically through the study of filters (Dagum 1982.a; 1982.b and 1983) and corroborated by several empirical studies, among others, in Kuiper (1978), Kenny and Durbin (1982), Dagum and Morry (1984) that the use of one year ahead ARIMA extrapolations reduces the revisions of the concurrent seasonal factors obtained from the X-11-ARIMA program (Dagum, 1980).

A set of four ARIMA models $((0,1,1)(0,1,1)_{12}$, $(0,1,2)(0,1,1)_{12}$, $(0,2,2)(0,1,1)_{12}$, and $(2,1,0)(0,1,1)_{12})$ identified by Chiu, Higginson and Huot (1985) fitted and forecasted well a large variety of economic time series. These models were ranked on the basis of the one year ahead forecast error as well as some seven other criteria including fit and parsimony. The inclusion of the four models in the X-11-ARIMA program in the order of their ranking is to ensure that a good set of extrapolated values can be produced even if the user has no expertise in ARIMA model identification.

From the viewpoint of minimizing the filter revisions of the X-11-ARIMA method, the maximum forecast horizon should not be restricted to one year only for all series. A study by Huot, Chiu, Higginson and Gait (1986) showed that for the four ARIMA models and certain parameter values,

forecast horizons different from 12 months still produce a significant reduction of the filter revisions. Thus, it is necessary to investigate which factors would determine the length of the forecast horizon to be used in order to produce current seasonally adjusted estimates with the smallest revision. Furthermore, it is also of interest to know how the varying forecast horizon affects the ranking of the models.

The ultimate objective of a larger study in preparation by the authors is to produce guidelines based on certain characteristics of the series regarding the number of forecast values necessary to yield seasonally adjusted estimates of minimal revision. It is obvious that the revisions in these estimates will depend on the forecast error. (A perfect three-year forecast would result in estimates that do not get revised at all). Thus, as a first stage in this direction, this paper examines the relationship between the forecast error at different time horizons and certain characteristics of the series namely, the amount of irregular variation present and, the pattern of the trend-cycle component.

Section 2 describes the design of the experiment; section 3 looks into the performance ranking of the four ARIMA models according to the global structure of the series. In order to see why some series are forecasted with greater accuracy than others, the performance ranking of the models is evaluated as a function of the noise and the trend-cycle in the series. Section 4 deals with the relationship between the forecast error and the noise in the series and section 5 with the relationship between the forecast error and the pattern of the trend-cycle component. Finally, section 6 gives the conclusions of this study.

2. THE DESIGN OF THE EXPERIMENT

In designing the experiment our main objectives are: (1) to have a

representative sample of economic time series in terms of the amount of irregular fluctuations present in them; (2) to eliminate the dependence of the forecast error on the particular month of the year chosen as forecast origin; and (3) to minimize the number of ARIMA fitting and forecasting necessary in order to reduce costs.

To achieve these objectives, the following design is adopted. A sample of 120 series is selected from five sectors of the economy (labour, external trade, manufacturing, finance and agriculture).

All the series begin in January 1970 and for each series two sets of forecasts up to 24 steps ahead are generated from different point origin (one year apart). The last time point for which a forecast is made corresponds to May 1981 coming from a series that ends in May 1979. May 1981 is chosen as the last forecast time point to avoid the effect of the atypical 1981 recession on the forecast errors.

Another (simpler) alternative would be to produce 12 sets of forecasts per series each starting at a different month of the year. This design, however, requires a much larger number of ARIMA fits and would increase substantially the cost since the number of series cannot be reduced without jeopardizing the representativeness of the sample.

The four ARIMA models described before are applied to each series and forecasts are generated from two time origins; i.e. eight sets of forecasts are obtained from each series.

The information obtained for each series includes the forecast errors for 24 time points for all eight sets, the ARIMA parameter values, the goodness of fit statistic, the forecasting origin, the class to which the series belong according to the amount of irregularity and the series identifier. This information is merged with another collected previously from each series during the X-11-ARIMA seasonal adjustment run concerning the amount of irregular and trend-cycle variations in the series.

The series fall into five classes (of 24 series each) according to the amount of irregular variation present as identified by the X-11-ARIMA program in table F.2.B. The five classes are:

Class 1	0.0% - 5.0%	irregular	variation
Class 2	5.1% - 10.0%	"	"
Class 3	10.1% - 20.0%	"	"
Class 4	20.1% - 30.0%	"	"
Class 5	30,1% - 50.0%	"	"

The amount of irregular variation in a series, however, is often measured as a function of the residuals left after fitting a model. We therefore calculated the amount of irregular variation both ways, as identified by the X-11-ARIMA program and from the ARIMA model fitting. Table 1 shows the average irregular variation estimated from the X-11-ARIMA and from the ARIMA models. Except for class 1 where the two values are close, the noise estimated from the ARIMA models is always much smaller than the one obtained from the X-11-ARIMA program.

Place table 1 about here.

