ARMA MODELS AND THE BOX JENKINS METHODOLOGY

by

S. MAKRIDAKIS*
AND
M. HIBON**

95/45/TM
(Revised version of 95/33/TM)

* Research Professor of Decision Sciences and Information Systems, at INSEAD, Boulevard de Constance, Fontainebleau 77305 Cedex, France.

* Research Associate, at INSEAD, Boulevard de Constance, Fontainebleau 77305 Cedex, France.

ARMA MODELS AND THE BOX JENKINS METHODOLOGY

Spyros Makridakis and Michèle Hibon, INSEAD

Abstract

The purpose of this paper is to study the Box-Jenkins methodology to ARIMA models and determine the reasons why in empirical tests it is found that the post-sample forecasting accuracy of such models is worse than much simpler time series methods. It is concluded that the major problem is the way of making the series stationary in its mean (i.e., the method of differencing) that has been proposed by Box and Jenkins. If alternative approaches are utilized to remove and extrapolate the trend in the data, ARMA models outperform the corresponding methods involved in the great majority of cases. In addition it is shown that using ARMA models to seasonally adjusted data slightly improves post-sample accuracies while simplifying the use of ARMA models. It is also confirmed that transformations slightly improve post-sample forecasting accuracy, particularly for long forecasting horizons. Finally, it is demonstrated that AR(1) and AR(2), or their combination, produce as accurate post-sample results as those found through the application of the Box-Jenkins methodology.

Keywords

Time-Series Forecasting, ARMA Models, Box-Jenkins, Empirical Studies, M-Competition

ARMA MODELS AND THE BOX JENKINS METHODOLOGY

Spyros Makridakis and Michèle Hibon, INSEAD

AutoRegressive (AR) models were first introduced by Yule in 1926. They were consequently supplemented by Slutsky who in 1937 presented Moving Average (MA) schemes. It was Wold (1938), however, who combined both AR and MA schemes and showed that ARMA processes can be used to model all stationary time series as long as the appropriate order of p, the number of AR terms, and q, the number of MA terms, was appropriately specified. This means that any series x_t can be modelled as a combination of past x_t values and/or past e_t errors, or

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \ldots + \phi_p x_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \ldots - \theta_q e_{t-q} \qquad (1)$$

Using (1) for modelling real life time series requires four steps. First the original series x_t must be transformed to become stationary around its mean and its variance. Second, the appropriate order of p and q must be specified. Third, the value of the parameters $\phi_1, \phi_2, \ldots, \phi_p$ and/or $\theta_1, \theta_2, \ldots, \theta_q$ must be estimated using some non-linear optimization procedure that minimizes the sum of square errors or some other appropriate loss function. Finally, practical ways of modelling seasonal series must be envisioned and the appropriate order of such models specified.

The utilization of the theoretical results suggested by Wold, expressed by equation (1), to model real life series did not become possible until the mid 1960s when computers, capable of performing the required calculations to optimize the parameters of (1), became available and economical. Box and Jenkins (1976, original edition 1970) popularized the use of ARMA models through the following: (a) providing guidelines for making the series stationary in both its mean and variance, (b) suggesting the use of autocorrelations and partial autocorrelation coefficients for determining appropriate values of p and q (and their seasonal equivalent P and Q when the series exhibited seasonality), (c) providing a set of computer programs to help users identify appropriate values for p and q, as well as P and Q, and estimate the parameters involved and (d) once the parameters of the model were estimated, a diagnostic check was proposed to determine whether or not the residuals e_t were white noise, in which case the order of the model was considered final (otherwise another model was entertained in (b) and steps (c) and (d) were repeated). If the diagnostic check showed random residuals then the model developed was used for forecasting or control purposes assuming of course constancy, that is that the order of the model and its non-stationary behavior, if any, would remain the same during the forecasting, or control, phase.

The approach proposed by Box and Jenkins came to be known as the Box-Jenkins methodology to ARIMA models, where the letter "I", between AR and MA, stood for the word "Integrated". ARIMA models and the Box-Jenkins methodology became highly popular in the 1970s among academics, in particular when it was shown through empirical studies (Cooper, 1972; Nelson, 1972; Elliot, 1973; Narasimham et al., 1974; McWhorter, 1975; for a survey see Armstrong, 1978) that they could outperform the large and complex econometric models, popular at that time, in a variety of situations.

Empirical Evidence

The popularity of the Box-Jenkins methodology to ARIMA models was shattered when empirical studies (Groff, 1973; Geurts and Ibrahim, 1975; Makridakis and Hibon, 1979; Makridakis et al., 1982; Huss, 1985; Makridakis et al., 1993; Fildes et al., 1995), using real data, showed that simple methods were equally or more accurate than Box-Jenkins (the short name attributed to the methodology proposed by Box and Jenkins) when post-sample comparisons were made. Today, after acrimonious arguments and considerable debates, it is accepted by a large number of researchers that in empirical tests Box-Jenkins is not an accurate method for post-sample time series forecasting, at least in the domains of business and economic applications where the level of randomness is high and where constancy of pattern, or relationships, cannot be assured. At the same time there is a sizeable number of those who oppose the findings of empirical studies and do not accept the conclusion that the

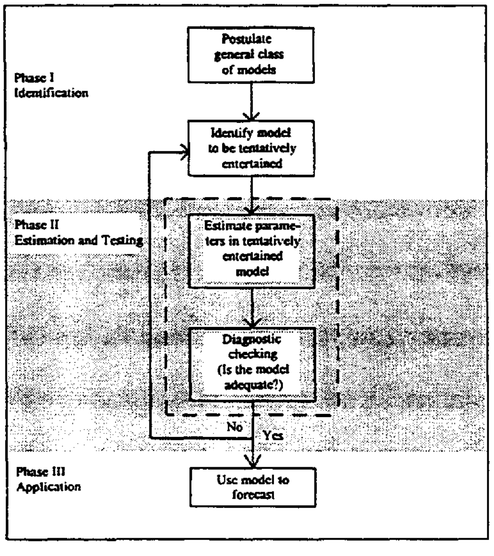
Box-Jenkins methodology is not the most appropriate approach for modelling time series (Jenkins, 1982; Newbold, 1983).

The purpose of this paper is to examine the post-sample forecasting performance of ARIMA models in order to determine the contribution of each of its elements. That is, dealing with the seasonality in the series, achieving stationarity, specifying an appropriate ARMA model and doing diagnostic checking. It is concluded that the major problem with the Box-Jenkins methodology is the "I" (that is the way that the series are made stationary in their mean). When alternative ways of dealing with and extrapolating the trend are provided, ARMA models are more accurate than the corresponding time series methods that extrapolate the trend in the same way. These encouraging results suggest that statistical theory and empirical results are in agreement and that more appropriate ways of dealing with the trend in ARMA, or any other type of time series models, must be envisioned beyond those available today.

The Steps (Elements) of the Box-Jenkins Methodology

Figure 1 presents the four steps of the Box-Jenkins methodology. This section examines each of these steps and discusses its possible contribution to post-sample forecasting accuracy and therefore its need and value.

FIGURE 1 SCHEMATIC REPRESENTATION OF THE BOX-JENKINS APPROACH
(Source: Makridakis et al., 1983)



Stationarity: Before equation (1) carr be used the series ought to be stationary in its mean and variance. The Box-Jenkins approach suggests short and seasonal (long) differencing to achieve stationarity in the mean, and logarithmic or power transformation to achieve stationarity in the variance. The value of both differencing and transformations have been questioned. Pierce (1977) argued that differencing was not an appropriate way of making the data stationary and instead he proposed linear de-trend. Nelson and Plosser (1982) argued that some series could be better made stationary through differencing while others through linear de-trending. Others (Parzen, 1982; Newton and Parzen, 1984; Meese and Geweke, 1984) have used a pre-filter consisting of a long memory AR model to capture possible non-stationarity in the series before using a regular ARMA model.

Box and Jenkins suggest logarithmic or power transformations to achieve stationarity in the variance. The value of such transformations to improve post-sample forecasting accuracy has also been debated and no agreement has been reached as to whether or not transformations are helpful (Chatfield and Prothero, 1973). At the same time it is clear that transformations require personal judgment and the possibility of making errors, even when utilized by high level academic experts (see comments on paper by Chatfield and Prothero, 1973). At the empirical level there is also no evidence that logarithmic or power transformations improve post-sample forecasting accuracy (Granger and Nelson, 1978; Makridakis and Hibon, 1979; Meese and Geweke, 1984).

