J.R. Statist. Soc. B (1986)
48, *No.* 1, *pp.* 39–47

Distinguishing Outlier Types in Time Series

By C. R. MUIRHEAD†

Imperial College, London

[Received March 1983. Final revision June 1985]

SUMMARY

Distinguishing an outlier in a time series arising through measurement error from one arising through a perturbation of the underlying system can be of use in data validation. In this paper a method of testing for the presence of an outlier of unknown type is proposed. Then the properties of a rule based on the likelihood ratio which attempts to distinguish the two types of outlier are examined and compared with those of the corresponding Bayes rules. An example involving data from an industrial production process is studied.

Keywords: TIME SERIES; TYPE I (OBSERVATION) OUTLIER; TYPE II (INNOVATION) OUTLIER; MIXED OUTLIER; LIKELIHOOD RATIO RULE; BAYES RULES; PITMAN EFFICIENCY

1. INTRODUCTION

Fox (1972) considered a time series $\{X_t\}$ satisfying

$$X_t = \sum_{i=1}^{p} \alpha_i X_{t-i} + \eta_t \tag{1.1}$$

for all t, where η_t's are independently distributed, in particular with $\eta_t \sim N(0, \sigma^2)$ for $t \neq q$. The series $\{Y_t\}$ is observed, where

$$Y_t = X_t \quad \text{for} \quad t \neq q. \tag{1.2}$$

At some time q however, an outlier of one of the two following types is observed in $\{Y_t\}$.

Type I (observation) outlier

$$Y_q = X_q + \Delta, \quad \eta_q \sim N(0, \sigma^2).$$

This outlier occurs only in the observed series $\{Y_t\}$ and not in the underlying series $\{X_t\}$. Usually this would arise through measurement or recording error.

Type II (innovation) outlier

$$\eta_q \sim N(\Delta, \sigma^2), \quad Y_q = X_q.$$

Here an outlier occurs in the innovation series associated with $\{X_t\}$. In the case of, say, an industrial production process, this corresponds to a large perturbation of the underlying system.

Note that the observations following time q are disturbed by the occurrence of a type II outlier, but not by that of a type I outlier. In the terminology of Huber (1972), the former outlier gives rise to a 'bump' in the series, whereas an 'isolated' outlier arises in the latter case.

If data from a continuously operating production process are being monitored, the occurrence of a type I outlier indicates that action is required, possibly to adjust the measuring instrument or at least to print an error message on the data base. However, if a type II outlier occurs, no adjustment of the measurement operation is required. Thus distinguishing between these two types of outlier can be helpful in data validation.

† *Present address*: National Radiological Protection Board, Chilton, Didcot, Oxon, OX11 0RQ, UK.

0035–9246/86/48039 $2.00

The detection of outliers may be carried out sequentially or after a fairly large amount of data has been collected. For the former case, some control chart procedures were suggested by Muirhead (1983) (see also the comment in Section 2). For the case where a given set of data is being analysed, Fox (1972) considered tests, related to the likelihood ratio, of the null hypothesis that no outlier is present at time q against the alternative hypothesis than an outlier of a given type, either I or II, is present. This was done both when q is specified *a priori* and, as would usually be the case in practice, when q is selected after examining the data and noting a suspicious observation.

In order to test H_0: $\Delta = 0$ when the type of the outlier is unknown, it is suggested here that the alternative hypothesis be based on the mixed outlier model of Fox (1970), under which outliers of both types occur at time q. Note that within this mixed model the simple type I and type II models are of primary interest. The likelihood function associated with the mixed outlier model is obtained in Section 2, and it is shown that the maximum likelihood estimates of σ^2 and $\{\alpha_i\}$ under this model do not depend on the presence or type of any outlier at time q. A statistic to test for the presence of an outlier of unknown type is also suggested in this section.

If it has been established that an outlier is present, a rule can then be applied to decide whether it is of type I or of type II. One possible decision rule involves selecting the outlier type having the larger (maximized) likelihood. The power properties of this likelihood ratio rule are examined in Section 3 and compared with those of the corresponding Bayes rules in Section 4. In Section 5 the likelihood ratio rule and one of the Bayes rules are applied to data from a chemical production process.

