Outliers in Time Series

By A. J. Fox†

Imperial College, London

[Received January 1970. Final revision February 1972]

SUMMARY

Two models are considered for outliers and their effects in time series. Likelihood ratio and approximate likelihood ratio criteria are derived for these models and the power functions are compared with that of the approach generally applied in the past.

Keywords: TIME SERIES; OUTLIERS; MAXIMUM LIKELIHOOD RATIO TESTS; SIMULATION OF POWER CURVES

1. Introduction

The detection of outliers has mainly been considered for single random samples, although some recent work deals also with standard linear models; see, for example, Anscombe (1960) and Kruskal (1960). Essentially similar problems arise in time series (Burman, 1965) but there seems no published work taking into account correlations between successive observations.

In the past, the search for outliers in time series has been based on the assumption that the observations are independently and identically normally distributed. This assumption leads to analyses which will be called random sample procedures.

Two types of outlier that may occur in a time series are considered in this paper. A Type I outlier corresponds to the situation in which a gross error of observation or recording error affects a single observation. A Type II outlier corresponds to the situation in which a single "innovation" is extreme. This will affect not only the particular observation but also subsequent observations. For the development of tests and the interpretation of outliers, it is necessary to distinguish among the types of outlier likely to be contained in the process. The present approach is based on four possible formulations of the problem: the outliers are all of Type I; the outliers are all of Type II; the outliers are all of the same type but whether they are of Type I or of Type II is not known; and the outliers are a mixture of the two types.

Since more practical solutions than those given by likelihood ratio methods are often obtained from simplifications of likelihood ratio criteria, some simpler criteria are derived. These criteria are of the form $\hat{\Delta}/\hat{\sigma}_{\hat{\Delta}}$, where $\hat{\Delta}$ is the estimated error in the observation tested and $\hat{\sigma}_{\hat{\Delta}}$ is the estimated standard error of $\hat{\Delta}$.

Throughout this paper, trend and seasonal components are assumed either negligible or to have been eliminated. The method adopted to remove these components might affect the results in some way.

† Present address: Medical Services Division, Department of Employment.

2. A TEST FOR TYPE I OUTLIERS

A Type I outlier is such that the outlier is the only observation affected. A model to represent this type of outlier is

$$u_t = \sum_{r=1}^{p} \alpha_r u_{t-r} + z_t \quad (t = p+1, \ldots, n), \tag{2.1}$$

where the α_r are autoregressive parameters and the $\{z_t\}$ are independently $N(0, \sigma_z^2)$. The observations, $\{y_t\}$, are such that

$$y_t = \begin{cases} u_t & (t \neq q), \\ u_q + \Delta & (t = q). \end{cases}$$

Since it is assumed that any trend in the series has been removed, the α_r are taken to be such that the process $\{u_t\}$ is stationary. The order p of the regression is assumed known.

It is possible either to test whether y_q, for a particular value of q, is an outlier or to test all values $\{y_t\}$ to see if any are outliers. The former and simpler problem will be dealt with first. It is desired to test the hypothesis, H_0: $\Delta = 0$, against the alternative, H_1: $\Delta \neq 0$.

Model (2.1) gives a stationary autoregressive process whose covariance matrix is

$$\mathbf{V} = \sigma_z^2 \mathbf{W},$$

where $\mathbf{W}$ is an $n \times n$ Laurent matrix. The elements of $\mathbf{W}$,

$$W_{ij} = w_{|i-j|}, \tag{2.2}$$

depend only on the autoregressive parameters p and α_r $(r = 1, \ldots, p)$ (see Durbin, 1959, p. 311, equation (19) and Walker, 1967, p. 49). Durbin (1959, p. 310) shows that the elements of the inverse of $\mathbf{W}$ are given asymptotically by

$$W^{i,i\pm r} = \sum_{i=0}^{p-|r|} \alpha_i \alpha_{i+|r|} = w^{(r)}. \tag{2.3}$$

Let $\hat{W}^{ij}$ be the elements of $\mathbf{W}^{-1}$, which is the inverse of $\mathbf{W}$, estimated under H_0. Then $\hat{W}^{ij}$ are obtained by substituting the maximum likelihood estimates of the autoregressive parameters under H_0 into (2.3). Similarly, if $\tilde{W}^{-1}$ is the inverse of $\mathbf{W}$ estimated under H_1, its elements, $\tilde{W}^{ij}$, are obtained by substituting the maximum likelihood estimates of the autoregressive parameters derived under H_1 into (2.3).

Maximization of the likelihoods under the two hypotheses leads to the following likelihood ratio criterion:

$$\lambda_{q,n}^{(I)} = \frac{(\mathbf{y}-\tilde{\boldsymbol{\Delta}})' \tilde{\mathbf{W}}^{-1} (\mathbf{y}-\tilde{\boldsymbol{\Delta}})}{\mathbf{y}' \hat{\mathbf{W}}^{-1} \mathbf{y}}, \tag{2.4}$$

where $\tilde{\boldsymbol{\Delta}} = \tilde{\Delta}(0, \ldots, 0, 1, 0, \ldots, 0)'$ is the estimate of the displacement in the qth observation. When there are n observations in the series, $\lambda_{q,n}^{(I)}$ is the criterion for testing that the qth observation is a Type I outlier. In the later analysis of the power curves, the test based on $\lambda_{q,n}^{(I)}$ is called Test I.