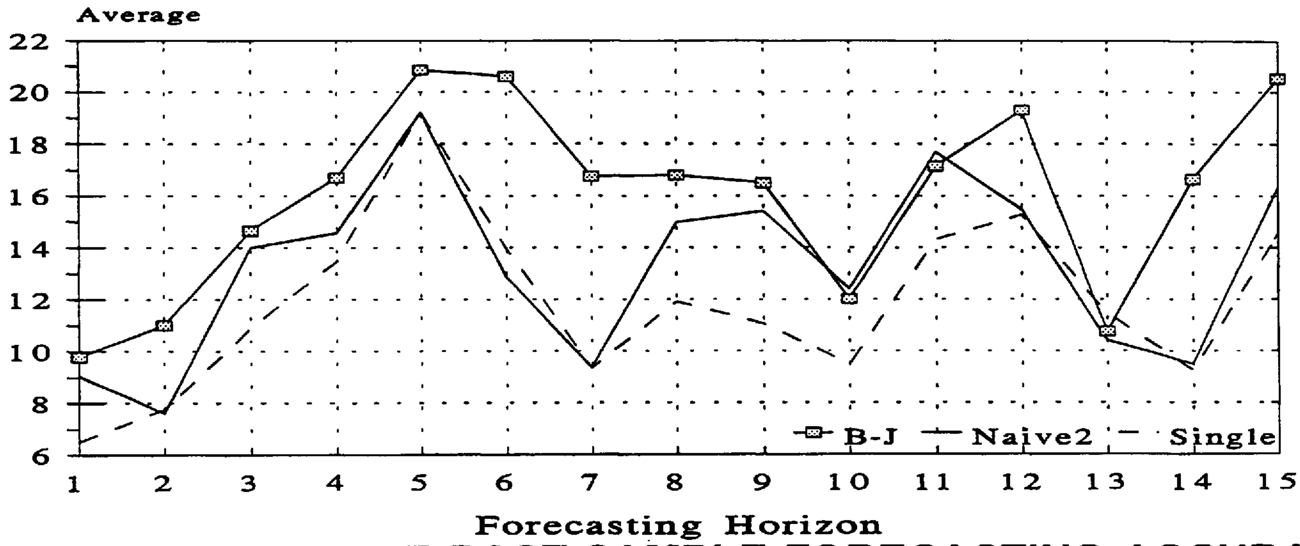
Seasonality: In case the series are seasonal, the Box-Jenkins methodology proposes multiplicative seasonal models coupled with long-term differencing, if necessary, to achieve stationarity in the mean. The difficulty with such an approach is that there is practically never enough data available to determine the appropriate level of the seasonal ARMA model with any reasonable degree of confidence. Users therefore proceed through trial and error in both identifying an appropriate seasonal model and also in selecting the right long-term (seasonal) differencing. In addition, seasonality complicates the utilization of ARMA models as it requires using many more data while increasing the modelling options available -- and making the selection of an appropriate model more difficult. There are no studies showing the value of the seasonal part of the ARIMA models. Moreover, there has not been any empirical work to test whether or not deseasonalizing the data first, using a decomposition procedure (a suggestion made by Durbin, 1979), and subsequently using the Box-Jenkins method on the seasonally adjusted data improves post-sample forecasting accuracy.

Order of ARMA Model: The order of the ARMA model is found by examining the autocorrelations and partial autocorrelations of the stationary series. Box and Jenkins (1976) provided both a theoretical framework and practical rules for determining appropriate values for p and q as well as their seasonal counterparts P and Q. The only difficulty is that often more than one model could be entertained, requiring the user to choose one of them without any knowledge of the implications of his or her choice on post-sample forecasting accuracy. In terms of the Box-Jenkins methodology, any model which results in random residuals is an appropriate one. There is no empirical work testing the effects of selecting various p and q (and P and Q) orders, some of which might result in non-random residuals, or determining the implications of such a choice on post-sample forecasting accuracy. Furthermore, Box and Jenkins recommend the principle of parsimony meaning that a simpler (having fewer parameters) model should be selected in case more than one model is possible. There has also not been any work to determine if this suggestion results in improvements in post-sample forecasting accuracy.

Estimating the Model's Parameters: This part of the Box-Jenkins methodology is the most straightforward one. The non-linear optimization procedure, based on the method of steepest descent (Marquardt, 1963), is used to estimate the parameter values of p and/or q (and their seasonal equivalent P and/or Q). Apart from occasional problems when there is no convergence (in which case another model is entertained) the estimation provides no special difficulties except for its inability to guarantee a global optimum (a common problem of all non-linear algorithms). The estimation is completely automated requiring no judgmental

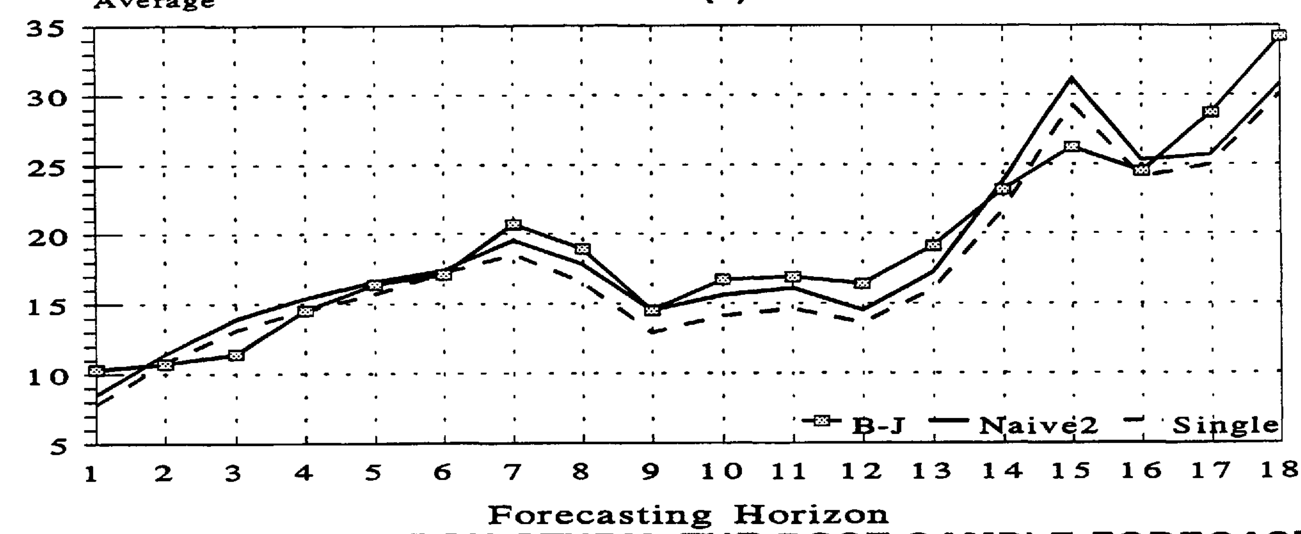
FIGURE 2(a)

M2-COMPETITION: THE POST-SAMPLE FORECASTING ACCURACY OF B-J, NAIVE2 AND SINGLE EXPONENTIAL SMOOTHING



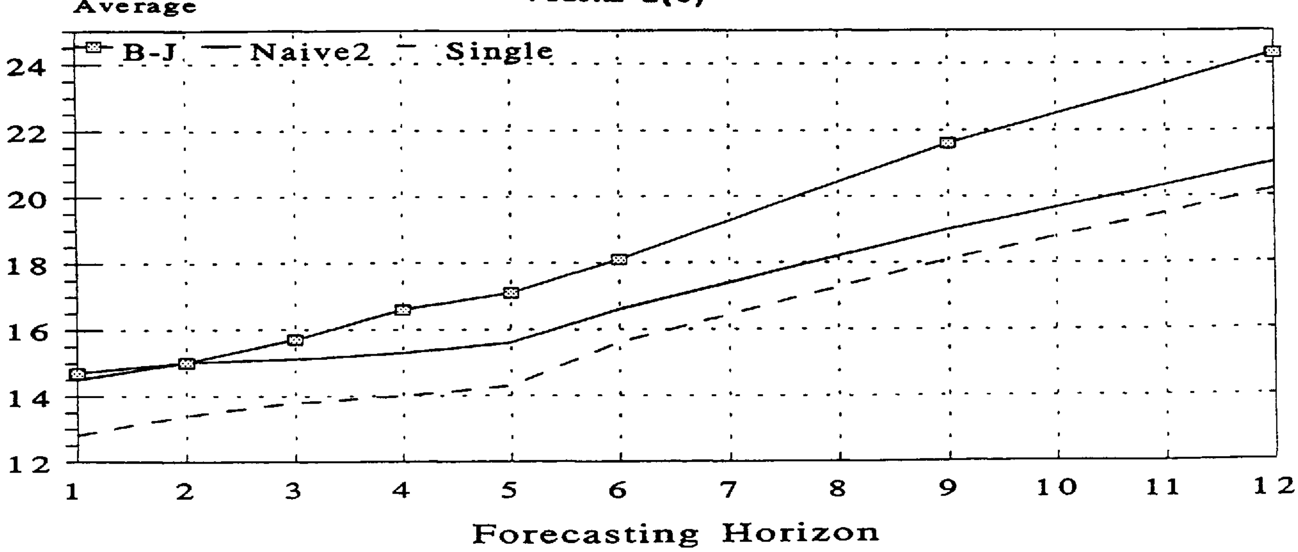
M-COMPETITION: THE POST-SAMPLE FORECASTING ACCURACY OF B-J, NAIVE2 AND SINGLE EXPONENTIAL SMOOTHING

FIGURE 2(b)



MAKRIDAKIS AND HIBON STUDY: THE POST-SAMPLE FORECASTING ACCURACY OF B-J, NAIVE2 AND SINGLE EXPONENTIAL SMOOTHING

FIGURE 2(c)



inputs, and therefore testing, as all computer programs use the same algorithm in applying the Marquardt optimization procedure.