If visual examination of the data or control charts indicate that more than one outlier is present, the methods described here can be adapted to allow each to be studied individually. However, it will be assumed that there is not a population of outliers in the data, at least not over too short a period of time. An alternative model, such that at each time point there is a small probability that an outlier of a given type will occur, has been considered by Abraham and Box (1979), using a Bayesian approach, and by Martin (1979). In the latter paper it was suggested that different estimates of $\{\alpha_i\}$ could be compared in an attempt to identify the type of the outliers when this is unknown.

2. MIXED OUTLIER MODEL

Suppose that (1.1) and (1.2) hold, with

$$Y_q = X_q + \Delta_1, \eta_q \sim N(\Delta_2, \sigma^2)$$

It will be assumed in this section that $Y = (Y_1, Y_2, \ldots, Y_n)^T = y$ has been observed, where n is fixed. Suppose also that $1 \leqslant p < q$ and $v = n - q \geqslant 1$, where q is provisionally assumed to be known. Let $\alpha = (\alpha_1, \ldots, \alpha_p)^T$ and

$$U_t(\alpha) = Y_t - \sum_{i=1}^{p} \alpha_i Y_{t-i} \; (t > p). \tag{2.1}$$

If $\sigma^2 \; \Sigma(\alpha)$ is the covariance matrix associated with $Y_{(p)} = (Y_1, \ldots, Y_p)^T$, then the log-likelihood function has the form $l = l_a + l_b$ where, in terms of the random variables $Y_{(p)}$ and $\{U_t(\alpha)\}$,

$$-2l_a(\alpha, \sigma^2) = (q-1)\log\sigma^2 + \log|\Sigma(\alpha)| + \sigma^{-2}\{Y_{(p)}^T \Sigma^{-1}(\alpha) Y_{(p)} + \sum_{i=p+1}^{q-1} U_i^2(\alpha)\},$$

$$-2l_b(\Delta_1, \Delta_2, \alpha, \sigma^2) = (v+1)\log\sigma^2 + \sigma^{-2}\left[\{U_q(\alpha) - \Delta_1 - \Delta_2\}^2 + \sum_{i=1}^{v}\{U_{q+i}(\alpha) + \alpha_i \Delta_1\}^2\right].$$

In particular, $l_b = l_b^{(1)} + l_b^{(2)}$, where

$$-2l_b^{(1)}(\Delta_1, \Delta_2, \alpha, \sigma^2) = \sigma^{-2}\left[\{U_q(\alpha) - \Delta_1 - \Delta_2\}^2 + \left\{\gamma(\alpha)\Delta_1 + \frac{1}{\gamma(\alpha)} \sum_{i=1}^{v} \alpha_i U_{q+i}(\alpha)\right\}^2\right], \quad (2.2)$$

$$-2l_b^{(2)}(\alpha, \sigma^2) = (v+1)\log \sigma^2 + \sigma^{-2} \sum_{i=2}^{v} \{U_{q+i}^*(\alpha)\}^2.$$

Here

$$\gamma(\alpha) = \left(\sum_{i=1}^{v} \alpha_i^2\right)^{1/2} \quad (2.3)$$

and $\{U_{q+i}^*(\alpha): 2 \leqslant i \leqslant v\}$ are the recursive residuals (Brown *et al.*, 1975) obtained by regressing U_{q+i} on α_i for $i = 1, 2, \ldots, v$ ($\alpha_i = 0$ for $i > p$).

Note that the likelihood l has been decomposed into a term $l_a + l_b^{(2)}$ involving only (α, σ^2) and a term $l_b^{(1)}$ which vanishes when (Δ_1, Δ_2) attain their maximum likelihood estimates $(\hat{\Delta}_1, \hat{\Delta}_2)$. Thus the maximum likelihood estimates $(\hat{\alpha}, \hat{\sigma}^2)$ of (α, σ^2) do not depend on Δ_1 or Δ_2.

