ERNESTINI GIZIAKI

TESTS FOR DETECTING OUTLIERS IN TIME SERIES

PLAN

ABSTRACT

The problem of testing a set of data for outliers is not new in statistics. Methods have been proposed for the general linear model. These methods are not appropriate for detecting outliers in time series data. However, it seems to be particularly important to be able to detect outliers in time series, especially if these data are to be used for forecasting purposes.

In this paper certain tests for the presence of outliers are described. These tests are then applied, using simulation methods, to several ARIMA models, and results are compared. An illustration of the procedure is given with published U. K. economic data.

1. INTRODUCTION

In recent years much attention has been given to the detection of outliers in data arising in various areas of statistics. An extensive survey of the various techniques for detecting outliers has been done by Barnett and Lewis, 1978 and D. M. Hawkins, 1980. In general these techniques are applicable to data comprising observations which are supposed to be independent and identically distributed.

Such methods are not, however, appropriate for examining outliers which may arise in time series analysis, because typical data sets encountered in practice will be strongly correlated and this means that not only the successive observations are autocorrelated, but also strong seasonal effects occur. In time series, an outlier is not necessarily an extreme value, but it can be a change or a break in the pattern of the series. However, although occasional outliers are unlikely to affect the estimation procedures, provided they are few in number and the sample size is sufficiently large, they have a disproportionate effect on the forecasts, particularly when an irregular observation or a recording error lies in a very near the end of the sample time observation (E. Giziaki, 1993). Thus the detection of outliers in time series data is important in forecasting work, such as in many applications of time series analysis in business and economics.

An examination of the problem of outliers for non-seasonal autoregressive time series processes has been made by A. J. Fox, 1972. Fox proposed two tests for detecting errors in observations in autoregressive time series, based on the principles of likelihood ratio and direct evaluation of the suspected outlier. However, Fox did not consider moving average, mixed and seasonal ARMA models. He used simulations to compare his test with a test based on a random sample procedure, which assumes that the time series observations are independently and identically normally distributed.

In this paper we extend Fox's approach to cover several other time series models. To this end estimates of the value of the suspected outlier and its variance are presented. A third criterion based on the one step ahead forecast error is included. The three tests are then compared using simulations and the results are commented.

In Section 4 a set of U. K. economic data (The Iron and Steel Production Index) was examined for the presence of outliers using the tests described in Section 2. The presence of outliers was tested in particular data points and an economic interpretation was then suggested for the anomalous behaviour of the series.

The tests considered rely on certain results obtained which are shown in the Appendix.

2. DESCRIPTION OF TESTS

2.1. General

We assume that a given time series

$$z' = (z_1, z_2, \ldots, z_n)$$

is generated by a model of the seasonal autoregressive integrated moving average (ARIMA) class, i.e.

$$\varphi(B)\, \Phi(B^s)\, \nabla^d \nabla_s^D z_t = \theta(B)\, \Theta(B^s)\, a_t \quad (1)$$

or

$$\varphi(B)\, \Phi(B^s)\, w_t = \theta(B)\, \Theta(B^s)\, a_t \quad (2)$$

This is a general or that covers seasonal and non-seasonal models. In (1) $\{a_t\}$ denotes a sequence of uncorrelated random variables, normally distributed with zero mean and common variance σ_a^2. B is the backward shift operator, such that $Bz_t = z_{t-1}$. ∇ and ∇_s denote the ordinary and seasonal difference operators, whilst φ, Φ, θ, Θ are polynomials in real coefficients of degree p, P, q, Q respectively. The sequence z is known as a multiplicative ARIMA (p, d, q) (P, D, Q)$_s$ process (Box and Jenkins, 1976 – K. Giziakis, 1979).

The observations are such that:

$$x_t \begin{cases} = w_t & \text{for} \quad t \neq r \\ = w_t + \delta & \text{for} \quad t = r \end{cases}$$

We test whether x_r for a particular value of r is an outlier, i.e. a spurious or aberrant observation (AO). Thus for t = r there is an outlier, if δ is non-zero, and our testing hypothesis is as follows:

$$H_o: \quad \delta = 0$$

against

$$H_A: \quad \delta \neq 0 \quad (3)$$

If there is an outlier, x_r is adjusted accordingly.