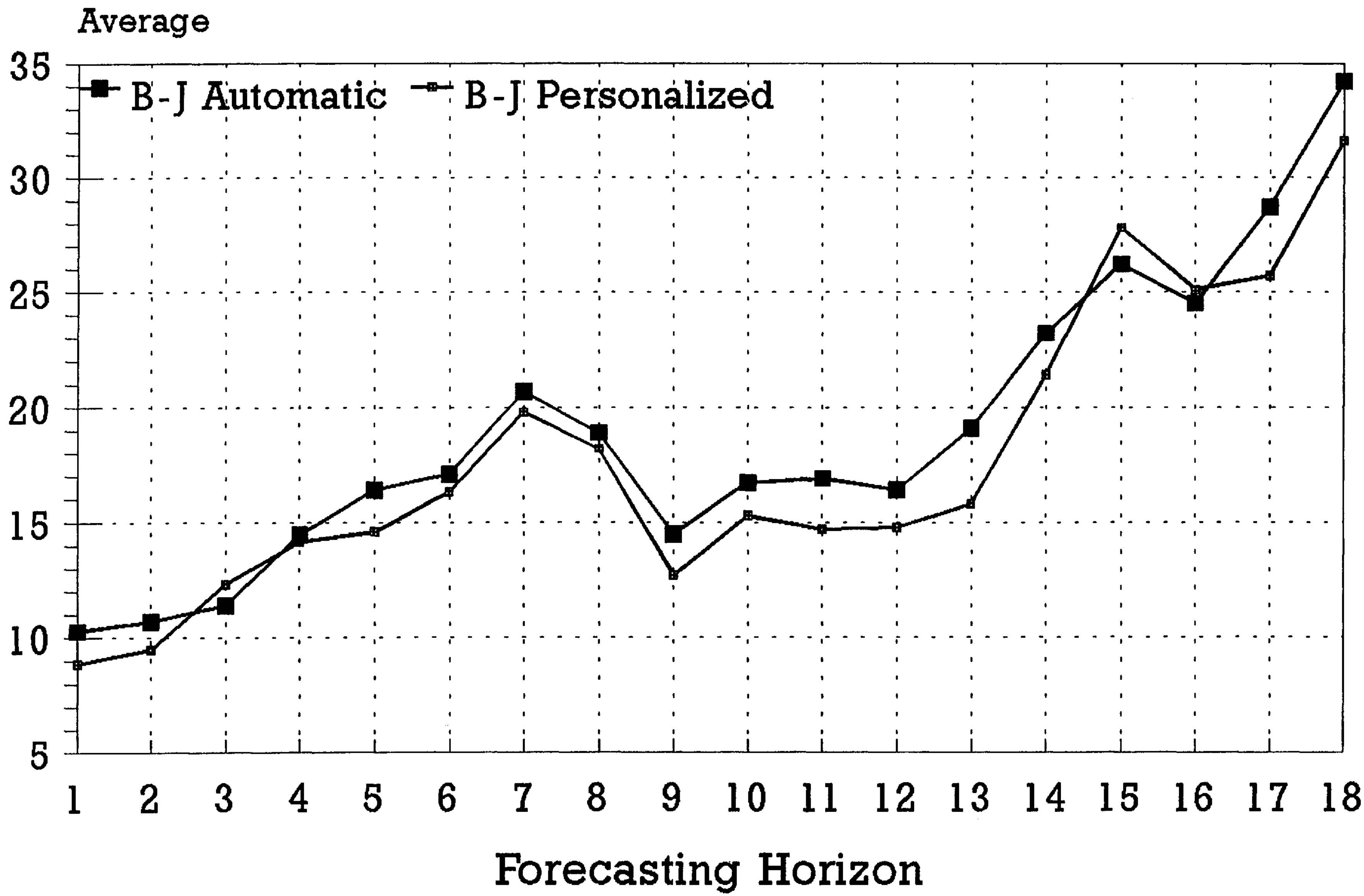
Diagnostic Checks: Once an appropriate model had been entertained and its parameters estimated, the Box-Jenkins methodology required examining the residuals of the actual values minus those estimated through the model. If such residuals are random, it is assumed that the model is appropriate. If not, another model is entertained, its parameters estimated, and its residuals checked for randomness. In practically all instances a model could be found to result in random residuals. Several tests (e.g., the Box-Pierce Statistic, Box and Pierce, 1970) have been suggested to help users determine if overall the residuals are indeed random. Although it is a standard statistical procedure not to use models whose residuals are not random, it might be interesting to test the consequences of lack of residual randomness on post-sample forecasting accuracy.

Post-Sample Forecasting Accuracy: Personalized vs Automatic Box-Jenkins

The Makridakis and Hibon (1979) study, the M-Competition (Makridakis et al., 1982), the M2-Competition (1993) as well as many other empirical studies (Schnaars, 1986; Koehler and Murphree, 1988; Geurts and Kelly, 1986; Watson et al., 1987; Collopy and Armstrong, 1992) have demonstrated that simple methods such as exponential smoothing outperform, on average, the Box-Jenkins methodology to ARMA models. Figures 2(a), (b) and (c) show the MAPE (Mean Absolute Percentage Errors), for various forecasting horizons, of Naive 2 (a deseasonalized random walk model), Single exponential smoothing (after the data has been deseasonalized, if necessary, and the forecasts subsequently re-seasonalized) as well as those of the Box-Jenkins methodology found in the Makridakis and Hibon (1979) study, the M-Competition (Makridakis et al., 1982) and the M2-Competition (Makridakis et al., 1993). The results indicate that Single smoothing outperformed "Box-Jenkins" overall and in most forecasting horizons, while Naive 2 also does better than "Box-Jenkins", although by a lesser amount. The results of Figure 2 are surprising since it has been demonstrated that Single exponential smoothing is a special case of ARMA models (Cogger, 1974; Gardner and

FIGURE 3

B-J: MEAN ABSOLUTE PERCENTAGE ERROR (MAPE)
Automatic vs Personalized (M-Competition: 111 Series)



McKenzie, 1985). Moreover, it makes no sense that Naive 2, which simply uses the latest available value, taking seasonality into account, does so well in comparison to the Box-Jenkins methodology, a statistically sophisticated and theoretically correct method.

In the M-Competition (Makridakis et al., 1982) the "Box-Jenkins" method was run on a subset of 111 (one out of every 9) series from the total of the 1001 series utilized. The reason was that the method required personal judgment, making it impractical to use all 1001 series as the expert analyst had to model each series individually, following the various steps described in the last section, and spending, on average, about one hour before a model could be confirmed as appropriate for forecasting purposes (see Andersen and Weiss in Makridakis et al., 1984).

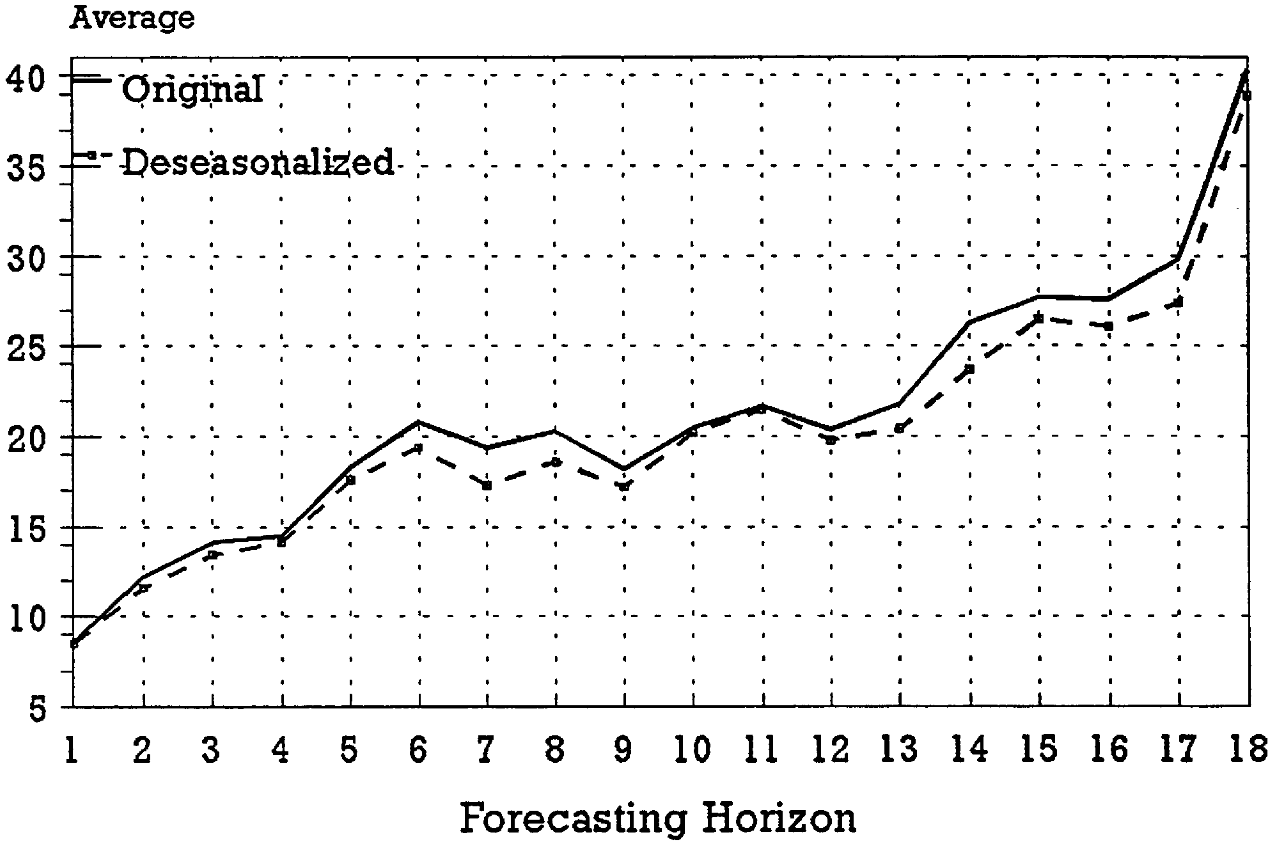
Since the M-Competition was completed, several empirical studies have shown that automatic Box-Jenkins approaches (Hill and Fildes, 1984; Libert, 1983; Texter and Ord, 1989) performed about the same or better in terms of post-sample accuracy as the personalized approach followed by Andersen and Weiss (1984). Figure 3 shows the results of a specific automatic Box-Jenkins program (Stellwagen and Goodrich, 1991) together with that of the personalized approach utilized by Andersen and Weiss in the M-Competition. Figure 3 illustrates that the post-sample accuracies of the automatic and personalized approaches are about the same, allowing us to use an automatic Box-Jenkins version for the remainder of this study.