Consider a test of H_0: $E(Y_q) = \Delta_1 + \Delta_2 = 0$, i.e. the observation at time q is not an outlier, against H_A: $E(Y_q) \neq 0$. By (2.2) the associated log-likelihood ratio statistic is equivalent to the statistic

$$T = \frac{U_q(\hat{\alpha})}{\hat{\sigma}}. \quad (2.4)$$

The simulations of Fox (1972) suggest that, under H_0, the distribution of T is well approximated by Student's t on $n - 2$ degrees of freedom for n not too small. If q is not known *a priori* but is selected after looking for any suspicious observations in the data, the significance level for the test based on T can be approximated using Bonferroni's inequality, i.e. as $nP(|T^*| \geqslant |T_{\text{obs}}|)$, where T^* is t_{n-2}. Note that a control chart based on $\{U_t(\tilde{\alpha})\}$, where $\tilde{\alpha}$ is an estimate of α which is robust with respect to outliers of both types (such as the generalized M-estimate of Martin, 1979), may be useful in detecting outliers. (When the data are monitored sequentially and an estimate α^0 of α is available from past data, Muirhead, 1983, suggests that a control chart based on $\{U_t(\alpha^0)\}$ be operated.)

If the model for $\{X_t\}$ contains moving-average terms, so that $p = \infty$ in (1.1), the likelihood function and residuals can be constructed using the Kalman filtering procedure described by Harvey and Phillips (1979). Provided that the effect of using a finite amount of data to construct the residuals has died out before time q, the results given in Sections 3 and 4 still apply.

3. LIKELIHOOD RATIO RULE

Suppose that it has been established that an outlier has occurred at time q. Here the outlier will be thought to be of a single type, i.e. arising either through measurement error or through a perturbation of the underlying process. Let the observations made up to and including time $q + v$ be used in attempting to identify the outlier type, where $v \geqslant 1$. As mentioned in Section 2, v would be taken equal to $n - q$ if a given set of data is being analysed. If, however, the observations are being monitored sequentially, v would be chosen *a priori* to represent the time between the outlier occurring and a decision regarding its type being made.

For convenience of notation, let the dependence of U_t and γ (given by (2.1) and (2.3) respectively) on α be suppressed. Using (2.2) it can be shown that, under $\Delta_1 = 0$, the maximum value of $l_b^{(1)}$ with respect to Δ_2 is such that

$$-2\sigma^2 l_b^{(1)}(\Delta_1 = 0, \Delta_2 = \tilde{\Delta}_2, \alpha, \sigma^2) = \left(\gamma^{-1} \sum_{i=1}^{v} \alpha_i U_{q+i}\right)^2 = S_I, \quad (3.1)$$

while, under $\Delta_2 = 0$, the maximum value of $l_b^{(1)}$ is such that

$$-2\sigma^2 l_b^{(1)}(\Delta_1 = \tilde{\Delta}_1, \Delta_2 = 0, \alpha, \sigma^2) = \frac{1}{1+\gamma^2}\left(\gamma U_q + \gamma^{-1} \sum_{i=1}^{v} \alpha_i U_{q+i}\right)^2 = S_{\text{II}}. \tag{3.2}$$

Hence, assuming for the moment that α is known, the likelihood ratio rule, which selects the outlier type having the larger (maximized) likelihood, is as follows:

L.r. rule: Decide on type I/II if and only if

$$W = S_{\text{I}} - S_{\text{II}} > / < 0. \tag{3.3}$$

The statistic W is the difference of two correlated $\sigma^2\chi_1^2$ random variables, one of which (S_i) is non-central and the other ($S_{\text{III}-i}$) central under a type i outlier (i = I, II). Hence, if a test of the adequacy of the single outlier model is required, the F-statistic min $(S_{\text{I}}, S_{\text{II}})/\hat{\sigma}^2$ can be examined.