It is here assumed that any trend and seasonality in $\{z_t\}$ have been removed by differencing, and that the process $\{w_t\}$ is stationary. The order of the process is also assumed known.

The covariance matrix of the process $(w_t\}$ is $\sigma_a^2 M_n$, where M_n is an nxn Laurent matrix expressed in terms of the parameters of the model (Durbin, 1959).

2.2. Test I (Likelihood ratio)

This involves maximization of the likelihood under the two hypotheses stated in (3) and (as shown in the Appendix) leads to the likelihood ratio statistic:

$$\lambda^{I} = \frac{\left(x-\tilde{\delta}\right)' \tilde{M}_n^{-1}\left(x-\tilde{\delta}\right)}{x'\hat{M}_n^{-1}x} \tag{4}$$

where

$\hat{M}_n^{-1}$, $\tilde{M}_n^{-1}$ are the estimates of the inverses of M_n under the two hypotheses and

$\tilde{\delta} = \tilde{\delta}$ (0,0,0....0,1,0,0....0) is the estimate of the displacement in the r^{th} observation.

A result which proved computationally useful here is that the inverse covariance matrix of a stationary autoregressive moving average process of order (p, q) is given approximately by the covariance matrix of an autoregressive moving average of order (q,p). (Shaman, 1975 – Shaman, 1976 – Godolphin, 1980).

Differentation of the log likelihood, formed under the alternative hypothesis, with respect to δ provides an estimate of δ. These estimates and the variance of δ for certain ARIMA models are presented in the Appendix. A detailed coverage including more ARIMA models can be found in E. Giziaki, 1987.

Following standard procedures (as for example described in Kendall and Stuart, 1968, Vol. II, Ch. 24) when M_n is known, a linear transformation could be applied to $\{x_t\}$ to yield a set of uncorrelated variables, and we obtain under the H_o:

$$(\lambda^{I})^{-1} \sim 1 + \frac{1}{n-k} F_{1,n-k} \tag{5}$$

where k is the number of parameters to be estimated and n is the number of observations. Since M_n is usually unknown, Fox verified, using simulated series that the F– distribution provides a good approximation even in the case of unknown M_n (Fox, 1972).

Under the alternative hypothesis $(\lambda^{I})^{-1}$ has a non-central t-distribution (Kendall and Stuart, *op. cit.*, pp. 254-255). An approximation of the non-central t distribution is given by Scheffe, 1959 (Hays, 1981). This approximation provides the cumulative probability given the non-central distribution with parameters g and c, where g is the degrees of freedom and c the non-centrality parameter. The expression is presented in the Appendix[1].

2.3. Test II (δ test)

Fox also considered the δ-test, which is simpler than the likelihood ratio method. This is denoted by λ^{II} and it is

$$\lambda^{II} = \frac{\tilde{\delta}}{\tilde{\sigma}_{\delta}} \tag{6}$$

Assuming that δ is a linear combination of normally distributed variables, as shown in the examples given in the Appendix, this statistic has a t-distribution. Its standard error has been found (using spectral methods) for simple non-seasonal autoregressive processes (Grenander and Rosenblatt, 1966, p. 83). An alternative derivation is given in the Appendix, which can easily be extended to seasonal models too.

2.4. Test III (forecast error)

Another possible test is the one that uses the one-step ahead forecast error, i.e.

$$e(1) = z_{t+1} - \hat{z}_t(1) = a_{t+1} \tag{7}$$

where $\hat{z}_t(1)$ is the forecast made at time t for the period t+1.

If the model is correct and the true parameter values are used these forecast errors must be uncorrelated for a minimum mean square error forecast. In practice, when the model for a series must be identified and the parameters estimated, the e (1)'s will in general be autocorrelated.

If the model is adequate, it is possible to show that

$$\hat{a}_t = a_t + 0\left[\frac{1}{\sqrt{n}}\right] \tag{8}$$

[1] The approximation given by Scheffe, 1959 is based upon the normal distribution. The approximation is found by use of the expression:

$$\Pr(t'_{g,c} \le y) = \Pr\{z \le (y-c)(1 + y^2/2g)^{-1/2}\}$$

where z is a standard normal variable (Hays, 1981).

where $\hat{a}_t$ are the estimated residuals and as the series length increases, $\hat{a}_t$ become close to the white noise a_t (Box and Jenkins, 1976, p. 289).