Attributing the Differences in Post-Sample Forecasting Accuracies

Since we found no substantive differences between the personalized and automatic versions of Box-Jenkins (see Figure 3), we have run all the 1001 series of the M-Competition using an automatic Box-Jenkins procedure (Stellwagen and Goodrich, 1991) in order to have a large sample and be able to attribute more reliably differences in post-sample forecasting accuracy to the various aspects of the Box-Jenkins methodology.

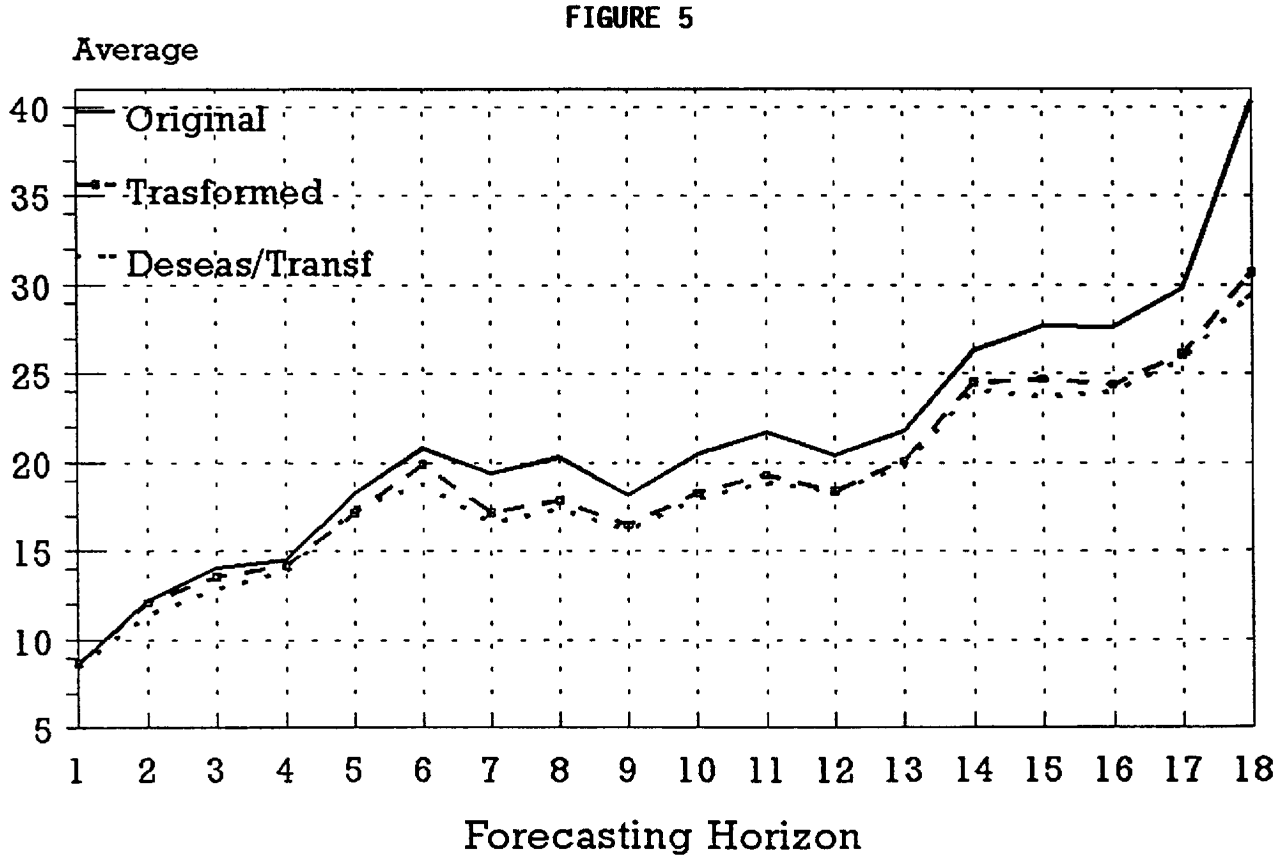
FIGURE 4

MEAN ABSOLUTE PERCENTAGE ERROR (MAPE): ORIGINAL vs SEASONALLY ADJUSTED DATA



MEAN ABSOLUTE PERCENTAGE ERROR (MAPE): ORIGINAL, DESEASONALIZED AND TRANSFORMED DATA

FIGURE 5



Deseasonalizing the Series

In the discussion of the Makridakis and Hibon (1979) paper it was suggested (Durbin, 1979) that the Box-Jenkins methodology should also be applied to the seasonally adjusted data to determine the effect of seasonality in post-sample accuracy. This suggestion is being tested in this study.

X_t, the original data, can be deseasonalized by dividing it by the seasonal index, S_j, computed through the classical decomposition method (Makridakis et al., 1983), or

$$X_t' = X_t / S_j \tag{2}$$

where S_j is the seasonal index corresponding to the jth month, if the data is monthly, or the jth season if it is quarterly. If the data is not seasonal, all indices are set to equal 1.

Once the forecasts have been computed using the automatic Box-Jenkins program, they can be reseasonalized by multiplying them by the corresponding seasonal index, or

$$\hat{X}_t = \hat{X}_t' S_j$$

Figure 4 shows the MAPE of the original and deseasonalized versions of the automatic Box-Jenkins. Using ARIMA models on the deseasonalized data results in more accurate post-sample forecasts, although the differences between the two approaches are small and, although consistent, not statistically significant. This finding suggests slight but steady improvements in post-sample forecasting accuracy, albeit not statistically significant, between deseasonalizing the data, or using ARIMA models to the original series as suggested by Box and Jenkins. As it is easier and much simpler to apply ARIMA models to the deseasonalized series, this study suggests that at least for the 1001 series of the M-Competition it is preferable to use ARIMA models to seasonally adjusted data.

Log or Power Transformations

Figure 5 shows the MAPEs when log or power transformations were employed, when necessary, to achieve stationarity in the variance of the original data and the seasonally adjusted data. There is a small improvement when logarithmic or power transformations are applied to the original and deseasonalized data, but the differences are not statistically significant except for horizon 18. However the differences are consistent and increase as the forecasting horizon becomes longer. This finding is not in agreement with previous ones claiming that power or log transformations do **not** improve at all post-sample forecasting accuracy (when the transformation is applied to the deseasonalized data there is an additional but very slight improvement, see Figure 5). As transformations improve forecasting accuracy it must be determined if the extra work required to make these transformations justifies the small improvements found, and whether or not such statistically insignificant improvements (except for horizon 18) will be also found with other series than those of the M-Competition.

Transformations for Achieving Stationarity in the Mean

To the approach of differencing suggested by Box and Jenkins (1976) for achieving stationarity in the mean there are several alternatives employing various ways to remove the trend in the data.