The estimate of Δ used in (2.4) is asymptotically

$$\begin{aligned} \tilde{\Delta}^* &= \sum_j \tilde{W}^{qj} y_j \Big/ \tilde{W}^{qq} \\ &= \{\tilde{w}^{(0)}\}^{-1} \{\tilde{w}^{(0)} y_q + \tilde{w}^{(1)} (y_{q-1} + y_{q+1}) + \ldots + \tilde{w}^{(p)} (y_{q-p} + y_{q+p})\}. \end{aligned} \tag{2.5}$$

In the first-order autoregressive case this estimate is

$$\tilde{\Delta}^* = y_q - \frac{\tilde{\alpha}_1}{1+\tilde{\alpha}_1^2}(y_{q-1}+y_{q+1}).$$

Exactly the same estimate of Δ would be obtained if y_q were estimated by

$$\hat{y}_q = \hat{\gamma}_1(y_{q-1}+y_{q+1}),$$

where γ_1 is the least squares estimate of the regression parameter γ_1 in the formal regression through the origin of y_q on y_{q-1} and y_{q+1}. In the pth-order case, $\tilde{\Delta}^*$ is the same estimate as would be obtained by using

$$\hat{y}_q = \sum_{r=1}^{p} \hat{\gamma}_r(y_{q-r}+y_{q+r}),$$

where $\hat{\gamma}_r$ $(r = 1, \ldots, p)$ are the least squares estimates of the parameters $\hat{\gamma}_r$.

As noted in Section 1, a solution which is more practical than the likelihood ratio solution is often obtained by considering a simpler criterion of the form

$$\lambda_{q,n} = \tilde{\Delta}/\hat{\sigma}_{\tilde{\Delta}}. \tag{2.6}$$

The variance of $\tilde{\Delta}$ can be estimated by spectral methods (Grenander and Rosenblatt, 1966, p. 83) or by substituting the usual estimates of σ_z^2 and α_r $(r = 1, \ldots, p)$ into

$$\begin{aligned} \operatorname{var}(\tilde{\Delta}) &= 2\pi\sigma_z^2 \left(\int_{-\pi}^{\pi} \left| \sum_{r=0}^{p} \alpha_r e^{irw} \right|^2 dw \right)^{-1} \\ &= \sigma_z^2/(1+\alpha_1^2+\ldots+\alpha_p^2). \end{aligned} \tag{2.7}$$

Criterion (2.6) is asymptotically equivalent to the likelihood ratio criterion (2.4).

The usual conditions for the asymptotic χ^2 distribution of the log likelihood ratio criterion are not applicable because the parameter Δ enters in only a single observation. When $\mathbf{W}$ is known, a linear transformation to the elements $\{y_t\}$, which yields a set of uncorrelated random variables, $\mathbf{U}$ say, can be applied. In terms of $\mathbf{U}$, this is the standard testing situation for the normal linear hypothesis. The denominator of $\lambda_{q,n}^{(\mathrm{I})}$ is the general sum of squares about the mean and the numerator is the residual sum of squares after fitting Δ. Hence

$$\{\lambda_{q,n}^{(\mathrm{I})}\}^{-1} = 1+(n-2)^{-1}(\text{mean square ratio for testing } \Delta = 0)$$

is distributed as

$$1+(n-2)^{-1} F_{1,n-2} \tag{2.8}$$

when $\Delta = 0$. If we assume $E(y_t) = 0$, the number of degrees of freedom for the mean square ratio is increased to $n-1$. Under the alternative hypothesis, $\{\lambda_{q,n}^{(\mathrm{I})}\}^{-1}$ has a non-central t-distribution.

Since $\mathbf{W}$ is in general unknown, the possibility of using the distribution based on known $\mathbf{W}$ as an approximation to the true distribution was investigated. The results obtained by using the following criteria were compared:

$$\text{(i)} \quad \frac{(\mathbf{y}-\tilde{\mathbf{\Delta}})'\,\hat{\mathbf{W}}^{-1}(\mathbf{y}-\tilde{\mathbf{\Delta}})}{\mathbf{y}'\,\hat{\mathbf{W}}^{-1}\,\mathbf{y}}, \tag{2.9(i)}$$

which is the exact criterion, (2.4);

$$\text{(ii)} \quad \frac{(\mathbf{y}-\tilde{\boldsymbol{\Delta}})'\,\hat{\mathbf{W}}^{-1}(\mathbf{y}-\tilde{\boldsymbol{\Delta}})}{\mathbf{y}'\,\hat{\mathbf{W}}^{-1}\,\mathbf{y}}, \tag{2.9(ii)}$$

which is the same as (2.4) but with the estimate of $\mathbf{W}^{-1}$ in the numerator replaced by the estimate made under H_0; and

$$\text{(iii)} \quad \frac{(\mathbf{y}-\tilde{\boldsymbol{\Delta}})'\,\mathbf{W}^{-1}(\mathbf{y}-\tilde{\boldsymbol{\Delta}})}{\mathbf{y}'\,\mathbf{W}^{-1}\,\mathbf{y}}, \tag{2.9(iii)}$$

which is the criterion based on known $\mathbf{W}^{-1}$.

Simulation was used to examine differences between these three criteria. The autoregression parameters and the value of Δ were varied, the series length being fixed at 100 observations. The values of each criterion were obtained for each of 25 series based on a particular set of α_r $(r = 1, \ldots, p)$ and Δ. There were no notable differences in the values of the criteria obtained for individual series. In particular, if the 5 per cent significance points of the distribution based on known $\mathbf{W}$ are used as a criterion of significance, different decisions were only obtained in 4 out of 600 runs. The investigation was repeated with only 30 observations in the series and similar results were obtained. This is good confirmation that (2.4) can be considered to be distributed as the criterion based on known $\mathbf{W}$.

A more complex situation is that in which the position of the outlier is unknown. The results obtained for testing a particular observation can be extended to cover this situation.