Hence, if the sample to which an adequate model is fit is moderately large and we build our forecasts from the beginning of the series, at time t+1, where t is large, the $\hat{a}_t$'s will approach the white noise a_t, which is a random series distributed normally with mean 0 and variance σ_a^2.

The variance of the one-step ahead forecast error is an underestimate of the true variance, since it assumes that the coefficients of the forecasting model are known, where as in fact they must be estimated leading to a corresponding decrease in accuracy in the resulting forecasts. However, for moderately long series, this factor will be of relative small importance (Newbold and Granger, 1977, pp. 155, 91-93, 161).

This test criterion is:

$$\frac{\hat{a}_t}{\hat{\sigma}_a} \tag{9}$$

and it is assumed to have a t-distribution. σ_a is estimated from the residuals. In the introduction, it was pointed out that in the time series analysis context, an outlier is not always an extreme value as it happens in other areas of statistics. Therefore even a value, as extreme as two standard deviations, should be examined carefully. In the next section, where the power of the tests is considered, the effectiveness of this test is discussed.

3. COMPARISON OF THE TESTS

Simulated series were used to compare the power of the three tests. The models employed are AR(2), MA(1) and SAR (1,0,0) $(1,0,0)_{12}$.

Twenty series of 100 observations each were employed for each one of the sets of the parameter values tried. The evaluation of the power considered was for values of δ, the error in the r-th observation, ranging from $\pm \frac{1}{2}\sigma_a$ to $\pm 3\sigma_a$ at the 5% level of significance.

Test I and test II are asymptomatically equivalent. Test I and II are more powerful than test III in all the cases examined.

Test I is less powerful in the case of a MA(1) model than it is the same test in the case of an AR(2) or a seasonal autoregressive model. This loss in power may be due to the computational formula of δ for this model. That is, the weight given to each observation depends on the position of the outlier and the variance of δ is also dependent on the position of the outlying observation. The same is not true in the case of an autoregressive model. In the case of MA(1), test III approaches test I.

In seasonal autoregressive models, there is a great loss in power of test III as compared to test I, which is very powerful. Test III is less powerfull than the other two tests for all the models considered, but in many practical situations this relatively simple procedure may well prove adequate.

4. AN APPLICATION: U. K. IRON AND STEEL PRODUCTION INDEX FOR 1952-1979

A plot of the 112 quarterly observations of this series appears in figure 1. The model which appeared to fit the data best was $(0,1,0)\ (0,1,1)_4$ i.e.

$$(1-B)\ (1-B^4)\ z_t = (1-\Theta B^4)\ a_t$$

or

$$w_t = (1-\Theta B^4)\ a_t \tag{10}$$

The parameters were estimated to be $\Theta = 0.9266$ and $\sigma_a = 5.20$.

Observations at times 78, 82 and 89 were tested for the presence of outliers. The reason for selecting these observations is explained at the end of this section. Observations at times 94 and 110 were also considered since test III has values of three standard deviations and above. Observations numbered 86, 93 and 105 were also examined.

Estimates of δ and its sampling variance were first obtained from the relationships:

(i)
$$\tilde{\delta} = x_r + \sum_{k=1}^{i} \Theta^k \left[x_{r-4k} + x_{r+4k}\right] \tag{11}$$

where the x_t are as before the observed (differenced) observations and

$$i = \max\left[\frac{r-1}{4}, \frac{n-r}{4}\right]$$

(ii)
$$\text{var}\ \tilde{\delta} = (1-\Theta^2 + \Theta^{2v})\ \sigma^2 \tag{12}$$

where $v = \min\left[\frac{n-r+1}{4}, \frac{r}{4}\right]$

and n = number of observations employed in this estimation.

The hypothesis $\delta = 0$ was tested against its alternative at various levels of significance, using the three test procedures described earlier on the differenced stationary series x_t.