The trend, T_t, can be modeled as:

$$T_t = f(t) \tag{3}$$

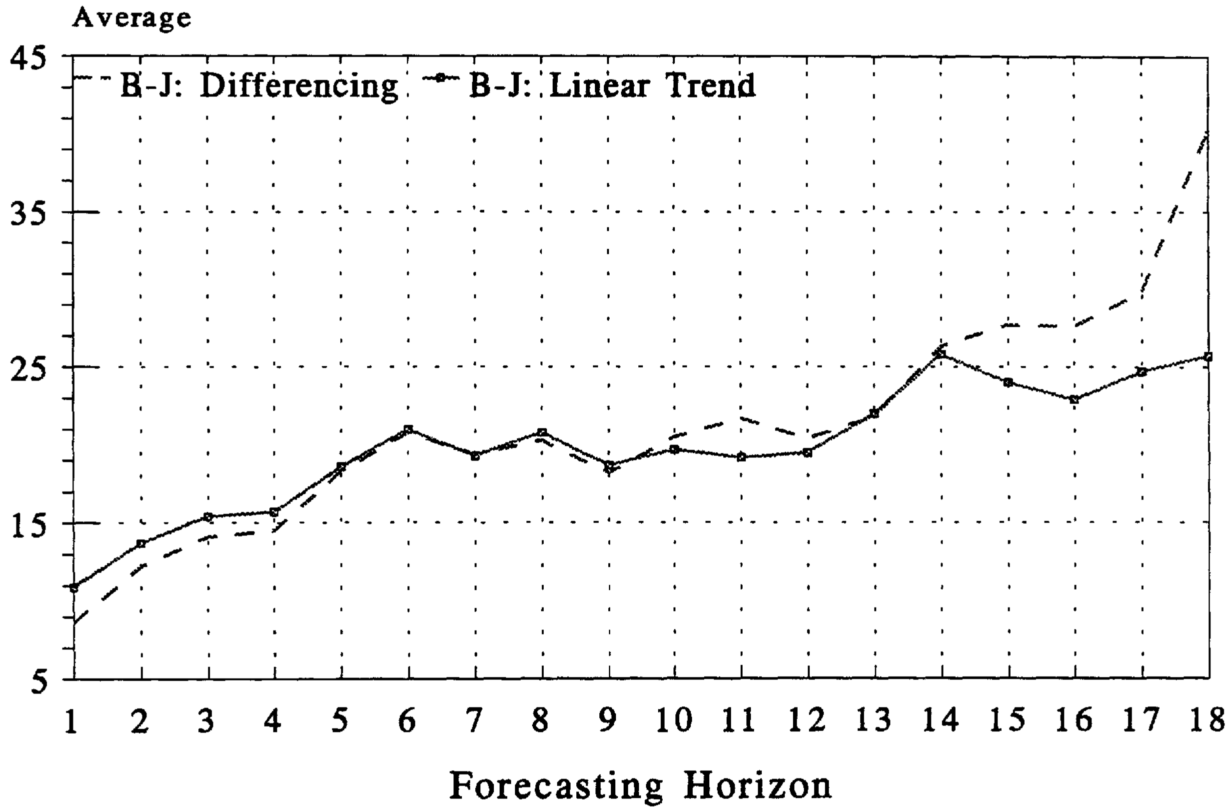
where $t = 1, 2, 3, \ldots, n$

In the case of a linear trend (3) becomes

$$T_t = a + bt \tag{4}$$

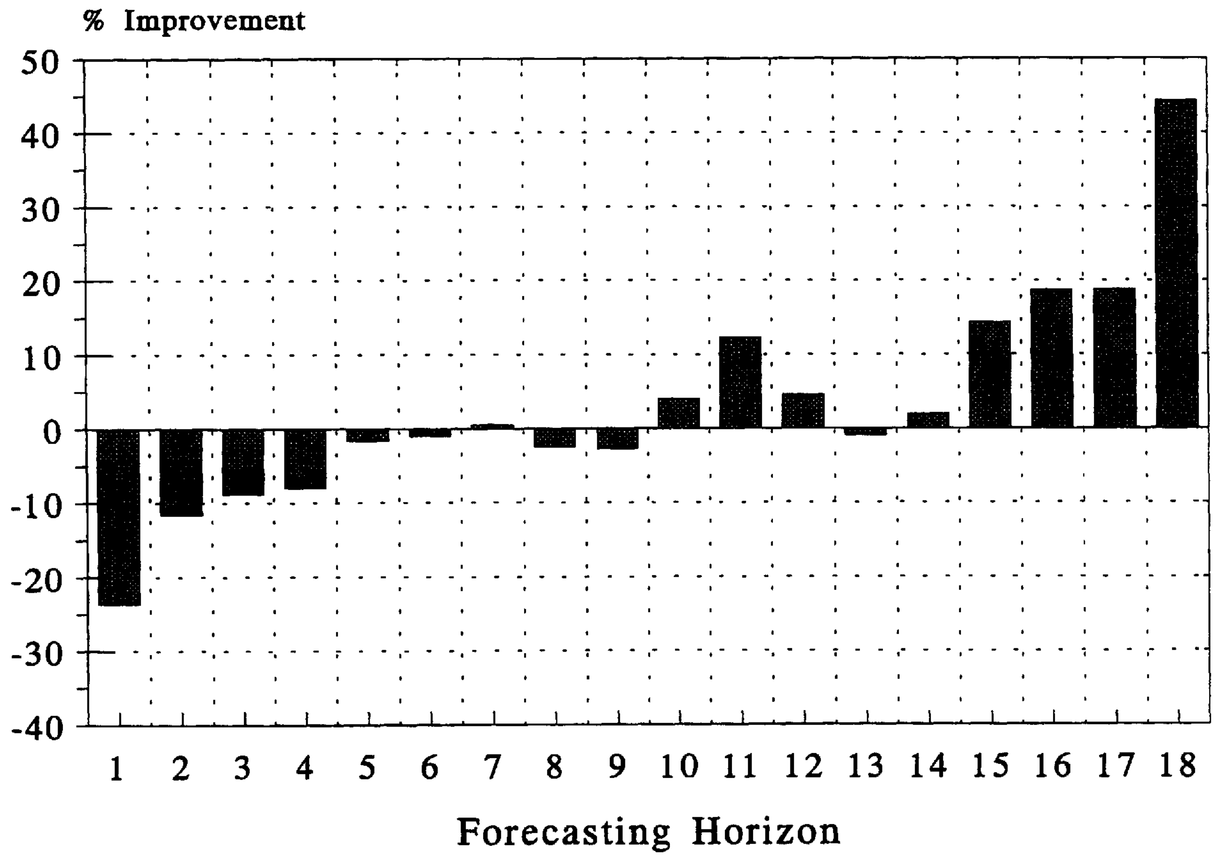
FIGURE 6

MEAN ABSOLUTE PERCENTAGE ERROR (MAPE): ACHIEVING STATIONARITY: LINEAR TREND vs DIFFERENCING



IMPROVEMENT IN THE MAPE: LINEAR TREND vs DIFFERENCING

FIGURE 7



where a and b are sample estimates of the linear regression coefficients α and β in

$$T_t = \alpha + \beta t + u_t$$

where u_t is an independent, normally distributed error term with zero mean and constant variance.

Alternatively, other types of trends can be assumed, or various pre-filters can be applied for removing the trend.

Whatever the approach being followed, T_t can be computed and subsequently used to achieve stationarity assuming an additive

$$x_t = X_t - \hat{T}_t \tag{5}$$

or multiplicative trend

$$x_t = \frac{X_t}{\hat{T}_t} \tag{6}$$

Figure 6 shows the forecasts of the data made stationary through differencing (the approach suggested by Box-Jenkins) and that through linear de-trending using expression (4). Figure 6 shows that the linear trend is slightly better, in terms of post-sample forecasting accuracy, than the method of the first differences. The results of the linear trend improve for longer forecasting horizons (see Figure 7) although overall the differences between the approach of differencing and that of the linear trend are small and for most horizons non-statistically significant. This finding suggests that the two approaches produce equivalent results with an improvement of differencing for short forecasting horizons and the opposite holding true for long ones (see Figure 7). This makes sense as differencing better captures short-term trends while linear regression long-term ones.

Dampening the Trend

In reality few trends increase or decrease consistently making differencing and linear extrapolation not the most accurate ways of predicting their continuation. For this reason the forecasting literature recommends dampening the extrapolation of trends as a function of their randomness (Gardner and McKenzie, 1985). In this study this dampening is achieved in the following four ways:

1. Damped Exponential Trend:

$$T'_{t+\ell} = S_t + \sum_{i=1}^{\ell} \phi^i T'_t$$

$\ell = 1, 2, 3, \ldots, m$

Where $S_t = \alpha X_t + (1-\alpha)(S_{t-1} + T'_{t-1})\phi$

and $T'_t = \beta(S_t - S_{t-1}) + (1-\beta)T'_{t-1}\phi$

and where α, β and ϕ are smoothing parameters found by minimizing the sum of the square errors between the actual values and those predicted by the model forecasts.