The criterion, developed for the case of a known outlier position, is

$$k_{q,n} = \{\lambda_{q,n}^{(\mathrm{I})}\}^{-1}, \tag{2.10}$$

which is distributed as $1+\{1/(n-2)\}\,F_{1,n-2}$. When the position of the outlier is unknown, the maximum likelihood statistic is equivalent to

$$\max_{q=p+1,\ldots,n-p} (k_{q,n}). \tag{2.11}$$

The distribution of the maximum of a set of n correlated F values is not known. Estimates of the relevant significance points can, however, be found by simulating the situation. Noting

$$(n-2)\,(k_{q,n}-1) \sim F_{1,n-2}, \tag{2.12}$$

the cumulative distribution function of

$$\max_{q=p+1,\ldots,n-p} \{(n-2)\,(k_{q,n}-1)\} \tag{2.13}$$

can be simulated for particular series lengths and, by repeating the simulation, estimates of the significance points of the test based on (2.13) can be obtained. The estimated significance points shown in Table 1 were found for models up to and including third-order models using various values of α_1, α_2 and α_3. The significance points were found, to a good approximation, to be unaffected by changes in these values. This has some theoretical justification since, if the difference between $\tilde{w}^{(r)}$ and $w^{(r)}$ is neglected, the random variables Δ^* defined by equation (2.5) form a stationary p-dependent normal process (see Watson, 1954). The observed proportions of values larger than a test value for particular series lengths can be considered as estimated binomial parameters with variance $p^*(1-p^*)/n^*$, where p^* is the cumulative probability for the particular test value and n^* is the number of simulations. Confidence limits can thus be put round the estimated proportions. From these confidence limits, approximate confidence regions are obtained for the cumulative distribution function.

TABLE 1

Significance points for criterion (2.13) *testing for an outlier at an unknown point*

n \ $\alpha\%$	10%	5%	$2\frac{1}{2}\%$	1%
50	9·7	10·8	12·3	14·8
100	11·3	12·8	14·3	15·8
200	11·9	13·2	14·7	16·3
500	12·4	13·5	15·0	16·8

An example, for $n = 200$ and 95 per cent confidence limits, is: 10 per cent point, 11·2–12·8; 5 per cent point, 12·4–14·6; 25 per cent point, 13·5–16·0 and 1 per cent point, 14·5–17·5.

The behaviour of the power curves has been investigated by simulating first-order models with $\alpha_1 = \pm 0{\cdot}1, \pm 0{\cdot}3, \pm 0{\cdot}5, \ldots, \pm 0{\cdot}9$. The following probabilities were estimated as functions of Δ, the error inserted into a particular observation;

(i) $$P\left\{\max_{q=2,\ldots,99} (k_{q,100}) > C\right\},$$

where C is the significance point that was chosen;

(ii) $$P\left\{k_{p,100} = \max_{q=2,\ldots,99} (k_{q,100}) \,\middle|\, \max_{q=2,\ldots,99} (k_{q,100}) > C\right\};$$

and

(iii) $$P\left\{k_{p,100} = \max_{q=2,\ldots,99} (k_{q,100})\right\}.$$

Since no marked dependence on the autoregressive parameter was found, the curves plotted are the means of the curves obtained for each value of α_1. Although the simulations are based on first-order models, Watson's result, referred to above, suggests that the effect of p should be small for large n.

The 5 per cent significance point, found from Table 1, is $C = 1{\cdot}129$. Fig. 1 shows the power of the test deciding whether or not the series contains an outlier, which is (i) above. Fig. 2 shows the power of the test selecting the known outlier given that the criterion is significant at the 5 per cent significance level, which is (ii) above. The probability that the value used for the test is the value known to be an outlier is shown in Fig. 3. The test based on $\lambda_{q,n}^{(\mathrm{I})}$, when compared with the random sample procedure, is shown to be more powerful for testing for an outlier in known position. The power of (2.13) is, therefore, presumably greater than that obtained by extending the random sample procedure to deal with the case of unknown outlier position.

Certain approximations have been made, namely in the use of **W**, in the derivation of the test criterion's distribution. The simulations giving the curves shown in Figs. 1, 2 and 3 were, however, based on criterion (2.9(i)). The results confirm that the approximations made do not introduce serious errors into the analysis.

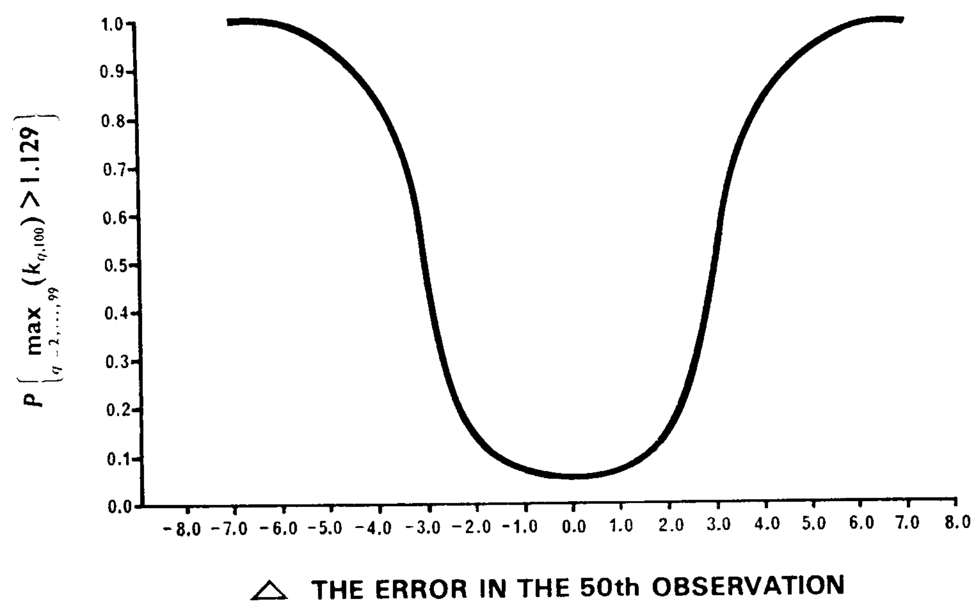

FIG. 1. The probability, as a function of Δ, of $\max_{p=2,\ldots,99}(k_{q,100})$ being significant at the 5 per cent significance level.