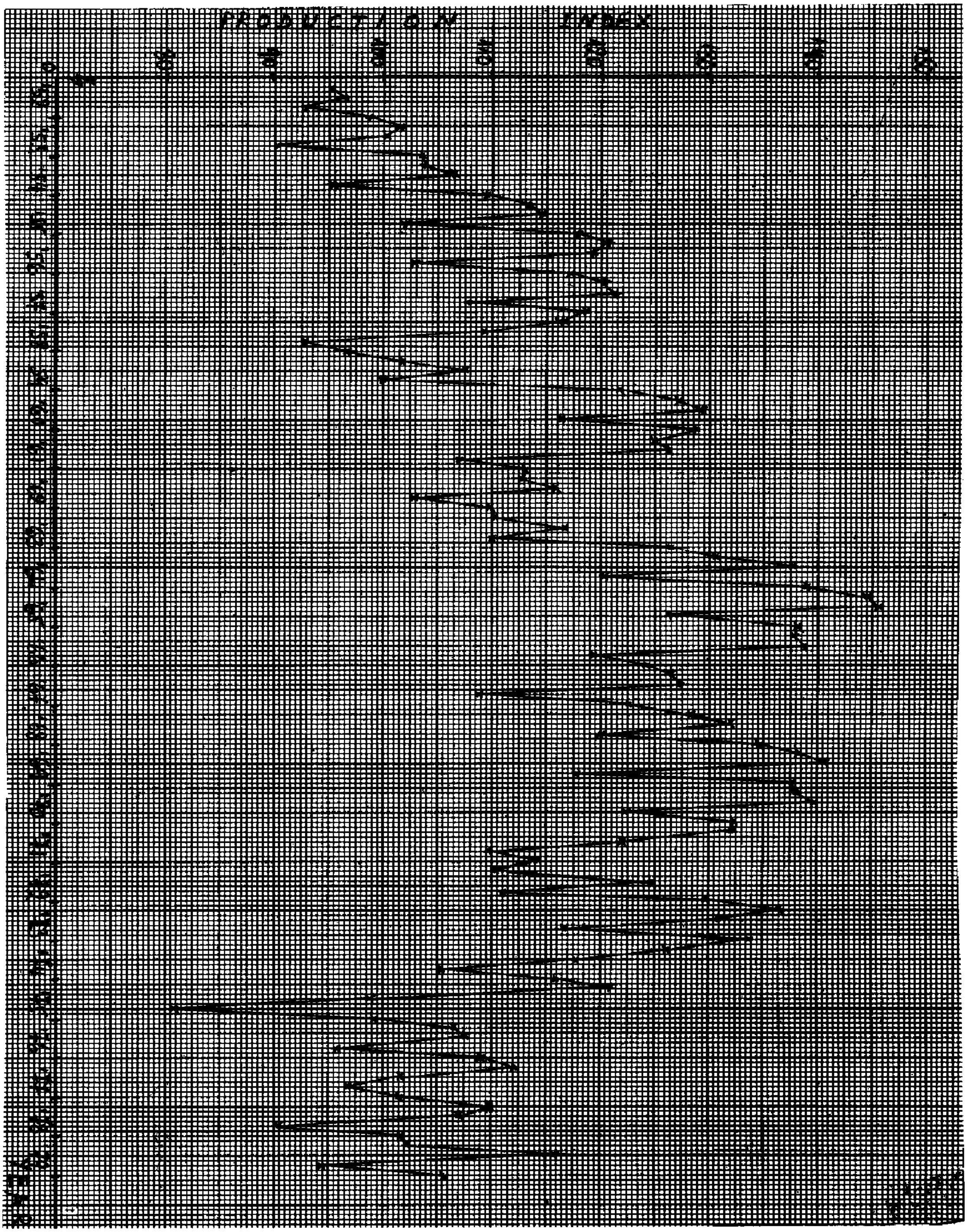

Figure 1. U. K. Iron and Steel Production Index for 1952-1979.

The test results were as under:

Observation number	*Test I*	*Test II*	*Test III*
78 (1971, II)	1.36	–6.0	–2.11
82 (1972, II)	1.37	8.0	2.63
86 (1973, II)	1.23	–2.21	–1.31
89 (1974, I)	1.92	–9.2	–1.92
93 (1975, I)	1.16	2.2	0.61
94 (1975, II)	1.68	–8.9	–4.28
110 (1979, II)	1.38	7.39	2.9

The corresponding critical values of the F and t-statistics indicated outliers at points 78, 82 and 89, except for the weaker test III which showed a marginal significance at observation 78 and no statistical significance at the observation 89. Observations 86 and 93 are of marginal significance for test II, and of no statistical significance for test III.