The probability of deciding that the outlier is of type I, namely $P(W > 0)$, depends only on the standardized magnitude of the outlier (i.e. $|\Delta|/\sigma$, where $\Delta = E(Y_q)$), on the outlier type (either I or II) and on γ, which reflects the degree of auto-correlation in $\{X_t\}$ (see (2.3)). Given these parameters, the joint distribution of S_{I}/σ^2 and S_{II}/σ^2 can be obtained and $P(W > 0)$, expressed as a convolution, can be computed fairly easily by numerical integration, using for example NAG routine D01BAF (Numerical Algorithms Group, 1983). This probability is plotted against $\lambda = |\Delta|/\sigma$ in Fig. 1 for both outlier types and for various values of γ.

Let C denote the event that the outlier type is correctly identified. Then it can be seen from Fig. 1 that $P(C)$ increases with γ, i.e. with the amount of auto-correlation in $\{X_t\}$, and with λ. In particular, for γ and λ not too small, this probability is fairly high (e.g. $P(C) > 0.82$ for $\lambda \geqslant 4$ and $\gamma \geqslant 0.5$). Fig. 1 also shows that, for given values of λ and γ, $P(C)$ is roughly the same under both types of outlier. However, this probability is slightly larger under type I than under type

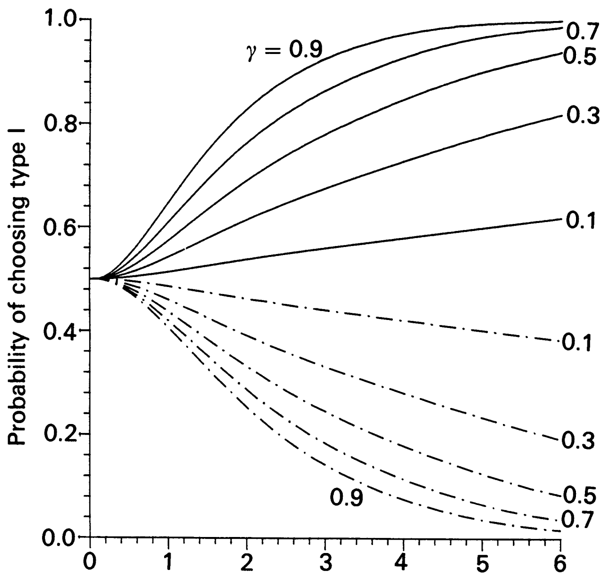

Fig. 1. Plot of the probability of deciding in favour of a type I outlier against λ, under the likelihood ratio rule with $\lambda = 0.1$ (0.2) 0.9. —— The outlier is of type I. -·-·- The outlier is of type II.

II, since the non-centrality parameter associated with $S_{\rm I}$ in the former case is larger than that associated with $S_{\rm II}$ in the latter case. In particular,

$$E(W) = \begin{cases} \Delta^2\gamma^2 & \text{for type I,} \\ -\dfrac{\Delta^2\gamma^2}{1+\gamma^2} & \text{for type II.} \end{cases}$$

From examination of probabilities computed over a wider range of values of λ and γ than those considered in Fig. 1, it appears that, for λ large relative to 1, $P(C)$ can be approximated solely as a function of $\delta = |E(W)|^{1/2}/\sigma$ in the manner shown in Table 1.

TABLE 1

Approximation to the probability of correctly identifying the outlier type using the likelihood ratio rule

δ	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5
Approx. $P(C)$	0.69	0.78	0.85	0.90	0.94	0.96	0.98	0.99

Consider now the effect of substituting an estimate of α into W when α is unknown. If a given set of data is being analysed, the maximum likelihood estimate under the mixed outlier model, namely $\hat{\alpha}$, can be used. Alternatively, if the observations are being monitored sequentially, an estimate of α obtained using past data can be employed. Then, if the degree of auto-correlation in $\{X_t\}$ has been over (under)-estimated, the probability of deciding that outlier is of type I (II) will be increased. As an illustration, consider the case where $\{X_t\}$ is $AR(1)$ with $\alpha = \alpha_1 = 0.5$. For an outlier of either type having $\lambda = 5$, Table 2 gives $P(C\,|\,\alpha_1^*)$ as a function of α_1^*, the estimate of α_1 substituted into W.