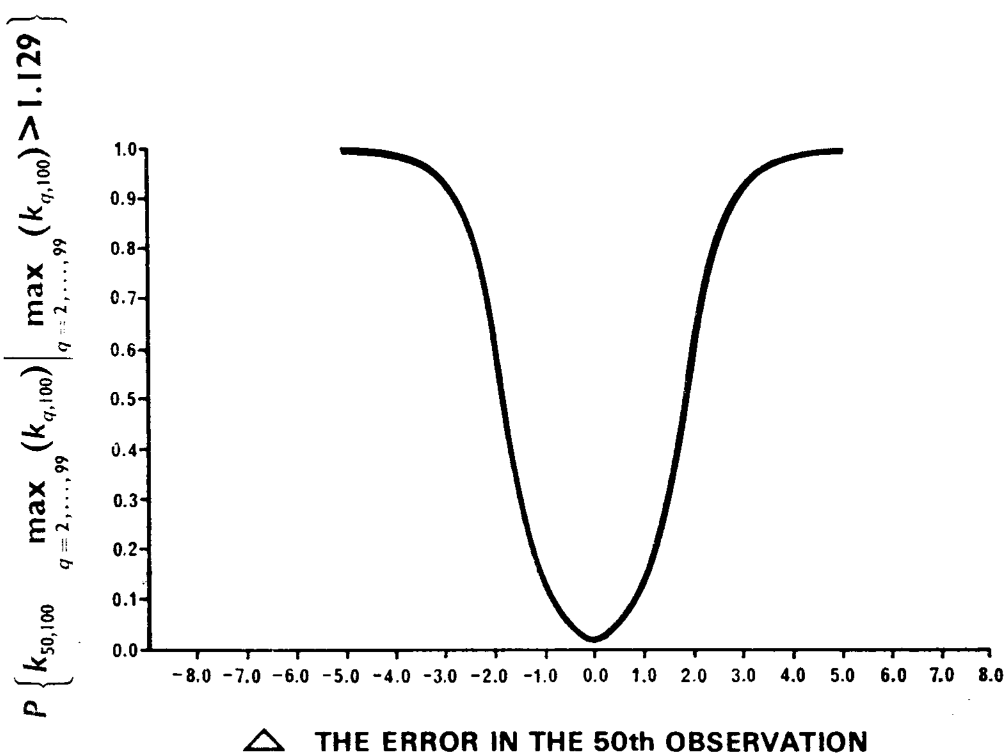

FIG. 2. The probability, as a function of Δ, of selecting the value known to be an outlier once the test has decided at the 5 per cent significance level that there is an outlier.

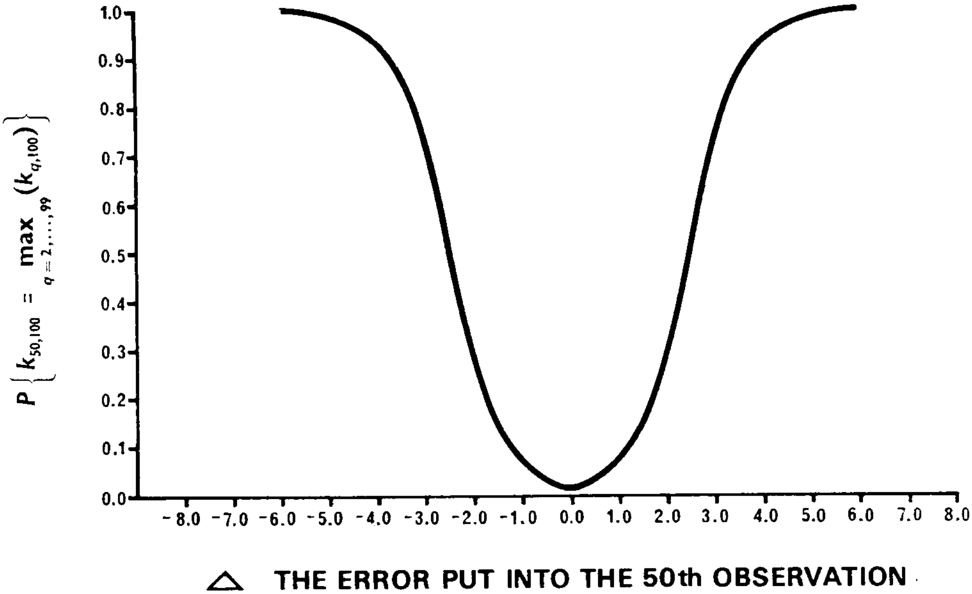

FIG. 3. The probability of using the known outlier as the basis for the test for whether or not an outlier exists is plotted as a function of Δ.

3. A TEST FOR TYPE II OUTLIERS

Suppose now that

$$y_t = \sum_{r=1}^{p} \alpha_r y_{t-r} + \Delta_t + z_t, \tag{3.1}$$

where

$$\Delta_t = \begin{cases} 0 & (t \neq q), \\ \Delta & (t = q), \end{cases}$$

and the α_r $(r = 1, \ldots, p)$ and $\{z_t\}$ are defined as for (2.1). Then the outlier Δ affects y_q and through it subsequent observations $y_{q+1}, \ldots, y_n$. Suppose first that a particular observation is to be tested, i.e. the hypothesis, H_0: $\Delta = 0$, is to be tested against the alternative, H_1: $\Delta \neq 0$, for known q. Maximization of the likelihoods under the two hypotheses leads to the likelihood ratio criterion

$$\lambda_{q,n}^{(\mathrm{II})} = \left\{ \sum_{t=p+1}^{n} \left(y_t - \sum_{r=1}^{p} \hat{\alpha}_{r1} y_{t-r} - \hat{\Delta}_t \right)^2 \Big/ \sum_{t=p+1}^{n} \left(y_t - \sum_{r=1}^{p} \hat{\alpha}_{r0} y_{t-r} \right)^2 \right\}^{\frac{1}{2}(n-p)}, \tag{3.2}$$

where $\hat{\alpha}_{r0}$ and $\hat{\alpha}_{r1}$ are the estimates of α_r $(r = 1, \ldots, p)$ under H_0 and H_1 respectively and $\hat{\Delta}_t$ is the estimate of Δ_t under H_1.