The prolonged coal strike in early 1972, and the miners' overtime ban in late 1973 followed by the "three-day week" and a further miners' strike in early 1974 are factors underlying the unusal observations numbered 78, 82 and 89. Tests criteria I and II have indicated presence of outliers at the particular points.

The later outliers may suggest a change in the level of the series after early 1975, when a general decline in the industry began to be apparent. However the testing criteria presented in this paper are not appropriate for testing changes in the level of a time series (E. Giziaki, 1994).

The seasonal nature of the series tends to obscure the detection of points where changes seem to be occuring, but the test procedure outlined above, in conjunction with Box-Jenkins methodology, may be useful in the location of such changes.

5. CONCLUDING REMARKS

Since it may be difficult to detect aberrant observations in a time series merely by inspecting a plot of the data, forecasting work is often hazardous. The tests for outliers examined can therefore be of practical use in detecting anomalies. The criteria developed herewith test whether a particular observations is an outlier or not. For many purposes the relative simple test III described above will prove adequate.

APPENDIX

1. Likelihood ratio test

Assuming normality for a's, the estimated likelihood under the two hypotheses given in (3) is:

$$L_o = \frac{1}{(2\pi\hat{\sigma}_a^{\,2})^{n/2}\left|\hat{M}_n\right|^{1/2}}\exp\left\{-\frac{1}{2\hat{\sigma}_a^{\,2}}x'\hat{M}_n^{-1}x\right\} \tag{13}$$

and

$$L_A = \frac{1}{(2\pi\tilde{\sigma}_a^{\,2})^{n/2}\left|\tilde{M}_n\right|^{1/2}}\exp\left\{-\frac{1}{2\tilde{\sigma}_a^{\,2}}(x-\tilde{\delta})'\tilde{M}_n^{-1}(x-\tilde{\delta})\right\} \tag{14}$$

When M_n is known, a linear transformation to $\{x_t\}$, which gives a set of uncorrelated random variables, can be used and the likelihood ratio reduces to

$$\lambda = \frac{(\tilde{\sigma}_a^{\,2})^{n/2}\left|\tilde{M}_n\right|^{1/2}}{(\hat{\sigma}_a^{\,2})^{n/2}\left|\hat{M}_n\right|^{1/2}}$$

Since δ affects only one observation, the effect on the value of the parameters will be negligible (E. Giziaki, 1987). Hence $\left|\hat{M}_n\right| \approx \left|\tilde{M}_n\right|$ and

$$\lambda^{2/n} \approx \frac{(x-\tilde{\delta})'\tilde{M}_n^{-1}(x-\tilde{\delta})}{x'\hat{M}_n^{-1}x} \tag{15}$$

For Known M_n (15) is distributed as in (5). Since M_n is usually unknown, the possibility of using the distribution based on known M_n as an approximation to the true distribution was examined. The comparison has shown that there were no notable differences in the values of the criterion based on known M_n and criterion (15) (Fox, 1972).

2. Estimation of δ and its standard error

Differentiation of the log likelihood with respect to δ in (14) provides an estimate of δ and hence its variance. In the following, we examine particular cases.

AR(1)

If $\{x_t\}$ is an AR(1) process, then M_n^{-1} in (14) will be approximately equal to the covariance matrix of an MA(1) process (Shaman, 1975). The differentiation then yields

$$(x_r - \tilde{\delta})\ (1+\varphi^2) - \varphi\ (x_{r+1} + x_{r-1}) = 0$$

from which $$\tilde{\delta} = x_r - \frac{\varphi}{1+\varphi^2}(x_{r+1} + x_{r-1})$$

which may be written as

$$\tilde{\delta} = \frac{1}{1+\varphi^2}(a_r - \varphi a_{r-1})$$

Hence $$\mathrm{var}(\tilde{\delta}) = \frac{\sigma_a^2}{1+\varphi^2}$$

In a similar fashion by employing the M_n^{-1} approximation for ARMA models (Shaman, 1975, 1976 – Godolphin, 1980) we may obtain the following results (reference E. Giziaki, 1987):