2. Horizontal Extrapolation of the Trend through Single Exponential Smoothing:

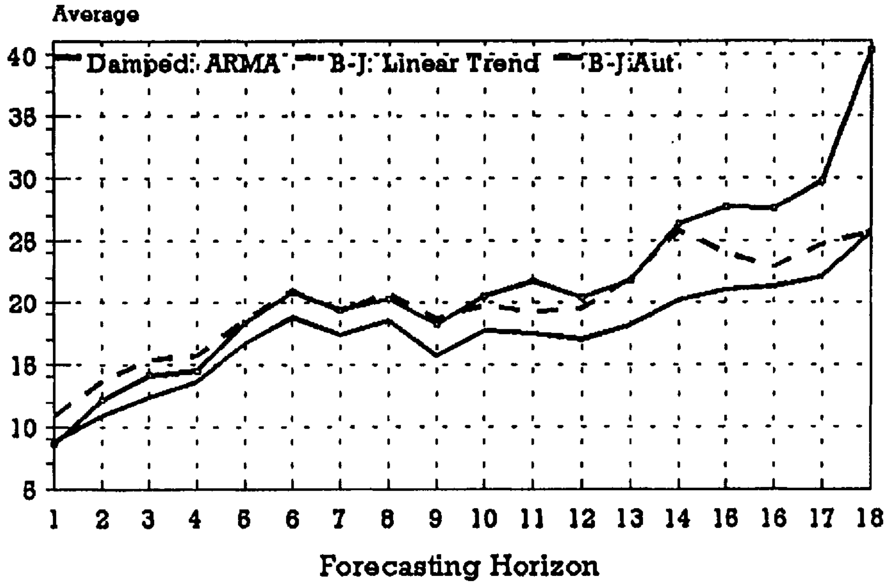
$$T'_{t+\ell} = \alpha X_t + (1-\alpha)T'_t$$

$\ell = 1, 2, 3, \ldots, m$

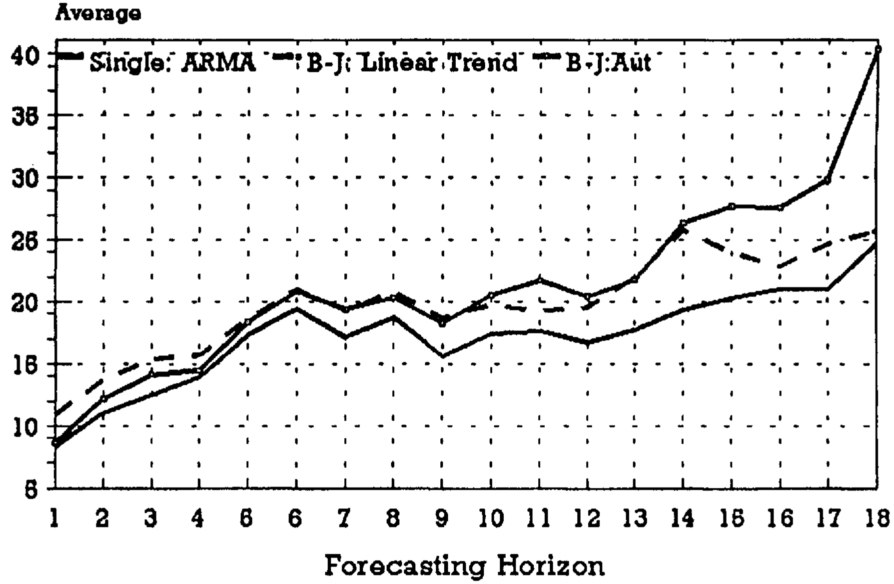
where α is a smoothing parameter found by minimizing the sum of square errors.

FIGURE 8

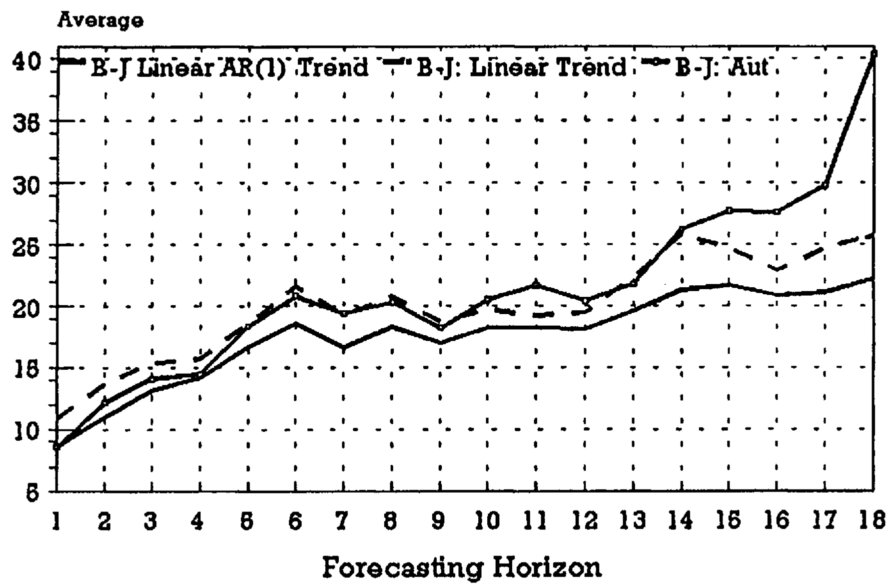
MAPE: ARMA WITH TREND OF DAMPED SMOOTHING, B-J AUTOMATIC, AND DAMPED SMOOTHING



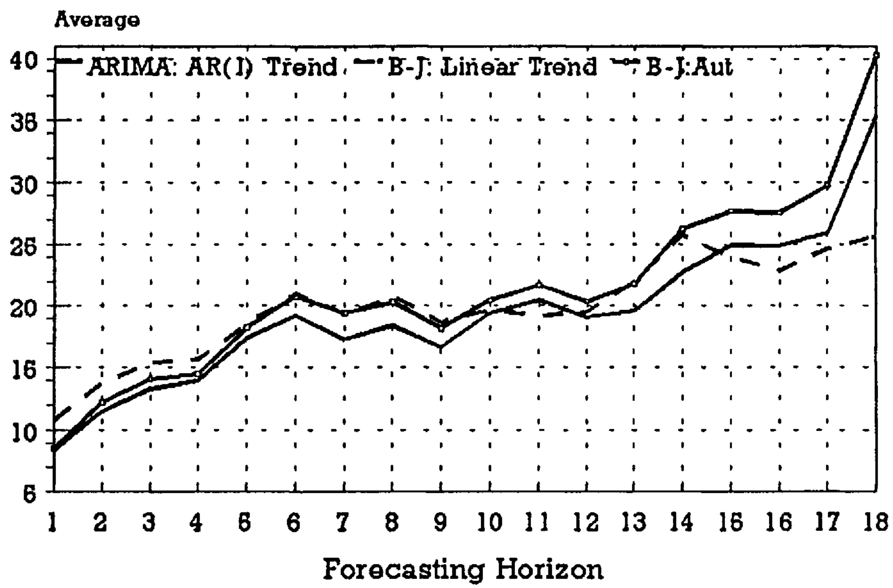
MAPE: ARMA WITH TREND OF SINGLE SMOOTHING, B-J WITH LINEAR TREND AND B-J AUTOMATIC (DIFFERENCING)



MAPE: ARMA WITH LINEAR AR(1) TREND, B-J AUTOMATIC, AND ARMA LINEAR TREND



MAPE: ARMA WITH TREND OF AR(1), ARMA WITH LINEAR TREND AND B-J AUTOMATIC (DIFFERENCING)



3. **The AR(1) Extrapolation of the Linear Trend:** Instead of extrapolations the linear trend of expression (4) as

$$\hat{T}_{t+\ell} = a + b(n + \ell) \tag{7}$$

where $\ell = 1, 2, 3, \ldots, m$

We can instead use

$$\hat{T}'_{t+\ell} = a + b(n + \ell)\phi^{\ell} \tag{8}$$

where ϕ is the AR(1) parameter calculated from the available data. As the value of ϕ is smaller than 1, the trend in (8) is dampened depending upon the value of the autoregressive coefficient ϕ.

4. **The AR(1) Extrapolation of the ARIMA Trend:**

The trend of the ARIMA forecasts can be damped by multiplying it by ϕ^{ℓ} where ϕ is the AR(1) parameter calculated from the available data. In such a case the trend becomes:

$$\hat{T}'_{t+\ell} = \phi^{\ell}\hat{T}_{t+\ell-1} \tag{9}$$

Figure 8 shows the results of the four ways of dampening the trend versus the method of differencing advocated by Box and Jenkins and the linear trend suggested by Pierce (1977). All four ways of damped trend outperform the method of differencing and that of the linear trend consistently and sometimes substantially. This finding leaves little doubt that the question of achieving stationarity in the mean is crucial and that neither the method of differencing nor that of linear trend are the most accurate ways for doing so, at least as far as economic and business series are concerned. Instead more effective ways of extrapolating the trend must be found either through dampening it, or through other alternatives using prefilters.