A re-parameterization of (3.1), for example, by considering

$$y_t = \sum_{r=1}^{p} \alpha_r y_{t-r} + \Delta_t^* z_t,$$

where $\Delta_t^* = 1$ $(t \neq q)$ and $\Delta_t^* \neq 1$ $(t = q)$ leads to

$$\lambda_{q,n}^* = \left(y_q - \sum_{r=1}^{p} \hat{\alpha}_{r1} y_{q-r}\right) \Big/ \left\{(n-p)^{-1} \sum_{t=p+1}^{n} \left(y_t - \sum_{r=1}^{p} \hat{\alpha}_{r0} y_{t-r}\right)^2\right\}^{\frac{1}{2}} \tag{3.3}$$

and

$$-2\log \lambda_{q,n}^{(\mathrm{II})} \sim (\lambda_{q,n}^*)^2.$$

This criterion arises naturally since, from (3.2),

$$\hat{\Delta} = y_q - \sum_{r=1}^{p} \hat{\alpha}_{r1} y_{q-r}$$

and $\operatorname{var} \hat{\Delta} \simeq \sigma_z^2$. The maximum likelihood estimate of σ_z^2 is

$$(n-p)^{-1} \sum_{t=p+1}^{n} \left(y_t - \sum_{r=1}^{p} \hat{\alpha}_{r0} y_{t-r}\right)^2.$$

Criterion (3.3) is the analogue of $\lambda_{q,n}$ in Section 2. In the analysis of the power curves, the test based on $\lambda_{q,n}^*$ is called Test II.

By ignoring the differences between $\hat{\alpha}_{r0}$, $\hat{\alpha}_{r1}$ and α_r,

$$(\lambda_{q,n}^*)^{-2} = \left\{(n-p)^{-1} \sum_{t=p+1}^{n} \left(y_t - \sum_{r=1}^{p} \hat{\alpha}_{r0} y_{t-r}\right)^2\right\} \Big/ \left(y_q - \sum_{r=1}^{p} \hat{\alpha}_{r1} y_{q-r}\right)^2$$

is found to be asymptotically distributed as

$$(n-p)^{-1}\{1 + (n-p-1) F_{n-p-1,1}\}. \tag{3.4}$$

Hence

$$\frac{(n-p-1)(\lambda_{q,n}^*)^2}{\{n-p-(\lambda_{q,n}^*)^2\}} \sim F_{1,n-p-1}. \tag{3.5}$$

Under H_1, the distribution of the criterion is a non-central t distribution.

A similar extension to that applied in Section 2, leads to a solution when the position of the outlier is unknown, but this has not been studied.

4. Simulated Power Curves for a Known Position of Outlier

The power curves, which were generated when the position of the outlier is assumed to be unknown, are discussed in Section 2. This section contains the results of a more detailed investigation into the power curves when the position of the outlier is known.

The power curves of Test I and Test II are now compared with the commonly applied random sample procedure, which is based on the assumption that the $\{y_t\}$ are independently and identically distributed as $N(\mu, \sigma^2)$, where μ and σ^2 are unknown.

Many of the ensuing remarks, made about the power of the tests, are explained by Table 2, which compares the different criteria for a first-order autoregressive model. The approximate criteria, $\lambda_{q,n}$ and $\lambda_{q,n}^*$, demonstrate how, if the outlier is of Type I, the analysis is based on a two-sided regression, whereas if the outlier is of Type II the analysis is based on a one-sided regression.

The significance level of the random sample procedure may be shown to be asymptotically unaffected by the use of the inappropriate null model, but the power does depend on the autoregressive parameters. The degrees of freedom in the random sample procedure would be reduced by a factor $2\alpha^2 n/(1+\alpha^2)$ for small sample size.

Particular models were considered in order to generate power curves. The first family considered is the first-order autogressive process given by

$$y_t = \alpha_1 y_{t-1} + z_t \quad (t = 2, \ldots, n), \tag{4.1}$$

where z_t is distributed as $N(0, 1)$. The parameter α_1 was selected to take values between $\pm 0{\cdot}9$.

TABLE 2

The test statistics used and the criteria derived when the approximate criteria are applied to the different types of error for a first-order autoregressive model with parameter α_1

		Type I outlier test based on $\lambda_{q,n}$	*Type II outlier test based on* $\lambda^*_{q,n}$	*Test based on a random sample procedure*
Test statistic		$y_q - \dfrac{\alpha_1}{1+\alpha_1^2}(y_{q-1}+y_{q+1})$	$y_q - \alpha_1 y_{q-1}$	y_q
Type I outlier shift	Mean	Δ	Δ	Δ
	Variance	$\sigma_z^2/(1+\alpha_1^2)$	σ_z^2	$\sigma_z^2/(1-\alpha_1^2)$
Non null mean/$\sqrt{}$variance		$\{\Delta\sqrt{(1+\alpha_1^2)}\}/\sigma_z$	Δ/σ_z	$\{\Delta\sqrt{(1-\alpha_1^2)}\}/\sigma_z$
Type II outlier shift	Mean	$\Delta/(1+\alpha_1^2)$	Δ	Δ
	Variance	$\sigma_z^2/(1+\alpha_1^2)$	σ_z^2	$\sigma_z^2/(1-\alpha_1^2)$
Non null mean/$\sqrt{}$variance		$\Delta/\{\sigma_z\sqrt{(1+\alpha_1^2)}\}$	Δ/σ_z	$(\Delta/\sigma_z)\sqrt{(1-\alpha_1^2)}$

The power curves, estimated when Test I is compared with the random sample procedure for Type I errors, are given in Fig. 4. These curves show dependence on the absolute value of α_1. This dependence can be explained by considering the test based on $\lambda_{q,n}$, Table 2, where the mean is Δ and the variance is $\sigma_z^2/(1+\alpha_1^2)$. As α_1 increases the variance decreases and so greater power is obtained. Incidentally, if the known value of α_1 is used to estimate the power of the test based on $\lambda^{(\mathrm{I})}_{q,n}$, curves which are nearly the same as those in Fig. 4 are obtained. Fig. 4 shows that, for small absolute value of α_1, almost exactly the same power curves are obtained for Test I as for the random sample procedure. As the absolute value of α_1 increases the power of Test I increases. The power of the random sample procedure, however, is observed to decrease because the variance that is estimated increases with $|\alpha_1|$ whereas the true variance of the process remains constant.