AR(2)

$$\tilde{\delta} = x_r - \frac{\varphi_1(1-\varphi_2)}{1+\varphi_1^2+\varphi_2^2}(x_{r+1}+x_{r-1}) - \frac{\varphi_2}{1+\varphi_1^2+\varphi_2^2}(x_{r+2}+x_{r-2})$$

$$\mathrm{Var}(\tilde{\delta}) = \frac{\sigma_a^2}{1+\varphi_1^2+\varphi_2^2}$$

SAR(1,0,0)(1,0,0)$_s$

$$\tilde{\delta} = x_r - \frac{\varphi(1+\Phi_S^2)}{(1+\varphi^2)(1+\Phi_S^2)}(x_{r+1}+x_{r-1}) + \frac{\varphi\Phi_S}{(1+\varphi^2)(1+\Phi_S^2)}(x_{r-(S-1)}+x_{r+S-1}) +$$
$$\frac{\varphi\Phi_S}{(1+\varphi^2)(1+\Phi_S^2)}(x_{r+S+1}+x_{r-S-1}) - \frac{\Phi_S(1+\varphi^2)}{(1+\varphi^2)(1+\Phi_S^2)}(x_{r+S}+x_{r-S})$$

$$\mathrm{Var}(\tilde{\delta}) = \frac{\sigma_a^2}{(1+\varphi^2)(1+\Phi_S^2)}$$

MA(1)

$$\tilde{\delta} = x_r + \sum_{i=1}^{k} \theta^i (x_{r+i} + x_{r-i})$$

$$\mathrm{Var}(\tilde{\delta}) \approx (1 - \theta^2 + \theta^{2v})\, \sigma_a^2 \quad \text{for } v \neq 1$$

where $v = \min\{n-r+1, r\}$

and $\mathrm{var}(\tilde{\delta}) \approx (1-\theta^2)\, \sigma_a^2 \quad \text{for } v = 1$

ARMA(1, 1)

$$\tilde{\delta} = x_r + \frac{(1-\theta\varphi)(\theta-\varphi)}{1-\varphi^2-2\varphi\theta} \sum_{i=1}^{k} \theta^{i-1} (x_{r+i} + x_{r-i})$$

where $k = \max\{(r-1), (n-r)\}$

Acknowledgement

The author would like to thank Mr B. Ethell, former Senior Lecturer of the City of London Polytechnic and Ed. Godolphin, Senior Lecturer, Royal Holloway College, University of London for their helpful comments.

REFERENCES

Barnett V. & Lewis T. (1978), *Outliers in statistical data,* John Wiley and Sons.

Box G. E. P. & Jenkins G. M. (1976), *Time Series Analysis Forecasting and Control. Revised Edition,* San Francisco, Holden Day.

Durbin J. (1959), «Efficient estimation of parameters in moving average models», *Biometrika,* Vol. 46, pp. 306-316.

Fox A. J. (1972), «Outliers in time series», *J. R. Statist. Soc.,* B, Vol. 34, pp. 350-363.

Giziaki E. (1987), *Tests for uncharacteristic changes in time series data,* Ph. D. thesis, London, U. K.

Giziaki E. (1993), «The effect of outliers on the predictive performance of univariate economic time series models», Paper presented in the *International Symposium on Economic Modelling, University of Piraeus,* Piraeus, Greece, June.

Giziaki E. (1994), «A test for detecting changes in the level of a time series», *Volume issued by the University of Piraeus to honour Prof. Stavropoulos.*

Giziakis K. (1979), *An application of the Box-Jenkins approach to management forecasting,* M. A. dissertation, University of Kent, Canterbury, Kent, U. K.

Godolphin E. J. (1980), «Estimation of Gaussian Linear Models», *Cahiers C.E.R.O.,* Vol. 22, 3-4, pp. 243-254.

Grenander U. & Rosenblatt M. (1966), *Statistical analysis of stationary time series,* Wiley and Sons, N.Y.

Hawkins D. M. (1980), *Identification of outliers,* Chapman and Hall.

Hays W. L. (1981), *Statistics for the Social Sciences,* 3rd Edition, New York, Holt, Rinehart and Winston.

Kendall M. & Stuart A. (1968, 1976), *The advanced theory of Statistics,* London, Griffin.

Newbold P. & Granger C. W. J. (1977), *Forecasting economic time series,* Academic Press.

Shaman P. (1975), «An approximate inverse for a covariance matrix of MA and AR processes», *Annals of Statistics,* Vol. 3, pp. 532-538.

Shaman P. (1976), «Approximations for stationary covariance matrices and their inverses with application to ARMA models», *Annals of Statistics,* Vol. 4, pp. 292-301.