TABLE 2

Probability of correctly identifying the type of an outlier having $\lambda = 5$ using the likelihood ratio rule, for an $AR(1)$ model with estimate α_1^* of $\alpha_1 = 0.5$

α_1^*	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70
$P(C\,|\,\text{I}, \alpha_1^*)$	0.763	0.803	0.840	0.872	0.900	0.926	0.943	0.958	0.970
$P(C\,|\,\text{II}, \alpha_1^*)$	0.949	0.938	0.921	0.900	0.875	0.843	0.805	0.761	0.711

If α_1 is estimated efficiently using a series of length m, then, for m large, the distribution of α_1^* can be approximated as $N(\alpha_1, (1-\alpha_1^2)/m)$. Hence for such an estimator with $m = 100$ say, it can be shown using a finer version of Table 2 that unconditionally $P(C)$ is approximately equal to 0.896 under type I, compared with a value of 0.900 arising from the use of the true value of α_1 in this rule. Similarly, for a type II outlier, $P(C)$ is approximately equal to 0.869 unconditionally when α_1^* is used, compared with 0.875 for the case where $\alpha_1 = 0.5$ is used in the rule. Thus the unconditional probability of correctly identifying the outlier type when an estimate of α is used is similar to that based on the true value of α in this case. In general, however, the former probability may be somewhat smaller than the latter if the length of the series used to estimate α is not large.

4. BAYES RULES

The use of Bayes decision rules will now be considered. As in the previous section it will be assumed that the occurrence of an outlier at time q has been established and this outlier will be thought to be of a single type.

Let $\Delta = E(Y_q) \sim N(0, a\sigma^2)$ *a priori*, where a is known, and independently let the prior probability of a type i outlier be $P(i)$, $i = \mathrm{I}, \mathrm{II}$ ($P(\mathrm{I}) + P(\mathrm{II}) = 1$). The loss associated with choosing the correct type will be taken as 0, while that associated with an incorrect decision will be $l(\mathrm{i})$ when type i occurs. Then, with reference to (2.2) and for α, σ^2 known, the Bayes rule (Wald, 1950, p. 17) for distinguishing the outlier types is to decide in favour of type I/II if and only if

$$\frac{E_\Delta[\exp\{l_b^{(1)}(\Delta_1 = \Delta, \Delta_2 = 0, \alpha, \sigma^2)\}]}{E_\Delta[\exp\{l_b^{(1)}(\Delta_1 = 0, \Delta_2 = \Delta, \alpha, \sigma^2)\}]} > / < r,$$

where

$$r = \frac{l(\mathrm{II})P(\mathrm{II})}{l(\mathrm{I})P(\mathrm{I})}. \tag{4.1}$$

Using (2.2) and the result that

$$-2 \log E_\Delta\left[\exp\left\{\frac{-k(\Delta - z)^2}{2\sigma^2}\right\}\right] = \frac{kz^2}{\sigma^2(1 + ak)} + \log(1 + ak),$$

then the above rule can be re-expressed as follows:

Bayes rule: Decide on type I/II if and only if

$$Q(a) > / < d. \tag{4.2}$$

Here

$$Q(a) = \left(\gamma^{-1} \sum_{i=1}^{v} \alpha_i U_{q+i}\right)^2 - \frac{(1 + a)}{1 + a(1 + \gamma^2)}\left(\frac{a\gamma U_q}{1 + a} + \gamma^{-1} \sum_{i=1}^{v} \alpha_i U_{q+i}\right)^2, \tag{4.3}$$

$$d = \sigma^2 \log\left[r^2 \frac{\{1 + a(1 + \gamma^2)\}}{(1 + a)}\right]. \tag{4.4}$$

If α and σ^2 are unknown they can be replaced by estimates obtained either by fitting the mixed outlier model or by using past data. However, in order to study the properties of Bayes rules, these parameters will be assumed here to be known.