3. PERFORMANCE RANKING OF THE ARIMA MODELS AT DIFFERENT FORECAST HORIZONS

Each of the 120 series in the sample is fitted using the four ARIMA models. For a given series each model either ranks first, second, third or fourth according to the mean absolute percentage error (MAPE) of the

forecasts. Figure 1 shows the performance ranking of the four models when they ranked first. At a 6 months forecast horizon, model $(0,2,2)(0,1,1)_{12}$ ranked first 38% of the time followed by models $(0,1,1)(0,1,1)_{12}$, $(0,1,2)(0,1,1)_{12}$ and $(2,1,0)(0,1,1)_{12}$ with 27%, 19% and 16% respectively. Figure 1 also indicates that the ranking depends on the forecast horizon. The dispersion of the forecasting performance of the four models decreases with an increasing forecast horizon. For instance, the dispersion in the forecast performance which ranges from 16% to 38% at a 6 months forecast horizon reduces to 21% to 29% at 24 months. Moreover the ranking of the first three models changes. Model $(0,2,2)(0,1,1)_{12}$ now comes third at a 24 months forecast horizon ranking first only 22% of the time while model $(0,1,2)(0,1,1)_{12}$ which was third now ranks first.

Place figure 1 about here.

The percentage of times each of the four models ranked second is about 5% for model $(0,2,2)(0,1,1)_{12}$ and 31% for the remainders with small dispersion in their forecast performance. The $(0,2,2)(0,1,1)_{12}$ model ranked third in 4% of all the cases, the $(2,1,0)(0,1,1)_{12}$ model in 44% and each of remaining two models ranked third in 26% of the cases.

Figure 2 shows the relative ranking of the models when they ranked fourth. At a 6 months forecast horizon, model $(0,2,2)(0,1,1)_{12}$ ranked fourth in 50% of all the cases. This precentage increased to 68% at the 24 months horizon, indicating a dererioration forecasting performance. This suggests that model $(0,2,2)(0,1,1)_{12}$ fits and forecasts well a class of series that is not adequately picked up by the other three models. In fact

the $(0,2,2)(0,1,1)_{12}$ model either performed excellently or very poorly from the view point of the MAPEs.

Place figure 2 about here.

The relationship between the MAPEs of the forecasts and the forecast horizon in figure 3 shows that when the four models ranked first (as in figure 1), the upper limit for the MAPEs is 7% only and model $(0,2,2)(0,1,1)_{12}$ performed best. Figure 4 shows the MAPEs of the forecasts of all the 120 series for each ARIMA model. As in figure 2, model $(0,2,2)(0,1,1)_{12}$ is outperformed by the other three models which are almost equivalent with respect to the MAPEs.

Place figures 3 and 4 about here.

4. THE RELATIONSHIP BETWEEN THE MAPEs OF ARIMA EXTRAPOLATIONS AND THE NOISE IN THE SERIES

Several authors have discussed the relationship between the amount of irregular variation in the series and predictability of the series (see e.g. Makridadis and Hibon, 1979; Granger and Newbold, 1978 and Nelson, 1976). This relationship is analysed here for ARIMA models. Figure 5 shows the mean absolute percentage error (MAPE) of the forecasts from an $(0,1,1)(0,1,1)_{12}$ ARIMA model for four time horizons against five classes of irregular variation.

Place figure 5 about here.

It is apparent that the MAPE increases with the amount of noise present for each of the four time horizons of 6, 12, 18 and 24 months. The increase is very large as we move from class 2 to 3, that is for series with a maximum of 10% of irregularity to a maximum of 20% and similarly from 30% to 50%. On the other hand, a decrease is observed between classes 3 and 4. This unexpected behaviour could only be explained by the characteristics of the sample series that fell in class 4 given the relative small size of the sample.

If the amount of irregularity is fixed, the dispersion of the MAPEs is very small among the four time horizons of 6, 12, 18 and 24 months. A similar pattern was observed for the MAPEs of forecasts from the ARIMA models $(0,1,2)(0,1,1)_{12}$ and $(2,1,0)(0,1,1)_{12}$.

On the other hand, figure 6 shows a different pattern for the MAPEs of the extrapolations from a $(0,2,2)(0,1,1)_{12}$ model. The dispersion of the MAPEs of the forecasts for the four time horizons is large within each class of irregular variation. However, similarly to the other 3 models, the MAPE increases with increasing amount of noise in the series.

Figure 7 shows the predictive performance of the four ARIMA models for a time horizon of 12 months (which is currently the only one included in the X-11-ARIMA program). It can be seen that the MAPEs produced by the $(0,2,2)(0,1,1)_{12}$ model are much higher than those of the other three remaining models for each class of irregular variation. In fact, the MAPEs of the $(0,2,2)(0,1,1)_{12}$ ARIMA model ranked first in the highest proportion

(33%) of the sample series with MAPEs smaller than 5% but when it ranked fourth, the MAPEs of the forecasts ranged between 50% and 70%, much higher than the MAPEs of forecasts from the other models.

Place figure 7 about here.

5. RELATIONSHIP BETWEEN MAPEs OF ARIMA EXTRAPOLATIONS AND THE PATTERN OF THE TREND-CYCLE COMPONENT OF THE SERIES

In order to evaluate the effect of the trend-cycle on the forecasting performance of the four ARIMA models, we compare here the MAPEs of the forecasts with a measure M that provides information on the pattern of the trend-cycle component in the series. The measure M is defined by

$$M = \frac{\sum_{i=2}^{n} \dfrac{(T_i - T_{i-1})}{T_{i-1}}}{\sum_{i=2}^{n} \left| \dfrac{T_i - T_{i-1}}{T_{i-1}} \right|} \qquad (1)$$

where T_i is the annual total of the original series for year i. M can take values between -1 and +1. When M is equal to one, the series has a mono-

tonically increasing annual trend. For values between 0 and 1, the series has reversals of direction in its annual rates of change (negative values). The series is then assumed to be affected by the business cycle.