FIGURE 9
MAPE: B-J AUTOMATIC WITH AR(1), AR(2)

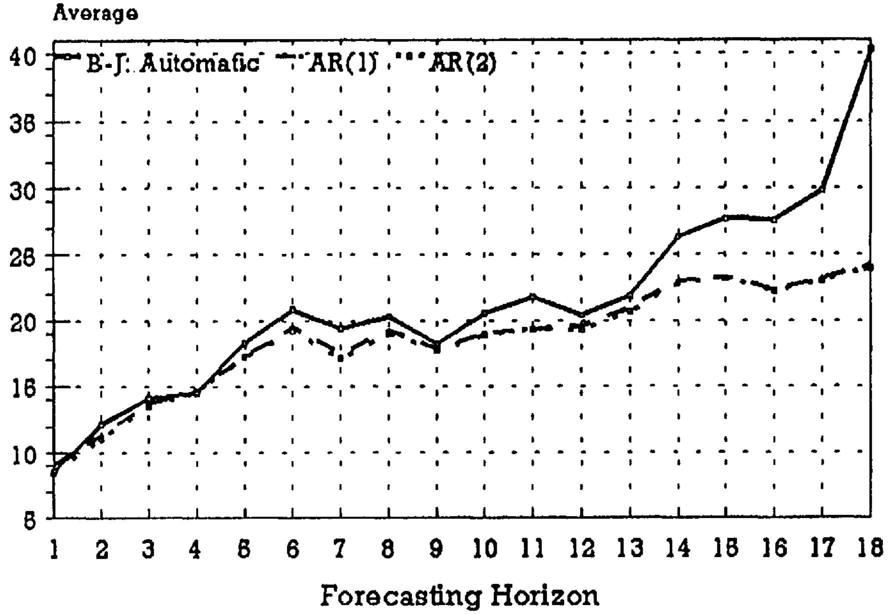


FIGURE 10
MAPE: B-J AUTOMATIC, COMBINATION OF AR(1) AND AR(2), AND MODEL WITH MINIMUM SSE

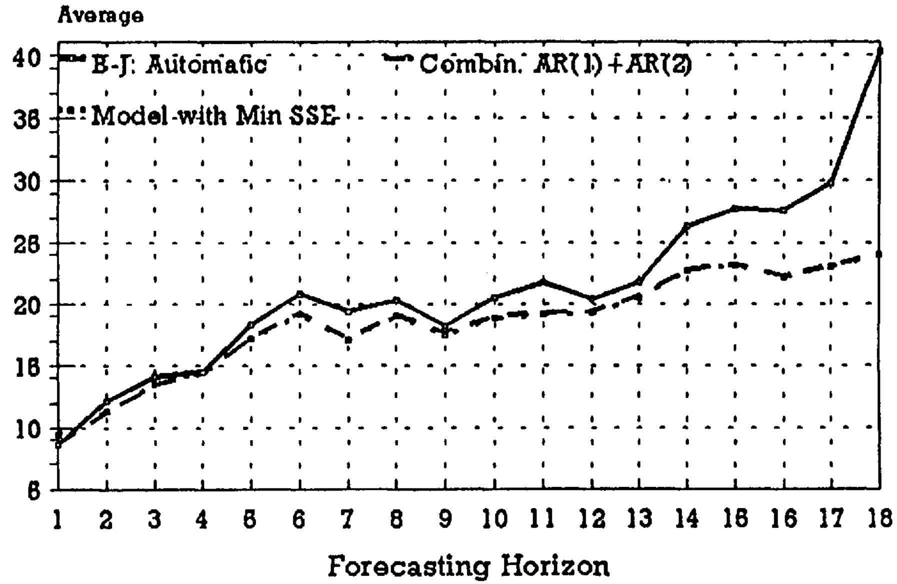


FIGURE 11
MAPE: B-J AUTOMATIC WITH MA(1) AND MA(2)

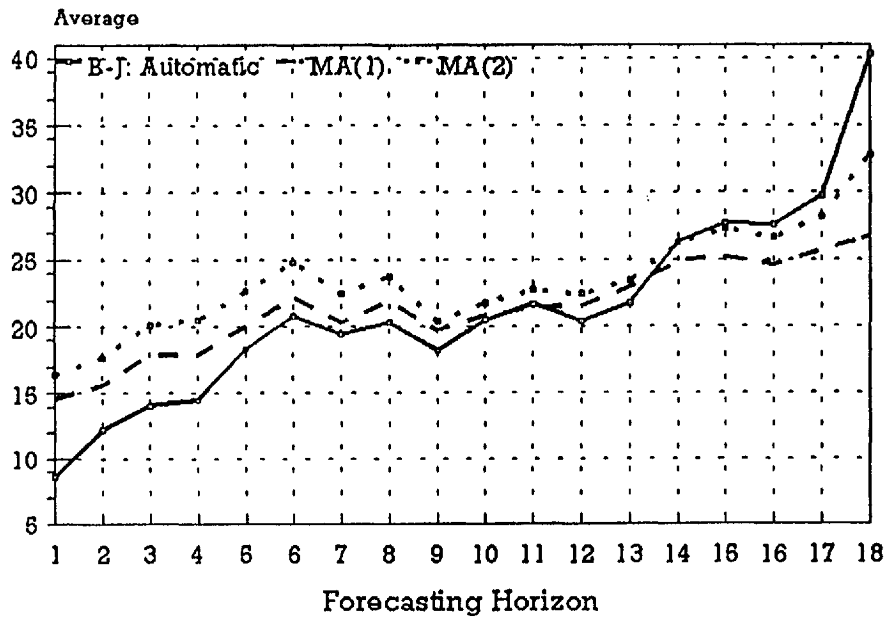
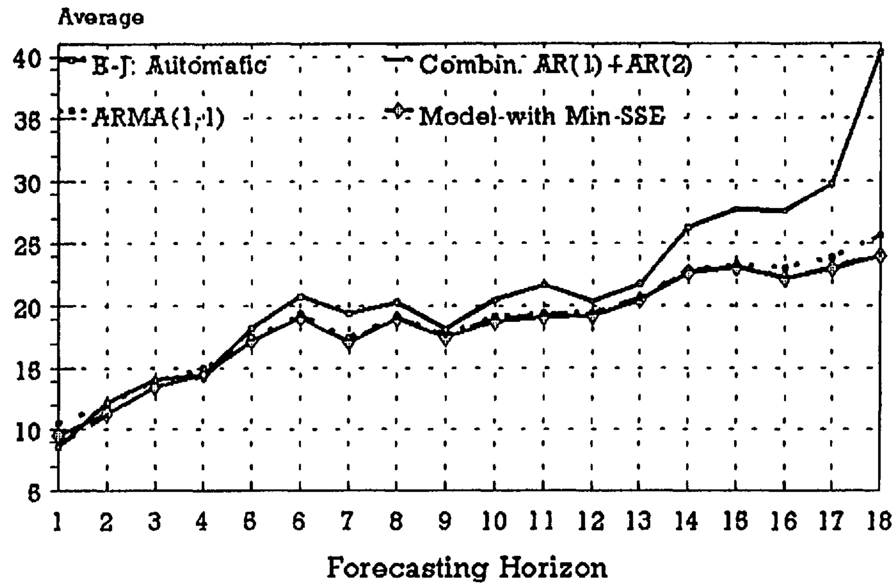


FIGURE 12
MAPE: B-J AUTOMATIC, COMBINATION OF AR(1) AND AR(2), ARMA(1,1) AND MODEL WITH MINIMUM SSE



This finding suggests that the key to more accurate post-sample predictions is the "I" of the Box-Jenkins methodology to ARIMA models. As statistical theory requires stationarity for applying ARMA models it cannot be blamed for the poor performance, accuracy way, of such models which is due to the way the trend in the data is extrapolated. Once the trend in ARMA models has been extrapolated the same way as that of the most accurate of time series methods, then their post-sample accuracy is superior to those methods, although by a small amount.

Using a Restricted Class of ARMA Models

By deseasonalizing the data first we can restrict the class of models being used to five major ones: AR(1), AR(2), MA(1), MA(2) and ARMA (1,1). An alternative to Box-Jenkins methodology is to run all five of them and select the one which minimizes the Sum of Square Errors (SSE) for each series. Figure 9 shows the post-sample forecasting accuracies for AR models and suggests that AR(1) and AR(2) are more accurate than those selected through the Box-Jenkins methodology. Figure 10 shows the combination of AR(1) and AR(2), and the model among the five (AR(1), AR(2), MA(1), MA(2) and ARMA (1,1)) that minimizes the sum of square errors in model fit. Figure 11 shows the post-sample accuracy of the MA(1) and MA(2). It suggests that the two MA models are worse than those selected by the Box-Jenkins methodology for shorter horizons and more accurate for long ones. Finally, Figure 12 shows the post-sample accuracy of ARMA (1,1),the combination of AR(1) and AR(2) and the model that minimizes the SSE (taken from Figure 10). These three types of models produce very similar results whose post-sample forecasting accuracy is superior to that of the model selected through the automatic version using the Box-Jenkins methodology. This means that the insistence of the Box-Jenkins methodology of achieving random residuals before a model is considered appropriate is not necessarily the only alternative to achieving the most accurate post-sample forecasts through ARMA models. AR(1) and AR(2) models applied to seasonally adjusted data, their combination, and ARMA (1,1) models provide at least as accurate post-sample results as those achieved through the Box-Jenkins methodology. The extra advantage of such models is that they are much easier to apply as they require less effort and computer time. It may be worthwhile, therefore, to study the theoretical properties of AR models to determine why their post-sample accuracies match those of the wider class of ARMA ones. It may also be interesting to determine why the post-sample accuracy of strictly MA models is less accurate than those of AR at least for short and medium-term forecasting horizons.