Consider a few special cases of (4.2).

Rule L: $a = \infty$, $d = 0$. The choice of a corresponds to taking an (improper) uniform prior distribution for Δ, as employed by Abraham and Box (1979). In addition $Q(\infty) = W$ (see (3.1)–(3.3)). Consequently, by also taking $d = 0$ in (4.2), this gives the likelihood ratio rule (3.3). Here $l(\mathrm{I})P(\mathrm{I}) > l(\mathrm{II})P(\mathrm{II})$ for $\gamma > 0$, the contribution to the expected prior loss from the occurrence of a type I outlier is greater than that associated with type II.

Rule B: $a = \infty$, $r = 1$. The log-likelihood ratio statistic $W = Q(\infty)$ is used, but here $d > 0$ for $\gamma > 0$. In particular, $l(\mathrm{I})P(\mathrm{I}) = l(\mathrm{II})P(\mathrm{II})$, i.e. both outlier types make the same contribution to the expected prior loss. It would be hoped that this choice of d would correct for the slight imbalance towards selecting type I present in rule L.

Since $a = \infty$ for both these rules, they will be compared with the following examples of rules based on finite prior variances for Δ but with similar values of d to rule L and B respectively.

Rule B': $a = 4$, $d = 0$. *Rule B''*: $a = 4$, $r = 1$.

The probability of deciding that the outlier is of type I, i.e. $P\{Q(a) > d\}$, conditional on $\lambda = |\Delta|/\sigma$, the outlier type and γ can be computed *via* numerical integration in a similar manner to before. In Fig. 2 this conditional probability is plotted against λ for both types of outlier and for rules L, B, B', B'' in the case $\gamma = 0.5$.

It can be seen from Fig. 2 that the choice of a has only a slight effect on this probability when λ is close to 0. However, for λ large, the behaviour of each rule is characterized mainly by the value of a, rather than by that of d (and hence rather than by those of r—see (4.1)—and σ^2 used in (4.4)).

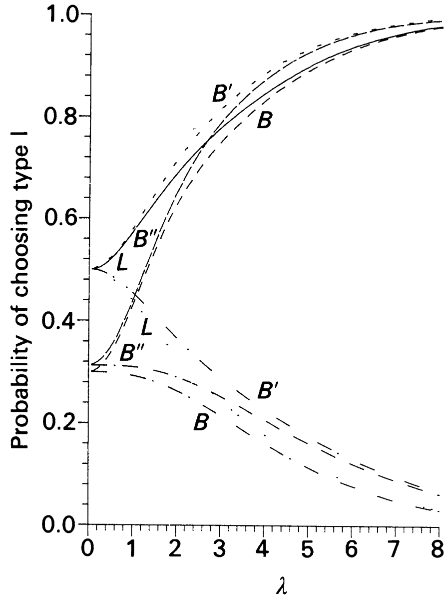

Fig. 2. Plot of the conditional probability of deciding in favour of a type I outlier against λ, under rules L, B, B', B'' and for $\gamma = 0.5$. The presence (absence) of dots denotes that the outlier is of type II (I).

It can also be seen from this Figure that, for $\lambda > 4$ say, the ratio

$$P(C\,|\,\mathrm{I}, \lambda)/P(C\,|\,\mathrm{II}, \lambda) \tag{4.5}$$

is close to 1 when $a = \infty$ (i.e. rules L and B) but is somewhat larger than 1 when $a = 4$ (i.e. for rules B' and B''). Examination of Pitman efficiencies, as carried out by Muirhead (1983), suggests that this imbalance towards selecting type I for λ large is minimized over $\{Q(a)$: $0 < a \leqslant \infty\}$ (see (4.3)) by taking $a = \infty$. In addition, the Pitman efficiency of $Q(a)$ at distinguishing the outlier types, as given by

$$P.E.\{Q(a)\} = \lim_{\Delta \to 0} \frac{[E\{Q(a)\,|\,\mathrm{I}, \Delta\} - E\{Q(a)\,|\,\mathrm{II}, \Delta\}]^2}{(\Delta/\sigma)^4 \operatorname{var}\{Q(a)\,|\,\Delta = 0\}}$$

$$= \frac{\{a^2\gamma(1 + \tfrac{1}{2}\gamma^2)\}^2}{(1 + a^2)\{1 + a^2(1 + \gamma^2)\}},$$

is a maximum for $a = \infty$. This suggests that the log-likelihood ratio statistic $W = Q(\infty)$ (see (3.1)–(3.3)) should be used in the rule.