Values between -1 and 0 will have a similar interpretation but the sample of series chosen for this study has very few cases where M is smaller than .60. Table 2 shows the frequency distribution of the sample according to the values of M.

Place table 2 about here.

Figure 8 shows the MAPE of the forecasts from an $(0,1,1)(0,1,1)_{12}$ ARIMA model for the four time horizons versus several values of M. It is apparent that the average forecast error decreases with increasing M. For the case of M = 1, a perfectly monotonic trend, the MAPE is very small. There are no major differences for the four time horizons when M is close to 1. Series with more cyclical movements ($.6 < M < .8$) however, tend to have higher MAPEs at longer time horizons i.e. the models predictive power diminishes if the series are affected by cycles. The other models showed similar forecast error pattern according to the M measure.

Place figure 8 about here.

Finally, we looked at the relationship between the MAPE of the ARIMA extrapolations and the various amount of irregularity in the series for fixed values of M. Figure 9 shows the MAPEs for the $(0,1,1)(0,1,1)_{12}$ model with a 12 month time horizon and M = 1 and $M < .99$.

Place figure 9 about here.

When $M = 1$, the MAPEs of the forecasts are small and change very little with increasing values of the irregular component. On the other hand, when the series have been cyclically affected such as for $M < .99$, the MAPEs increase with increasing irregularity in the series. In fact, when the amount of irregularity in the series is larger than 10%, the presence of the business cycle seems to be the main cause that affects the forecasting performance of the ARIMA models.

6. CONCLUSIONS

This study analyzed the predictive performance of the four ARIMA models to be incorporated into a new version of the X-11-ARIMA seasonal adjustment computer program. The predictive performance of each model is evaluated globally (the standard approach) for 120 macroeconomic series and also as a function of both the amount of irregularity in each series and the pattern of the trend. For each case the mean absolute percentage error (MAPE) of the forecasts are calculated for various time horizons. The results show that the MAPEs for these four ARIMA models are more sensitive to the presence of the business cycle in the series than to either the amount of irregularity or the length of the forecast horizon. This is particularly evident when the contribution of the irregulars to the total variance of the series is larger than 10%.

ACKNOWLEDGEMENT

The authors would like to acknowledge the valuable computer support provided by Helen Fung and Katherine Pashley.

REFERENCES

CHIU, K., HIGGINSON, J. and HUOT, G. (1985): "Performance of ARIMA Models in Time Series" Survey Methodology Journal, Vol. 11, No. 1, pp: 51-64.

DAGUM, E.B. (1980): The X-11-ARIMA Seasonal Adjustment Method; Ottawa: Statistics Canada, Cat. No. 12-564E.

DAGUM, E.B. (1982.a): "Revision of Time Varying Seasonal Filters", Journal of Forecasting, Vol. 1, pp: 173-187.

DAGUM, E.B. (1982.b): "The Effects of Asymmetric Filters on Seasonal Factor Revisions"; Journal of the American Statistical Association, Vol. 77, No. 380, pp: 732-738.

DAGUM, E.B. (1983): "Spectral Properties of the Concurrent and Forecasting Linear Filters of the X-11-ARIMA Method"; The Canadian Journal of Statistics, Vol. 11, No. 1, pp: 73-90.

DAGUM, E.B. and MORRY, M. (1984): "Basic Issues on the Seasonal Adjustment of the Canadian Consumer Price Index"; Journal of Business and Economic Statistics, Vol. 2, No. 3, pp: 250-259.

GRANGER, C.W.J. and NEWBOLD, P. (1977): Forecasting Economic Time Series, Academic Press, New York.

HUOT, G., CHIU, K., HIGGINSON, J. and GAIT, N. (1986): "Analysis of Revisions in the Seasonal Adjustment of Data Using X-11-ARIMA Model-Based Filters"; International Journal of Forecasting, Vol. 2, No. 2, pp: 217-229.

KENNY, P. and DURBIN, J. (1982): "Local Trend Estimation and Seasonal Adjustment of Economic Time Series"; Journal of the Royal Statistical Society, Series A, 145, Part 1, pp: 1-41.

KUIPER, J. (1978): "A Survey and Comparative Analysis of Various Methods of Seasonal Adjustment"; in Seasonal Analysis of Economic Time Series (Arnold Zellner, Editor), Washington, D.C.: U.S. Government Printing Office, pp: 59-76.

MAKRIDAKIS, S. and HIBON, M. (1979): "Accuracy of Forecasting: An Empirical Investigation" (with discussion); Journal of the Royal Statistical society, Ser. A. 142, Part 2, pp: 97-145.

NELSON, C.R. (1976): "The Interpretation of R^2 in Autoregressive-Moving Average Time Series Models"; The Americal Statistician, Vol. 30, No. 4, pp: 175-180.