The investigation was extended to second-order autoregressive models with essentially the same conclusions, the improvement over the random sample procedure being maintained. The average power curve for the second-order models considered corresponds almost exactly to the average of the curves shown in Fig. 4.

The power curves, obtained by comparing the approximate likelihood-ratio criterion $\lambda^*_{q,n}$, Test II, with the random sample procedure, when both are applied to model (3.1), demonstrate that the power of Test II is not dependent on the absolute value of the autoregressive parameter. In Fig. 5, the average power curve for Test II is plotted and compared with the random sample procedure power curves for various values of α_1. Test II gives a significant improvement over the random sample procedure. The random sample procedure power curves are almost identical to those obtained when the error was superimposed after the series was generated; that is those curves shown in Fig. 4.

The effects of selecting the wrong null model, when the errors are assumed to be only of one type, are investigated because it is important to have an idea of the power lost if the wrong type of error is assumed.

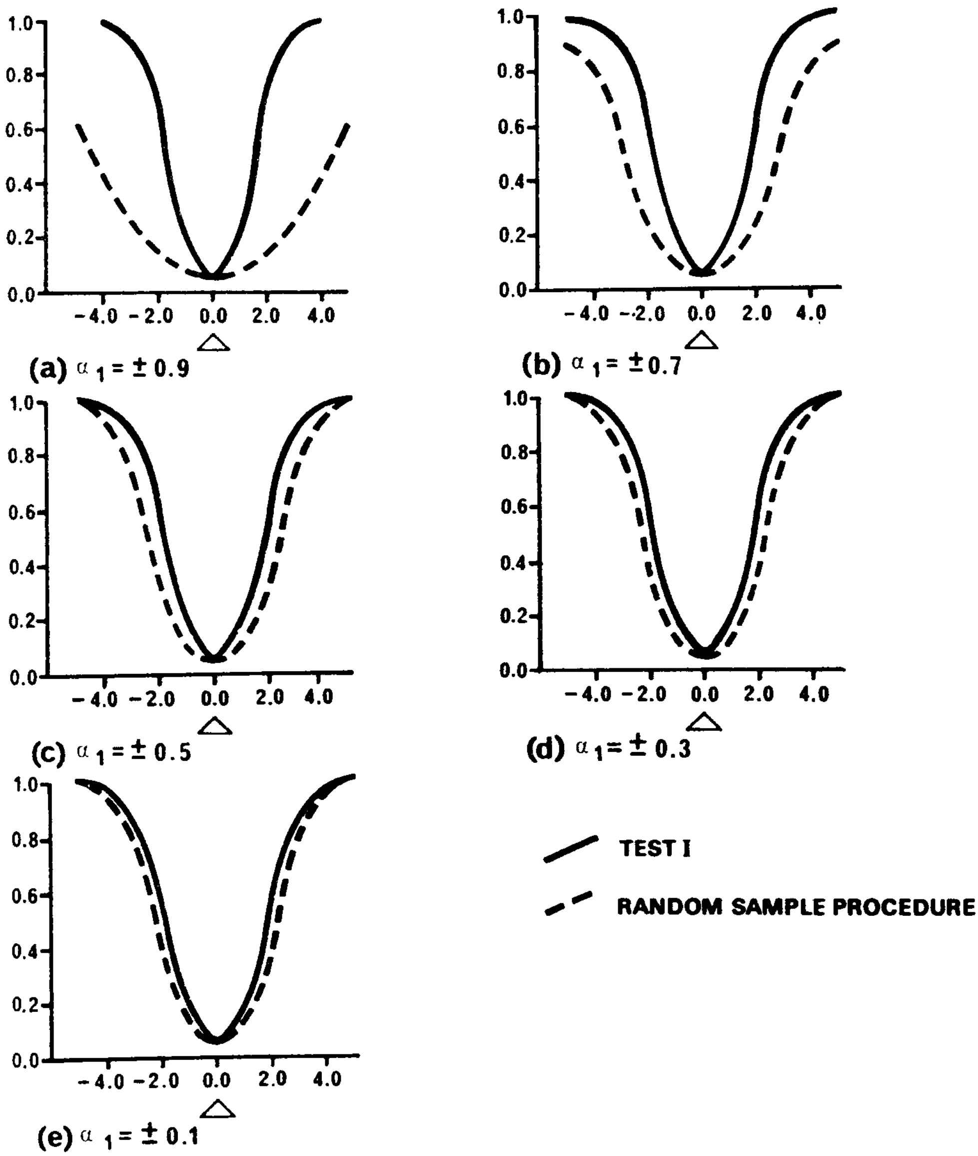

FIG. 4. The power, as a function of Δ, of Test I compared with that of the random sample procedure. The true model is (2.1) with $p = 1$.

The results of investigations along these lines show that Test II, when applied to Type I outliers, gives slightly less power than Test I. The curves are however not dependent on the absolute value of α_1. Although the wrong model is assumed the power is still higher than that for the random sample procedure. In the converse situation similar results are obtained but with Test II being the more powerful. Figs 6 and 7 show the power curves, obtained by simulation, as functions of Δ. It is interesting to note that, when the error is one which affects subsequent observations,