As regards the choice of d in (4.2), Fig. 2 shows that for λ large the imbalance towards selecting type I, as measured by (4.5), is slightly smaller under rule B than under rule L.

However, for small or moderate values of λ (say $\lambda < 3$), there is a fairly large imbalance towards selecting type II under the former rule, whereas (4.5) remains close to 1 under the latter.

5. EXAMPLE

Fig. 3 shows a series of 110 observations corresponding to the measurement (on a standardized scale) of a variable in an Imperial Chemical Industries (I.C.I.) production process. The analysis considered here is preceded by the collection of all these data, i.e. is non-sequential. Since the observation at time 88 seems to be extremely small compared with the others, let $q = 88$. (Note that analysis of the small observation at time 32 does not suggest that it is an outlier or that it has a large effect on the parameter estimates for the time series model.)

Examination of the data over $t = 1$–80 suggests that an AR(1) model should be fitted. On fitting this model by maximum likelihood over $t = 1$–110 under the assumption that there is a possible outlier of unknown type at time 88, the estimate of the auto-regressive parameter is $\hat{\alpha}_1 = 0.486$ (est. $S.E. = 0.084$).

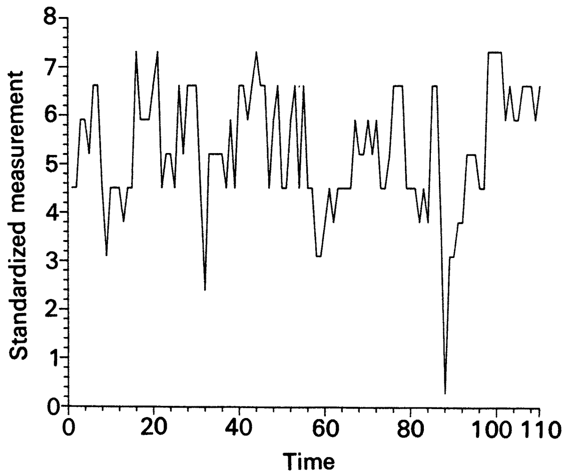

Fig. 3. Series from an industrial production process (units unspecified).

The observed value of the statistic T (see (2.4)), which tests for the presence of an outlier at time $q = 88$, is -4.30. If no outlier were present, T would be approximately distributed as Student's t on 107 d.f. (i.e. near enough $N(0, 1)$), where here the mean of the series had to be estimated. Using Bonferroni's inequality to make allowance for selection, the approximate significance level is 0.4%, i.e. there is strong evidence that an outlier has occurred. Note that provided the outlier is of a single type, T estimates Δ/σ.

Consider now the use of the likelihood ratio rule L and Bayes rule B in identifying the outlier type. With reference to (3.1)–(3.3), (4.2)–(4.4), the following values are observed, after replacing α_1 by $\hat{\alpha}_1$:

$$\frac{S_{\mathrm{I}}}{\hat{\sigma}^2} = 0.04, \quad \frac{S_{\mathrm{II}}}{\hat{\sigma}^2} = 2.90, \quad \frac{W}{\hat{\sigma}^2}\left(= \frac{Q(\infty)}{\hat{\sigma}^2}\right) = -2.86, \quad \frac{d}{\hat{\sigma}^2} = \begin{cases} 0 \ \text{, for rule } L, \\ 0.21, \text{for rule } B. \end{cases}$$

Thus, by (4.2), both rules decide in favour of a type II (innovation) outlier, i.e. that the outlier arose through a perturbation of the underlying process. In addition the small observed value of $S_{\mathrm{I}}/\hat{\sigma}^2$, which would be approximately distributed as χ_1^2 if only an innovation outlier were present, indicates that this model provides a good fit. This can also be seen from Fig. 3 by the gradual, rather than immediate, return of the observations to the equilibrium following time $q = 88$.