Table 1. The average irregular variations in the series as identified by the X-11-ARIMA program and ARIMA models

Classes	$(1 - \text{average } R^2) \times 100$ (ARIMA models)	$\bar{I}$ (X-11-ARIMA)
1	2.56	2.79
2	4.18	7.50
3	5.58	14.13
4	10.35	23.50
5	28.65	40.79

Table 2. Frequency Distribution of the Sample According to the Measure M

Intervals	Frequency	% of Total
-1.0	0	0.00
-0.99 to -0.80	1	0.83
-0.79 to -0.60	0	0.00
-0.59 to -0.40	1	0.83
-0.39 to -0.20	3	2.50
-0.19 to 0.00	2	1.67
0.01 to 0.20	1	0.83
0.21 to 0.40	7	5.83
0.41 to 0.60	13	10.83
0.61 to 0.80	11	9.17
0.81 to 0.99	34	28.33
1.00	47	39.18
Total	120	100.00

AT FOUR DIFFERENT FORECAST HORIZONS
WHEN THE MODEL RANKED FIRST

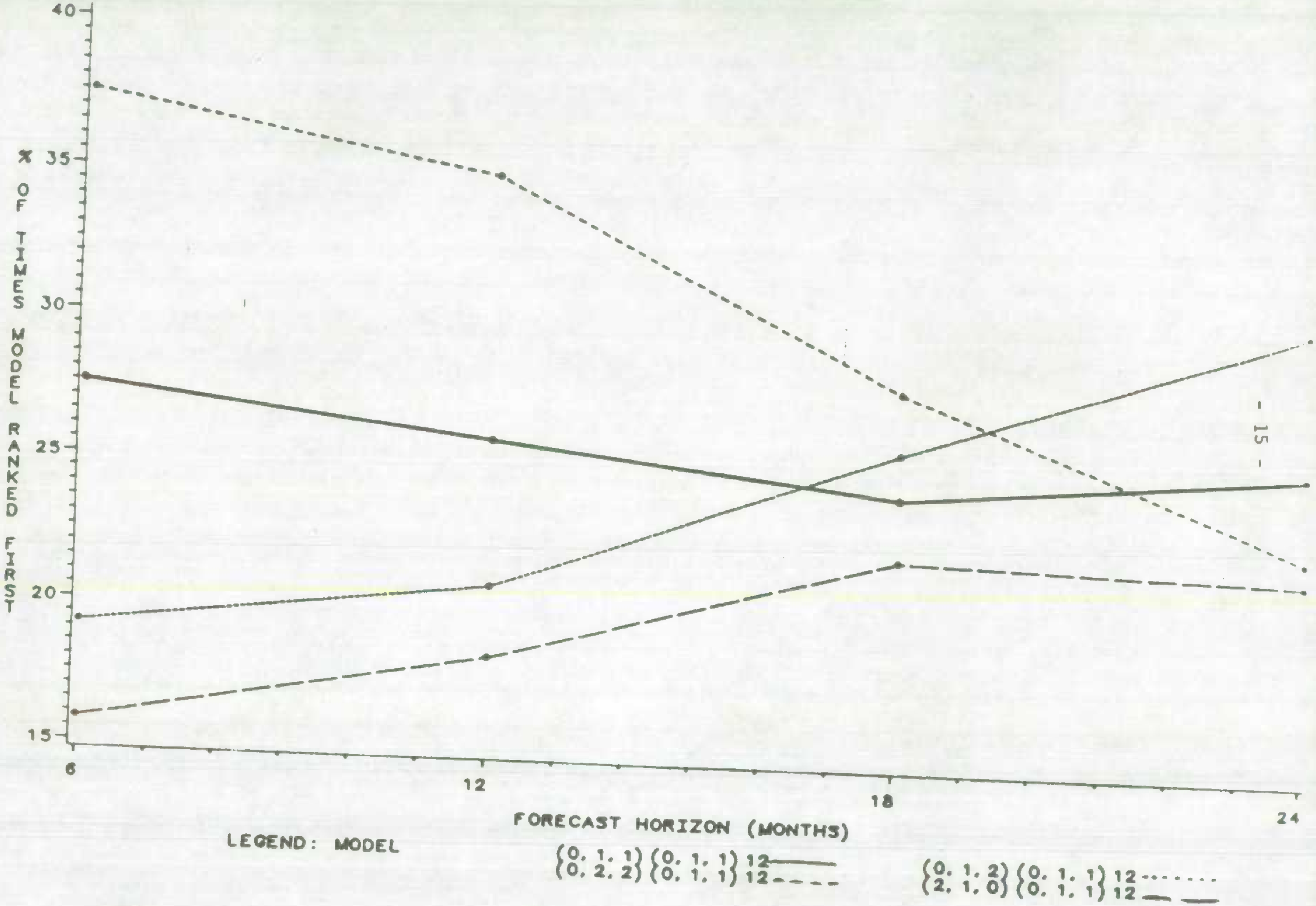