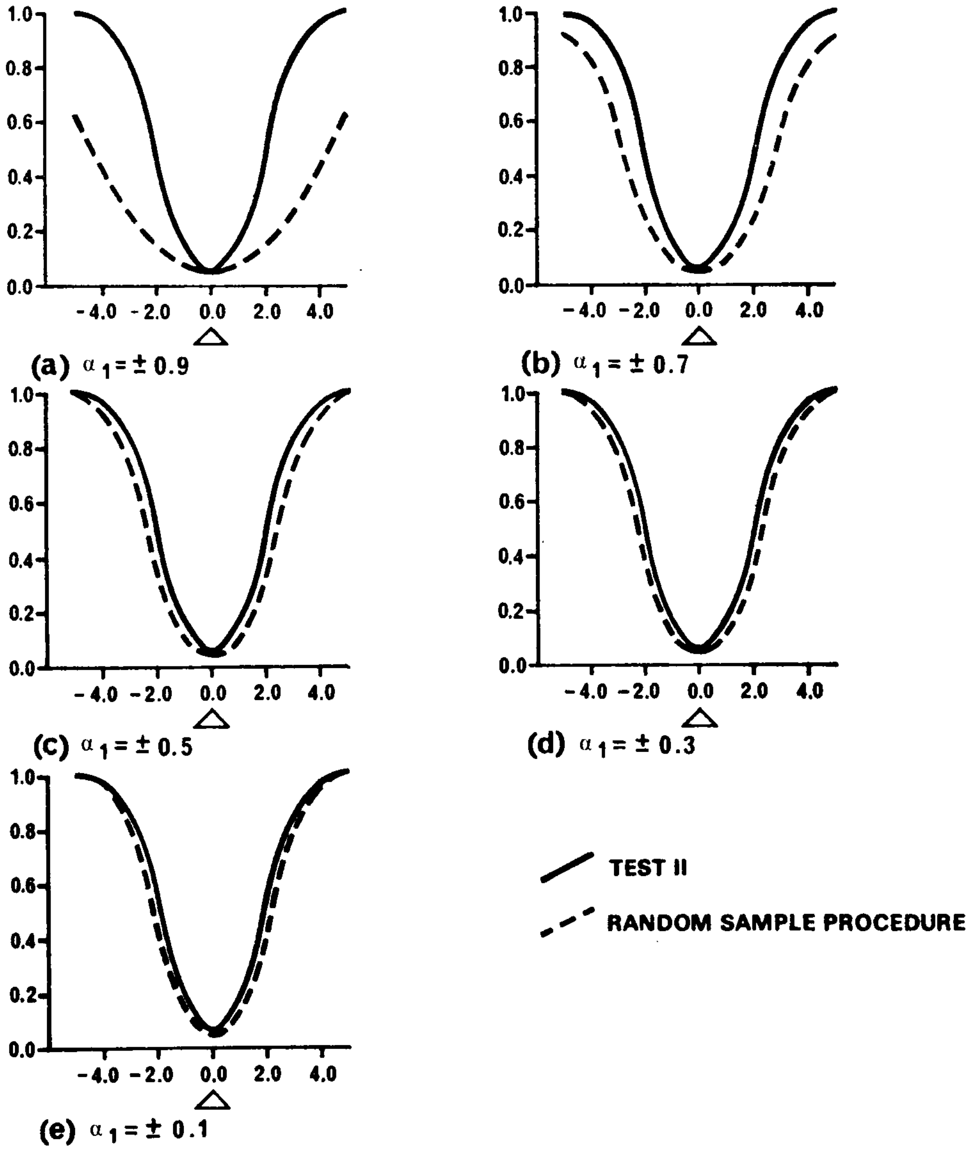

FIG. 5. The power, as a function of Δ, of Test II (averaged over all α_1) compared with that of the random sample procedure, for model (3.1) with $p = 1$.

the power of the wrong approach rises as the absolute value of α_1 decreases and it remains higher than the power of the random sample procedure until the absolute value of α_1 is less than about 0·3.

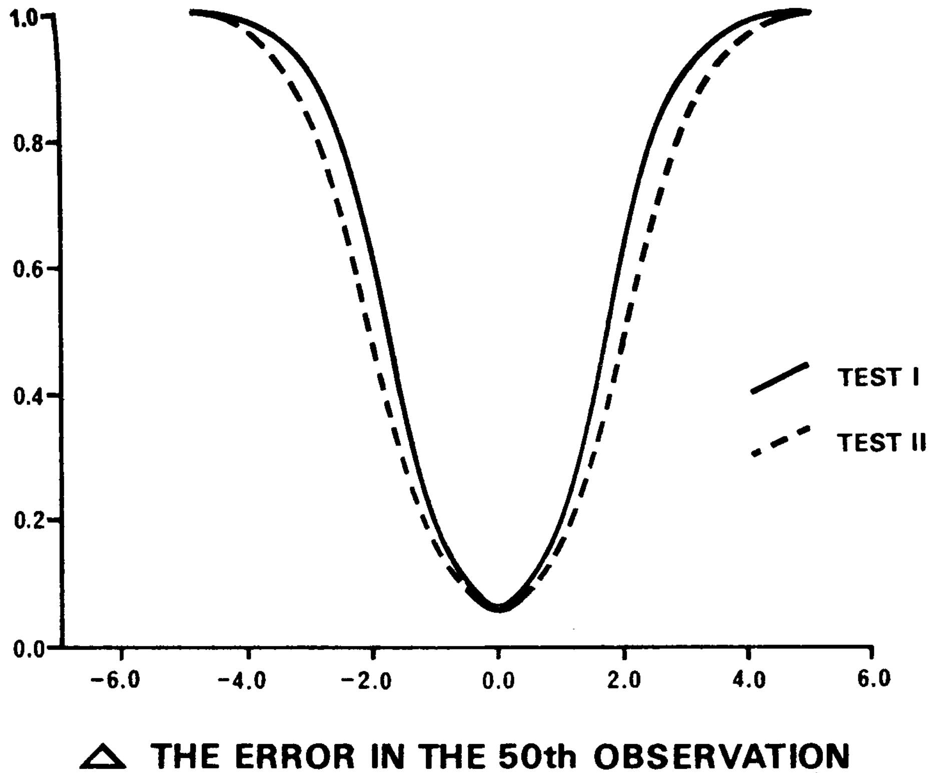

FIG. 6. The power, as a function of Δ, of Test I compared with that of Test II for model (2.1) with $p = 1$.

5. DISCUSSION OF THE MODELS

The previous sections suggest that the two models, based on categorizing an outlier as of Type I or of Type II, provide the criteria that should be used to test for outliers. The loss of power, because the wrong procedure has been applied, was investigated. Whichever model is assumed the results are shown to be superior to those obtained by using the random sample procedure for almost any value of the autoregressive parameter.