Conclusions

This paper has studied the various aspects of the Box-Jenkins methodology to ARMA models. The major conclusion has been that the way that the data is made stationary in its mean is the most important factor determining post-sample forecasting accuracies. Most importantly, when the trend in the data is identified and extrapolated using the same procedure as other methods that have been found to be more accurate in empirical studies then ARMA models perform consistently better than these methods, although the differences are small and non-statistically significant. In addition it was concluded that using seasonally adjusted data improves post-sample accuracies in a small but consistent manner, and that log and power transformations also contributed to small improvements in post-sample accuracies which become more pronounced for long forecasting horizons. Finally, it was concluded that AR(1) or AR(2) models, or their combination, produced as accurate post-sample predictions as those found by applying the automatic version of the Box-Jenkins methodology suggesting that it is neither necessary, as far as post-sample accuracy is concerned, to study the autocorrelations and partial autocorrelations to determine the most appropriate ARMA model nor to make sure that the residuals of such a model are necessarily random.

REFERENCES

Andersen, A., and Weiss, A., (1984) "Forecasting: The Box-Jenkins Approach" in *The Forecasting Accuracy of Major Time Series Methods*, Makridakis et. al, John Wiley & Sons, Chichester, England.

Armstrong, J.S., (1978) "Forecasting with Econometric Methods: Folklore versus Fact with Discussion", *Journal of Business,*, **51**, 549-600.

Box, G.E.P., and Jenkins, G., (1970) *Time Series Analysis, Forecasting and Control*, Holden-Day, San Francisco.

Box, G.E.P., and Pierce, D.A., (1970) "Distribution of the Residual Autocorrelations in Autoregressive-Integrated Moving-Average Time Series Models", *Journal of the American Statistical Association*, **65**, 1509-1526.

Chatfield, C. and Prothero, D.L., (1973) "Box-Jenkins Seasonal Forecasting: Problems in a Case Study", *Journal of the Royal Statistical Society, A,* **136**, 295-336.

Cogger, K.O., (1974) "The Optimality of General-Order Exponential Smoothing", *Operations Research,* **22**, 858-867.

Collopy, F. and Armstrong, J.S., (1992) "Rule-Based Forecasting", *Management Science,* **38**, 1394-1414.

Cooper, R.L., (1972) "The Predictive Performance of Quarterly Econometric Models of the United States". In *Econometric Models of Cyclical Behavior* (B.G. Hickman, ed.) New York: National Bureau of Economic Research.

Durbin, J., (1979) "Comment on: Makridakis, S. and Hibon, M., "Accuracy of Forecasting: An Empirical Investigation", *Journal of the Royal Statistical Society (A)*, **142**, 133-134.

Elliot, J.W., (1973) "A Direct Comparison of Short-Run GNP Forecasting Models", *Journal of Business*, **46**, 33-60.

Fildes, R., Hibon, M., Makridakis, S., and Meade, N., (1995) "The Accuracy of Extrapolative Forecasting Methods: Additional Empirical Evidence", INSEAD Working Paper.

Gardner, E.S. Jr., and McKenzie, E., (1985) "Forecasting Trends in Time Series", *Management Science,* **31**, 1237-1246.

Geurts, M.D. and Ibrahim, I.B., (1975) "Comparing the Box-Jenkins Approach with the Exponentially Smoothed Forecasting Model Application to Hawaii Tourists", *Journal of Marketing Research,* **12**, 182-188.

Geurts, M.D. and Kelly, J.P., (1986) "Forecasting Retail Sales Using Alternative Models", *International Journal of Forecasting,* **2**, 261-272.

Granger, C.W.J., and Nelson, H., (1978) "Experience Using the Box-Cox Transformation when Forecasting Economic Time Series", Working Paper 78 (1), Department of Economics, U.C.S.D.

Groff, G.K., (1973) "Empirical Comparison of Models for Short-Range Forecasting", *Management Science*, **20**(1), 22-31.

Hill, G. and Fildes, R., (1984) "The Accuracy of Extrapolation Methods: An Automatic Box-Jenkins Package SIFT", *Journal of Forecasting*, **3**, 319-323.

Huss, W.R., (1985) "Comparative Analysis of Company Forecasts and Advanced Time Series Techniques in the Electric Utility Industry", *International Journal of Forecasting,* **1**, 217-239.

Jenkins, G.M., (1982) "Some Practical Aspects of Forecasting in Organizations", *Journal of Forecasting*, **1**, 3-21.

Koehler, A.B. and Murphree, E.S., (1988) "A Comparison of Results from State Space Forecasting with Forecasts from the Makridakis Competition", *International Journal of Forecasting*, **4**, 45-55.

Libert, G., (1983) "The M-Competition with a Fully Automatic Box-Jenkins Procedure", *Journal of Forecasting*, **2**, 325-328.

Makridakis, S., and Hibon, M., (1979) "Accuracy of Forecasting: An Empirical Investigation", (with discussion), *Journal of the Royal Statistical Society*, Series A, Vol. 142, Part 2, pp. 79-145 (lead article).

Makridakis, S., et al., (1982) "The Accuracy of Extrapolative (Time Series Methods): Results of a Forecasting Competition", *Journal of Forecasting*, Vol. 1, No. 2, pp. 111-153 (lead article).

Makridakis, S., Wheelwright, S., and McGee, V., (1983), *Forecasting Methods and Applications* (2nd Edition), Wiley and Sons, New York.

Makridakis, S. et al., (1993) "The M2-Competition: "A Real-Time Judgmentally-Based Forecasting Study", *International Journal of Forecasting*, Vol. 9, No. 1, pp. 5-23.

Marquardt, D.W., (1963) "An Algorithm for Least Squares Estimation of Non-Linear Parameters", *Soc. Indust. Appl. Math.,* **11**, 431-441.

McWhorter, A. Jr., (1975) "Time Series Forecasting Using the Kalman Filter: An Empirical Study", *Proc. Amer. Statist. Ass.,* Business and Economics Section, pp. 436-446.

Meese, R. and Geweke, J., (1984) "A Comparison of AutoRegressive Univariate Forecasting Procedures for Macroeconomic Time Series", *Journal of Business and Economic Statistics*, **2**, 191-200.

Narasimham et al., (1974) "On the Predictive Performance of the BEA Quarterly Econometric Model and a Box-Jenkins Type ARIMA Model", *Proceedings of the American Statistical Association: Business and Economics Section*, pp. 501-504.

Nelson, C.R., (1972) "The Prediction Performance of the FRB-MIT-PENN Model of the US Economy", *American Economic Review*, **5**, 902-917.

Nelson, C.R., and Plosser, C.I., (1982) "Trends and Random Walks in Macroeconomic Time Series: Some Evidence and Implications", *Journal of Monetary Economics*, **10**, pp. 139-162.

Newbold, P., (1983) "The Competition to End all Competitions" in Armstrong, J. Scott and Lusk, Edward J., (eds), Commentary on the Makridakis Time Series Competition (M-Competition), *Journal of Forecasting*, **2**, 276-279.

Newton, H.J., and Parzen, E., (1984) "Forecasting and Time Series Model Types of 111 Economic Time Series" in Makridakis, S. et. al., *The Forecasting Accuracy of Major Time Series Methods*, Wiley, Chichester, England.

Parzen, E., (1982) "ARARMA Models for Time Series Analysis and Forecasting", *Journal of Forecasting*, **1**, pp. 67-82.

Pierce, D.A., (1977) "Relationships - and the lack thereof - between economic time series, with special reference to money and interest rates", *Journal of the American Statistical Association*, **72**, pp. 11-26.

Schnaars, S.P., (1986) "A Comparison of Extrapolation Models on Yearly Sales Forecasts", *International Journal of Forecasting*, **2**, 71-85.

Slutsky, E., (1937) "The Summation of Random Causes as the Source of Cyclic Processes", *Econometrica*, **5**, 105-146.

Stellwagen, E.A., and Goodrich, R.L., (1991) *Forecast Pro-Batch Edition*, Business Forecast Systems, Inc.

Texter, P.A. and Ord, J.K., (1989) "Forecasting Using Automatic Identification Procedures: A comparative Analysis", *International Journal of Forecasting*, **5**, 209-215.

Watson, M.W., Pastuszek, L.M. and Cody, E., (1987) "Forecasting Commercial Electricity Sales", *Journal of Forecasting*, **6**, 117-136.

Wold, H., (1938) "A Study in the Analysis of Stationary Time Series", Almgrist & Wiksell, Stockholm.

Yule, G.U., (1926) "Why Do We Sometimes Get Nonsense-Correlations between Time Series? A Study in Sampling and the Nature of Time Series", *Journal of Royal Statistical Society*, **89**, 1-64.