FIGURE 1

PERFORMANCE RANKING OF THE FOUR ARIMA MODELS
AT FOUR DIFFERENT FORECAST HORIZONS
WHEN THE MODEL RANKED FOURTH

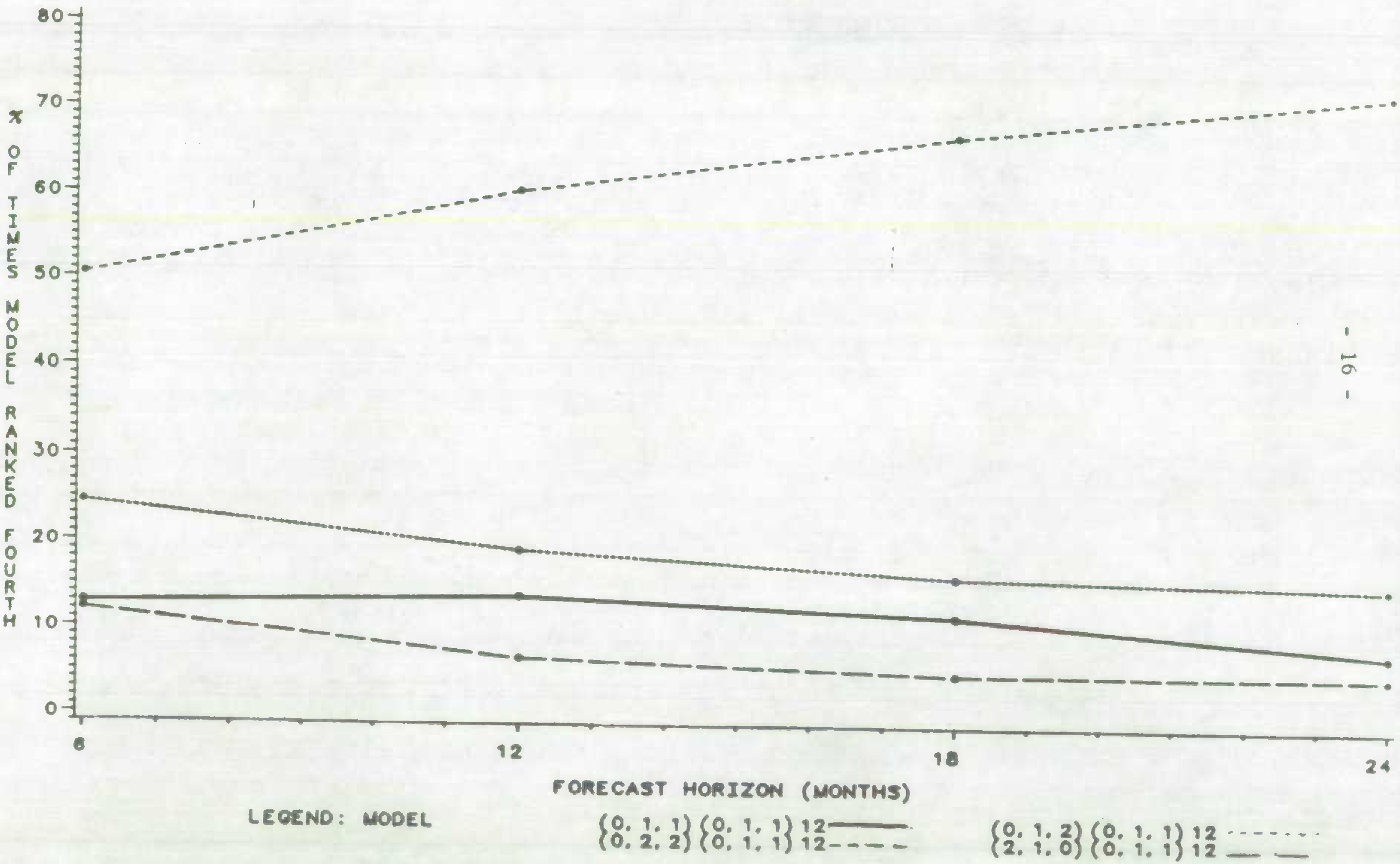


FIGURE 2

M.A.P.E.S FOR THE SERIES FORECAST USING THE ARIMA MODEL WHICH HAS THE SMALLEST FORECAST ERRORS

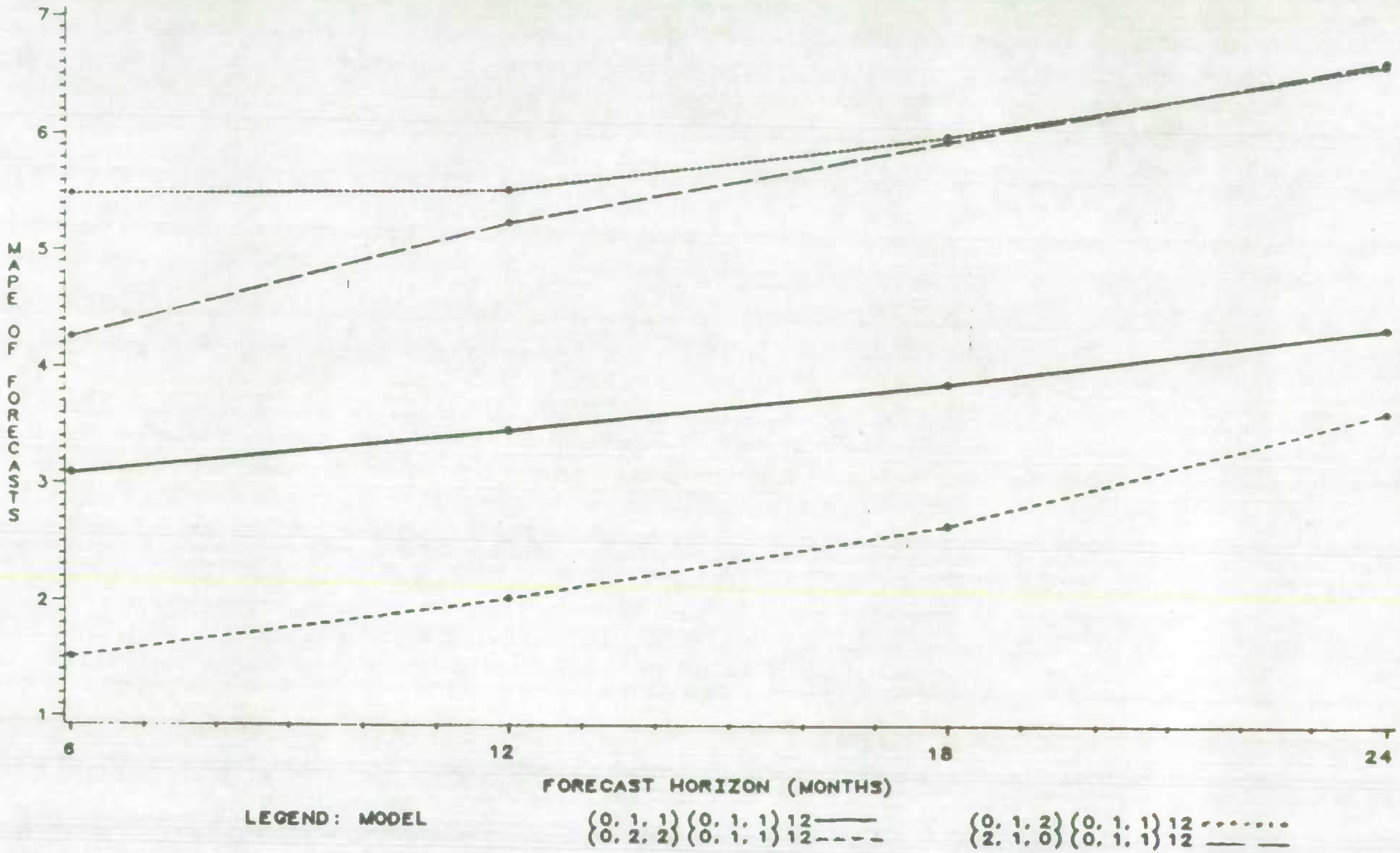
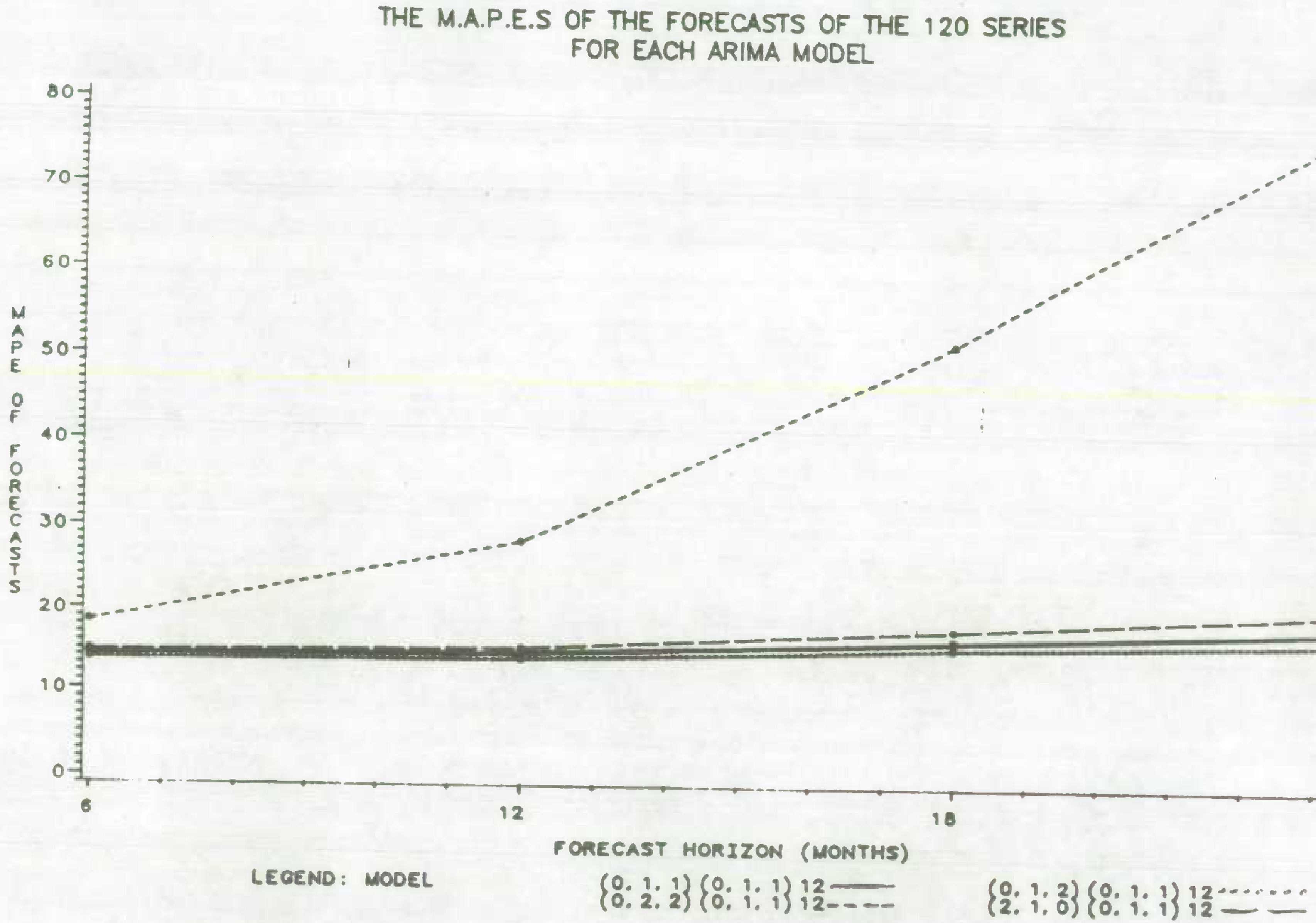