When in doubt whether the observation, that is suspected of being an outlier, is a Type I or a Type II outlier, one can compare the likelihood of

$$y_t = \sum_{r=1}^{p} \alpha_r (y_{t-r} - \Delta_1 \delta_{t-r,q}) + \Delta_1 \delta_{t,q} + z \tag{5.1}$$

with that of

$$y_t = \sum_{r=1}^{p} \alpha_r y_{t-r} + \Delta_2 \delta_{t,q} + z_t, \tag{5.2}$$

where

$$\delta_{t,q} = \begin{cases} 1 & (t=q), \\ 0 & (t \neq q), \end{cases}$$

and Δ_1 and Δ_2 are the Type I and Type II outliers, respectively. Comparing the likelihoods will be equivalent to performing both tests and selecting the model by which

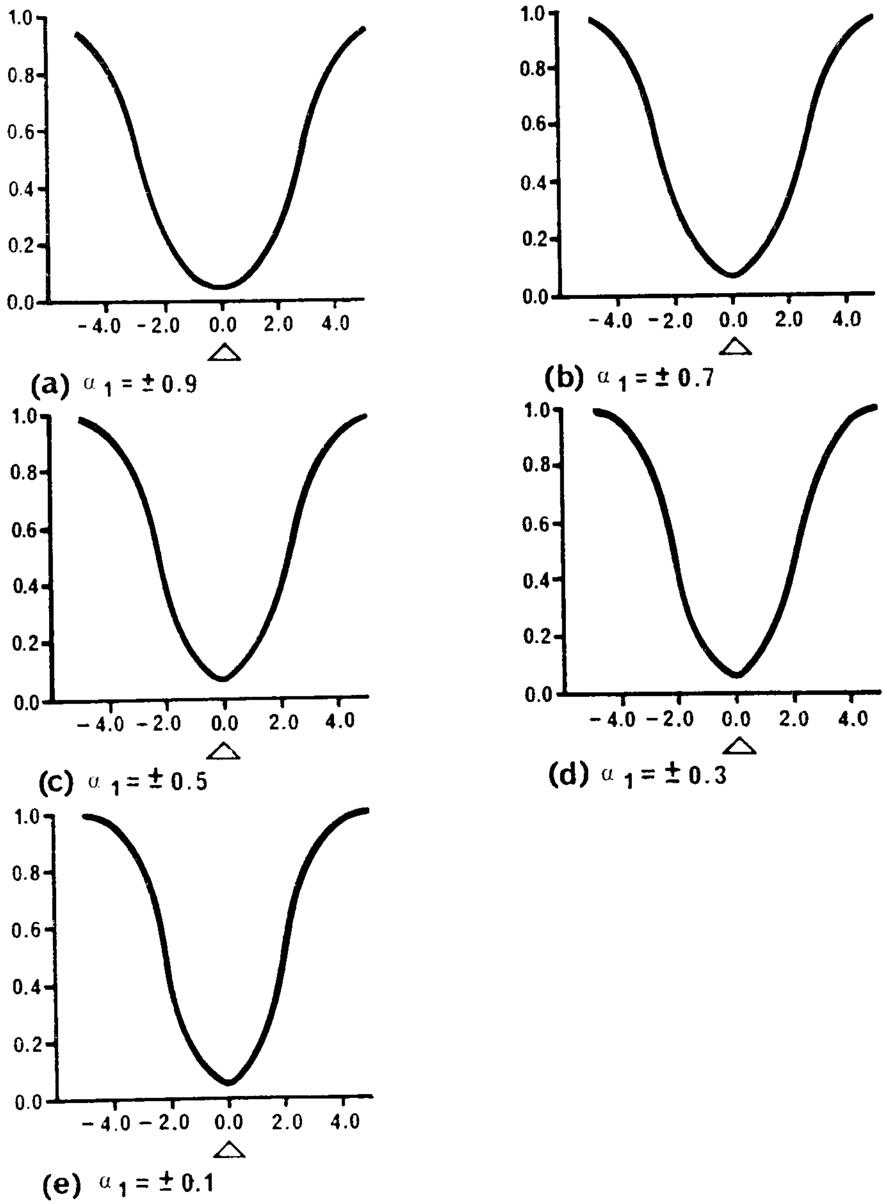

FIG. 7. The power, as a function of Δ, of Test I for model (3.1) with $p = 1$.

the observation is found more extreme. Since the distributions of the two criteria are related, a decision can be made without evaluating the individual likelihoods.

In practice, the most convenient approach would be to see if one can detect the effect the observation is having on subsequent observations. If the effect is detectable then Test II should be used; if not, then $\lambda_{q,n}$ should be used.

The criterion for testing whether the observation is made up of both types of outlier has also been derived. It is felt that the practical value of this criterion is restricted by the difficulty encountered in partitioning the outlier into its component parts. Because of this difficulty, neither the form of this criterion nor its distribution are presented.

Acknowledgements

The author is indebted to Professor D. R. Cox for his advice in the preparation of this paper and to the referee for many helpful comments and criticisms.

References

Anscombe, F. J. (1960). Rejection of outliers. *Technometrics*, **2**, 123–146.

Burman, J. P. (1965). Moving seasonal adjustment of economic time series. *J. R. Statist. Soc.* A, **128**, 534–558.

Durbin, J. (1959). Efficient estimation of parameters in moving average models. *Biometrika*, **46**, 306–316.

Grenander, U. and Rosenblatt, M. (1966). *Statistical Analysis of Stationary Time Series*. New York: Wiley.

Kruskal, W. H. (1960). Some remarks on wild observations. *Technometrics*, **2**, 1–3.

Walker, A. M. (1967). Some tests of separate families of hypotheses in time series analysis. *Biometrika*, **54**, 39–68.

Watson, G. S. (1954). Extreme values in samples from *m*-dependent stationary stochastic processes. *Ann. Math. Statist.*, **25**, 798–800.