If the estimates $\gamma = |\hat{\alpha}_1| = 0.486$ and $\hat{\lambda} = 4.30$ for γ and λ respectively held exactly, then it can be shown that $P(C)$ would be quite close to 0.85 under both the above rules and for both outlier types.

6. CONCLUDING REMARKS

In attempting to distinguish between the two outlier types, the results of Section 4 suggest that a decision rule based on W, the log-likelihood ratio statistic ((3.1)–(3.3)), should be used. In particular, it would probably be best to use either the likelihood ratio rule L or the Bayes rule, B, for which both types make the same contribution to the expected prior loss. As can be seen from Figs 1 and 2, the probability of correctly identifying the outlier type using either rule can be quite high in a moderately auto-correlated series (say with $\gamma \geqslant 0.5$ — see (2.3)).

As with rule L, substituting estimates of α and σ^2 into the Bayes rules should have little effect on the unconditional probability of correct identification when the length of the series used to construct these estimates is reasonably large (say 100 or more). However, further examination could be made of the effect on $P(C)$ of using parameter estimates based on a relatively short series. Consideration could also be given to robust methods of identifying the time series model.

To deal with outliers in multivariate time series and those arising under more complex error structures, the problem can be extended to systems following a state space model. Here an outlier arising through a perturbation of the underlying 'state' of the process would be of type II, while an outlier arising through measurement error would be of type I. Muirhead (1983) has shown that, by using the Kalman filter to construct the likelihood function, it is possible to implement rules for distinguishing outlier types in this situation. However, the properties of these rules need to be studied further.

ACKNOWLEDGEMENTS

I wish to thank my Ph.D. supervisor, Professor D. R. Cox, for his guidance with this work. I would also like to thank Mr M. J. Weeks of I.C.I. Mond Division for supplying the data used in the example and Dr A. Baines (University of Leeds), Mr D. W. Brown (I.C.I. Petrochemicals and Plastics Division) and Mr G. Thackray (I.C.I. Mond Division) for related discussions. In addition, I am grateful to the referees for their comments. This research was funded by the S.E.R.C. and I.C.I. plc.

REFERENCES

Abraham, B. and Box, G. E. P. (1979) Bayesian analysis of some outlier problems in time series. *Biometrika*, **66**, 229–236.

Brown, R. L., Durbin, J. and Evans, J. M. (1975) Techniques for testing the constancy of regression relationships over time (with Discussion). *J.R. Statist. Soc.* B, **37**, 147–192.

Fox, A. J. (1970) *The Development and Application of a Forecasting Model.* Ph.D. Thesis, Department of Mathematics, Imperial College, University of London.

——(1972) Outliers in time series. *J.R. Statist. Soc.* B, **34**, 350–363.

Harvey, A. C. and Phillips, G. D. A. (1979) Maximum likelihood estimation of regression models with autoregressive-moving average disturbances. *Biometrika*, **66**, 49–58.

Huber, P. (1972) Robust statistics: a review. *Ann. Math. Statist.*, **43**, 1041–1067.

Martin, R. D. (1979) Robust estimation for time series autoregressions. In *Robustness in Statistics* (R. L. Launer and G. N. Wilkinson, eds), pp. 147–176. New York: Academic Press.

Muirhead, C. R. (1983) *Statistical Analysis of Data from On-line Computers.* Ph.D. Thesis, Department of Mathematics, Imperial College, University of London.

Numerical Algorithms Group (1983) *Fortran Mark 10 Library Manual.* Oxford: NAG Ltd.

Wald, A. (1950) *Statistical Decision Functions.* New York: Wiley.