FIGURE 3



THE M.A.P.E.S OF THE FORECASTS OF THE 120 SERIES
FOR EACH ARIMA MODEL
MAPE OF FORECASTS
80
70
60
50
40
30
20
10
0
6
12
18
24
FORECAST HORIZON (MONTHS)
LEGEND: MODEL
(0, 1, 1) (0, 1, 1) 12
(0, 2, 2) (0, 1, 1) 12
(0, 1, 2) (0, 1, 1) 12
(2, 1, 0) (0, 1, 1) 12

FIGURE 4

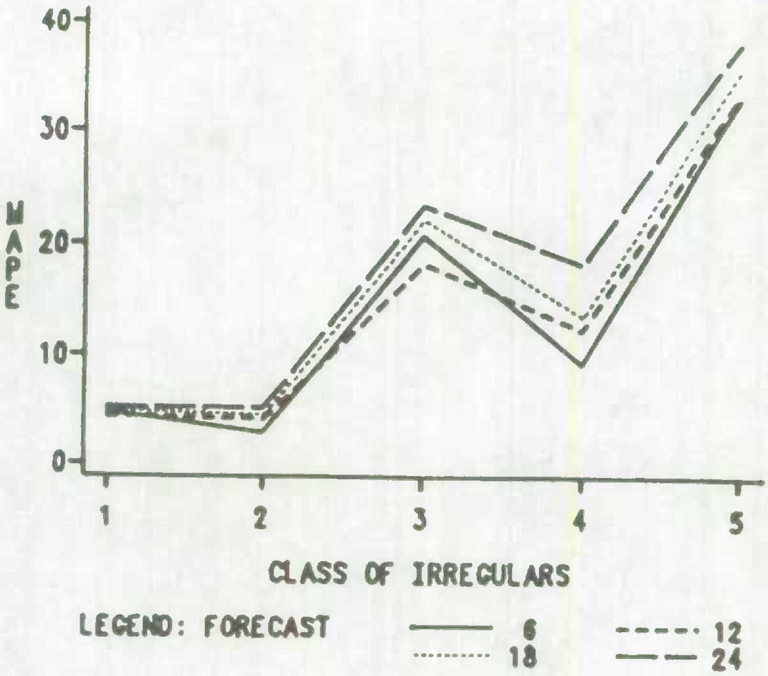


Figure 5. Mape of forecasts from a (011)(011)12 ARIMA model for different time horizons for various amounts of irregularity in the series

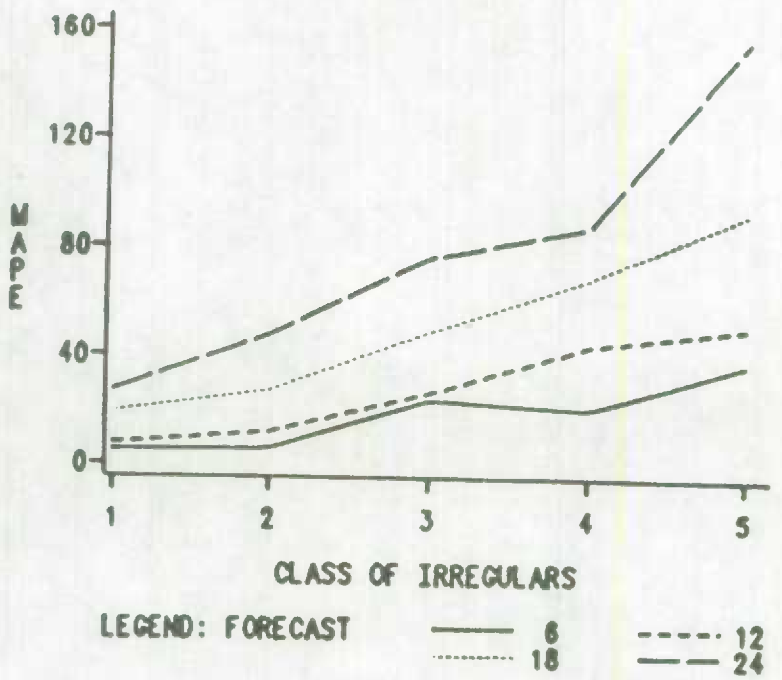


Figure 6. Mape of forecasts from a (022)(011)12 ARIMA model for different time horizons for various amounts of irregularity in the series

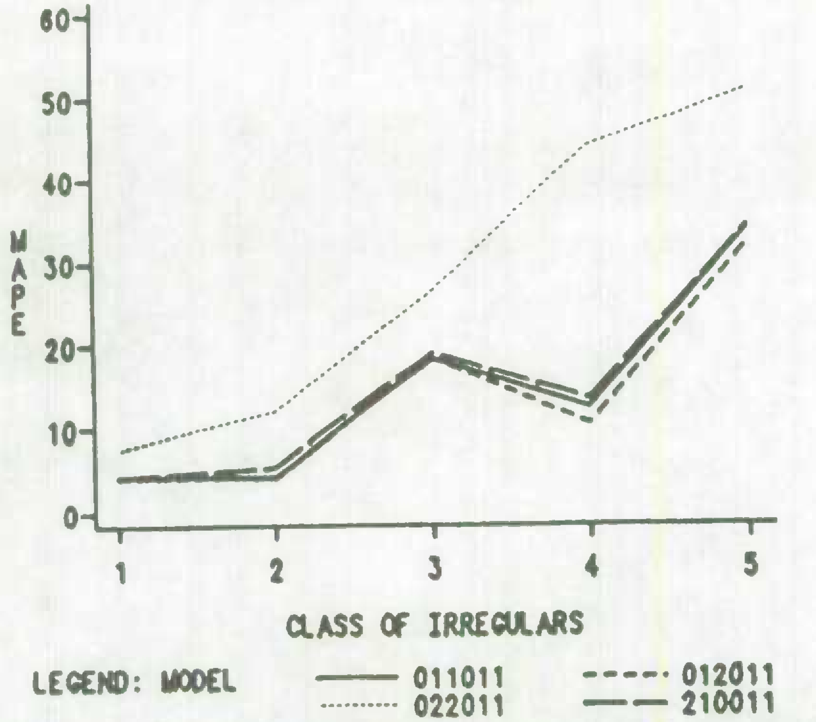


Figure 7 Mape of the 12 month ahead forecasts from four ARIMA models for various amounts of irregularity in the series

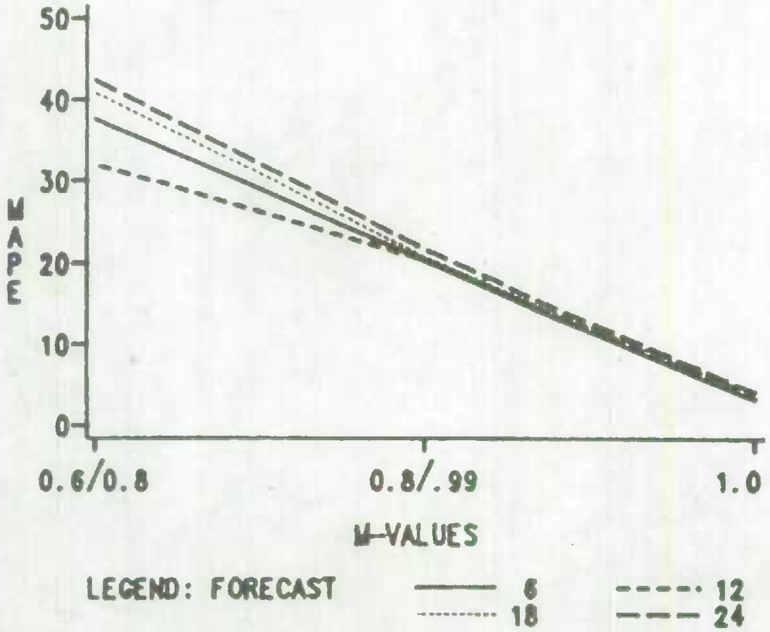


Figure 8 Mape of forecasts from a (011)(011)12 ARIMA model for different time horizons as a function of the M-values of the series

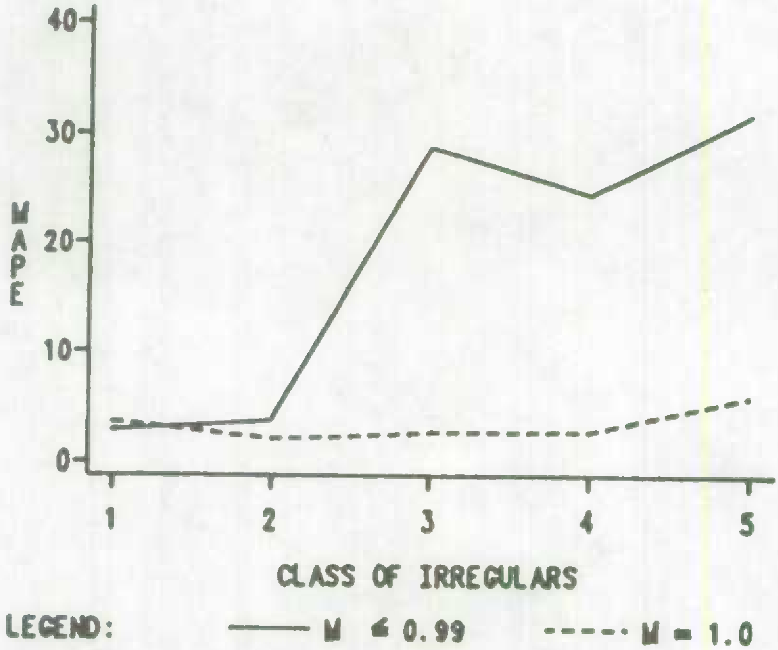


Figure 9 Mape of the 12 month ahead forecasts from a (011)(011)12 ARIMA model for various amounts of irregularity, M ≤ 0.99 and M = 1.0

76018

c. 1